

From System Performance to Strategic Interaction

Artificial Intelligence, Reciprocal Adaptation, and Crisis Stability

Author: Zijun Fu

Produced by Civilization Leap Research Group

Public website edition based on TNSR S0.2

CL-ACA-019 · Chinese baseline V1.2.1 · English TNSR S0.2

This English edition follows the version identified above. It is not a translation of the October 10 V1.3 Chinese edition.

FROM SYSTEM PERFORMANCE TO STRATEGIC INTERACTION

Artificial Intelligence, Reciprocal Adaptation, and Crisis Stability

Zijun Fu | Independent Researcher | Shenzhen, China | ORCID 0009-0003-5160-0321

Abstract

An AI system can give one state more time and leave both states with less time to correct. This article develops a conditional theory of how unilateral gains in prediction or speed can be translated through organizational procedures and adversary inference into a jointly produced decision-time architecture. It separates four defeasible hypotheses: AI arrangements may change perception; perception may generate plural responses; responses may feed back across actors and reduce retained correction time; and such losses may affect crisis stability only under additional strategic and institutional conditions. The article operationalizes a non-fungible reciprocal correction-time budget comprising verification time (T_V), communication time (T_C), and political re-entry time (T_P). Alliance adoption, nuclear and other high-irreversibility systems, and experiments or wargames are used as structured plausibility probes. Existing evidence supports early links and competing pathways, not the completed causal chain in real interstate crises. The central question is therefore not whether AI accelerates decisions, but whether reciprocal adaptation preserves enough time to discover and correct error before consequences become irreversible.

Keywords: artificial intelligence; strategic stability; crisis decision-making; deterrence; reciprocal adaptation; verification time; political re-entry; shared survival

1. Introduction: The Paradox of Faster Decisions

An AI system can give one state more time and leave both states with less time to correct. Earlier detection may help one organization verify a warning or prepare a defense. Once embedded in doctrine, authority, alert practices, alliance procedures, and visible operational arrangements, however, the same gain becomes an object of adversary inference. A rival may answer with automation, preauthorization, dispersal, or its own faster process. The strategic effect is therefore not located in model performance alone. It is produced by how organizations allocate the gain and by the interaction that follows.¹

Speed has no fixed strategic sign. Faster warning can enlarge verification; faster defensive coordination can strengthen survivability; faster communication can clarify intent. But a shorter technical cycle can also remove review, normalize action on warning, or punish pause. If a rival reads the change as a first-mover advantage or a weakening political firebreak, it may respond in kind. Each actor can rationally seek speed for safety while the feedback leaves both with less usable time to detect error, communicate limits, or reverse course.²

The familiar question—who can decide faster?—is therefore too narrow. The harder question is what happens to the time retained for correction after organizations and rivals adapt. Technical gains may be reassigned to cross-checking, dissent, and communication, or used to remove review and advance commitment. This article asks: How

1 Avi Goldfarb 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; Michael C. Horowitz, “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>.

2 Michael C. Horowitz, “When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability,” *Journal of Strategic Studies* 42, no. 6 (2019): 764–88, <https://doi.org/10.1080/01402390.2019.1621174>; Michael C. Horowitz, “Artificial Intelligence and the Future of Strategic Stability,” *Texas National Security Review* 9, no. 2 (2026): 66–83, <https://doi.org/10.1353/tns.00034>.

do organizational adoption and reciprocal adaptation translate advanced AI capabilities into changes in the time available for verification, communication, and political re-entry during strategic crises?

The argument is conditional. Advanced AI does not directly or unidirectionally determine strategic stability. Its effects travel through a sequence: capabilities are adopted by organizations; organizations distribute information and authority; those distributions affect posture and decision procedures; rivals infer meanings from visible or suspected changes; responses feed back across multiple rounds; and only then do consequences appear in deterrence, crisis stability, or escalation control. Every link can branch or fail. A rival may retaliate, imitate, verify independently, harden defenses, demand oversight, reject automation, accommodate, or seek a guardrail. A system may shorten one portion of a process while enlarging another. Political leaders may recover time that military organizations appear to have lost. A theory that admits only acceleration will miss restraint; a theory that treats every use of AI as stabilizing will miss jointly produced pressure.

The article makes three contributions. First, it separates system performance from strategic interaction and specifies the organizational and perceptual bridges between them. Second, it states four externally legible hypotheses—H1a through H1d [Canonical P3a through P3d]—that can fail separately: AI arrangements may change perception; perception may generate plural responses; responses may feed back and reduce retained correction time on more than one side; and reduced time may raise crisis risk only under additional conditions. Third, it operationalizes a non-fungible reciprocal correction-time budget, (T_V , T_C , T_P), for verification, communication and attribution correction, and political re-entry.

These contributions are narrower than claims already present in the literature. Research has long examined action–reaction dynamics, the security dilemma, military innovation, speed, autonomy, nuclear command and control, and the possibility that AI changes deterrence or crisis decision-making. Recent work directly addresses adoption incentives and escalation in U.S.–China military AI competition, the sociotechnical mediation of deterrence effects, and the non-linear relationship between AI and strategic stability. This article does not claim to discover those domains. Its proposed addition is a testable architecture that follows performance gains through organizations and adversary inference, specifies what counts as reciprocal rather than unilateral compression, and gives stabilizing or null outcomes the same standing as escalation evidence.³

The evidence supports theory building, not a declaration that reciprocal compression has occurred in an interstate crisis. Official materials establish adoption and design intent; experiments identify bounded micro-mechanisms; policy sequences suggest mutual attention; wargames and simulations reveal possible escalation, restraint, exit, and command burdens. Missing is an AI-specific trace of one actor’s organizational change, a rival’s interpretation and procedural response, feedback to the first actor, changes in retained T_V , T_C , or T_P on multiple sides, and a credible crisis outcome. The article makes that missing chain explicit and testable.

Section 2 organizes existing work around performance, organizational mediation, and interactive feedback. Section 3 defines the causal architecture, scope, and probe-based design. Section 4 states H1a–H1d. Section 5 develops the reciprocal correction-time budget. Section 6 tests the architecture through alliance, high-irreversibility, experimental, and wargaming evidence. Section 7 asks what would make the argument wrong. Section 8 gives limited design implications. The conclusion returns to the central paradox: faster machines may be valuable chiefly when they return time to verify, communicate, and stop.

EVIDENCE BOUNDARY — AUTHOR CONTROL: The complete reciprocal-compression chain is not claimed to have been observed. Section 1 establishes the puzzle, research question, narrow contributions, and inferential gap.

3 Michael C. Horowitz, “Artificial Intelligence and the Future of Strategic Stability,” *Texas National Security Review* 9, no. 2 (2026): 66–83, <https://doi.org/10.1353/tns.00034>; Lindsay Rand, “Products of Patrons and Producers: A Sociotechnical Approach to Technological Innovation and Nuclear Deterrence,” *Journal of Global Security Studies* 11, no. 3 (2026): ogag018, <https://doi.org/10.1093/jogss/ogag018>; Shaoyu Yuan, “World and War in the Age of AI: Cyber Risk, Deterrence, and Decision-Making in U.S.-China Competition,” *Defence Studies*, published online August 15, 2026, <https://doi.org/10.1080/14702436.2026.2718345>.

2. Existing Explanations and the Remaining Gap

2.1 Security, crisis, and the organizational inheritance

The starting point is the security dilemma. Security is not a possession that one actor can calculate from its own inventory alone. Measures intended to reduce vulnerability may appear threatening because capabilities are visible while intentions are uncertain, offensive and defensive preparations are difficult to distinguish, and future restraint cannot be guaranteed. The mechanism does not require aggressive preferences. It can arise among actors seeking protection. For AI, this means that the relevant question is not whether a new capability is labeled defensive by its owner, but whether rivals can distinguish its mission, limits, and authority and whether adoption changes their expected cost of waiting. A system that improves warning may be stabilizing if it makes retaliation more survivable; the same system may appear destabilizing if it is believed to improve counterforce targeting. The strategic property is relational.⁴

Crisis bargaining and organizational theories add two further cautions. Information may be private, incentives to misrepresent may persist, and commitment problems can make negotiated outcomes difficult even for rational actors. States are also not frictionless unitary calculators. Perception, bureaucratic routines, standard operating procedures, and normal-accident dynamics shape which signals enter a decision, how anomalies are treated, and whether a warning is escalated or questioned. AI may improve parts of this process, but it can also automate the institutional habits that produced error before its arrival. A model that processes more data does not remove incentives to deceive; a faster staff process does not dissolve organizational rivalry; an interface that presents a ranked option does not guarantee that omitted alternatives remain politically recoverable.⁵

Strategic stability is likewise multidimensional. Peacetime competition, crisis stability, deterrence credibility, wartime escalation, and command-and-control resilience can move in opposite directions. A technology may strengthen retaliation and reduce incentives for a disarming strike while increasing accidental-use risk through automation. It may improve tactical discrimination but widen the set of targets that can be struck quickly. It may generate stable effects at one stage and unstable effects at another. This literature is a direct warning against an aggregate “AI stability score.” It also explains why a time-budget approach must keep verification, communication, and political re-entry separate: stability can fail through different mechanisms, and gains in one dimension do not erase the loss of another.⁶

2.2 What the military-AI literature already explains

The military-AI literature has already moved beyond autonomous weapons as the sole object. Horowitz treats AI as an enabling technology whose strategic consequences depend on diffusion and organizational absorption. Goldfarb and Lindsay distinguish prediction from judgment, arguing that cheaper prediction can increase the

4 David A. Baldwin, “The Concept of Security,” *Review of International Studies* 23, no. 1 (1997): 5–26, <https://doi.org/10.1017/S0260210597000053>; John H. Herz, “Idealist Internationalism and the Security Dilemma,” *World Politics* 2, no. 2 (1950): 157–80, <https://doi.org/10.2307/2009187>; Robert Jervis, “Cooperation under the Security Dilemma,” *World Politics* 30, no. 2 (1978): 167–214, <https://doi.org/10.2307/2009958>.

5 Graham T. Allison and Philip Zelikow, *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd ed. (New York: Longman, 1999); James D. Fearon, “Rationalist Explanations for War,” *International Organization* 49, no. 3 (1995): 379–414, <https://doi.org/10.1017/S0020818300033324>; Charles Perrow, *Normal Accidents: Living with High-Risk Technologies* (New York: Basic Books, 1984); Scott D. Sagan, *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons* (Princeton, NJ: Princeton University Press, 1993); Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966).

6 Ben Garfinkel and Allan Dafoe, “How Does the Offense-Defense Balance Scale?” *Journal of Strategic Studies* 42, no. 6 (2019): 736–63, <https://doi.org/10.1080/01402390.2019.1631810>; Todd S. Sechser, Neil Narang, and Caitlin Talmadge, “Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War,” *Journal of Strategic Studies* 42, no. 6 (2019): 727–35, <https://doi.org/10.1080/01402390.2019.1626725>; Caitlin Talmadge, “Emerging Technology and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today,” *Journal of Strategic Studies* 42, no. 6 (2019): 864–87, <https://doi.org/10.1080/01402390.2019.1631811>.

relative importance of human judgment rather than eliminate it. King's studies of digital targeting show AI and data entering military intelligence and target-development processes through concrete organizational arrangements. These works establish the first bridges in the present chain: capabilities are not strategic effects; they must be translated into practices, and those practices depend on institutions, data, skills, and task structures.⁷

Research on human-machine decision-making makes the translation problem still clearer. Horowitz and Lin-Greenberg show that identical or comparable crisis information can produce different judgments when attributed to AI rather than human analysis, but not in a uniformly hawkish direction. Horowitz and Kahn demonstrate across nine countries that task difficulty, knowledge, trust, self-confidence, and perceived competence influence automation bias and algorithm aversion. Shandler, Gross, and Shereshevsky find that AI labeling can reduce approval for a strike while failing to shorten decision time in high-fidelity military decision-support experiments. Together, these findings undermine two opposite simplifications: humans will not automatically surrender to AI, and the presence of caution does not mean AI has no strategic influence. Organizational protocols and beliefs determine whether machine output becomes advice, default, veto, or ignored signal.⁸

Alliance scholarship identifies a separate arena in which time and authority are jointly produced. Lin-Greenberg argues that uneven AI capabilities, data-sharing barriers, divergent trust, and adversary deception can complicate multinational operations and decision-making. AI can accelerate the tempo of war while alliance decisions still require negotiation among governments with different legal authorities, risk tolerances, and political obligations. The result is not simply slower or faster decision-making. A common operating picture may increase technical coordination while the political coalition loses time to debate, or disparities in trust may slow collective action even as machine processes accelerate. Alliance interaction therefore provides a natural setting in which T_V, T_C, and T_P may diverge.⁹

Nuclear-focused work raises the irreversibility stakes. SIPRI studies describe potential benefits in warning, maintenance, intelligence, surveillance, reconnaissance, and decision support, alongside concerns about false alarms, cyber compromise, opacity, counterforce perceptions, and arms-racing. Herbert Lin argues for meaningful human control, posture-level risk assessment, self-restraint, and uses of AI that strengthen oversight rather than replace judgment. Horowitz's 2026 analysis emphasizes that AI's relationship with strategic stability is non-linear and conditioned by learning, data, use cases, and second-strike vulnerability. None of these works supports a blanket statement that AI is either stabilizing or destabilizing. They show why the same technical capacity can alter incentives differently across postures and over time.¹⁰

A second group of studies directly narrows the room for originality. Rand's sociotechnical framework rejects technological determinism by locating deterrence effects in the interplay of technological characteristics, strategic

7 Anthony King, "Digital Targeting: Artificial Intelligence, Data, and Military Intelligence," *Journal of Global Security Studies* 9, no. 2 (2024): ogae009, <https://doi.org/10.1093/jogss/ogae009>.

8 Michael C. Horowitz and Lauren Kahn, "Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts," *International Studies Quarterly* 68, no. 2 (2024): sqae020, see "Research Design," figs. 2–7, and tables 1–2, <https://doi.org/10.1093/isq/sqae020>; Michael C. Horowitz and Erik Lin-Greenberg, "Algorithms and Influence: Artificial Intelligence and Crisis Decision-Making," *International Studies Quarterly* 66, no. 4 (2022): sqac069, see the two survey experiments and findings, <https://doi.org/10.1093/isq/sqac069>; Ryan Shandler, 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), two experiments involving 2,015 Israeli military personnel, <https://doi.org/10.1177/00220027261463443>.

9 Erik Lin-Greenberg, "Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making," *Texas National Security Review* 3, no. 2 (2020): 56–76, <https://doi.org/10.26153/tsw/8866>; Glenn H. Snyder, "The Security Dilemma in Alliance Politics," *World Politics* 36, no. 4 (1984): 461–95, <https://doi.org/10.2307/2010183>.

10 Vincent Boulanin et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (Stockholm: SIPRI, 2020), <https://www.sipri.org/publications/2020/policy-reports/artificial-intelligence-strategic-stability-and-nuclear-risk>; Herbert S. Lin, "Artificial Intelligence and Nuclear Weapons: A Commonsense Approach to Understanding Costs and Benefits," *Texas National Security Review* 8, no. 3 (2025): 98–109, <https://doi.org/10.26153/tsw/60739>.

logics, and social actors. Johnson, Zatsepina, and Yuan discuss feedback, speed, adoption, divergent deterrence logics, and action–reaction pressure. Lamberth and Scharre examine the feasibility and limits of AI arms control. These contributions mean that no responsible article can claim that reciprocal adaptation, organizational mediation, or AI-related speed risks are new subjects. The remaining gap is more specific: the literature lacks a common, non-fungible measurement architecture for the time retained to verify, communicate, and politically intervene across reciprocal organizational changes, coupled with propositions that allow each link to fail without protecting the overall thesis.¹¹

2.3 The remaining analytical and empirical gap

The gap also has an empirical side. Official strategies, acquisition announcements, and international resolutions show that states and alliances have recognized the problem and begun to design rules. They do not establish operational effects. Experiments identify micro-mechanisms but do not reproduce real leaders, secrecy, repeated interaction, or institutional responsibility. Wargames expose plausible pathways while remaining sensitive to scenarios, adjudication, player composition, and rules. Model-on-model simulations reveal properties of models in constructed worlds, not state behavior. The literature is therefore richest at the beginning and middle of the causal chain and thinnest at the point where reciprocal organizational changes become an observed stability outcome. The correct response is not to abandon theory, but to make the inferential stopping points explicit.

The proposed contribution can now be stated without inflation. Existing research already examines AI-enabled speed, military adoption, human-machine judgment, strategic stability, alliance friction, nuclear risk, and reciprocal competition. This article asks how those elements might be assembled into a decision-time architecture that can be observed and falsified. It does not claim that reciprocal compression exists as a general law. It specifies the evidence that would justify saying it occurred, the stabilizing outcomes that would prevent it, and the facts that would require the argument to be reduced or abandoned. That is the remaining analytical task.

The remaining gap is partly a problem of level. Security-dilemma research usually begins with states and their reciprocal interpretations; military-AI research often begins with technical functions or adoption; organizational research begins with routines, authority, and failure. The strategic effect under study appears only when these levels are followed in sequence. A prediction system does not confront an adversary. An organization uses it, defines what counts as warning, decides who may challenge the output, and assigns authority to act. The adversary then observes only part of that arrangement and may infer intent from capability, deployment pattern, doctrine, or silence. The missing middle is therefore not a new label for competition. It is the translation process through which a technical affordance becomes an organizational signal and then an interactive constraint.

A second gap concerns temporal decomposition. Existing work often distinguishes fast and slow systems, but an interstate decision process contains several clocks: machine processing, staff interpretation, source verification, alliance consultation, legal or operational authorization, adversary communication, and political reconsideration. These clocks may run in parallel, sequentially, or not at all. A system can reduce machine latency while increasing the queue of unresolved warnings. It can accelerate target generation while slowing coalition approval. It can provide leaders with more notice but fewer meaningful alternatives. Without separating these functions, researchers may mistake a faster component for a faster decision, or a faster decision for a smaller opportunity to correct error. The article’s time budget is designed to make those distinctions observable rather than rhetorical.

A third gap is normative and empirical at once. Strategic studies rightly avoid treating delay as an unconditional good: delay can leave civilians or forces exposed, allow an attack to succeed, or prevent timely defense. Yet organizations also require time for challenge, attribution, consultation, and accountable authority. The analytical task is not to praise slowness. It is to identify which delays are friction, which are safeguards, and which technical gains can be converted into additional correction rather than earlier commitment. That distinction links

11 James Johnson, “Revisiting the ‘Stability–Instability Paradox’ in AI-Enabled Warfare,” *Review of International Studies*, First View (2024): 1–19, <https://doi.org/10.1017/S0260210524000767>; Megan Lamberth and Paul Scharre, “Arms Control for Artificial Intelligence,” *Texas National Security Review* 6, no. 2 (2023), <https://tnsr.org/2023/05/arms-control-for-artificial-intelligence/>; Luba Zatsepina, “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>.

established debates on speed, control, and stability to a researchable question about the allocation of saved time. It also prevents the article’s normative concern for responsibility from being smuggled into the evidence as a presumed causal effect.

EVIDENCE BOUNDARY — AUTHOR CONTROL: Originality is limited to the mediating architecture, separately defeasible H1a–H1d [Canonical P3a–P3d], and the non-fungible reciprocal correction-time budget. The article does not claim to originate AI speed risks, action–reaction dynamics, or sociotechnical mediation.

3. Theory, Scope Conditions, and Research Design

The unit of analysis is strategic interaction among states, alliances, and military organizations under conditions of time pressure, incomplete information, and potentially irreversible consequences. The article does not analyze whether a particular model is conscious, aligned in a philosophical sense, or legally responsible. Nor does it assume that artificial general intelligence has arrived. “Advanced AI” refers here to systems capable of materially affecting high-consequence processes such as intelligence fusion, warning, targeting, cyber defense, option generation, command support, logistics, autonomous execution, or the coordination of distributed assets. The theory applies only when those capabilities enter organizations and become relevant to adversaries’ expectations or choices.

The causal architecture contains seven distinct transitions: capability to organizational adoption; adoption to authority and strategic posture; posture to decision-time allocation; allocation to adversary perception; perception to response; response to reciprocal, multi-round feedback; and feedback to crisis-stability outcomes. The transitions are analytical bridges, not assumptions. An acquisition announcement may establish adoption without revealing authority. A lower system latency may establish technical speed without showing a shorter authorization cycle. A rival’s public concern may establish perception without proving procedural change. A bilateral arms race may show competition without demonstrating that verification, communication, or political re-entry have declined. Each transition therefore requires evidence appropriate to its level.

Transition	Current evidentiary reach	Prohibited inference
AI capability → organizational adoption	Relatively strong factual baseline	Adoption does not establish operational performance or strategic effect
Adoption → authority / posture / decision time	Moderate mechanism evidence	Technical latency does not equal organizational decision time
Decision process → adversary perception	Conditional experimental support	Experimental perception does not establish state behavior
Perception → response	Conditional and directionally plural	Response cannot be coded as acceleration only
Response → reciprocal time change	Partial / structural hypothesis	Parallel adoption is not reciprocal compression
Time change → crisis stability	Open / conditional hypothesis	No current basis for a general net-effect claim

The theory is most relevant under five scope conditions. First, the mission is time-sensitive enough that delay has strategic cost. Second, consequences are sufficiently severe or difficult to reverse that error correction matters independently of task performance. Third, organizations can alter authority, review, or alert procedures in response to perceived advantages. Fourth, rivals can observe, infer, or imagine those changes, even imperfectly. Fifth, actors remain capable of adaptation across more than one round. Where tasks are routine, consequences easily reversible, AI use invisible and strategically irrelevant, or organizations unable to change, the reciprocal mechanism should be weak or absent. These conditions prevent every use of AI from being relabeled as crisis compression.

Four argumentative identities govern the article. A claims are established facts or formal institutional materials. B claims are unsettled mechanisms and competitive interpretations. C claims are cross-domain structural

hypotheses whose complete sequence is unverified. F claims are explicit normative commitments. The labels prevent category errors: a policy proves a rule exists, not that control works; an experiment supports a bounded mechanism, not a crisis outcome; shared survival states a normative floor, not an empirical result.

The research design combines mechanism-centered theory with structured plausibility probes. The probes are not selected to prove the theory. They represent different points at which the proposed chain should confront difficult evidence: alliance AI exposes multi-government authority and interoperability; nuclear and other high-irreversibility systems expose the cost of lost firebreaks; experiments and wargames expose perception, response, uncertainty, and repeated interaction. Each probe is evaluated by the same questions: What is the factual baseline? Which transition can the material support? What can it not support? Which alternative explanation competes with the proposed mechanism? What observation would weaken the article? Calling these materials probes rather than case studies is deliberate because the available evidence lacks a complete event chain, transparent time series, and causal identification.

The design also distinguishes AI from adjacent processes. General digitization, intelligence reform, precision strike, commercial cloud adoption, organizational centralization, common-threat perception, and shifts in doctrine can all alter decision time. If those factors explain an observed change better than AI, the AI variable must be downgraded. The article therefore treats AI as an enabling and mediating condition, not a single sufficient cause. The aim is to identify the incremental role of AI-related capability, visibility, trust, or automation in a broader sociotechnical process. This distinction is essential because otherwise any acceleration in a modern military could be counted as evidence for the argument.

Evidence is assessed by an inferential reach rule. Formal policy or procurement can reach organizational intent and adoption, but not effectiveness. Randomized experiments can reach individual judgment under defined manipulations, but not repeated interstate adaptation. Wargames can reach plausible interactive mechanisms and reveal unanticipated branches, but not event probabilities. Process tracing can reach reciprocal organizational change only if it observes actions, interpretations, procedural responses, feedback, and timing. Stability effects require comparison or a credible counterfactual that separates AI's incremental role from common threats and pre-existing reforms. The current literature reaches the first three levels unevenly and does not reach the last two for the full AI-specific chain.

The theory is intentionally vulnerable. H1d can fail while perception and procedure remain worth studying; H1c can fail while the time budget remains useful for unilateral design. But the central claim about jointly produced correction time is defeated if repeated bilateral evidence shows no reciprocal feedback or no loss in any matching time dimension. The framework is valuable only if it clarifies where the evidence stops.

The primary unit of observation is a decision-process episode nested within an interaction. An episode begins when a relevant anomaly, warning, adversary signal, or policy change enters an organization and ends when the organization commits, revises, delays, communicates, or terminates a consequential path. The interaction begins when another actor perceives or infers that episode and adjusts its own procedure. This unit avoids two common errors. It does not treat a model output as a state decision, and it does not require an entire war to become the case. It permits researchers to trace smaller but strategically meaningful changes in warning, review, authorization, communication, and political control while retaining the requirement that reciprocal compression must cross organizational boundaries and return through feedback.

Each transition has distinct observables. Capability is recorded through task performance and technical limits. Adoption requires evidence that an organization integrates the capability into a real or exercised workflow. Authority change requires rules, interfaces, or practice showing who may recommend, approve, delegate, suspend, or reverse. Strategic posture requires deployment, alert, doctrinal, or operational change. Perception requires statements, experiments, intelligence assessments, interviews, or behavior that links the other actor's arrangement to an interpretation. Response requires a subsequent procedural or posture change. Feedback requires the first actor to react again. Crisis effect requires an outcome such as corrected attribution, avoided action, escalation choice, accidental engagement, successful communication, or preserved reversibility. Missing evidence at one level cannot be replaced by abundance at another.

The theory is least applicable where decisions are low consequence, easily reversible, insulated from adversary observation, or governed by abundant time and robust appeal. It is also weaker where AI functions are purely administrative and do not alter information, authority, posture, or external inference. By contrast, the theory should be most visible when consequences become difficult to reverse; warning and authorization are tightly coupled; actors fear a closing window; procedures are partly observable or inferable; and each side can alter the other's incentives. Stating these negative and positive scope conditions protects the article from turning every military use of AI into evidence. It also creates a sampling rule: select episodes because the proposed mechanism could operate and include cases in which it does not.

The three probes follow a structured, focused logic rather than a most-likely-case claim. Each is asked the same questions: What AI-related function entered the process? What organizational rule or dependence mediated it? What could another actor observe or infer? Which response branches were available? Which time dimension could change? What evidence would demonstrate restraint, null effect, or reversal? Alliances expose consultation and shared dependence; nuclear and other high-irreversibility systems expose survivability, warning, and political authority; experiments and wargames expose perception, response plurality, and interaction under controlled or constructed conditions. Their diversity is deliberate. Agreement across unlike methods strengthens a mechanism; disagreement identifies a scope condition rather than an inconvenience to be averaged away.

A future causal test would require paired or sequential process evidence. One design could compare the same organization before and after an AI-supported workflow while separately timing verification, communication, and political review. Another could randomize information about an adversary's AI arrangement and observe both immediate judgment and subsequent procedural choice. A bilateral exercise could allow each team to adapt over multiple rounds while recording when saved time is used for independent checks, earlier authorization, or communication. Historical process tracing could compare publicly observable reforms with internal records, interviews, or declassified materials. None of these designs is available in complete form here. The present article therefore earns only mechanism-building status and sets out the data architecture that would be required to move beyond it.

EVIDENCE BOUNDARY — AUTHOR CONTROL: This is a mechanism-building article using structured plausibility probes, not a comparative case study or causal-effect estimate. A/B/C/F identities define argumentative responsibility, not a single evidence ranking.

4. Four Separately Defeasible Propositions

4.1 H1a [Canonical P3a]: AI arrangements and adversary perception

H1a concerns perception. When an adversary becomes aware that AI or automation is involved in a high-consequence process, does that knowledge alter judgments of threat, credibility, accident risk, responsibility, or willingness to act? The proposition is not that an “AI” label always heightens fear. It is that provenance, visibility, perceived competence, and degree of automation can become strategically meaningful attributes. The expected direction depends on the task and narrative. AI support may be interpreted as superior analysis, less accountable action, greater accident risk, reduced political control, or a merely technical improvement. H1a is therefore a B claim: it is empirically supported in some settings but remains conditional across actors and contexts.

Horowitz and Lin-Greenberg provide direct evidence for the mechanism. In randomized scenarios involving U.S. national-security professionals, attributing analysis or an accident to AI changed support for action, but in different directions: AI-provided intelligence reduced support for a preemptive strike in one design, while an adversary's automated accident could generate greater retaliatory preference than a comparable human-caused accident. The strategic implication is not that AI is pacifying or escalatory. It is that the source and perceived agency of a decision can alter how an event is interpreted. The experiment reaches perception and response preference; it does not reach real leaders, repeated interaction, organizational time allocation, or crisis stability.

Schwartz and Horowitz's experiments on automated nuclear threats reinforce the conditionality. Automation could affect credibility, perceived accident risk, and willingness to cross or avoid a red line, while some outcomes were null and some responses involved accommodation rather than escalation. Lushenko's conjoint evidence

similarly indicates that complete autonomy and higher risk can reduce trust, while the identity of other adopters does not automatically make military professionals more accepting of AI. Shandler and colleagues find algorithm aversion in parts of the process and no general reduction in decision time. These studies make H1a defensible and an unconditional automation-bias story indefensible.¹²

4.2 H1b [Canonical P3b]: Perception and the response set

H1b concerns the response set. A perception change does not specify what an adversary will do. Responses may include retaliation, accommodation, independent verification, emulation, countermeasures, greater human oversight, refusal to automate, dispersal, defensive hardening, or a proposal for reciprocal limits. A theory that codes only acceleration builds its conclusion into the dependent variable. H1b therefore requires a competitive response design: multiple outcomes must be measured together, and the conditions that make each more likely must be stated. The mission, level of risk, doctrine, alliance commitments, confidence in the system, domestic politics, and whether the AI use is visible as defensive or offensive can all alter the branch.

Open-source evidence suggests that strategic perception can enter organizational attention and procurement. Zeng's analysis of Chinese military-AI discourse shows the United States functioning as a reference point, source of anxiety, and stimulus for discussion of adaptation. U.S. officials have likewise tied the urgency of organizational integration and decision advantage to a perceived pacing challenge. These materials support the possibility of perception-to-adoption pressure, but they do not identify a single cause or reveal secret decision processes. General military modernization, bureaucratic interests, commercial technology, and common threat environments remain plausible explanations. H1b should therefore be tested through paired policy sequences rather than inferred from parallel spending.¹³

The same competitive relationship can also generate restraint. In November 2024, official U.S. and Chinese readouts recorded agreement on the need to maintain human control over decisions to use nuclear weapons and to develop military AI in a prudent and responsible manner. That statement is not a treaty, a verification regime, or proof that command procedures changed. It does demonstrate that reciprocal concern can produce a guardrail rather than symmetric acceleration. Any empirical test of H1b that ignores communication, restraint, and jointly expressed limits would misdescribe the range of strategic response.¹⁴

4.3 H1c [Canonical P3c]: Reciprocal compression

H1c concerns reciprocal compression and carries the heaviest evidentiary burden. It is not satisfied when one actor uses AI, when two actors modernize, or when an arms race is declared. It requires a process in which one

12 Paul Lushenko, "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, see "Research Design" and "Findings," table 2 and figs. 1–2, <https://doi.org/10.1017/cfl.2025.10019>; Joshua A. Schwartz and Michael C. Horowitz, "Delegating Destruction: Coercive Threats and Automated Nuclear Systems," International Organization 80, no. 1 (2026): 179–204, see "Data and Methods," "Results," and fig. 1, <https://doi.org/10.1017/S0020818325101215>; Ryan Shandler, 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>.

13 Kathleen Hicks, "Remarks by Deputy Secretary of Defense Kathleen Hicks Keynote on 'The Global AI Contest' at the Advantage DOD 2024: Defense Data & AI Symposium in Washington, D.C. (As Delivered)," U.S. Department of Defense, February 21, 2024, <https://www.defense.gov/News/Speeches/Speech/Article/3683202/remarks-by-deputy-secretary-of-defense-kathleen-hicks-keynote-on-the-global-ai/>; Jinghan Zeng, "The US Factor in Chinese Perceptions of Militarized Artificial Intelligence," International Affairs 101, no. 2 (2025): 677–89, <https://doi.org/10.1093/ia/iaae323>.

14 Ministry of Foreign Affairs of the People's Republic of China, "President Xi Jinping Meets with U.S. President Joe Biden in Lima," November 17, 2024, https://www.mfa.gov.cn/eng/xw/zyxw/202411/t20241117_11527672.html; The White House, "Readout of President Joe Biden's Meeting with President Xi Jinping of the People's Republic of China," November 16, 2024, archived at <https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/11/16/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-3/>.

actor's organizational change is interpreted by another, the second actor changes its own procedure, that response feeds back to the first, and the multi-round sequence reduces retained opportunities for correction. At least two rounds are necessary because reciprocal adaptation is a feedback process, not a correlation. The relevant change must be organizational: shorter verification, weaker review, earlier preauthorization, reduced communication, or less political capacity to intervene. A lower sensor latency by itself is not reciprocal compression.

Current evidence leaves H1c as a C proposition. Policy competition, warfighting concepts, and alliance reforms show mutual attention and adaptation. Tactical or operational studies show counter-adaptation. Yet the open record does not contain a complete AI-specific process trace in which both sides repeatedly alter procedures and each loses a matching form of correction time relative to its own baseline. The difference between reciprocal adoption and reciprocal compression is decisive. Two states can adopt more AI while expanding verification; one state can accelerate while the other refuses; both can automate tactical functions while political authorization becomes more deliberate. H1c must remain an invitation to measurement, not a description of an observed general pattern.

4.4 H1d [Canonical P3d]: Time loss and crisis risk

H1d asks whether a verified loss of correction time raises crisis risk. Even that connection is conditional. Reduced T_V, T_C, or T_P should be most dangerous where second-strike confidence is low, doctrine is opaque, warning is noisy, authorization is difficult to reverse, and crisis communication is fragile. The same reduction may matter less where defenses are resilient, forces are survivable, independent checks are strong, leaders can delay or cancel, and adversaries have credible channels for clarification. H1d therefore rejects a universal relation between shorter time and instability. It predicts an interaction among time, vulnerability, uncertainty, and institutional control.

The four propositions also impose a sequence of defeat. If AI attribution does not alter relevant perceptions across settings, H1a narrows. If perception changes do not produce systematic procedural responses, H1b narrows. If responses do not feed back across actors or do not reduce retained time, H1c is withdrawn. If lower retained time does not increase misperception, accidental action, or unintended escalation—or if stabilizing arrangements reliably offset those effects—H1d is reduced. The chain cannot be rescued by combining weak findings across levels. A mechanism that fails at one bridge may redirect the theory, but it cannot be crossed by rhetoric.

Proposition	Identity	Current status	Mandatory stopping point
H1a [Canonical P3a]: AI / automation arrangements change perception	B	Conditionally supported	No claim of a uniform threatening effect
H1b [Canonical P3b]: perception produces a response	B–C	Conditionally supported; direction plural	Acceleration, restraint, verification, refusal, and guardrails are coequal outcomes
H1c [Canonical P3c]: responses produce reciprocal time compression	C	Direct evidence insufficient	Requires bilateral / multilateral, multi-round process evidence
H1d [Canonical P3d]: reciprocal compression changes crisis risk	C	Unverified and conditional	Net effect depends on vulnerability, doctrine, verification, communication, and reversibility

EVIDENCE BOUNDARY — AUTHOR CONTROL: H1a–H1d [Canonical P3a–P3d] may fail separately. Evidence supporting an early bridge cannot be used to complete later bridges. H1c and H1d remain structural hypotheses.

5. The Reciprocal Correction-Time Budget

Strategic analysis treats weapons, survivability, information, and communication as resources. This article adds a reciprocal correction-time budget: time that remains usable for correction after organizations and rivals adapt.

“Reciprocal” does not mean pooled, equal, or centrally allocated time. It means that interaction can expose several actors to related losses of correction opportunities. Gross cycle time obscures this structure because machine processing, verification, communication, and political authority are controlled by different actors and can move in opposite directions.

The common time budget is represented as (T_V, T_C, T_P). T_V, verification-retention time, is the interval and procedural opportunity remaining after an anomaly, warning, or competing interpretation appears for independent cross-checking, counterexample generation, source comparison, dissent, and authorization review. It is not simply the time a model uses to process data. Observable indicators include whether a second source is consulted, whether an objection can be raised, how long a warning remains revisable, how false positives are reviewed, and whether cancellation or rollback remains technically and organizationally possible.

T_C, communication-retention time, is the time and channel capacity still available before an action becomes irreversible for actors to check intent, correct attribution, authenticate messages, communicate limits, and activate crisis mechanisms. It includes more than the existence of a hotline. A channel can formally exist but be politically unusable, technically disrupted, distrusted, or activated too late. Relevant indicators include accessibility, authentication, time from activation to confirmed receipt, opportunity for a response before action, and whether the explanation can alter the operational path. AI may enlarge T_C by detecting anomalies earlier or compress it by making action occur before clarification can arrive.

T_P, political re-entry time, is the time and real authority available for politically responsible decision-makers to redefine the problem, reject a recommendation, delay, downgrade, or reverse a path already shaped by machines and organizations. Political re-entry is not synonymous with a final human signature. If leaders see only a small set of machine-ranked options, receive information after key assumptions have been fixed, or face procedures in which refusal is treated as unacceptable delay, formal authorization can coexist with little substantive control. Indicators include veto points, opportunities for a new briefing, reversibility of predelegation, records of delay or cancellation, and whether leaders can demand a different problem formulation rather than select among predetermined actions.

The dimensions are non-fungible. More T_V cannot compensate for the disappearance of T_P if no responsible authority can stop action. A fast, authenticated hotline cannot redeem the absence of independent verification. Additional political deliberation cannot correct an action that has already become technically irreversible. The vector therefore resists a single strategic-correction score. This is not an objection to measurement. It is a claim about construct validity: the components perform different functions, and aggregation would allow an abundance in one to hide a fatal absence in another. Research should report the vector and its distribution across actors, not a total.

Dimension	Function	Illustrative observable indicators	Defeating observation
T_V — verification-retention time	Independent checks, counterexamples, dissent, authorization review	Second-source checks; challenge nodes; warning revision; false-positive review; rollback	AI adoption consistently increases verification time and lowers error
T_C — communication-retention time	Intent clarification, attribution correction, authenticated crisis channels	Hotline availability; authentication; confirmed receipt; reply before irreversible action	AI consistently creates earlier, trusted communication opportunities
T_P — political re-entry time	Politically responsible authority can redefine, delay, reject, or reverse	Veto nodes; rebriefing; cancellation; reversibility of predelegation; authority to demand alternatives	Political leaders retain ample, usable authority and time throughout

A minimum finding of reciprocal compression requires five observations. First, an organizational process changes for one actor. Second, another actor perceives or infers the change. Third, that actor changes its own process. Fourth, the response feeds back and alters the first actor’s process in a subsequent round. Fifth, at least one matching time dimension declines for each actor relative to its own baseline, and the decline is not restored by new verification, effective communication, or political review. This rule excludes single-state acceleration, one-round

reactions, pure technical latency, and total workflow duration without functional decomposition. It also requires restraint, refusal, oversight, and guardrails to be coded as outcomes that can stop the sequence.

Baselines present a practical challenge. Organizations differ in doctrine, classification, alliance consultation, command structure, and the kinds of crises they face. Cross-case comparison should therefore begin with within-organization or within-scenario baselines: the same institution before and after an adoption, the same exercise under different authority rules, or paired teams given different information about an adversary's AI use. Machine processing, staff analysis, verification, authorization, communication, and political reconsideration should be timestamped separately. Without such decomposition, a claim that AI "saved time" is analytically empty because it does not say whose time was saved, for what purpose, or whether the saving remained available when correction was needed.

Consider a stylized warning process. An AI-enabled system detects a possible attack thirty minutes earlier than a legacy system. One organization uses the gain to consult an independent sensor, convene a dissenting team, and call the adversary; T_V and T_C expand even if the final decision occurs at the same clock time. Another organization uses the gain to move forces automatically and predelegate responses; technical awareness increases while T_P contracts. A rival observing the second process may raise alert levels, which feeds back and further reduces both sides' communication time. The same thirty-minute performance gain can therefore create stability, instability, or no change. The theory concerns the allocation and interaction of the gain, not the gain alone.

The vector is recorded separately for every actor and round, followed by a relational account of how one change altered the others. One state may lose verification time while another loses political re-entry; one ally may retain communication while another is already committed. Common exposure implies neither common authority nor common information, and it creates no supranational allocator of time.

Time is also distributed across organizational nodes. A nominal ten-minute window may contain nine minutes of machine processing and one minute of political confirmation, or one minute of processing and nine minutes of contested interpretation. Parallel processing can expand usable time if independent teams genuinely work from different sources; it can merely create an illusion of redundancy if every team depends on the same model, data pipeline, or assumption. Sequential review can protect against correlated error, but it can also become ceremonial if later reviewers lack authority or information. Measurement must therefore record not only duration but independence, information access, authority, and whether an intervention can still change the outcome.

A practical coding protocol would timestamp six functions separately: detection or generation; interpretation; independent verification; authorization; external communication; and political reconsideration. For each function, researchers should record the controlling actor, information available, uncertainty shown, permitted alternatives, and the last moment at which reversal remained effective. The protocol should also register time returned by AI. If earlier detection creates an extra hour but doctrine automatically advances the next step, the organization did not retain the hour. If it uses the gain for a second sensor, a dissent team, or an authenticated message, retained correction time has increased. The dependent variable is therefore organizational use of saved time, not technical saving itself.

Reciprocal compression can be partial and asymmetric. Both sides need not lose all three dimensions, and equal clock reductions are not required. The minimum test asks whether each side loses at least one matching function through feedback; stronger findings would show repeated decline across several dimensions. Researchers should also code restoration. A new hotline may recover T_C, an independent sensor may recover T_V, and a reversible delegation rule may recover T_P. A sequence that initially compresses time but then produces a durable safeguard is not failed evidence. It is evidence that reciprocal adaptation can generate institutional learning rather than only escalation.

The common time budget is an author construct and remains a C-level research object. It is not a validated scale, and the present article does not estimate it for real national command systems. Its value at this stage is diagnostic: it forces analysts to identify the correction function they believe is changing, the actor who controls it, the baseline against which change is measured, and the institutional mechanism that makes the change consequential. If future research shows that the vector cannot be measured reliably, or that its dimensions do not

predict any relevant crisis outcome, the construct should be revised or abandoned. Precision must be earned rather than simulated.

EVIDENCE BOUNDARY — AUTHOR CONTROL: (T_V, T_C, T_P) is an author construct and a C-level research object, not a validated scale. The dimensions must be reported separately and cannot compensate for one another.

6. Structured Plausibility Probes

6.1 Probe A: Alliance AI and shared decision infrastructure

The alliance probe begins with adoption. NATO's July 2024 revised AI strategy prioritized responsible use, testing, data, interoperability, compute, energy, and dual-use innovation. SHAPE reported in April 2025 that acquisition of the Maven Smart System NATO had been finalized on March 25. In August 2026, NATO's Joint Warfare Centre reported that the system had achieved full technical operational capability on June 22 and connected command-and-control systems across the alliance. These are strategy, procurement, and capability facts—not evidence of operational accuracy, faster political decisions, or crisis-stability effects.¹⁵

The stabilizing logic is straightforward. Shared data standards and decision support can reduce information fragmentation, improve a common operating picture, help allies coordinate defense, and detect anomalies earlier. Where alliance deliberation is slow because members lack the same facts, better fusion can enlarge the effective time available for political choice. AI-assisted translation and scenario comparison can also reduce cognitive burden. In time-budget terms, a system can increase T_V by exposing inconsistencies, increase T_C by giving allies a common factual baseline, and increase T_P by presenting alternatives early enough for governments to consult before military options harden.

The destabilizing logic is different from a simple machine error. Alliances contain governments with unequal capabilities, trust, legal authorities, and risk tolerances. If one member provides the model, cloud, data pipeline, or analytical architecture, other members may depend on judgments they cannot fully inspect while still bearing political responsibility. Shared infrastructure can create correlated failure: the same data contamination, model bias, update, or commercial interruption reaches several members at once. Interoperability can therefore produce common understanding, but it can also produce common blindness. T_V may fall if every member verifies through the same dependency; T_P may fall if technically weaker members are presented with a decision already structured elsewhere.

Alliance consultation makes T_C and T_P especially visible. A national command process can sometimes act through a single political authority; a coalition must negotiate among members, interpret commitments, and manage fears of abandonment or entrapment. AI-enabled tempo can strain this process if fleeting opportunities become the measure of alliance effectiveness. Yet a slower coalition process may also be a firebreak rather than a defect. The empirical question is whether AI reduces redundant delay while preserving political authorization, or whether operational speed converts allied deliberation into a residual formality. Measuring only the time from detection to a common recommendation would miss the difference.

This probe reaches the organizational middle of the theory. It confirms shared-infrastructure adoption and identifies interoperability, trust, unequal capability, and authority as mediators. It does not show how adversaries

15 NATO, "Summary of NATO's Revised Artificial Intelligence (AI) Strategy," July 10, 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>; NATO Joint Warfare Centre, "NATO Maven Smart System Achieves Full Technical Operational Capability," August 12, 2026, reporting achievement on June 22, 2026, <https://www.jwc.nato.int/article/maven-achieves-ftoc/>; Supreme Headquarters Allied Powers Europe, "NATO Acquires AI-Enabled Warfighting System," April 14, 2025, reporting acquisition finalized March 25, 2025, <https://shape.nato.int/news-archive/2025/nato-acquires-ai-enabled-warfighting-system>. These are strategy, procurement, and technical-capability records, not evidence of strategic-stability effects.

interpret the change, whether responses feed back across rounds, or whether T_V, T_C, or T_P ultimately rise or fall.

Alliance adoption also changes the audience of an AI-enabled decision. A national system may be designed for one command culture yet become a shared informational substrate for governments with different legal obligations, threat perceptions, and tolerance for uncertainty. A common picture can reduce disagreement about facts while increasing disagreement about what those facts authorize. If one member treats machine-ranked options as operationally mature and another treats them as provisional intelligence, the alliance can appear technically integrated while remaining politically fragmented. The relevant question is therefore not only whether allies can exchange data, but whether they can identify the point at which a shared analytical product becomes a recommendation, a commitment, or an externally visible signal.

Correlated dependence is the hardest failure mode. Common models, data, updates, and contractors make coordination easier and errors more synchronized. Independent national checks may look inefficient yet protect against common-mode failure. The relevant test is whether interoperability preserves substitute analysis and nationally accountable veto points, not whether every member uses the same interface.

6.2 Probe B: Nuclear and other high-irreversibility decision systems

The second probe concerns nuclear and other high-irreversibility systems. It does not assume AI has launch authority. Its value is that weak correction can matter before final authorization: AI may enter warning, sensor fusion, intelligence, maintenance, cyber defense, conventional counterforce analysis, simulation, or communication support and shape what leaders see and how much time remains.

The positive case is substantial. Better anomaly detection and predictive maintenance can reduce failures. Improved fusion may identify false alarms earlier or strengthen the survivability of retaliatory forces, lowering incentives for preemption. Defensive cyber tools can reveal compromise, and simulation can help leaders compare consequences and identify off-ramps. Herbert Lin's analysis stresses applications that enhance human oversight and evaluates AI within the broader nuclear posture rather than treating it as a standalone technology. These pathways would enlarge the common time budget if machine speed is used to create more evidence, more resilient communication, and more informed political choice.

The risk case also begins before a launch decision. An AI-supported warning system may create false confidence in sensor data or in a model trained on simulated attacks. Better intelligence and conventional precision may be interpreted as a threat to second-strike survivability. Cyber compromise or suspected data manipulation may generate uncertainty about which warning channels remain trustworthy. If organizations respond by predelegating authority, moving to launch-on-warning, or reducing review to preserve a closing window, T_P and T_V can contract even while the final command formally remains human. The strategic issue is not whether a machine presses a button. It is whether the machine-organizational system leaves politics with a real opportunity to define, question, and stop the path.

Open sources require discipline. Saalman documents different public evidence and discussion across Chinese, Indian, and U.S. nuclear postures without reconstructing classified command procedures. UN General Assembly Resolution 80/23 places potential AI integration into nuclear command, control, and communications on a multilateral agenda and calls for human control and oversight. These sources establish recognition and transparency gaps, not transfer of use authority, operational compliance, or a treaty obligation.¹⁶

Behavioral evidence suggests that automation changes nuclear signaling without dictating a single response. Schwartz and Horowitz find that automated nuclear threats can affect credibility, accident perceptions, and willingness to yield, including restraint and null effects. The result matters for H1a and H1b: the automated character of a threat can change how it is received, yet the branch may be accommodation rather than escalation.

16 Lora Saalman, *AI in Chinese, Indian and US Nuclear Postures, Norms and Systems* (Stockholm: SIPRI, 2026), <https://doi.org/10.55163/QZBL7397>; 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 December 1, 2025, A/RES/80/23, <https://docs.un.org/en/A/RES/80/23>.

The study cannot establish reciprocal time compression, but it demonstrates why an adversary's beliefs about control, accident, and responsibility belong inside strategic analysis.

The non-fungibility of the time budget is clearest here. Redundant sensors may preserve T_V while absent crisis channels erase T_C; both may coexist with lost T_P under predelegated authority. Conversely, political control cannot compensate for warning that leaves no time to authenticate an apparent attack. Meaningful human control therefore requires information, time, authority, and organizational support—not a signature alone.

No conclusion about nuclear stability follows from this probe alone. The material identifies plausible mechanisms, public policies, experimental effects, and high-cost failure modes. It does not reveal complete NC3 processes or a net AI effect across crises. The probe supports the claim that political re-entry is an analytically distinct resource and that second-strike confidence, doctrinal opacity, verification, and reversibility should be modeled as scope conditions. It cannot support the claim that reciprocal compression is already occurring or that AI integration will necessarily increase nuclear danger.

Analysis must follow the entire warning-to-decision chain. AI can affect which sensors are prioritized, how tracks are fused, which hypotheses are displayed, how confidence is represented, and which options reach leaders. None is launch authority, yet each can narrow the practical choice set. The question is whether leaders retain information, alternative interpretations, and effective authority to redefine the event rather than approve a machine-shaped menu.

Survivability changes the value of time. Where decision-makers trust that retaliatory forces and command links can survive, additional verification is less likely to appear fatal. Where they fear decapitation, counterforce success, cyber compromise, or loss of communication, delay can be interpreted as forfeiting the ability to act. AI may improve survivability through anomaly detection, maintenance, dispersal, and defensive coordination; it may also undermine confidence by improving tracking or by creating uncertainty about what an adversary can see and disable. The same technical function can therefore expand or compress the time budget through its effect on second-strike confidence, not through speed alone.

Governance signals remain design evidence. National directives, alliance strategies, UN resolutions, and bilateral statements show concern, intent, and proposed restraint; they do not show behavior under stress. Outcome evidence would include exercises in which leaders delay or cancel after an AI warning, independent sources change a recommendation, fail-safe modes work, communication arrives before commitment, and responsibility remains attributable after system or supplier failure.

6.3 Probe C: Perception experiments, wargames, and exit pathways

The third probe combines perception experiments, human wargames, and model simulations. These materials are methodologically heterogeneous, but together they are useful for a narrow purpose: they reveal branches that a linear speed narrative would omit. The probe does not aggregate them into a single evidentiary score. Experiments can identify how a manipulated label or feature changes judgment. Wargames can expose interactive and organizational pathways under controlled scenarios. Model simulations can test how particular systems behave inside artificial worlds. Their results should be compared for mechanism, not treated as interchangeable estimates of real-world probability.

The central experimental result is directional plurality. Horowitz and Lin-Greenberg find both reduced support for preemption under AI-provided analysis and greater retaliatory preference after an adversary's automated accident. Schwartz and Horowitz find credibility, accident, and accommodation effects; Shandler finds reduced strike approval in some conditions and no significant decision-time reduction; Lushenko finds trust shaped by risk and autonomy. These studies support H1a and parts of H1b while rejecting universal acceleration or aggression.

A small CSIS tabletop exercise adds an interactive but limited result. In lower-intensity rounds, whether players knew the adversary had integrated AI did not generate a clear difference; at higher intensity, uncertainty about the opponent's integration was associated with more preventive theater strikes. The non-linearity is more important than the direction. It suggests that the strategic meaning of AI can remain dormant until combined with threat level and uncertainty. The exercise cannot identify a population effect, and it did not measure the time

budget. It does, however, show why crisis intensity and visibility should be explicit moderators rather than background assumptions.¹⁷

RAND's Deterrence in the Age of Thinking Machines offers another pathway. In an exploratory East Asia matrix game, an automated defensive system could trigger retaliatory fire without a new human order, while autonomy also signaled resolve. The game links authorization design and adversary interpretation, but it is a mechanism probe—not an estimate of escalation probability.¹⁸

Schneider, Schechter, and Shaffer's quasi-experimental cyber-nuclear wargame is especially valuable because it yields a counterintuitive result. Uncertainty about cyber vulnerabilities can promote restraint, while confidence in exploit capability can encourage more aggressive counterforce behavior; perceived vulnerabilities can also encourage predelegation or automation. The evidence concerns cyber rather than AI, so it cannot establish an AI effect. It nevertheless identifies a nearby causal structure: certainty and confidence, not uncertainty alone, can be destabilizing, and the same vulnerability can generate restraint at one point and risky authorization at another. This pattern supports the article's insistence on competing pathways and scope conditions.¹⁹

The 2026 Camp(ai)gn wargame likewise challenges linearity. Across ten iterations of a hypothetical 2035 U.S.–China conflict, autonomous systems could enable limited conflict and exit pathways while autonomous mass increased command burdens and possible escalation pressure. These are game-generated hypotheses, not established effects. Automation may enlarge operational choice while consuming human attention and political capacity.²⁰

Model simulations are farther from state behavior but useful for hypothesis generation. Rivera and colleagues observe arms-race and escalation dynamics, with rare nuclear use, in an artificial diplomatic-military environment. Payne's 2026 arXiv preprint reports signaling, counter-escalation, and sensitivity to time pressure and credibility. Models do not bear responsibility, inhabit institutions, or face material loss; prompts, scoring, and model selection shape their worlds. Their results should design human studies, not stand in for state evidence.²¹

The third probe ultimately strengthens the article by limiting it. Across methods, AI-related cues can alter perception; responses can be aggressive, cautious, accommodating, or null; autonomy can create accidental pathways and exit pathways; uncertainty can induce restraint; confidence can induce risk; machine speed need not shorten human decision time. What no material establishes is the full H1c–H1d chain in an actual interstate crisis. The combined evidence therefore justifies a research program centered on reciprocal process tracing and the common time budget. It does not justify a claim that an AI arms race has already compressed strategic correction time.

17 Benjamin Jensen, Yasir Atalan, and Jose M. Macias III, *Algorithmic Stability: How AI Could Shape the Future of Deterrence* (Washington, DC: Center for Strategic and International Studies, 2024), see “Would You Like to Play a Game?” and “What the Game Revealed,” tables 2–5, <https://www.csis.org/analysis/algorithmic-stability-how-ai-could-shape-future-deterrence>.

18 Yuna Huh Wong et al., *Deterrence in the Age of Thinking Machines* (Santa Monica, CA: RAND Corporation, 2020), see the East Asia exploratory matrix-game design and findings, <https://doi.org/10.7249/RR2797>.

19 Jacquelyn Schneider, Benjamin Schechter, and Rachael Shaffer, “Hacking Nuclear Stability: Wargaming Technology, Uncertainty, and Escalation,” *International Organization* 77, no. 3 (2023): 633–67, see “Results” and “Conclusion”; the study reports 115 teams and 580 participants, <https://doi.org/10.1017/S0020818323000115>.

20 Zachary Burdette et al., *Exploring AI's Operational Implications in a Future U.S.-China Conflict: Initial Observations from the Camp(ai)gn Wargame* (Santa Monica, CA: RAND Corporation, 2026), ten iterations involving more than twenty RAND researchers in a hypothetical 2035 Taiwan conflict, <https://doi.org/10.7249/PEA5022-1>.

21 Kenneth Payne, “AI Arms and Influence: Frontier Models Exhibit Sophisticated Reasoning in Simulated Nuclear Crises,” arXiv preprint arXiv:2602.14740, version 1, submitted February 16, 2026, <https://arxiv.org/abs/2602.14740>; Juan-Pablo Rivera et al., “Escalation Risks from Language Models in Military and Diplomatic Decision-Making,” in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 836–898, <https://doi.org/10.1145/3630106.3658942>.

Triangulation is necessary because each method fails differently. Experiments simplify organizations and consequences; high-fidelity interfaces still sit outside real crises; wargames depend on scenarios, adjudication, players, and incentives; simulations lack institutions and material stakes. Convergence should increase confidence only at the mechanism the methods genuinely share, not promote a finding beyond a missing causal bridge.

Null and reverse results are especially informative for the proposed theory. If an AI label produces no time change, the path from perception to procedure is blocked. If automation produces accommodation, independent verification, or a demand for human oversight, the response branch runs opposite to compression. If a wargame creates an exit pathway, the same capability may enlarge political re-entry. These findings should not be absorbed into a generic statement that effects are complex. They should be used to specify moderators: crisis intensity, visibility, task type, confidence, doctrine, organizational trust, survivability, and available communication. A conditional theory becomes stronger by predicting where its risk pathway should disappear.

Probe	Supports	Does not support
Alliance AI	Organizational adoption, interoperability, trust and authority as mediators	A demonstrated crisis-stability effect
Nuclear / high irreversibility	Importance of warning, survivability, human control, and political re-entry	Secret NC3 process claims or AI launch authority
Experiments / wargames / simulations	Perception effects, plural responses, interactive mechanisms, command burdens	Real-world event probabilities or the completed P3c–P3d chain

EVIDENCE BOUNDARY — AUTHOR CONTROL: All three are structured empirical illustrations / plausibility probes. Official adoption is not an effect estimate; experiments do not reproduce interstate crises; wargames and model simulations generate conditional mechanisms, not real-world probabilities.

7. What Would Make the Argument Wrong? Stabilizing Pathways, Alternatives, and Defeat Conditions

The theory should be judged as much by its stabilizing pathways as by its risk pathway. The first stabilizing mechanism is perceptual gain: earlier detection can increase T_V when organizations use the gain for independent checks, competing interpretations, and anomaly investigation. The second is resilience: distributed, degradable, and auditable command systems can reduce the fear that delay means losing the ability to respond. The third is communication: AI can organize information, authenticate or translate messages, and model alternative interpretations, provided that trusted human diplomatic channels remain available. The fourth is control: time saved by machines can be deliberately reassigned to dissent, authorization review, and political re-entry. These are not cosmetic balancing clauses; they are rival causal outcomes.

A second stabilizing pathway is diversity and redundancy. Common models and interfaces can coordinate, but dependence on one model, supplier, data stream, or doctrine can synchronize error. Independent sensors, manual fallbacks, protected dissent, and substitute analysis may appear inefficient in routine operations while preserving T_V and T_P under stress. The test is not maximal redundancy; it is whether an alternative can still challenge the dominant path in time.

The strongest alternative explanation is that AI is not an independent driver. Political objectives, pre-existing rivalry, force posture, digitization, precision strike, cyber vulnerability, and organizational reform may explain apparent acceleration. Talmadge's historical argument and Rand's sociotechnical framework make this alternative unavoidable. The article should attribute an AI increment only where a capability, label, visibility condition, or automation feature changes a process beyond the adjacent trend. If the same outcome occurs without AI, or if common threats explain both actors' reforms, AI should be treated as a secondary mediator or removed from the claim.

A further alternative is that speed stabilizes. Faster sensing may reduce false alarms, reveal attacks earlier, improve defensive coordination, or widen the period for choice. The theory predicts danger only when organizational and reciprocal adaptation reduce retained correction functions. If repeated research shows AI-assisted processes consistently enlarge T_V , T_C , and T_P while reducing error, the stabilizing pathway should become the main conclusion.

A third alternative is that humans resist rather than defer to machines. Algorithm aversion, professional norms, legal review, and high-risk caution can slow adoption or preserve judgment. A fourth is organizational inertia: rivals may observe one another without mirroring procedures because doctrine, alliance rules, command culture, or technical capacity imposes friction. A fifth is governance: testing, incident reporting, crisis communication, arms control, and voluntary constraints may mature quickly enough to block the feedback loop. None of these alternatives is a reason to remove AI from analysis. Each supplies a condition under which H1b, H1c, or H1d should fail.

The defeat conditions are mandatory. H1a narrows if AI attribution does not alter threat, credibility, accident, or responsibility judgments. H1b loses any acceleration tendency if imitation and preauthorization are no more common than verification, refusal, restraint, or guardrails. H1c is withdrawn if research finds no reciprocal multi-round feedback or no net decline in a matching time dimension. H1d is reduced if lower retained time does not predict relevant crisis risks or stabilizing mechanisms reliably offset them. The time-budget construct must also be abandoned if it cannot be measured validly.

Evidence that governance works must be outcome-based. Policies requiring human judgment, testing, legal review, or shutdown authority show design intent. They do not show that a person sees enough, has time, can refuse, or receives organizational support under pressure. A rule counts as stabilizing only when exercises, incidents, or audits show that it preserves correction. The same is true of hotlines: availability, authentication, credibility, and capacity to alter action matter more than formal existence.

The framework has two legitimate successes. It may identify conditions under which reciprocal AI adoption compresses correction time and increases risk. Or it may show that organizations use AI to expand verification, communication, and political judgment, making reciprocal adoption stabilizing. Both answer the research question. The purpose is to follow time and authority through interaction, not preserve a dramatic warning.

Stabilization also depends on whether capabilities are strategically legible. A defensive or resilience-enhancing application can still generate fear if rivals cannot distinguish it from preparation for rapid attack. Conversely, limited transparency about purpose, authority, testing, or geographical scope may reduce uncertainty without disclosing sensitive performance. Strategic explainability is therefore relational: it asks what another actor must understand to avoid the worst interpretation, not whether the internal model can produce a generic explanation. Hotlines, notifications, exercise protocols, declaratory limits, and observable human-control arrangements can all form part of this external legibility, although each creates its own security and verification costs.

A mature counterargument is that organizations will learn. Early adoption may create overconfidence or procedural confusion, but exercises, accidents, red teams, and professional norms may reintroduce safeguards. The theory accepts learning as a possible feedback outcome. Its concern is whether organizations have enough reversible experience to learn before high-consequence deployment becomes routine or locked in. Evidence of repeated correction—changed interfaces, restored review layers, narrowed authorities, independent evaluation, and successful refusal—would weaken the claim that competitive pressure steadily removes time. Evidence that lessons are recorded but overridden by readiness incentives would support it. Organizational learning must be observed in procedures, not inferred from the existence of lessons-learned documents.

The article also admits a stronger defeat: the common time budget may prove less explanatory than established variables. If force survivability, doctrine, political objectives, or crisis communication explain outcomes without reference to retained time, the new construct adds unnecessary complexity. If T_V , T_C , and T_P cannot be observed with acceptable reliability, or if organizations allocate time in ways that do not affect judgment and reversibility, the framework should be abandoned. The proper comparison is not between this theory and no theory, but between this architecture and simpler explanations. Novel terminology earns a place only when it changes what researchers can distinguish and test.

Claim	Observation that would force revision
H1a	AI attribution does not change threat, credibility, accident, responsibility, or action judgments in relevant settings
H1b	Acceleration / imitation is not more prevalent than verification, restraint, refusal, oversight, or guardrails
H1c	No reciprocal, multi-round procedural feedback or no net loss in a matching time dimension
H1d	Lower retained time does not increase relevant crisis risks, or stabilizing mechanisms reliably offset them
Reciprocal correction-time budget	The dimensions cannot be measured reliably or do not distinguish correction functions or predict relevant outcomes

EVIDENCE BOUNDARY — AUTHOR CONTROL: Positive, negative, reverse, and null evidence have equal authority to change the article. Policies count as stabilizing only when organizational outcomes show that correction functions are preserved.

8. Strategic Design Implications

Ask where the saved time goes. The first design implication is measurement: acquisition and exercises should separate machine processing, staff interpretation, verification, authorization, external communication, and political reconsideration. A system earns “decision advantage” only when the gain remains usable for correction—not merely when one stage ends sooner.

Second, meaningful human control must be built around information, time, authority, and support. DoD Directive 3000.09 already requires appropriate levels of human judgment, rigorous verification and validation under realistic and adaptive conditions, understandable interfaces, and governable disengagement from unintended behavior. The additional organizational claim here is that accountable actors must also see critical assumptions, possess usable time, be able to refuse or reformulate, and remain supported when caution conflicts with tempo. A confirmation step alone is not control.²²

Third, crisis communication is part of the technical architecture. Authentication, provenance, translation, anomaly reporting, and a pause for clarification can be designed and exercised. Complete opacity carries a strategic cost because rivals cannot distinguish routine integration, defensive automation, accident, and preauthorized attack. Confidence-building measures should protect T_C without requiring disclosure of every capability.

Fourth, authorization should be reversible wherever the mission permits. Predelegation and autonomous defensive response may be necessary, but scope, duration, geography, target class, and termination conditions should be explicit and testable. Political re-entry requires a real point at which leaders can redefine the problem, not approve action after organizational momentum has erased alternatives.

Fifth, alliances and high-irreversibility systems should preserve independent checks and substitute capacity. Common infrastructure improves interoperability but creates common-mode risk. Members should know which models, data, versions, assumptions, and suppliers they share, where independent verification exists, and whether a second path can produce a genuinely independent challenge before commitment.

Sixth, the normative boundary must remain limited. These implications do not require states to abandon competition, disclose every capability, accept unilateral vulnerability, or create a global sovereign. Shared survival is an explicit floor: reciprocal movement toward irreversible catastrophe cannot be treated as someone else’s

22 U.S. Department of Defense, DoD Directive 3000.09, Autonomy in Weapon Systems, January 25, 2023, sec. 1.2(a)(1)–(3), pp. 3–4; sec. 1.2(f)(1), (3)–(5), p. 6; sec. 2.9(b)(6)–(8), p. 11; and sec. 3, p. 14, <https://media.defense.gov/2023/Jan/25/2003149928/-1/-1/0/DOD-DIRECTIVE-3000.09-AUTONOMY-IN-WEAPON-SYSTEMS.PDF>.

external cost. The practical aim is to make advantage compatible with verification, communication, attributable responsibility, and the political capacity to stop.

These six implications yield three tests. Does an AI-enabled change make it easier to detect and correct a mistaken assumption, or merely easier to act on it? Does it make an adversary's posture more intelligible, or deepen worst-case inference? Does it return time to responsible decision-makers, or remove them earlier from the process? Tactical utility does not answer those questions and should not be mistaken for strategic stability.

Procurement and evaluation should therefore ask a question rarely captured by accuracy or latency metrics: what happens to saved time? Tests can require organizations to demonstrate that earlier detection produces an independent check, a dissent opportunity, a communication window, or a reversible political choice. A system that saves ten minutes but automatically advances the workflow has improved throughput, not necessarily strategic control. Acquisition authorities should record where the time goes under routine and degraded conditions, including model failure, data conflict, communication loss, and pressure from an apparent closing window. This turns correction time from an abstract caution into a testable system requirement.

Interface design is part of authority design. Decision support should expose uncertainty, source conflict, alternative hypotheses, and the consequences of waiting, not merely rank actions. It should permit users to request a fresh formulation rather than select among options generated inside one problem frame. Records should distinguish machine inference, staff judgment, political direction, and supplier intervention so that responsibility does not dissolve after failure. Under severe time pressure, the interface should make the last reversible moment visible. These requirements do not ensure good judgment, but they preserve conditions in which judgment can still occur and later be audited.

Exercises should be designed around reciprocal adaptation rather than unilateral performance. Red teams should be allowed to interpret a deployment in several ways, choose among acceleration, restraint, verification, refusal, and communication, and then force the first team to respond again. The exercise should record T_V, T_C, and T_P for each side after every round. It should also test whether a safeguard remains effective when incentives change: whether a hotline is used when leaders distrust one another, whether a veto survives an apparent attack, whether independent analysis remains independent under shared infrastructure, and whether reversible authority can actually be recalled. A safeguard that exists only in a calm scenario is not yet a crisis safeguard.

International measures can begin below the level of a comprehensive treaty. States can compare vocabulary for human control, notify selected exercises, preserve authenticated crisis channels, exchange methods for testing false alarms, define protected review intervals for specified high-consequence functions, or develop procedures for incident clarification. Such steps will not eliminate competition and may be vulnerable to deception. Their strategic value lies in creating observable opportunities to interrupt reciprocal compression. Any measure should be evaluated against both misuse and non-use: whether it exposes sensitive information or creates false reassurance, and whether the absence of the measure leaves actors with no credible way to correct interpretation before action becomes irreversible.

Commercial and dual-use dependencies must be included in the design boundary. Military organizations increasingly rely on private cloud services, model providers, data pipelines, chips, energy, and software-update chains. A supplier can affect availability or system behavior without holding political authority, while governments may remain responsible for consequences they cannot fully reconstruct. Contracts and certification should therefore specify version control, incident notification, access to audit evidence, degraded-mode operation, and the authority to suspend or substitute a service. These provisions are not a claim that infrastructure providers exercise sovereignty. They are safeguards against allowing a strategically consequential dependency to remain outside the visible responsibility chain.

Finally, strategic evaluation should include a correction-time stress test. The test would introduce conflicting sensor reports, an adversary explanation, a degraded communication channel, a disputed model update, and a politically significant alternative interpretation. Evaluators would ask whether independent verification can still occur, whether the other side can communicate before commitment, whether political leaders can change the frame, and whether actions remain technically reversible. Passing the test would not prove crisis stability, because a real adversary may behave differently. Failing it would reveal that the organization has converted technical speed into a

process that cannot safely absorb uncertainty. That is actionable knowledge before deployment, when correction is still comparatively inexpensive.

EVIDENCE BOUNDARY — AUTHOR CONTROL: Design implications are limited deductions from the proposed mechanism and include explicit F-level normative commitments. They do not establish world sovereignty, a universal governance blueprint, or unilateral disarmament duties.

9. Conclusion: Faster Machines and the Time to Correct

Artificial intelligence will not enter strategic affairs as a single weapon or a single decision-maker. It will enter as a layer across sensing, intelligence, planning, communication, alliance coordination, cyber operations, logistics, autonomous execution, and political briefing. Its strategic consequences will emerge through organizations and through the reactions of others. The question is therefore not whether one system is fast, accurate, or aligned in isolation. It is whether the interaction built around that system preserves the functions that allow error to be discovered and action to be reversed.

This article has proposed three tools: a mediating chain from capability through reciprocal feedback; four hypotheses that can fail separately; and a non-fungible reciprocal correction-time budget for verification, communication, and political re-entry. Existing evidence supports early bridges and several competing pathways. It does not establish that real states have entered reciprocal time compression or that any such compression has a uniform stability effect. Keeping that limitation visible prevents adoption facts, experiments, wargames, simulations, and policy statements from being silently added into a conclusion none can bear alone.

Future research should reconstruct the complete process: compare human and AI-supported workflows, time correction functions separately, observe multiple actors over repeated rounds, retain null and stabilizing outcomes, and vary survivability, doctrine, verification, communication, and reversibility. It should record who received what information, what authority changed, what an adversary inferred, which response followed, whether feedback returned, and which correction functions remained before commitment. If the theory fails, the design should show where; if it holds only narrowly, those conditions are the answer.

Classified and incomplete records will make that task difficult. The difficulty does not justify replacing organizational evidence with model benchmarks; it defines the evidence strategic analysis must seek.

The normative boundary is equally limited. Shared survival does not erase national security, authorize a world sovereign, or require a state to accept unilateral irreversible vulnerability. It requires strategic analysis to display catastrophic externalities that a local advantage can hide and to examine arrangements that make long-term self-protection more compatible with the survival of others. Responsibility must remain attributable; authority must remain intelligible; key assumptions must be challengeable; and termination, correction, and de-escalation must remain possible. These commitments do not prove the causal theory. They explain why the loss of correction time matters when the consequences cannot be repaired afterward.

The final claim is conditional but sharp. When every side uses AI to make itself safer, the shared world may become less safe—not because speed is evil, but because reciprocal adaptation can eliminate the time in which mistakes remain corrigible. Stronger intelligence earns strategic value when it helps actors see more uncertainty, test more interpretations, communicate before attribution hardens, and preserve political responsibility. When error may be irreversible, the most valuable product of faster machines may be the time they return to human beings to verify, reconsider, and stop.

EVIDENCE BOUNDARY — AUTHOR CONTROL: The conclusion remains conditional: the article does not assert that reciprocal time compression has already been observed or that AI has a uniform net effect on strategic stability.

Bibliography

Allison, Graham T., and Philip Zelikow. *Essence of Decision: Explaining the Cuban Missile Crisis*. 2nd ed. New York: Longman, 1999.

- Baldwin, David A. "The Concept of Security." *Review of International Studies* 23, no. 1 (1997): 5–26. <https://doi.org/10.1017/S0260210597000053>.
- Boulanin, Vincent, Lora Saalman, Petr Topychkanov, Fei Su, and Moa Peldán Carlsson. *Artificial Intelligence, Strategic Stability and Nuclear Risk*. Stockholm: SIPRI, 2020.
- Brenner, Jeffrey. "Artificial Intelligence and the Problem of Surprise." *International Security* 50, no. 3 (2026): 127–55.
- Burdette, Zachary, Barry Wilson, David R. Frelinger, Hiwot Demelash, and Julia Arnold. *Exploring AI's Operational Implications in a Future U.S.-China Conflict: Initial Observations from the Camp(ai)gn Wargame*. Santa Monica, CA: RAND Corporation, 2026. <https://doi.org/10.7249/PEA5022-1>.
- Fearon, James D. "Rationalist Explanations for War." *International Organization* 49, no. 3 (1995): 379–414. <https://doi.org/10.1017/S0020818300033324>.
- Garfinkel, Ben, and Allan Dafoe. "How Does the Offense-Defense Balance Scale?" *Journal of Strategic Studies* 42, no. 6 (2019): 736–63. <https://doi.org/10.1080/01402390.2019.1631810>.
- Geist, Edward, Aaron B. Frank, and Lance Menthe. *Understanding the Limits of Artificial Intelligence for Warfighters: Volume 4, Wargames*. Santa Monica, CA: RAND Corporation, 2024. <https://doi.org/10.7249/RRA1722-4>.
- 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.
- Herz, John H. "Idealist Internationalism and the Security Dilemma." *World Politics* 2, no. 2 (1950): 157–80. <https://doi.org/10.2307/2009187>.
- Hicks, Kathleen. "Remarks by Deputy Secretary of Defense Kathleen Hicks Keynote on 'The Global AI Contest' at the Advantage DOD 2024: Defense Data & AI Symposium in Washington, D.C. (As Delivered)." U.S. Department of Defense. February 21, 2024. <https://www.defense.gov/News/Speeches/Speech/Article/3683202/remarks-by-deputy-secretary-of-defense-kathleen-hicks-keynote-on-the-global-ai/>.
- 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>.
- Horowitz, Michael C., and Lauren Kahn. "Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts." *International Studies Quarterly* 68, no. 2 (2024): sqae020. <https://doi.org/10.1093/isq/sqae020>.
- Horowitz, Michael C. "Artificial Intelligence and the Future of Strategic Stability." *Texas National Security Review* 9, no. 2 (2026): 66–83. <https://doi.org/10.1353/tns.00034>.
- 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>.
- Horowitz, Michael C. "When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability." *Journal of Strategic Studies* 42, no. 6 (2019): 764–88. <https://doi.org/10.1080/01402390.2019.1621174>.
- Jensen, Benjamin, Yasir Atalan, and Jose M. Macias III. *Algorithmic Stability: How AI Could Shape the Future of Deterrence*. Washington, DC: Center for Strategic and International Studies, 2024.
- Jervis, Robert. "Cooperation under the Security Dilemma." *World Politics* 30, no. 2 (1978): 167–214. <https://doi.org/10.2307/2009958>.
- Jervis, Robert. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press, 1976.
- 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>.
- Johnson, James. "Revisiting the 'Stability–Instability Paradox' in AI-Enabled Warfare." *Review of International Studies*, First View (2024): 1–19. <https://doi.org/10.1017/S0260210524000767>.
- King, Anthony. "Digital Targeting: Artificial Intelligence, Data, and Military Intelligence." *Journal of Global Security Studies* 9, no. 2 (2024): ogae009. <https://doi.org/10.1093/jogss/ogae009>.
- Lamberth, Megan, and Paul Scharre. "Arms Control for Artificial Intelligence." *Texas National Security Review* 6, no. 2 (2023). <https://tnsr.org/2023/05/arms-control-for-artificial-intelligence/>.
- Lin-Greenberg, Erik. "Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making." *Texas National Security Review* 3, no. 2 (2020): 56–76. <https://doi.org/10.26153/tsw/8866>.
- Lin, Herbert S. "Artificial Intelligence and Nuclear Weapons: A Commonsense Approach to Understanding Costs and Benefits." *Texas National Security Review* 8, no. 3 (2025): 98–109. <https://doi.org/10.26153/tsw/60739>.
- 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>.
- Ministry of Foreign Affairs of the People's Republic of China. "President Xi Jinping Meets with U.S. President Joe Biden in Lima." November 17, 2024. https://www.mfa.gov.cn/eng/xw/zyxw/202411/t20241117_11527672.html.
- NATO Joint Warfare Centre. "NATO Maven Smart System Achieves Full Technical Operational Capability." August 12, 2026. <https://www.jwc.nato.int/article/maven-achieves-ftoc/>.
- NATO. "Summary of NATO's Revised Artificial Intelligence (AI) Strategy." July 10, 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>.
- Payne, Kenneth. "AI Arms and Influence: Frontier Models Exhibit Sophisticated Reasoning in Simulated Nuclear Crises." arXiv preprint arXiv:2602.14740, version 1, submitted February 16, 2026. <https://arxiv.org/abs/2602.14740>.
- Perrow, Charles. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books, 1984.
- Posen, Barry R. *Inadvertent Escalation: Conventional War and Nuclear Risks*. Ithaca, NY: Cornell University Press, 1991.

- Rand, Lindsay. “Products of Patrons and Producers: A Sociotechnical Approach to Technological Innovation and Nuclear Deterrence.” *Journal of Global Security Studies* 11, no. 3 (2026): ogag018. <https://doi.org/10.1093/jogss/ogag018>.
- Rivera, Juan-Pablo, Gabriel Mukobi, Anka Reuel, Max Lamparth, Chandler Smith, and Jacquelyn Schneider. “Escalation Risks from Language Models in Military and Diplomatic Decision-Making.” In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 836–898. <https://doi.org/10.1145/3630106.3658942>.
- Saalman, Lora. *AI in Chinese, Indian and US Nuclear Postures, Norms and Systems*. Stockholm: SIPRI, 2026. <https://doi.org/10.55163/QZBL7397>.
- Sagan, Scott D. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton University Press, 1993.
- Schelling, Thomas C. *Arms and Influence*. New Haven, CT: Yale University Press, 1966.
- Schneider, Jacquelyn, Benjamin Schechter, and Rachael Shaffer. “Hacking Nuclear Stability: Wargaming Technology, Uncertainty, and Escalation.” *International Organization* 77, no. 3 (2023): 633–67. <https://doi.org/10.1017/S0020818323000115>.
- 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>.
- 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): 727–35. <https://doi.org/10.1080/01402390.2019.1626725>.
- 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>.
- Snyder, Glenn H. “The Security Dilemma in Alliance Politics.” *World Politics* 36, no. 4 (1984): 461–95. <https://doi.org/10.2307/2010183>.
- Supreme Headquarters Allied Powers Europe. “NATO Acquires AI-Enabled Warfighting System.” April 14, 2025. <https://shape.nato.int/news-archive/2025/nato-acquires-ai-enabled-warfighting-system>.
- 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): 864–87. <https://doi.org/10.1080/01402390.2019.1631811>.
- U.S. Department of Defense. DoD Directive 3000.09, *Autonomy in Weapon Systems*. January 25, 2023. <https://media.defense.gov/2023/Jan/25/2003149928/-1/-1/0/DOD-DIRECTIVE-3000.09-AUTONOMY-IN-WEAPON-SYSTEMS.PDF>.
- 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 December 1, 2025. A/RES/80/23. <https://docs.un.org/en/A/RES/80/23>.
- The White House. “Readout of President Joe Biden’s Meeting with President Xi Jinping of the People’s Republic of China.” November 16, 2024. Archived at <https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/11/16/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-3/>.
- Wong, Yuna Huh, John M. Yurchak, Robert W. Button, Aaron Frank, Burgess Laird, Osonde A. Osoba, Randall Steeb, Benjamin N. Harris, and Soo Jeong Bae. *Deterrence in the Age of Thinking Machines*. Santa Monica, CA: RAND Corporation, 2020. <https://doi.org/10.7249/RR2797>.
- 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 August 15, 2026. <https://doi.org/10.1080/14702436.2026.2718345>.
- Zatsepina, 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>.
- Zeng, Jinghan. “The US Factor in Chinese Perceptions of Militarized Artificial Intelligence.” *International Affairs* 101, no. 2 (2025): 677–89. <https://doi.org/10.1093/ia/iaae323>.