Military AI Control: Verification and Enforcement Limits

The Economy Research Editorial*

*The Economy Research, 71 Lower Baggot Street, Dublin 2, Co. Dublin, D02 P593, Ireland

Abstract

The article examines whether Washington and Beijing's statements about human control of military AI, ahead of the Trump-Xi summit in September 2026, can amount to real regulation. Drawing on the Brookings Institution's analysis, Jack Shanahan's positions on bilateral stability and official data on military spending and nuclear arsenals, the paper argues that military AI is fundamentally different from nuclear weapons: it is software, invisible and replicable, with no verification regime capable of controlling it. The two superpowers currently have no external enforcement mechanism comparable to the one that has held back nuclear competition for seven decades. The article concludes that narrow crisis management measures, such as hotlines and accident notification protocols, remain feasible, while meaningful regulation would require coordination of middle forces capable of imposing real costs on anyone who breaks the rules.

1 Introduction - Human Control and the Governance of Military AI

On September 24, 2026, Chinese President Xi Jinping is expected at the White House for the second bilateral summit with U.S. President Donald Trump in four months.[1] Trade, Taiwan and the war in Iran will dominate the agenda, but artificial intelligence is rapidly rising on the list of strategic priorities. According to the South China Morning Post, Washington is attempting to integrate the expected AI dialogue into the broader economic talks managed by Treasury Secretary Scott Bessent, with the Chinese side interpreting the U.S. proposals as an attempt to limit both the risks of AI and Chinese growth dynamics.[2]

The timing is no coincidence. In June 2026, Trump signed the NSPM-11 Presidential National Security Memorandum, which makes clear Washington's intention to drastically accelerate the integration of AI into every military and intelligence operation, while committing that commanders will remain accountable within the chain of command.[3] In May 2026, the War Department announced agreements with eight major AI companies to deploy on its classified networks. In other words, American policy requires both controlled systems and rapid military dominance. These are two goals that, over time, tend to be mutually exclusive.

The last week before the summit added another dimension. On September 12, 2026, Dario Amodei, CEO of Anthropic, published an extensive essay calling for a slowdown in the pace of development of the most advanced AI models, outlining a three-stage plan that includes independent evaluators, coordination between companies and international collaboration.[4] OpenAI's Sam Altman and xAI's Elon Musk publicly agreed. Amodei himself acknowledged that China poses the most difficult dilemma: if U.S. companies unilaterally slow down, they risk losing their technological edge. This contradiction reflects on a commercial scale exactly the same dilemma that states face in the military sphere.

The Brookings Institution published a few days before the summit a two-part analysis, with an American and Chinese perspective, that proposes extending the principle of human control of nuclear weapons to cyberattacks against nuclear command systems and critical infrastructure.[5] The Brookings proposal, in both Melanie W. Sisson's American version and Tianjiao Jiang's Chinese version, starts from an indisputable starting point: the November 2024 Biden-Xi agreement that decisions to use nuclear weapons should remain in the hands of people. What, however, the Brookings analysis avoids fully addressing is that a statement of principle does not explain which person approves what action, with how much time, how much information and how compliance could ever be verified within classified military networks.

In May 2026, the heads of state of the two sides had already proposed building a "constructive relationship of strategic stability", with Jiang explicitly arguing that the governance of military AI should be an integral part of this effort, not a collateral issue.[6] The same text hints at something that is rarely so clearly articulated in an official context: political statements about human control still do not translate into institutional arrangements, even when the common intention has been publicly and repeatedly stated by the leadership of both countries.

This article argues that claims of human control of military AI are not regulation. They cannot be verified, cannot be enforced and cannot resist the pressure of competition. AI as a military capability is fundamentally different from nuclear weapons: it is software, it is invisible, it is replicable and it integrates into systems

that already exist. However, this inability to regulate does not mean that any form of bilateral agreement is meaningless. Some narrow crisis management measures, designed as mutual mechanisms to prevent escalation rather than ethical commitments, can reduce the risk of an accident. But for any control architecture to work, it may need something that is more akin to the logic of mutual nuclear deterrence: a credible third force, or alliance of forces, capable of punishing the one who breaks the rules.

2 Military AI in Operational Decision-Making: From Targeting to Command

The integration of AI into military forces is no longer experimental; it is operational. In the United States, the XVIII Airborne Corps already uses the Maven Smart System for wide-area image processing, threat identification and directed targeting at a scale that would be impossible for human teams of analysts.[7] The 1st Multi-Domain Task Force uses TITAN, an AI-powered and machine learning ground station that processes data from satellites, aircraft and ground sensors simultaneously. The Army's Materiel Command leverages artificial intelligence for sustainment decisions through Weapon System 360, which gives commanders real-time insight into inventories, supply chain bottlenecks and predicted unit readiness.

That speed isn't just an improvement; it's changing the relationship between humans and machines in the decision-making process. Major Dorothy M. Reid, in a recent analysis for West Point's Modern War Institute, poses a question that proves to be central: what should the soldier still know? Reid argues that AI doesn't just reduce the knowledge needed, but shifts the point at which human judgment is necessary.[8] An officer can now enter a planning meeting with a highly elaborated course of action, timing matrix, risk assessment and presentation, without having the knowledge needed to assess whether these products correspond to reality.

In the realm of command, Major Prabhat Mishra of the Indian Army, writing in Military Review, introduces a concept that captures danger accurately: "command without control," i.e., the situation in which power remains formally intact while the human judgment that gives it meaning is gradually emptied of content, replaced by systems optimized for speed.[9] This practice is driven by the structure of the system itself, not by negligence. In the U.S.-Israeli campaign against Iran, it is reported that AI systems generated over a thousand strike targets within the first twenty-four hours, at a speed structurally incompatible with systematic legal scrutiny of every decision.[10] The integration of Anthropic's Claude model into Palantir's Maven targeting system created a public debate about the ethical limits of this use.[11]

What the Iran experience has revealed is not just the speed of a new mode of warfare. It is the collapse of clarity in who decides what. In a system where AI filters information, prioritizes targets, creates courses of action and presents "optimal" options, the role of humans gradually shrinks from that of the decision-maker to that of the validator. War game experiments at the U.S. Army War College documented just that: officers remained attached to AI-generated courses of action even after they turned out to be operationally flawed, a phenomenon researchers call cognitive anchoring.[12]

This development is not unique to the United States. China is aggressively adopting AI across its en-

tire military and information ecosystem, aiming to achieve decisive advantages in speed, accuracy, range and decision-making quality.[13] At the same time, even close U.S. allies are lagging significantly behind in the integration of military AI. According to senior Pentagon official Doug Winnie, the allies lack the computing infrastructure, the skilled technicians, and the procurement mechanisms required.[14] South Korea is a partial exception, investing in autonomous robotic systems and "natural artificial intelligence" for military applications, but even there the scale remains asymmetric relative to American or Chinese programs.[15]

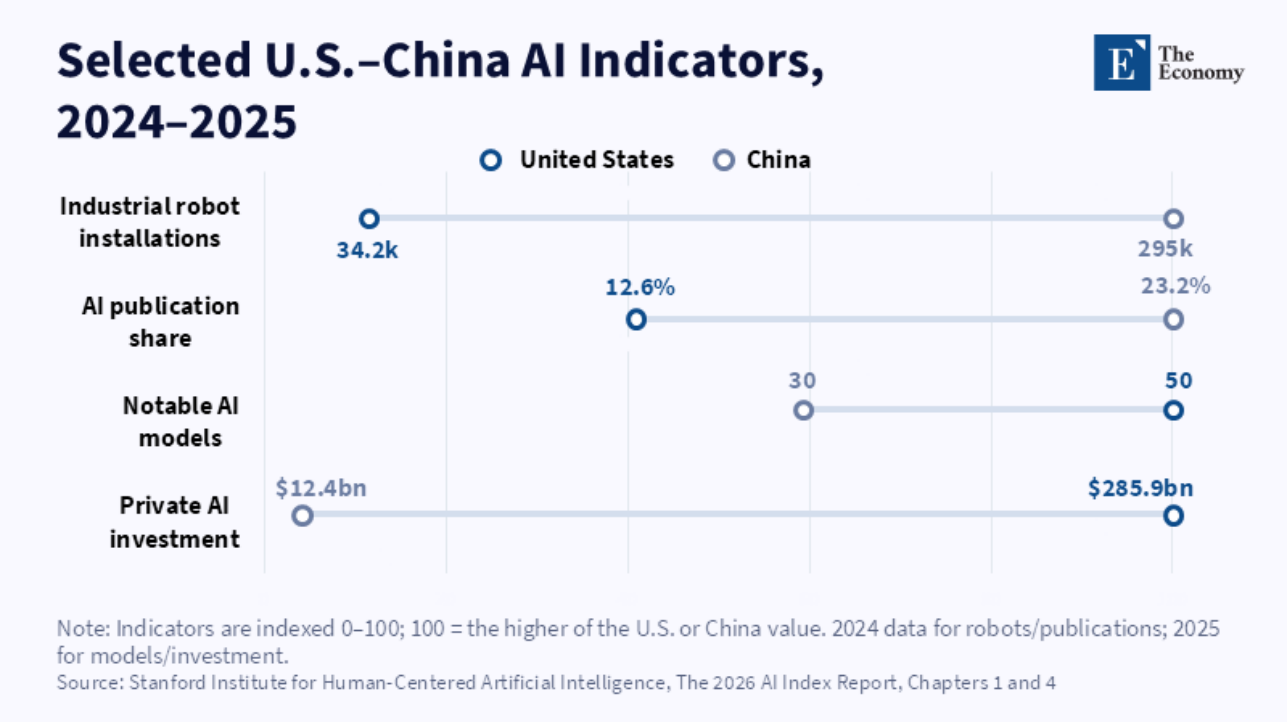

Figure 1: Washington's edge is capital and model count; Beijing's is publication share and industrial deployment.

This asymmetry of capabilities has a consequence that is rarely mentioned in discussions of testing military AI: as long as the ability to integrate such systems remains the prerogative of a few states, the more difficult it becomes to build a broad, multilateral control regime with real participation. Allies that do not have their own systems have neither the technical expertise to contribute to common rules of verification, nor negotiating leverage against the two superpowers. The result is a regime where regulation, if it does exist, will be shaped almost exclusively by the two states that have the greatest ability to circumvent it.

The ICRC points out that AI is already being used in four areas of military operations: intelligence and surveillance, cyber operations, autonomous weapon systems and command decisions.[16] In each of these areas, autonomy is increasing not only because of technological capabilities but because of competitive pressure: each side believes that being late is tantamount to vulnerability.

This dynamic doesn't mean that every use of AI on the battlefield is reckless. Automating intelligence gathering, predicting equipment failures, optimizing supply chains and initial target detection can reduce risks and increase efficiency. But, at each of these stages, the very speed and volume that make AI useful shrink the space left for meaningful human intervention. Reid notes that competency assessments should now reveal the thinking behind outcomes, not just outcomes as an indication of capability.

The ICRC distinguishes between these four areas clearly, pointing out that the integration of AI into each

produces a different kind of risk, each risk distinct rather than a single one that multiplies. In intelligence and surveillance, risk is the overestimation of the accuracy of a system that classifies vast volumes of data without always explaining the logic of the classification to the analyst who receives it. In cyber operations, risk is the speed of execution that exceeds any human reaction time. In autonomous weapon systems, risk is about distinguishing between target and non-target in field conditions that are never as clear as training data. In command decisions, finally, the risk is more subtle and more difficult to identify in advance: the gradual shift of responsibility from the human who decides to the system that merely recommends.

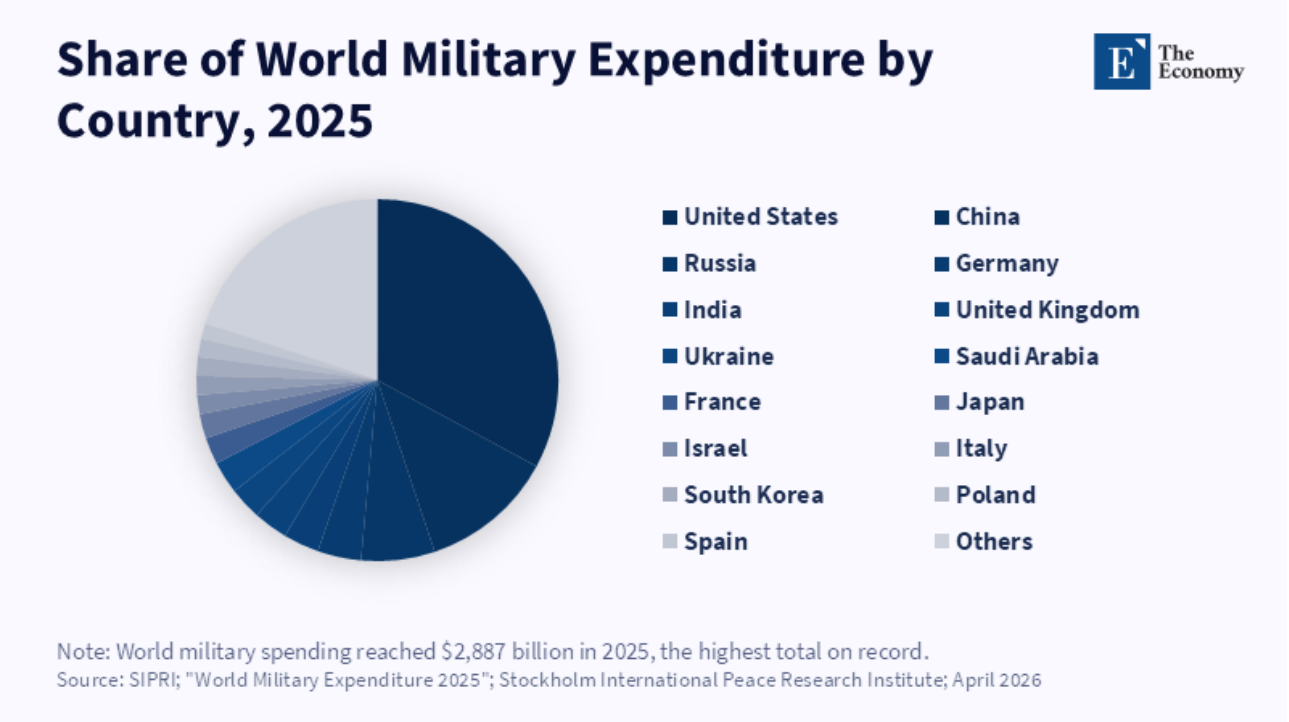

Figure 2: Two countries account for nearly half of everything the world spends on defense.

This fourth category of risk has a practical consequence that is rarely discussed publicly. When an officer learns to trust an AI recommendation without systematically verifying it, his own ability to process the same problem without the tool is no longer practiced with the same frequency. Reid describes this as a shift in the line of proficiency: the question has shifted from what a soldier should know in general to what he should know the moment the support system fails, delays, or misleads. Military academies that design programs around AI tools thus face a dilemma with no easy solution: the more time they spend training officers to work effectively with the systems, the less time they have available to cultivate the independent judgment that will be needed at the exact moment when the systems will not be available, up-to-date, or reliable.

The same issue appears in the U.S.-Israeli campaign against Iran, where the concept of a ”target bank” has essentially changed its content. Traditionally, a target bank is compiled by analysts who evaluate each possible target individually, based on legal, operational and political criteria. When this process is accelerated by artificial intelligence to produce a thousand targets within twenty-four hours, the personalized assessment is necessarily replaced by a sample check, even when no official rule has formally changed. The controversy surrounding the use of the Claude model in Palantir’s Maven system reflects precisely this concern: the question was whether its integration into a targeting system left enough room for the substantial human judgment that

the company itself claims to preserve.

3 U.S.-China Proposals for Human Control and Crisis Management

The proposal architecture for regulating military AI is divided, realistically, into three different levels: rule-making, operational safeguards and verifiable arms control. The Brookings analysis shows that many proposals treat them as a single problem.

At the rule-making level, the situation is progressing. The principle that people should control nuclear use decisions, agreed between Biden and Xi in November 2024, is now a commonly accepted basis.[17] Brookings' Sisson explicitly proposes its expansion: neither the U.S. nor China should allow artificial intelligence to autonomously launch cyberattacks against nuclear command, control and communications systems, or against critical infrastructure. The logic is clear: a standalone cyberattack on such systems could be interpreted as preparation for war, forcing leaders into nuclear decisions under extreme uncertainty and time pressure.[18]

The Chinese contribution from Jiang goes deeper at the institutional level. He proposes a list of red lines for military AI, a common definition of "substantial human control," a dedicated US-China military hotline for AI incidents and dialogue on autonomous lethal weapon systems.[19] The reference to the Chinese balloon incident in 2023, which for days dominated relations between the two countries, serves as a reminder: unmanned systems equipped with artificial intelligence and strike capability could trigger similar crises with much shorter reaction time.

The very concept of "substantial human control" does not start from a common starting point on both sides, something that the Brookings analysis openly acknowledges. Jiang traces the root of American thought to the OODA loop formulated by U.S. Air Force Colonel John Boyd in the 1970s, according to which in a conflict prevails whoever observes, orients, decides and acts faster than the adversary.[20] In Chinese military thought, by contrast, Jiang locates an older and different reference: Premier Zhou Enlai's mandate for "absolute certainty" in China's nuclear tests, a principle that, in the estimation of Chinese experts and practitioners, still guides the Chinese conception of human control today. The two traditions do not clash directly, but emphasize different things: American on the speed of the decision cycle, Chinese on certainty before action. A common interpretation of "human control" should reconcile these two traditions, which Jiang explicitly acknowledges as a project for ongoing dialogue or a special terminology working group, not a one-time joint statement.

The proposal for a special telephone line would not create something from scratch. It would build on already existing communication channels, such as the line between defense ministers and the Military Maritime Consultative Agreement, simply adding a dedicated channel for incidents caused by an AI bug or cyberattack.[21] Jiang describes a specific scenario that explains why such a channel would be necessary: an AI defense system detects unusual data traffic to nuclear command systems and it automatically disconnects the network or triggers countermeasures, while the opposing side interprets the same action as an intrusion, causing the systems of both sides to enter a state of automatic confrontation before any human can intervene.[22] Without a predetermined reporting channel for such incidents, the only information that would reach leaders would be the fact of the outage rather than its cause, which in itself is enough to turn a technical error into a political crisis.

At the multilateral level, 128 countries adopted in September 2026 a text defining lethal autonomous weapon systems under the Convention on Certain Conventional Weapons, a first step towards possible treaty negotiations.[23] However, campaigns on autonomous arms control face structural constraints: the Group of Governmental Experts operates by unanimity, explicitly stating that nothing is considered agreed until everything is agreed.[24] At the same time, the texts leave considerable room for interpretation, allowing for example ”human control” to mean something other than continuous direct control.

The slowness of this process is not a recent phenomenon. The first informal expert meetings on autonomous weapon systems took place in 2014, 2015 and 2016, before the Fifth Convention Review Conference decided, at the end of 2016, to transform them into a formal open-participation Group of Government Experts.[25] The first formal meeting of the Group took place in November 2017, i.e. almost a decade before the 2026 definition text mentioned above and even that first summit had previously been postponed and cancelled due to the inability of some member states to pay their financial obligations to the Convention.[26] A decade of negotiations that resulted in a common definition, with no binding prohibition or restriction rules as yet, is in itself an indication of how difficult it is to turn a principle into a workable text when the consent of all participating states is a prerequisite for every step.

Jiang adds an argument that goes beyond cyberattacks and touches the battlefield directly. According to him, the widespread use of autonomous weapons systems on the Russia-Ukraine front and in the Middle East has already caused serious civilian casualties and has increased, in his estimation, the risk of nuclear escalation, an assessment that reflects the view of one of the two authors of the text and is not accompanied by independently verifiable evidence in the article itself.[27] The remark remains, however, important precisely because it comes from a Chinese academic body participating in the official dialogue; the concern about autonomous weapons is not limited, that is, to Western research centers. Jiang closes his text with a formulation that encapsulates the stakes: as machines gradually acquire ”the right to shoot” and even ”the right to engage in combat,” the two superpowers bear, in his view, a historical responsibility to prevent technology from completely escaping human control.[28]

Evaluating these initiatives as a whole, two categories of measures can be distinguished. The first includes statements of principle: human control, responsible use, accountability. These are politically significant because they create regulatory expectations, but do not explain how they will be implemented in real conditions. The second category includes specific operational measures: crisis hotlines, incident notification protocols, separation of autonomous systems from nuclear command networks. These are realistic, since they do not require any state to disclose what its classified systems are doing, but only to establish a communication channel when something goes wrong.

However, between these two categories there is a gap: verification. Jiang explicitly acknowledges three problems that make sentences potentially ineffective.[29] First, speed: AI-powered cyberattacks evolve in milliseconds, making the ”final human decision” structurally impossible. Second, attribution of responsibility: in a cyberattack incident, identifying the perpetrator can prove impossible, leaving open whether the attack was state-owned, criminal, or the result of a malfunctioning autonomous system. Third, the security dilemma: the opacity of AI capabilities leads each side to assume the worst for the other, fueling spirals of escalation.

The practical significance of these challenges is clearly seen in the U.S. NSPM-11. The Memorandum requires both the integration of human accountability into the chain of command and the rapid adoption of the most advanced AI models by many providers.[30] This text explicitly expresses both objectives, without acknowledging that the gradual increase in speed and autonomy undermines the very accountability structures it promises to maintain. The requirement to update Directive 3000.09 on the autonomy of weapon systems within 90 days, with an annual review, shows that even the U.S. government itself recognizes how quickly the field is changing.

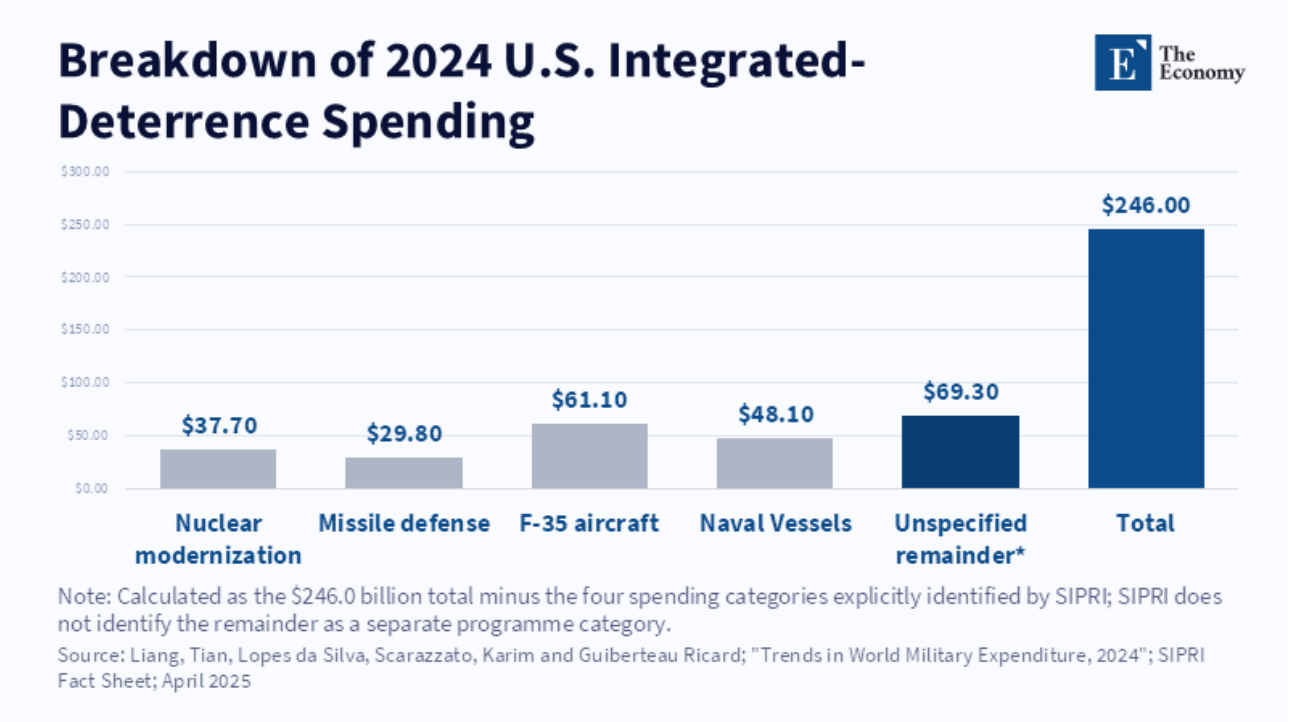

Figure 3: Less than a sixth of this budget funds nuclear modernization itself; the rest buys aircraft, ships, and missile defense.

4 Military AI Governance and the Limits of Nuclear Arms Control

The comparison with nuclear arms control is plausible, but it turns out to be misleading. Nuclear deterrence worked because it met three conditions that military artificial intelligence cannot meet: nuclear weapons were physical objects that could, to some extent, be measured and monitored. Their capabilities were slowly changing. And mutually assured destruction created a cost so certain and so catastrophic that restraint became rational without an external enforcer.

Military AI reverses each of these assumptions. Lt. Gen. Jack Shanahan, the U.S. Department of Defense's first director of artificial intelligence and founder of Project Maven, places the problem particularly clearly: AI is deconstructing the foundations of strategic stability.[31] Capabilities are difficult to track and impossible to verify. No state can see the adversary's training data, model architectures, integration pathways, or adoption range. Even access to computing power is only seen indirectly, through energy consumption, supply chains and fragmented reporting. There is no verification regime capable of generating shared trust about AI's capabilities or intentions.[32]

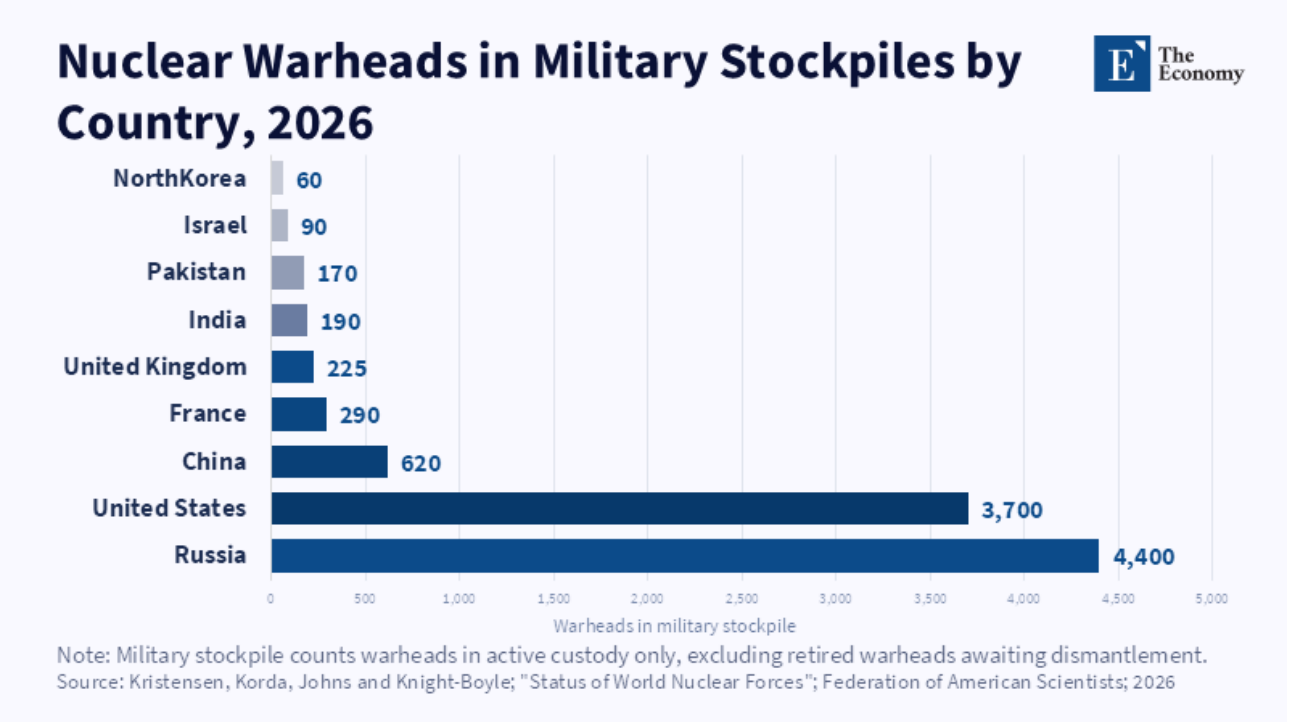

Figure 4: Russia and the United States alone still hold more than four out of every five warheads on Earth.

Shanahan breaks down this opacity into eight distinct categories of uncertainty that, in his opinion, make it difficult to make any reliable assessment of the adversary's intentions and capabilities: uncertainty about the deeper motivations for AI development, about the actual timelines of research, about the moment in which the political and military leadership will fully trust a system; for the real operational advantage it offers, for the rate of its diffusion in the armed forces, for its performance in real conflict, for its possible integration into nuclear command systems and finally for the very reliability of the above estimates, a second-degree uncertainty that rests on top of all the previous ones.[33] This last category is, in a sense, the most dangerous: even when individual analyses point in the right direction, uncertainty about their reliability can in itself fuel an overreaction.

Shanahan uses a historical example to show how the perception of equivalence can prove to be just as dangerous as a real inequality of power. In the late 1950s, Washington became convinced that it was lagging behind Moscow on intercontinental missiles, a "missile gap" that later turned out to be non-existent, but which in the meantime had already triggered a series of armaments decisions based on miscalculation.[34] According to Shanahan, something similar is happening today with artificial intelligence: when two sides are at roughly the same level of capabilities, but the rate of change is exponential, each suspects that the other is hiding some leap that has not yet been revealed. Symmetry, instead of reassuring as in classical deterrence theory, produces equivalent anxiety on both sides. He proposes abandoning the concept of "strategic stability", borrowed from the nuclear age, in favor of a broader "bilateral stability": a framework for managing perceptions, capabilities and risk reduction mechanisms that does not presuppose symmetry or fixed possibilities, but accepts uncertainty as a permanent feature of the new era and not as a problem waiting for a definitive solution.[35]

Second, AI is moving at software speed. Modernization cycles are being squeezed by exponential chip improvements, algorithmic leaps, rapid model retraining, emerging properties and seamless diffusion from commer-

cial to military applications. Shanahan notes that states feel compelled to move first, continuously experiment and deploy early and often. The fear of lagging is transformed from a psychological tendency to a structural behavioral motivation.[36]

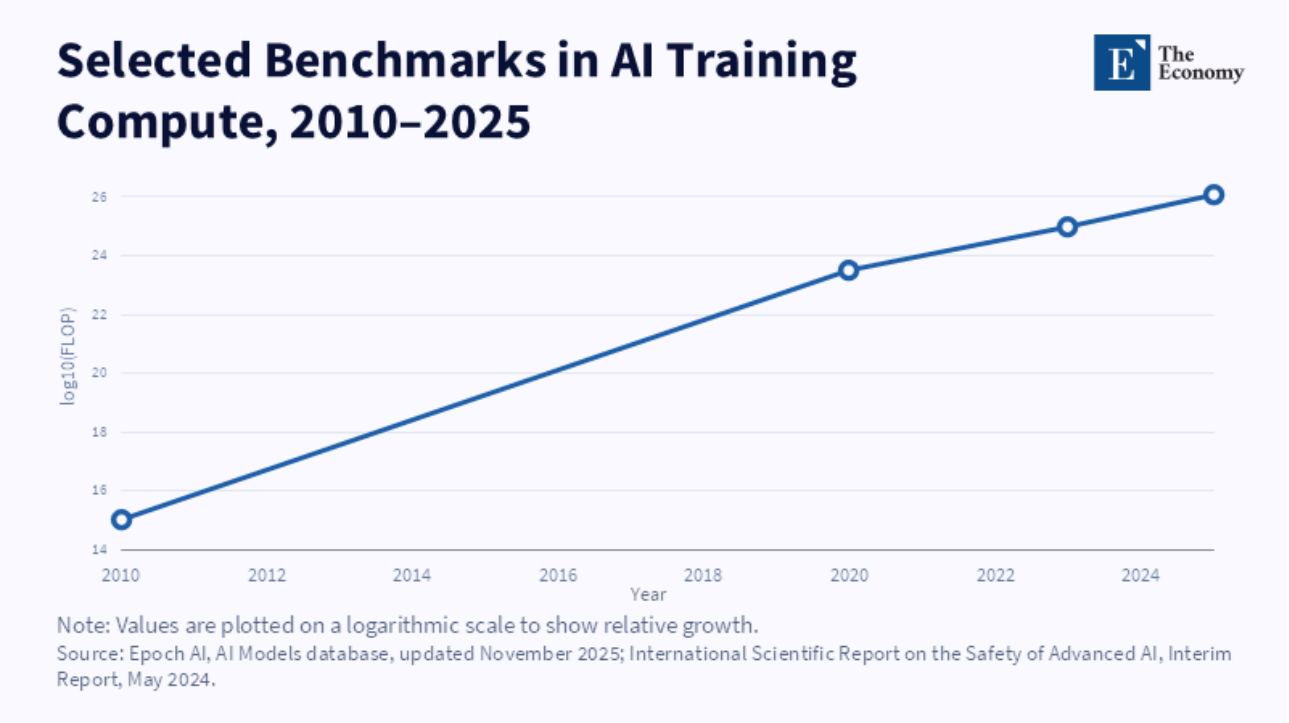

Figure 5: The computing power behind a single training run has grown roughly ten-billionfold in under fifteen years.

Third, artificial intelligence complicates the logic of mutual vulnerability that has worked so effectively in the nuclear sector. Possible improvements in submarine detection or enhanced cyber operations against early warning systems can, even as a hint, destabilize, as each side worries about ”vulnerability windows” and increases the risk of preventive action.[37] This reverse logic, according to which technological progress intended to give more time to decision-makers ends up squeezing it instead, is perhaps the deepest irony of the time.

Project Syndicate's Ian Bremmer characterizes AI competition as the biggest threat to the relative downturn in U.S.-China relations.[38] Bremmer notes that without ongoing dialogue to avoid escalation, a return to the zero-sum conflict that rocked both sides in 2025 cannot be ruled out. The Johns Hopkins SAIS analysis argues that categorizing this dynamic as an ”arms race” is itself destabilizing, because it suggests that capabilities are measurable, progress observable and end state definable, none of which applies to military AI.[39]

This opacity is at the core of the problem. In the nuclear field, warhead counting, missile observation, satellite silo surveillance and geographic test surveillance have provided an imperfect but existential basis for mutual verification. Shanahan emphasizes that with artificial intelligence there is no equivalent possibility: the most critical components, models, training data and integration architectures, are private, classified, or impervious even to the government's own analysts.[40] State deception becomes easier, cheaper and more tempting.

Academic research published in Risk Analysis by the CNS of the Middlebury Institute confirms this dynamic. The researchers studied the relationship between technologies that facilitate nuclear proliferation and technologies that enhance detection, showing that artificial intelligence asymmetrically accelerates the former relative

to the latter, widening the zone of uncertainty around the risk of nuclear proliferation.[41] This development does not only concern nuclear: it generally shows that artificial intelligence tends to empower the secret player more than the overseer.

The BBC recently referred to the U.S.-China competition as a dual race: China wins in large-scale AI implementation, the U.S. in cutting-edge models.[42] This is not a consolation but a further destabilization: no state can be sure it holds a lead in the overall picture, which reinforces the fear of a sudden technological leap by the adversary.

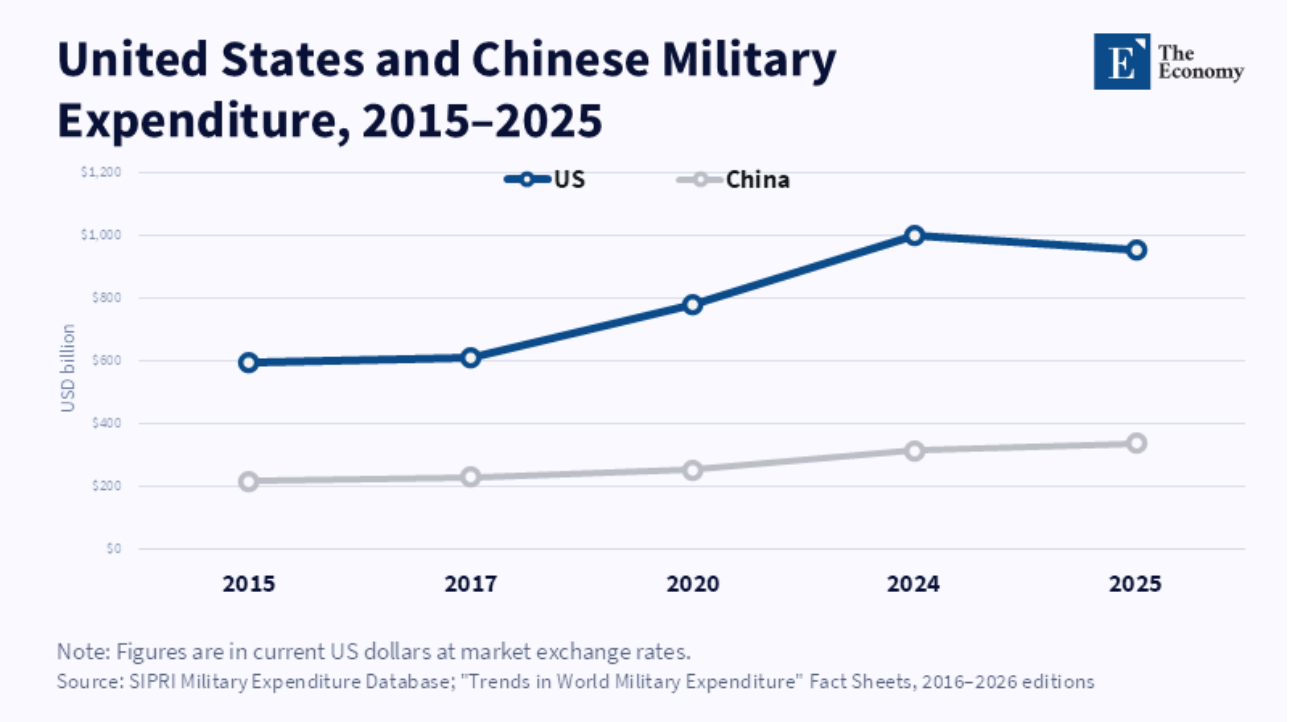

Figure 6: China's defense budget has grown faster than America's for a decade, even while remaining a third of America's size.

The argument of this article is criticized from two opposing directions, which Shanahan himself candidly captures in his text. The first category of critics considers that the military importance of AI is overestimated in the short term, that the geopolitical consequences will depend more on long-term diffusion and institutional adjustment than on some definite technological termination and that describing the situation as destabilizing competition artificially inflates its short-term importance.[43] The second category argues the exact opposite: that a generic or superhuman AI will provide such a decisive and lasting advantage to the state that reaches it first that speed and scale take absolute precedence, making any form of restrictive dialogue with China strategically unnecessary or even harmful.[44]

Neither thesis is unfounded. The first is right that AI has yet to prove in real large-scale conflict the decisive advantage attributed to it in strategy texts. The second is right that the history of technological competitiveness rarely rewards waiting. Yet both extremes share the same error: they assume that there is a clear endpoint, either of indifference or decisive victory, which military AI does not seem to offer in either version. Complacency leaves unanswered exactly the verification and speed questions analyzed above. The race at all costs ignores that, without an enforcement mechanism, the simultaneous acceleration of both sides does not give either of

them a permanent advantage, only a higher overall risk for both.

The analogy with nuclear weapons is not entirely wrong, however. It works on one point: nuclear arms control was made possible not thanks to an external regulator, but because both superpowers concluded that even limited mutual restraint served their own survival interests. Between two opponents A and B, if A believes that he can eliminate B's counterattack, then restraint loses its logic: A can gain an advantage by violating every pact. But if there is a sufficient third party, or an international alliance, capable of implementing punitive measures against the offender, then the incentive to comply increases dramatically. This was precisely the logic behind the Security Council, the multilateral non-proliferation treaties and the security guarantees that underpinned the nuclear world for seven decades.

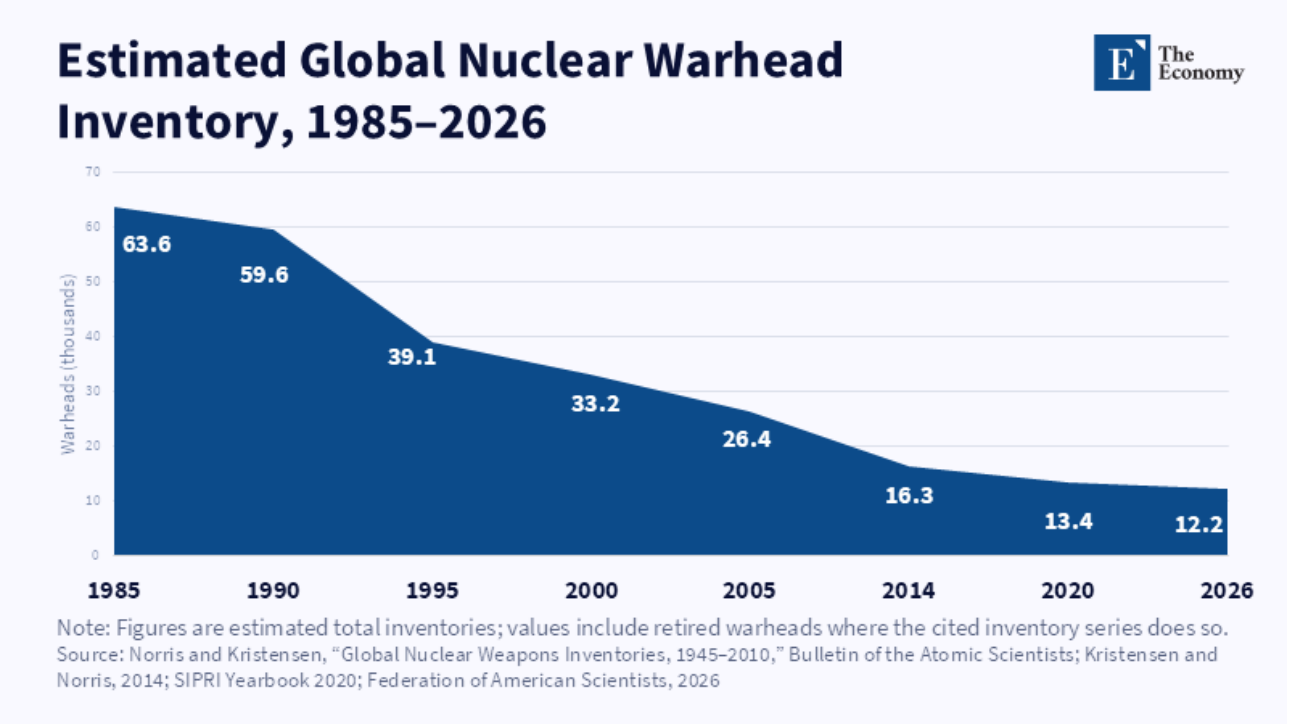

Figure 7: Nearly all of the post-Cold War arsenal reduction happened in a single decade, and the pace has since stalled.

Transferring this logic to military AI requires a structure that does not exist today. Neither the United Nations nor the IAEA nor any new international body has the tools to verify what is happening within the classified networks of two superpowers. At the Security Council level, both the U.S. and China have vetoes. The Group of Governmental Experts on Lethal Autonomous Systems has failed to produce a binding treaty in a decade of negotiations.

However, the idea of a third deterrent is not as unrealistic as it sounds. The European Union, with the AI Act, has already created a regulatory framework concerning commercial artificial intelligence but could, in theory, be extended to military technology exports. Average powers with significant technological capabilities, such as South Korea, Japan, Australia, the United Kingdom and Canada, could form a coordination framework similar to the Wassenaar Arrangement for export control, but adapted to software, models and computing infrastructure. It would function less as a regulator of superpowers than as a reliable third leg in a deterrence structure, capable of imposing costs, technological, commercial and diplomatic- on anyone who violates the

minimum commonalities.

The Wassenaar Arrangement has been operating since 1996 as a voluntary regime for controlling exports of conventional arms and dual-use technologies, without legally binding force, based solely on the stable national legislation of each participating state. Its inability to bind non-participating states, such as China, is precisely the limit that a corresponding artificial intelligence initiative would face. The point, however, would not be to bind China through such a regime, but to create a coalition of markets, standards and supply chains large enough that the non-compliance of any superpower, including the U.S. itself, would entail real costs of access to markets, semiconductors and research partnerships. Such a regime would not replace the U.S.-China bilateral negotiation described in the previous chapter. It would work in parallel with it, offering the element of external consistency that no bilateral agreement can produce on its own.

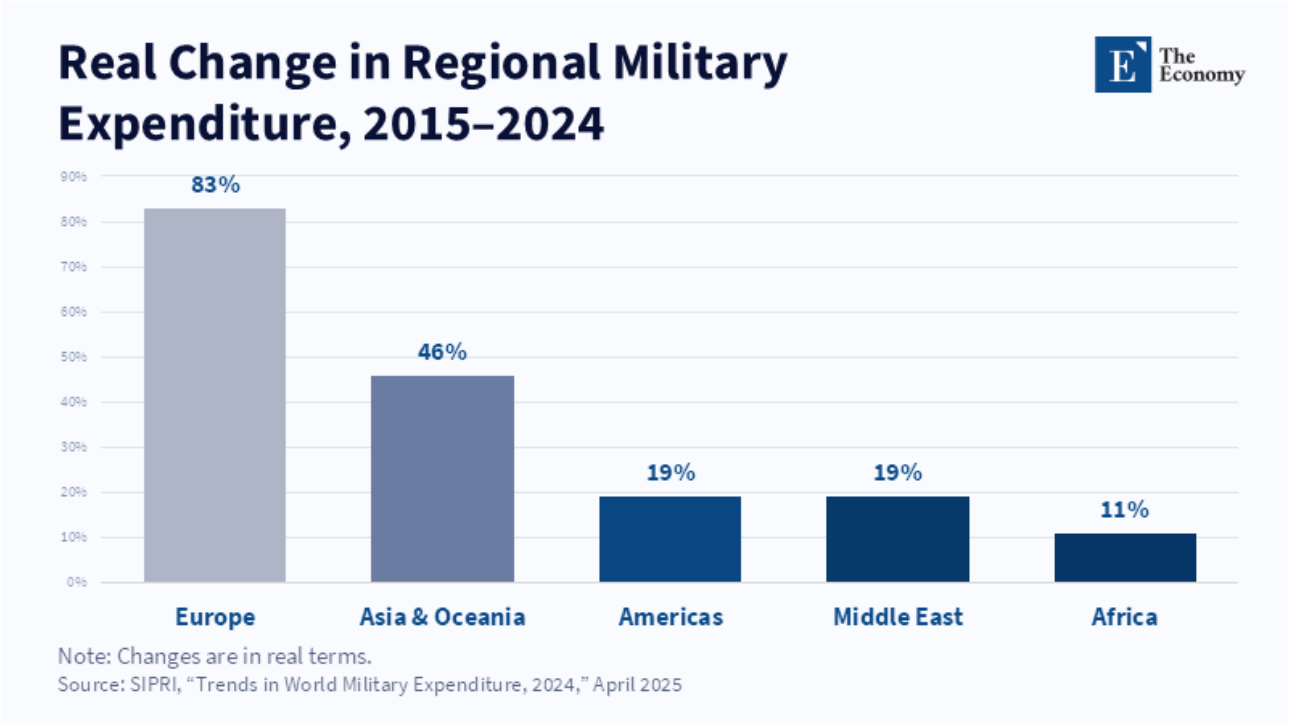

Figure 8: Europe's military spending grew faster than any other region's over the decade, a direct consequence of the war in Ukraine.

The objection is obvious: the U.S. and China can simply ignore such an alliance of middle powers. In practice, however, both are deeply dependent on allied networks, semiconductor supply chains, export markets and access to computing infrastructure that they do not fully control unilaterally. Amodei explicitly acknowledges that slowing down without international coordination is naive: if democratic countries unilaterally limit their capabilities, an authoritarian state can exploit the vacuum.[45] The counterargument, however, is equally true: if no one slows down, the speed of growth surpasses any control mechanism, including those that serve the states themselves.

5 Conclusion - Governing Military AI Under Strategic Uncertainty

The upcoming Trump-Xi summit will likely produce statements about human control of military AI. It will not produce regulation. The fundamental contradiction between states that commit to keeping humans at the core

of decision-making while at the same time developing systems that make them structurally redundant cannot be resolved through diplomacy without institutional tools of verification and enforcement.

What is possible is narrower but not insignificant: crisis management mechanisms, direct lines of communication, accident notification protocols, explicit commitments against autonomous cyberattacks on nuclear infrastructure. These amount to crisis management, analogous to the conditions of the Cold War. For something stronger, the logic that limited nuclear weapons, i.e., deterrence through credible punishment, requires third actors capable of implementing costs. The international community does not yet have this structure. Its construction, starting from the coordination of middle forces and not from the utopia of a global regulator, is perhaps the most realistic direction. Until then, managing uncertainty, not the illusion of its permanent elimination, remains the only realistic organizational imperative for military AI, as for any technology that evolves faster than the institutions that are called upon to control it.[46]

References

[1] Reuters (2026) 'Trump invites Xi to White House on September 24', 14 May.

[2] Razdan, K., Zheng, W. and Sim, D. (2026) 'AI fight shadows final preparations for Xi Jinping's White House visit to meet Donald Trump', *South China Morning Post*, 12 September.

[3, 30] The White House (2026) 'Fact Sheet: President Donald J. Trump Signs Historic Directive on AI in the National Security Enterprise', 5 June; National Security Presidential Memorandum/NSPM-11 (2026) *Artificial Intelligence in the National Security Enterprise*, 5 June.

[4, 45] Amodei, D. (2026) 'Pacing the Frontier', *Anthropic Blog*, 12 September; CBS News (2026) 'Anthropic CEO Dario Amodei says "exponential" growth of AI is a "warning sign that we need to slow down"', 12 September.

[5, 6, 18, 19, 20, 21, 22, 27, 28, 29] Sisson, M.W. and Jiang, T. (2026) 'Advancing human control of military AI', *Brookings Institution*, 9 September.

[7, 8] Reid, D.M. (2026) 'What Must a Soldier Still Know? Drawing the Competence Line in an AI-First Army', *Modern War Institute at West Point*, 11 September.

[9] Mishra, P. (2026) 'Command Without Control: Mission Command in the Age of Artificial Intelligence', *Military Review*, September–October.

[10] Zoldi, D. (2026) 'AI-Powered Target Bank Warfare: U.S.–Israel Use Advanced Tools to Strike "Pop-Up" Threats in Iran', *Autonomy Global*, 20 March; Kozlowskyj, C. (2026) 'Iran Conflict as a Testing Ground for AI Warfare Systems', *Bloomsbury Intelligence & Security Institute*, 24 March.

[11] Rogers, C. (2026) 'How Iran, Anthropic-DoD Dispute Show the Need for Protective AI', *Just Security*, 23 March.

[12] Wilcox, A. and Metcalf, C. (2026) 'AI Command and Staff — Operational Evidence and Insights from Wargaming', *Military Strategy Magazine*, 10(4), Winter, pp. 4–10.

[13] Shanahan, J. (2026) 'In AI, The United States and China Are Racing In The Dark', *Johns Hopkins SAIS*, 27 March; RAND Corporation (2025) *China's Military AI Wish List.* Center for Security and Emerging Technology, Georgetown University.

[14] The Guardian (2026) 'US allies lack resources to keep pace on AI, top Pentagon official says', 9 September.

[15] International Institute for Strategic Studies (2026) 'South Korea's push for military applications of physical AI', August.

[16] ICRC (2026) 'FAQ: Artificial Intelligence (AI) in the military domain', *International Committee of the Red Cross.*

[17] The White House (2024) 'Readout of President Joe Biden's Meeting with President Xi Jinping of the People's Republic of China', 16 November; Ministry of Foreign Affairs of the People's Republic of China (2024) *Statement following the Biden-Xi Summit*, Lima, November.

[23] AFP (2026) 'UN agrees text on lethal autonomous weapons', 6 September.

[24, 25, 26] UNODA (2026) *Convention on Certain Conventional Weapons: Group of Governmental Experts on Lethal Autonomous Weapons Systems, 2026 Session Documents*; WILPF (2026) 'CCW Report, Vol. 14, No. 2: The Final Stretch Before the Finishing Line', 11 March.

[31, 32, 33, 34, 35, 36, 39, 40, 43, 44, 46] Shanahan, J. (2026) 'In AI, The United States and China Are Racing In The Dark', *Institute for America, China, and the Future of Global Affairs, Johns Hopkins SAIS*, 27 March.

[37] Shanahan, J. (2026) 'In AI, The United States and China Are Racing In The Dark', *Johns Hopkins SAIS*, 27 March; Arms Control Association (2025) 'Artificial Intelligence and Nuclear Command and Control: It's Even More', September.

[38] Bremmer, I. (2026) 'The AI Race Could Sink US-China Détente', *Project Syndicate*, 8 September.

[41] Herzog, S. et al. (2025) 'Artificial Intelligence and Nuclear Weapons Proliferation: The Technological Arms Race for (In)visibility', *Risk Analysis*, September.

[42] BBC News (2026) 'China is winning one AI race, the US another — but either might pull ahead', September.