Artificial Intelligence and the Future of International Security: Emerging Challenges for the International System

Saira Akram [1*]

[1] Assistant Professor (OPS), department of Political Science, Bahauddin Zakariya University Multan, Punjab, Pakistan.
Email: Saira.akram29@yahoo.com

ABSTRACT: Artificial intelligence (AI) is now embedded in military targeting, intelligence analysis, cyber operations, information campaigns, and great-power industrial competition. Scholarship remains divided between accounts that treat AI as a revolutionary shift in power and accounts that stress its limits and uneven diffusion. This paper asks how AI is reshaping international security and under what conditions international institutions can manage its emerging risks. It combines a transparent narrative review of peer-reviewed research, official documents and policy analysis published mainly between 2018 and September 2026 with a structured comparison of three cases: United States–China competition over advanced semiconductors, AI-enabled drone warfare in the Russia–Ukraine war, and the multilateral effort to regulate lethal autonomous weapons through the Convention on Certain Conventional Weapons and the UN General Assembly. Drawing on the security dilemma and liberal institutionalism, it argues that AI's most consequential security effects arise less from "killer robots" than from three interacting mechanisms: the compression of decision time, the concentration of the inputs needed to build advanced AI, and the erosion of verifiable information. These mechanisms intensify competition but do not preclude cooperation. Institutions are most likely to succeed where they target narrow, observable behaviours, where leading states perceive shared exposure to accidents, and where developing countries gain material benefits rather than only obligations. The paper concludes with recommendations on human control, testing, incident reporting, confidence-building and inclusive capacity-building.

KEYWORDS: Artificial Intelligence, International Security, Strategic Stability, Security Dilemma, Liberal Institutionalism, Autonomous Weapons, AI Governance

Pages: 381 – 401
Volume: 5
Issue: 4 (April 2026)

Corresponding Author
Saira Akram
✉ Saira.akram29@yahoo.com

Cite this Article: Akram, S. (2026). Artificial Intelligence and the Future of International Security: Emerging Challenges for the International System. *The Regional Tribune, 5*(4), 380-401.
https://doi.org/10.55737/trt/v-iv.416

Introduction

In the rapid-fire world of tech, Artificial Intelligence has been hailed as one of the most transformative technologies to shape world politics. In 10 years, AI shifted from being a niche defence research topic to being a 'diplomatic leader level topic'. In November 2024, the presidents of the United States and China reaffirmed "the need to maintain human control over the decision to use nuclear weapons" (The White House, 2024) – the first-ever joint statement between the two countries. In December 2025, the UN General Assembly passed by 166 votes to 6, and a 7-strong abstention block, its third consecutive resolution on lethal autonomous weapons systems (United Nations General

Assembly [UNGA], 2025). In the meantime, battlefield reporting from Ukraine shows terminal guidance of mass-produced drones in the process of being automated with AI (Bondar, 2025), while even export restrictions on high-tech chips have become a key tool of great-power competition (Egan & Sanders, 2026).

As is well known to scholars of international relations, but is now more pressing: does a general purpose technology affect the distribution of power and the probability that conflict will occur, and can institutions influence the distribution of power? History indicates that technology has never been the only factor in change, from the railways to electricity to nuclear weapons to the internet. Its impact varies from organisation to organisation, as well as from doctrine to doctrine, the economic system, and political decision making (Ding & Dafoe, 2023; Horowitz, 2018). This is the same with AI. Alongside this is the realization that most current applications of AI in the military are narrow, brittle, and rely on human judgement (Goldfarb & Lindsay, 2022), while pronouncements of self-reliant superweapons and an unstoppable arms race abound.

Research problem/gap. Although the literature is expanding quickly, there are a number of problems with it, which are being tackled in this paper. Firstly, much of it looks at one domain, be it autonomous weapons, nuclear stability, or disinformation, in isolation, when there are underlying mechanisms that interact between the domains. Second, looking at governance literature, it can be observed that initiatives are often listed without the question being asked: under what circumstances do they influence the behaviour of the state? Third, the views of developing countries, which are primarily consumers, not producers, of frontier AI continue to be underrepresented in discussions of security, despite the fact that most states are missing from major governance efforts to date, such as the United Nations High-level Advisory Body on Artificial Intelligence (HLAB-AI) (United Nations High-level Advisory Body on Artificial Intelligence, 2024). An integrated, theory-based assessment that links the technical characteristics of various AI systems with specific security measures and the conditions for institutional management is missing.

Research questions. The main issue is: how is AI changing international security dynamics, and what are the possibilities for international institutions to control the new risks that AI poses? Three sub-questions follow. How does the use of various types of AI impact power, strategic stability, conflict, and cooperation dynamics? (2) In what way do these mechanisms show in three well-documented cases? (3) What governance schemes are possible with those mechanisms, and what are the challenges?

Objectives and contribution. The paper is designed to separate out the 'now' from the 'what if' of future developments, to describe the security impacts of AI using two different perspectives, and to present evidence-based policy recommendations. There are three aspects of its contribution. The idea behind it is that it structures nine challenges often talked about under three mechanisms: Decision compression, Input concentration, and Information erosion. Empirically, it contrasts three cases, based on the same criteria. Normatively, it is about what makes an institution work, not about the futility of cooperation or the magical quality of a new rule to be self-enforcing.

Literature Review

The literature on AI and international security can be structured into four debates: the first concerns
the redistribution of power by AI; the second is about the destabilisation of strategic relations; the third is on the effects of AI on the information and cyber environment; and the fourth is on the governance of AI. The fifth question that is behind all four is who gains and who takes risks from AI.

Revolution or Evolution? AI and the Power Sharing

AI has been viewed as a potential revolution in military affairs in the early scholarship. Horowitz (2018), however, believed that the best way to think about AI is as a general-purpose 'enabler' for which the impact on the balance of power is determined by the extent to which AI can be diffused and by the extent to which militaries can organisationally adapt. Payne (2018) argued that AI might disrupt the psychological underpinning of strategy, by taking human thinking out of the equation in important decisions. There is more recent work that is more sceptical. Since machine learning reduces the costs of prediction, but not of judgement, AI amplifies the need for human decision makers and is much less effective in the adversarial conditions of war, according to Goldfarb and Lindsay (2022). Similarly, Jensen et al. (2020) highlight data limitations and incoordination within organisations, and Ding and Dafoe (2023) demonstrate that general-purpose technologies, like electricity, change slowly and by widespread diffusion and not by weapons.

The argument is thus not so much about the effect AI has, but the how: battlefield supremacy or economic productivity and control of supply chains. This debate suffers from the fact that it tends to speak in generalities, that of the United States and China, with no consideration given to the experience of middle powers and developing states in the face of AI-enabled changes.

Strategic Stability, Escalation and Decision-making

The second group of works has to do with stability. War itself may be sped up, Johnson (2019, 2020) suggests, and AI can make it easier to monitor an opponent's retaliation, while seducing governments to give machines the responsibility for decision-making. A SIPRI study on nuclear-armed states finds that the most immediate threats are not with autonomous nuclear weapons but with the use of AI in early warning, intelligence, and command and control that may have the potential to create misperception (Boulanin et al., 2020). Micro-level evidence from experimental research. There is a Dunning–Kruger effect for automation bias, with individuals with limited knowledge of AI showing the highest levels of the bias, according to Horowitz and Kahn (2024). Rivera et al. (2024) find that, in simulated wargames, all five of the large language models (LLMs) tested exhibited escalatory behavior, including arms-race behavior. In fact, according to the results of wargames by Schneider et al. (2023), the fear of vulnerability itself didn't necessarily lead to a pre-emptive nuclear strike: the most serious threats were a lack of awareness about how vulnerable one's own nuclear systems were (which tempted players to conduct aggressive counterforce campaigns), and the temptation to pre-delegate or automate decisions.

There is a methodological gap between alarming theoretical scenarios and the limited empirical observation of the debate. Although simulations and surveys provide valuable insight, it is not clear how they apply to real crises, and few studies have explored how militaries actually use AI in procedures.

Autonomous Weapons, Law and Norms

The most advanced bodies of normative literature are on the subject of autonomous weapons.

However, Scharre (2018) not only records autonomy in current systems, but also asserts the importance of maintaining human control over the use of lethal force. Bode and Huelss (2018) demonstrate how such "appropriate" human control norms are being informally constructed by practice, prior to the formalisation of rules. Rosert and Sauer (2021) highlight the reasons for the failure of the campaign against "killer robots" to secure the same successes as campaigns in the past for IHL, including dual use and the lack of definition of the prohibited object, while Garcia (2018) addresses autonomous weapons as a threat to international peace and security, and not just IHL. There has

been ongoing debate between those who are proponents of a legally-binding instrument and those who believe that current IHL and national review is adequate.

Information, Cyber Operations and Societal Resilience

The fourth strand is taken up by information and cyber security. Deepfakes could "take a toll on the democratic discussion" and lead to a "liar's dividend" where accurate information is discredited as an impersonation, Chesney and Citron (2019) said. Goldstein et al. (2023) evaluate the potential of language models for reducing the influence operations costs and scaling up influence operations. There was, however, no empirical evidence that AI-generated disinformation impacted the outcome of the 2024 election cycle or that it impacted trust and amplified harmful narratives (Stockwell, 2024). In the field of cyber security, Buchanan (2016) demonstrated that it is possible to confuse defensive attacks with preparations for attacks, which could be exacerbated by AI automation. There is no consensus in the literature as to whether AI is a friend of an offense or a friend of a defence, and evidence is more often based on vendor reports or competitions than on independent longitudinal data.

Governance, Arms Control and Inequality

In the governance literature, a question that is raised is whether there is control of AI. Maas (2019) takes a lesson from the nuclear arms control field, contending that epistemic communities and norm institutionalisation can make a difference even in circumstances where verification is difficult. In the short term, Horowitz et al. (2020) suggest that treaties are too unrealistic and the implementation of confidence-building measures (CBMs) is more feasible, including information exchanges and incident communication. The political economy approach can also draw attention to how control of network hubs can be "weaponised" (Farrell & Newman, 2019), and semiconductor supply chains are viewed through this prism (Miller, 2022). De-colonial and critical approaches scholars believe that AI governance is simply perpetuating global asymmetries, and that data and labour are being harvested from the Global South, and rules are being developed elsewhere (Mohamed et al., 2020). The UN High-level Advisory Body (HLAB-AI, 2024) found that 118 member states were part of none of the seven prominent AI governance initiatives sampled by UN HLAB-AI, most of which are in the Global South.

Contributions of this Paper

In all these literatures, analysis is domain-specific; there is no explicit mechanism connecting technical aspects to political achievements, and the conditions for the success of governance are ignored. This paper provides a response to these debates by combining the debates in three cross-cutting ways, by means of comparative cases, and by two existing theories, which specify situations where competition or cooperation is likely to prevail.

Theoretical Framework

When two theories are combined it is because they address the two parts of the research question. Why there is competitive pressure in the case of states with defensive intentions is explained by the security dilemma. Liberal institutionalism, however, has offered an explanation as to how and when states still can cooperate. Combined, they dispense with the pessimism of a doomed inevitability of arms races and the optimism of a shared interest making for effective rules.

The Security Dilemma

Jervis (1978) refers to the security dilemma as a situation in which the actions of one state to enhance their own security decrease their own security and encourage other states to take countermeasures that make all states less

secure. Jervis recognized two factors that affect its intensity: the ability to differentiate the offensive and defensive abilities, and the ability of either the offence or the defence. The dilemma is most extreme when offence and defence are not distinguishable, and even harder when offence is preferred, but when defensive posture is easily signalled and defence is preferred, it is easier to cooperate. Buchanan (2016) expanded the idea to cyberspace, and intrusions required for defence are similar to those required for attack.

The security dilemma translates to three mechanisms and expectations when applied to AI:

- Indistinguishability. States often cannot draw conclusions about the intentions of their rivals from the capabilities which they share because of the dual use of models, chips, and data that are used for civilian, defensive and offensive purposes. Expectation: investment in AI will be seen as a threat to others, and it will lead to even more investments in AI, and to stricter restrictions on tech transfer.
- Perceived offence advantage. First-mover incentives increase if AI is thought to be beneficial for the attacker, such as if AI were to make cyber intrusions faster or to allow for tracking of retaliatory forces. Expectation: increased push for rapid deployment, and also to decrease human deliberation during crises.
- Uncertainty about capabilities. From the outside, it is difficult to observe and measure AI capabilities. Expectation: worst-case assessment and challenges in agreeing on verifiable limits.

Liberal Institutionalism

The liberal institutionalist perspective of the international system is anarchic, but it may believe that institutions can help to foster cooperation by lowering transaction costs, providing information, making things more transparent, extending the shadow of the future, and connecting issues (Keohane, 1984; Keohane & Martin, 1995). The importance of institutions is greatest in those circumstances in which states have overlapping interests, but cannot achieve them independently due to uncertainty and the worry of being cheated.

In its context of AI, it gives the following mechanisms and expectations:

- Information provision. Common assessments, incident reporting and scientific assessments help to minimize uncertainty of capability and risk. Expectation: Cooperation is more likely when there is credible, common knowledge, e.g., the UN Independent International Scientific Panel on AI.
- Focal points and norms. Institutions establish norms for acceptable conduct, including human control of decisions to use nuclear weapons that make it easy to see when they are not being followed and cost reputational consequences. Expectation: the lesser the scope and the more specific the commitment, the more likely it will be accepted.
- Side payments/issue linkage. Institutions can package the commitments together, including capacity building. Expectation: Developing states will have more positive support for governance processes if it has access to technology and skill.

Combining the Two Avenues of Thinking

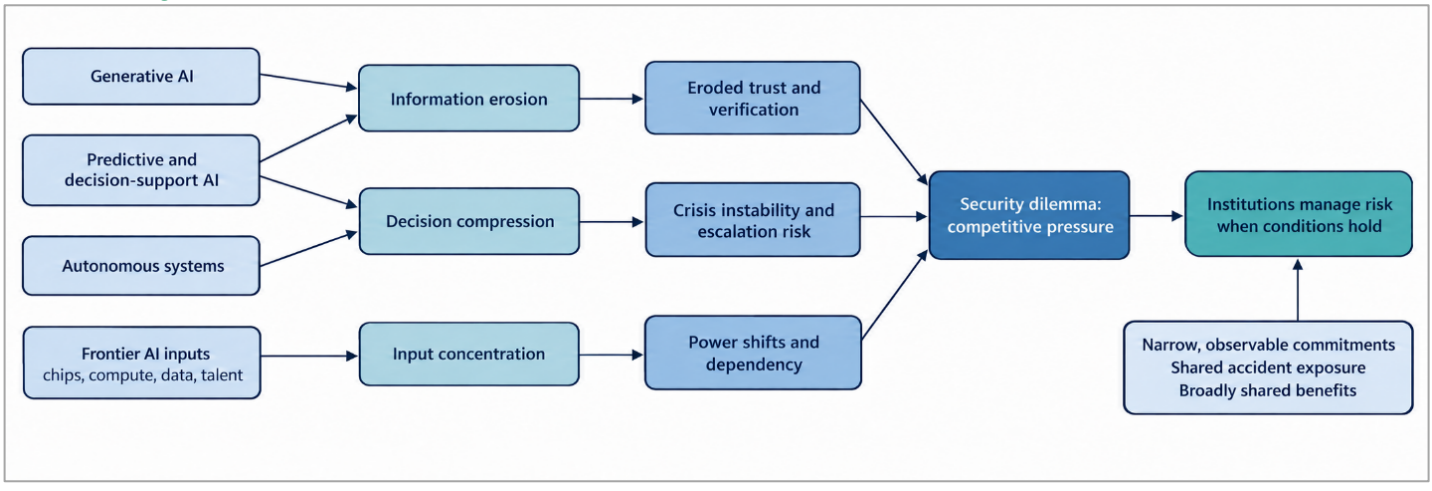
The two theories are not necessarily competing explanations. The security dilemma indicates what inhibits cooperation, and institutionalism indicates how the obstacles can be reduced. This combination creates a conditional proposition for analysis (Figure 1): international institutions are most likely to manage AI-related security risks when they address domains where they can reduce indistinguishability and uncertainty, major powers see themselves as being at risk for accidents or escalation, and there is a wide distribution of benefits to get them involved. Cooperation on the other hand, will be least where AI capabilities lie at the heart of relative power, are not easily seen, and are very concentrated within a few states or companies.

An aspect of both theories is that they adopt a state-centrism approach. Both were created for a technology that is mostly pushed by private firms. Firms, as mediators, are thus viewed as resources, disclosures, and commercial motives that affect the security dilemma and institutional choices of states.

Figure 1

Analytical Framework: from AI Categories to Security Mechanisms and the Conditions for Institutional Management

Methodology



Research Design

This study is qualitative, desk-based, and is a combination of a narrative literature review and structured, focused comparison of three cases. The choice of a narrative was made in preference to a systematic review for three reasons: (1) it covers a broad range of fields, including international relations, law, computer science, and policy analysis, (2) many of the most pertinent primary sources are academic and other grey literature that is not indexed in academic databases, and (3) it was also not possible to conduct a systematic review. The review is not systematic, meaning that the review was not carried out using a pre-registered review protocol, a 'search string' across the entire database, nor formal screening/exclusion procedures or processes, and the findings, therefore, should be interpreted in that context.

Source Selection and Review Period

The review period is from January 2018 to 24 September 2026, the date of the research. Previous foundational studies (Jervis, 1978 and Keohane, 1984) were included if they provided theoretical concepts. Leading international relations and security journals were selected and core peer-reviewed scholarship was identified. In September 2026, targeted web searches were then conducted to verify and update facts on the three cases, as well as on governance initiatives, such as UN General Assembly resolutions, the CCW Group of Governmental Experts, the REAIM summits, the UN Global Dialogue on AI Governance, the Council of Europe Framework Convention, US chip export controls, AI use in Ukraine, AI and elections, and AI-enabled cyber operations.

The following hierarchy was used in selecting sources: primary documents (UN records, treaty texts, official statements and regulations); peer-reviewed articles; reports from well-known research institutes (SIPRI, CSIS, CNAS and the Alan Turing Institute); and well-known journalists for dated events. Corporate reports were only used where an incident is the only public record of that event and are identified as such (as 'Self Reported'). For active conflicts, claims were treated as provisional, unless there was more than one independent source.

Case Selection

The selection of the three cases was based on three selection criteria: Each of the selected cases represents one of the mechanisms identified in the framework, each case is well documented in public sources, and the selected cases span the economic, battlefield, and institutional arenas. The cases are:

1. The input concentration and the security dilemma in technological competition are portrayed by the United States-China competition over advanced semiconductors (2022-2026).
2. AI-powered (drone) warfare in the Russia – Ukraine war (2022 – 2026) – decision compression and the loss of human control in a modern war.
3. Multilateral process on lethal autonomous weapons systems in the CCW and UN GA (2019- 2026): showing the potential and scope for institutional management.

Other cases, such as alleged AI-powered targeting in Gaza, were considered but excluded as primary cases due to a reliance on investigative reporting that is disputed in the public domain and the denial of governments, which would make it difficult to make comparisons.

Analytical Categories

Five criteria are pre-defined, and each of the cases is compared on the basis of the following: (a) principal actors; (b) AI applications involved; (c) observed security effects; (d) institutional responses; and (e) evidentiary limitations. This analysis subsequently examines each case with regard to the expectations based on the security dilemma and liberal institutionalism.

Limitations

There are five basic limitations of the design. It lacks representation of Chinese, Russian, and Global South scholarship and is based on English-language sources. It can't view classified military practice, so any information on military AI is based on what the states and researchers have made public. The majority of evidence from Ukraine is based on interviews and reporting which cannot be independently audited. The review is time-bound, meaning that at the time of the review, governance processes were in progress, such as the CCW Review Conference, which will take place in November 2026. Lastly, qualitative case comparison does not support causal generalisation, but plausibility and mechanism tracing.

Analysis and Discussion

There are nine challenges generally referenced in the literature, which can be considered as the manifestation of 3 mechanisms. Arranged in this manner, it is possible to see how issues in one arena exacerbate issues in another and which governance resources work best for each risk.

Decisions: Compressing Decisions

AI reduces the time and space for human decision-making in military actions, intelligence evaluation, and nuclear warning. The use of AI for terminal guidance in Ukraine has been reportedly shown to enhance the success of engagement from 10-20% to 70-80% after human target selection (Bondar, 2025). Autonomy is also susceptible to electronic jamming, raising a potential security problem, since countermeasures are likely to induce more autonomy. But there is no evidence that weapons are widely deployed that autonomously search for, select and attack targets.

Decision support adds risks by way of automation bias, opacity, and inaccurate prediction. When under pressure, operators might take algorithmic recommendations without proper analysis. Training is important to decrease

overreliance, and nominal human involvement does not ensure meaningful control. As well, biased training data and opaque models make it difficult to review and hold models to account legally (Goldfarb & Lindsay, 2022). Responsibility is still on the human beings and institutions, not machines (UNODA, 2025).

These pressures are especially threatening during times of nuclear crises. AI may amplify the weakness of the opponent, make false warnings, or induce premature delegation (Boulanin et al., 2020; Johnson, 2020). Unpredictable escalation and overconfidence thus need to be taken with caution. The Biden-Xi pledge on human control of nuclear employment is a timely solution to a critical juncture, but further measures are needed in warning and crisis management, intelligence, and across all other areas of human control to ensure deliberation, accountability and effective human oversight.

Mechanism Two: Input Concentration

The second is the fact that frontier AI is a technology that concentrates the inputs it needs—namely, advanced chips, compute, data, and expertise. Connects technological competition, inequality, and private power. Some companies are developing the most sophisticated chips, another is manufacturing the most cutting-edge lithography machines, and the manufacturing is concentrated in geographical regions (Miller, 2022). This gives rise to "weaponised interdependence" (Farrell & Newman, 2019) chokepoints. About three-quarters of the measured capacity for AI supercomputers is in the United States and 15 per cent in China in early 2025, which fuels the export-control conflict mentioned in Section 6.1.

In states that lack the ability to establish their own compute, they depend on other clouds and models for their security-related services. This puts them at risk for supply disruption, foreign data jurisdiction, and limited auditability. Overall, HLAB-AI (2024) identified 118 states that do not have any of the seven governance initiatives sampled. In these states, it's not so much an arms race, as an absence of rule-making and an absence from defensive technologies.

Security is designed by private companies – primarily American and Chinese – through their release and disclosure decisions and even provide the evidence base. Some claimed that its model was used in an AI-agent-driven espionage campaign by Anthropic (2025) that was attributed to a Chinese state sponsored group, but it is difficult to verify such self-reported uses. Open models also enable non-state actors to be empowered. This means that firms need to be regulated as regulated actors and as participants.

Mechanism 3: Information erosion

The third is the breaking down of verifiable information: what is true, who has done it, and what can be done. Connects disinformation with cybersecurity and arms-control verification. The use of AI for content creation has also reduced the costs of producing persuasive content (Goldstein et al., 2023), but there is a delay between the initial indications and the actual impacts. The elections in 2024 did not reveal any definite impact of AI on election results, but there was a decline in trust in the electoral process (Stockwell, 2024). The greater danger is the “liar's dividend” (Chesney & Citron, 2019) because of the ability to discredit any and all evidence that is actually authentic, making attribution, deterrence and accountability more challenging in crisis situations.

AI can also carry out automated intrusion at machine speed in cyberspace (Anthropic, 2025), and DARPA's AI Cyber Challenge showed the potential of automated patching of vulnerabilities. There can be a relative advantage to well-resourced defenders, with utilities and hospitals lagging behind, particularly in developing countries. There is also a lack of distinction between uses in attack and defence (Buchanan, 2016). In arms control, capabilities lie with

easy-to-copy and tweak weights of models. So plans for using observables like compute, tests and doctrine are increasingly being proposed (Maas, 2019).

Cross Cutting Challenge: Law, Sovereignty and Human Rights

The importance of all three mechanisms is that they lead to international law. The IHL principles of distinction, proportionality, and precaution apply (UNGA, 2025), and there is debate over whether there is a need for new rules for autonomous weapons. Surveillance AI is being exported, impacting human rights, and foreign platforms and infrastructure pose a threat to sovereignty. However, the EU AI Act is found to exclude military and national-security systems (Regulation (EU) 2024/1689, Art. 2(3)) as well as the Council of Europe Framework Convention excludes national defence (Council of Europe, 2024). This leads to governance shortfalls in the areas of highest security concerns.

Theoretical Evaluation

Much observed acceleration is explained as a result of war necessity and commercial competition as well as interstate fear, confirming the expectations of the security dilemma, but qualifying them. There is some support for liberal-institutionalism, as narrow and focal commitments have been able to bind even the great powers and broad commitments have gained wide but shallow backing. The evidence is summarized in Table 1.

Table 1

AI Applications, Security Opportunities, Risks and Supporting Evidence

AI application	Security opportunities	Principal risks	Supporting evidence and strength
Autonomous navigation and terminal guidance (drones)	Resilience to jamming; fewer munitions per target; reduced exposure of personnel	Erosion of human control; proliferation to non-state actors; arms-race pressure	Interview-based evidence from Ukraine (Bondar, 2025). Moderate: not independently audited
Decision support and target nomination	Faster processing of sensor data; potential reduction of some human errors	Automation bias; opacity; biased training data; diffused accountability	Experimental studies (Horowitz & Kahn, 2024); analytical work (Goldfarb & Lindsay, 2022). Moderate to strong for mechanism, limited for field practice
AI in nuclear-related early warning and ISR	Improved detection and data fusion; possible support for verification	False alarms; perceived vulnerability of second-strike forces; pressure to pre-delegate	Expert assessment (Boulanin et al., 2020); wargames (Schneider et al., 2023). Largely scenario-based
Generative models in strategic simulation	Training and red-teaming of decision-makers	Unpredictable escalation if used for real decisions	Simulation study (Rivera et al., 2024). Limited external validity
AI-enabled cyber operations	Automated vulnerability discovery and patching	Automated intrusion at scale; offence advantage for under-defended targets	DARPA (2025) competition results; developer-reported incident (Anthropic, 2025). Emerging, partly self-reported

AI application	Security opportunities	Principal risks	Supporting evidence and strength
Generative content and influence operations	Detection tools; faster fact-checking	Cheaper disinformation; liar's dividend; loss of trust	Election reviews (Stockwell, 2024); legal analysis (Chesney & Citron, 2019). Moderate: effects on outcomes not demonstrated
Compute and semiconductor supply chains.	Leverage to slow proliferation of frontier capabilities	Weaponised interdependence; retaliation; exclusion of developing states	Compute concentration estimates (Pilz et al., 2025); policy analysis (Egan & Sanders, 2026). Strong for concentration, contested for effectiveness of controls

Comparative Case Analysis

Each case is analysed based on the five criteria as outlined in Section 5.4, and then a cross case comparison is done.

Case One: United States–China Competition over Advanced Semiconductors

Actors. The key players are the US government, particularly the Commerce Department's Bureau of Industry and Security (BIS), the Chinese government, the top chip designers and manufacturers, and friendly countries where key parts of the supply chain are located. Applications. What is being competed for isn't a weapon, its computing power to train and deploy the leading edge AI models, both civilian and military applications.

Security effects. As of October 2022, the United States introduced broad export restrictions on advanced computing chips and chip-making equipment to China, due to worries about military modernisation and surveillance (Bureau of Industry and Security [BIS], 2022). The controls were tightened up in subsequent years. Policy then shifted. BIS shifted its review process for licenses to individual chips, such as Nvidia's H200 and AMD's MI325X, from a presumption of denial to a case-by-case basis in January 2026, and a presidential proclamation put a 25 percent tariff in place, which acts as an export fee; analysts estimated that China could still buy up to about 850,000 H200 chips under the volume cap under the presidential proclamation (Egan & Sanders, 2026). The case is a good example of the security dilemma. US controls, though supposedly defensive, were seen in Beijing as a way to limit China's development, and led to increased investment in domestic alternatives and China's own controls of critical materials. Commercial interests and domestic politics, not just security concerns drive technological competition, as US policy has been oscillating.

Institutional responses. AI chips are not regulated by a multilateral accord. Controls are national and informal with allies, and a challenge to the target. This lack is in line with liberal institutionalist theory: when a capability is crucial for relative power and benefits are concentrated, states opt for unilateral leverage rather than shared rules.

Evidentiary limitations. There is disagreement regarding the effectiveness of controls. Judgements of their impact on China's AI advancements are based on murky estimates of the extent of smuggling, stockpiling and domestic manufacturing, and policy has shifted quickly.

Case Two: AI-enabled Drone Warfare in the Russia–Ukraine War

Actors. The Ukrainian and Russian armed forces, a huge number of Ukrainian defence start-ups, foreign technology companies, and volunteer networks.

Applications. Examples of documented applications are the analysis of intelligence, surveillance and reconnaissance data, automatic target recognition, and autonomous navigation in the final phase of a drone's flight after target selection by a human (Bondar, 2025). Russian attempts to create a domestic drone ecosystem also include the development of Russia's drones for autonomous flight with the use of AI (Bondar, 2026).

Security effects. The case is the best evidence for decision compression publicly available. In an environment where electronic warfare is present, remote piloting is not reliable, and thus onboard autonomy is demanded for operation. The reported developments in hit rate and the fact that the unit cost of navigation modules is low compared to the unit cost of platforms indicates that the diffusion of autonomy is taking place via cheap and mass produced systems as opposed to expensive platforms (Bondar, 2025). There are two implications of this on a wider level. It brings down the bar for other states and non-state actors to adopt similar capabilities, and it also facilitates reductions in human involvement under pressure during wartime that Bode and Huelss (2018) call "the normalisation of incremental changes in human involvement before law. Meanwhile, the evidence doesn't indicate that the weapons are fully autonomous in terms of identifying targets themselves; humans are still involved with the target identification process in the case studies described.

Institutional responses. The use of AI in the conflict is not subject to a specific institutional restriction, apart from the general control of international humanitarian law (IHL). Ukraine has made a positive contribution by signing the REAIM 2026 "Pathways to Action" (Ministry of Foreign Affairs of Spain, 2026), while the Russian Federation voted against the 2024 General Assembly resolution on autonomous weapons (Perrin, 2025). The security dilemma would suggest that wartime incentives prevail over institutional ones.

Evidentiary limitations. Most evidence is from interviews with participants and manufacturers, which could have an interest in presenting effectiveness. Independent audit of figures is not possible, and Russian practice is not as transparent.

Case Three: the Multilateral Process on Lethal Autonomous Weapons

Actors. The REAIM summits, led by a group of middle powers, and other civil society coalitions, the UN General Assembly, the UN Secretary-General, and the International Committee of the Red Cross (ICRC) are all states parties to the CCW.

Applications. The process applies to weapons systems that target and engage without further human involvement and, more generally, the application of AI to the military.

Institutional response and security effects. Since 2017, the CCW Group of Governmental Experts (GGE) has been engaged in autonomous weapons. It was agreed that since 2023, a set of elements which will constitute the instrument must be formulated. It produced a ‘rolling text ', covering characterisation, application of IHL, prohibitions and restrictions, and accountability and risk mitigation, with states converging on the applicability of IHL, and on the principle that human responsibility for upholding IHL cannot be delegated to machines, while divergence remained over the scope of prohibitions and legal versus political measures (UNODA, 2025). Brazil et al. (2025) states that 42 states were willing to negotiate on this text. The European Union (2026) discussed the "agreed set of elements" as an important milestone at the preparatory committee for the CCW's Seventh Review Conference to be held in September 2026, but the nature of any instrument was left open. The Review Conference is on 16–20 November 2026.

Meanwhile, however, the General Assembly passed resolutions with majorities in 2023, 2024 and 2025 (Figure 2). Resolution 78/241 passed by 152 votes to 4 (UNGA, 2023), resolution 79/62 by 166 to 3 (Perrin, 2025), and

resolution 80/57 by 164 to 6 with 7 abstentions (UNGA, 2025). The commentators found that there was no mention of the beginning of negotiations in the 2025 text (Liu, 2025). The REAIM track exhibits the same trend of being "wide but not deep. The 2023 Call to Action is supported by both the U.S. and China, the 2024 Blueprint for Action by the U.S. but not China, and the 2026 Pathways to Action by about 40 states, but not by either superpower (Harjani & Liu, 2026). A resolution passed by the 2025 General Assembly on risks at the intersection of AI and nuclear weapons (80/23) was voted against or abstained from by all nuclear powers (Harjani & Liu, 2026), which is to remind us that even the limited nuclear-control norm has yet to be multilaterally institutionalised.

Figure 2

Recorded General Assembly votes on resolutions on lethal autonomous weapons systems, 2023–2025

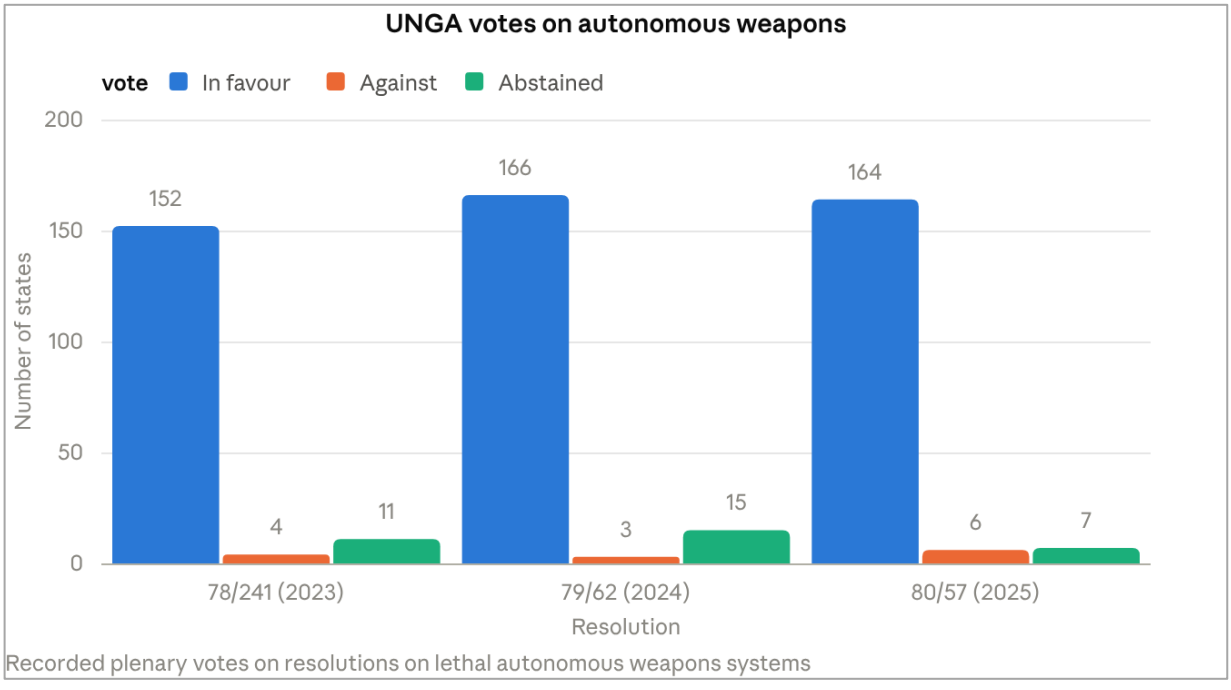


Figure 2 · UNGA (2023, 2025); Perrin (2025) · 3 resolutions

Note: Support has remained broad while the small group of opposing states has changed little, consistent with wide but shallow institutional endorsement.

Cross-case Comparison

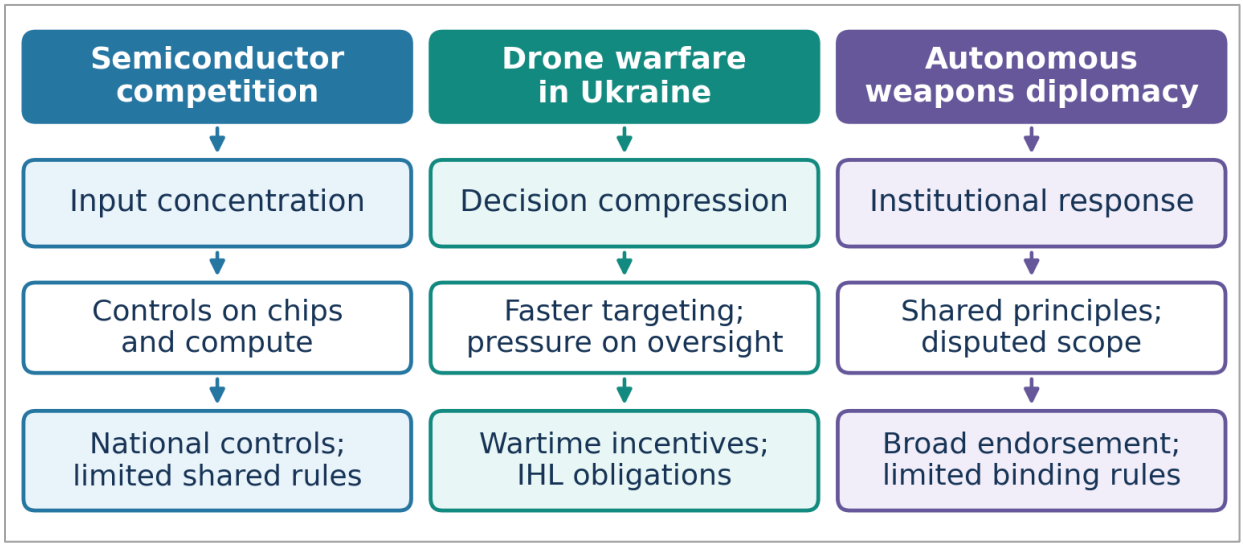
In the three cases, the mechanisms at various levels are observed. In the semiconductor case, the input concentration leads to a strong security dilemma and hardly any institutional management, as the stakes are related directly to relative power. In the Ukraine situation, the compression of the decision is due to the necessity of war, and institutions have little leverage during war. Institutions have developed common terminology and convergence on the principles of accountability and the applicability of IHL in the case of autonomous weapons, but have not been able to get binding commitments from the most capable states to develop such weapons.

The comparison is in support of the conditional proposition in the 4.3. Institutions advanced when commitments were minimal and great powers were equally vulnerable to catastrophic risk, and were least well advanced when capabilities were most central to relative power (chips), or incentives were most tied to active war (Ukraine). Another trend is the reduced involvement of the two AI superpowers in multilateral foray over 2024–2026, while middle powers have continued to carry out processes without the participation of the key actors. In each

of these three, the evidence is weakest at the highest risk. Figure 3 compares the mechanisms and institutional responses across the three cases.

Figure 3

Comparison of the Three Cases and their Principal Governance Constraints



Note: Qualitative synthesis of the case analysis in Section 6; the pathways are interpretive, not measured effects.

Benefits, Counterarguments and Conditions

A risk-only analysis would end up reiterating the technological determinism which it decries. AI does have security advantages, but only if and when this happens.

Potential Contributions to Security

Defensive cybersecurity. The window for critical infrastructure to be exposed to vulnerabilities could be shortened if they are automated with vulnerability discovery and patching. This capability was showcased in the DARPA AI Cyber Challenge in a competition environment, and finalist systems have been published as open-source software (DARPA, 2025). It is most likely to be a benefit in public and developing-country institutions where defenders are more likely to use such tools; it is least likely where defenders' access to these tools is limited to well-funded organizations.

Conflict early warning. Machine-learning models can be used on event data, satellite imagery and open-source reporting to identify patterns, like troop movements or displacement, earlier than a manual analysis. This is because the forecasts are not predictions to act on automatically, but rather only a suggestion to be investigated by a human: the predictive accuracy of rare events and of adversarial deception is limited (Goldfarb & Lindsay, 2022).

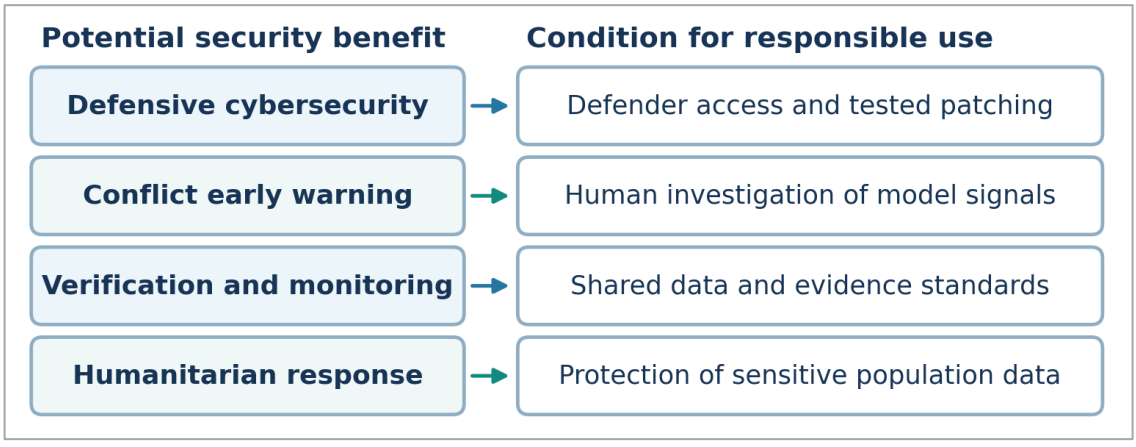
Verification and monitoring. Technology for analyzing satellite imagery, trade data, and sensor networks, using AI, could help with monitoring arms control/arms sanctions. Ironically, it is the technology that makes it difficult to verify AI that can make it easier to verify other weapons. The benefit is dependent on shared access to data and agreed standards of evidence (both of which require institutional cooperation).

Humanitarian response/crisis management. Given the right safeguarding of data related to vulnerable populations, AI can help in disaster damage assessment, logistics, and translation in disasters and conflicts. Similarly, Decision-support tools might enable more options to be considered in a time-constrained setting, but evidence of

escalation tendencies in language models (Rivera et al., 2024) and on automation bias (Horowitz & Kahn, 2024) suggests that this potential would have to be carefully tested and that designs would need to be used to expand rather than restrict human deliberation. Figure 4 summarises the conditions associated with these potential benefits.

Figure 4

Potential AI Contributions to Security and Conditions for Responsible Use



Note: These are conditional applications discussed in Section 7.1, not estimates of realised benefits.

Counterarguments to the Central Thesis

The issue of three counterarguments is of serious consideration. Firstly, the impact of AI might be exaggerated. The implications of the robustness of the mechanisms put forward here may be weaker than suggested if Goldfarb and Lindsay (2022) are correct in claiming AI will be complementary rather than supplanting human judgement and if its battlefield advantages are indeed unique to the extraordinary circumstances of the Ukraine war. This critique is invaluable, and is one of the reasons that the paper distinguishes established developments from scenarios. But revolutionary capabilities are not necessary for the mechanisms. The gradual shift toward more autonomy and decisions made in the presence of electronic warfare is reflected in decision compression; concentration of inputs is reflected in supply chains and export policy; erosion of information is reflected in the liar's dividend and in verification problems. All are visible at the present time.

Secondly, AI can enhance deterrence and stability. There is a possibility for better intelligence to decrease misperception and better defence to increase the costs of aggression. In such a case it is not impossible that both sides become transparent symmetrically. Where improvements are asymmetric, such as on tracking retaliatory forces, the security dilemma can lead to feeling exposed, which might result in a response of dispersing, automating or pre-delegating, thus reducing stability (Johnson, 2020; Schneider et al., 2023).

Thirdly, the institutions may not matter since big powers will simply renege. The engagement by the United States and China in recent schemes has been diminished, and there is evidence of realist scepticism. But the record also demonstrates that narrow commitments can draw in rivals, like in 2024, and that institutions influence norms and expectations among the multitude of states that are not great powers. The right conclusion, however, is not that institutions don't matter, but that they have effects that are sometimes and sometimes not.

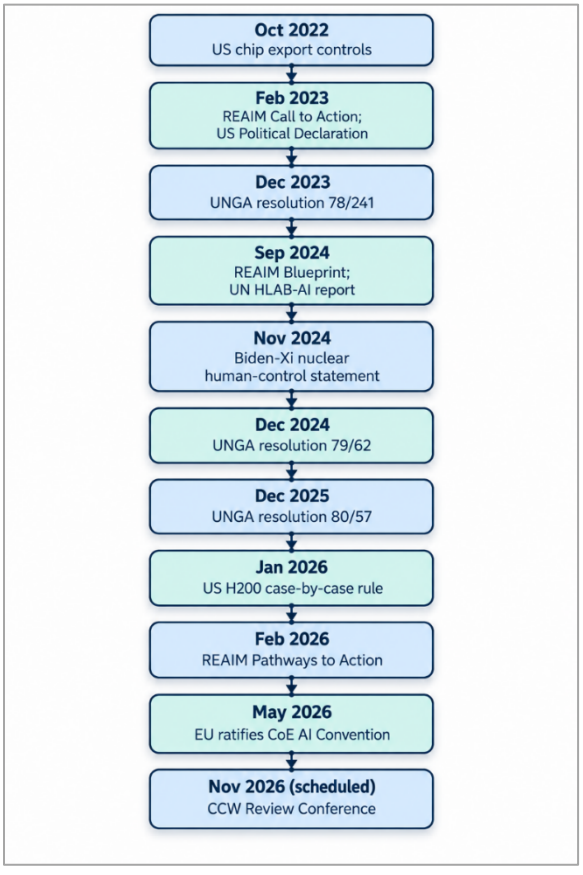
These counterarguments help to clarify, but do not necessarily bring the thesis down. AI is not on its own a cause for war or peace. It is the magnifier of existing political relations, and its security implications will vary according to the will and choices of states and firms in its development, deployment and governance.

Policy Recommendations

The recommendations are based on the conditional premise set out in Section 4.3 that they should be based on target narrow, observable behaviours, on the basis of shared exposure to accidents and on wide distribution of benefits. Recommendations need to be based on existing approaches, which is why the legal status of existing approaches is summarised in Table 2 and the instruments are listed in sequence in Figure 5 as of 24 September 2026.

Figure 5

Selected Milestones in the Governance of AI and International securIty, October 2022–November 2026



Note: Compiled from BIS (2022), U.S. Department of State (2023), UNGA (2023, 2025), HLAB-AI (2024), The White House (2024), Perrin (2025), Egan and Sanders (2026), Harjani and Liu (2026), Council of Europe (2024), and European Union (2026).

Table 2

Governance Approaches, Responsible Actors, Implementation Barriers and Limitations (Status as of 24 September 2026)

Governance approach	Legal status	Responsible actors	Implementation barriers	Limitations
International humanitarian law, including legal review of new weapons	Binding treaty and customary law	States; armed forces; legal advisers	Divergent interpretation; opacity of machine-learning systems complicates review	Does not specify required level of human control

Governance approach	Legal status	Responsible actors	Implementation barriers	Limitations
Council of Europe Framework Convention on AI (CETS 225)	Binding treaty for parties; ratified by the EU in May 2026	Parties; Conference of the Parties	Limited ratifications; national implementation	Excludes national defence matters
EU Artificial Intelligence Act (Regulation 2024/1689)	Binding EU law	EU institutions; member states; providers	Phased implementation; enforcement capacity	Excludes military, defence and national security uses
National export controls on AI chips	Binding domestic regulation	BIS and allied authorities	Smuggling; policy volatility; retaliation	Unilateral; contested legitimacy; limited inclusiveness
UNGA resolutions on autonomous weapons (78/241, 79/62, 80/57)	Non-binding	UN member states; Secretary-General	Opposition of key military powers	Recommendatory; 2025 text did not call for negotiations
CCW elements of an instrument on autonomous weapons	Under consideration; nature not yet decided	CCW states parties; GGE chair	Consensus rule; divergence on prohibitions	Outcome depends on Review Conference, 16–20 November 2026
Political Declaration on Responsible Military Use of AI and Autonomy (2023)	Voluntary political commitment	Endorsing states (47 by November 2023)	Continued engagement uncertain after 2025	No verification; China not a participant
REAIM Call to Action, Blueprint and Pathways to Action	Voluntary	Endorsing states; middle-power hosts	Declining endorsement; neither superpower endorsed 2026 text	Non-binding; no follow-up mechanism
US–China leaders' statement on human control of nuclear use decisions (2024)	Political statement	US and Chinese governments	No implementing measures reported	Bilateral; not institutionalised multilaterally
UN Scientific Panel on AI and Global Dialogue on AI Governance	Established by UNGA resolution 79/325; not regulatory	UN member states; panel experts; stakeholders	Funding; great-power commitment	Informational and deliberative only

Recommendations

1. Propose a measure to institutionalise human control over decisions on nuclear use. Building on the US–China statement in 2024, all nuclear-armed states should make unilateral declarations that decisions to use nuclear weapons will be under human control. These declarations could then be brought together in the form of a joint statement in the existing nuclear dialogues. Obstacles: mistrust that these statements are free and provide information about command systems. Trade-off: Verified commitment is harder to achieve than a declaratory norm, but a declaratory norm is a focal point for reputation.

2. Propose common testing and assessment guidelines for military AI (based on voluntary commitments). Defence ministries need to develop national legal reviews of new weapons to take into account machine-learning systems, such as testing in adverse situations, testing after software updates, and training the systems to reduce automation bias. Analysis procedures can be shared among states without revealing capabilities. Actors with responsibility: defence ministries, legal advisers and standards bodies. Obstacle: classification and competitionality and secrecy.
3. Put in place military AI incident reporting and confidence-building measures (proposed measure). States should establish communication pathways to communicate about accidental or unintended AI interactions and incidents and agree on procedures to be followed in the event of an incident involving AI systems, as suggested by Horowitz et al. (2020). A voluntary anonymous database of incidents could be managed by a third party. Transparency may show vulnerabilities; report from low-sensitivity categories.
4. Adopt a two-tiered instrument on autonomous weapons (proposed measure). Nov Review Conf: CCW states parties should use this Review Conference to initiate negotiations on the basis of the elements already developed: prohibitions on systems that can't be used in accordance with IHL or without context-appropriate human judgement and control, and regulation of others. Where there is no consensus, States that support negotiations should consider a General Assembly mandated process. Trade-off: There may be a process outside the CCW that would result in a text that is more robust but would potentially be more exclusionary of the major military powers.
5. Do more verification research (proposed measure). The UN Scientific Panel, national research organizations and independent research institutes should conduct research on observable proxies of AI capabilities (e.g., large-scale computing clusters and declared testing practices) and ways of verifying claims of AI model behaviour. Challenge: dual-use of the information involved.
6. Make inclusion material not only procedural (proposed measure). Governance mechanisms should be connected with tangible capacity development such as access to defensive cyber security tools, joint compute for public interest research and training for regulators as recommended by the HLAB-AI (2024). Such offers can be coordinated in a suitable forum, such as the UN Global Dialogue. Stakeholders: donor countries, multilateral development banks, UN agencies and companies. Trade-off: capacity-building can be accompanied by dual-use capabilities, which need safeguards.
7. Regulate and involve private security firms (proposed measure). States should ensure that providers of frontier models report any major misuse by state actors to national cyber security bodies; and provide for independent verification of any threat reports by companies. Obstacle: there may be resistance from companies to the idea of disclosure, and there are jurisdictional limits for cross-border cooperation.

Conclusion

This study questioned the current role of artificial intelligence in changing the international security and how international institutions will handle the new risks arising from artificial intelligence. It has two parts to its answer. First, AI is driving security change not as a one-shot wonder, but in a combination of three intertwined ways. In a world of decision compression, whether it's a drone terminal guidance system or nuclear early warning, there's less time and thinking room for human decision. Frontier AI supply chains put the chips, compute, data, and expertise required for frontier AI in the hands of a handful of states and companies, coercive supply chains, and the rest of states are dependent consumers. Information erosion makes it more difficult to establish facts, attribute actions, and verify capabilities; thus deterrence signalling and public trust and arms control are compromised.

Second, institutions can manage these risks, but not in all circumstances. Where capabilities can be used for multiple purposes, are important to relative power, and are difficult to monitor, like semiconductors and wartime drone development, the security dilemma prevails and competition wins out. However, liberal institutionalism can be helpful to account for the progress that has been made where commitments are narrow, benefits are focal, and institutions provide information and where all the major powers are exposed to catastrophic risk. The human-control norm for nuclear decision-making and the trajectory of the evolution of the CCW towards accountability and the applicability of IHL are examples of this. The recent de-escalation of involvement of the major AI actors on initiatives and the lack of defence in the most advanced binding instruments are examples of its limits.

The paper offers an integrated framework of the mapping between different categories of AI and specific security mechanisms, a structured comparison of three cases and recommendations based on the conditions for cooperation. It has the limitations of a desk based non-systematic review, including provisional wartime evidence, and self-reported corporate data, and a cut-off date before the November 2026 CCW Review Conference.

Four areas should be the focus of future research: empirical research into the actual implementation of AI decision support into military procedures; comparative analysis of Chinese, Russian and Global South perspectives on AI and security; ways to verify AI-related commitments; and systematic, independent datasets on AI-enabled cyber and influence operations without relying on vendor disclosures. The key takeaway lesson is that there is no determinate future for international security that depends on AI. Increases the stakes of political decision-making that is currently being undertaken by the states, companies and international institutions.

References

Anthropic. (2025). *Disrupting the first reported AI-orchestrated cyber espionage campaign* [Threat intelligence report]. https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf

Bode, I., & Huelss, H. (2018). Autonomous weapons systems and changing norms in international relations. *Review of International Studies, 44*(3), 393–413. https://doi.org/10.1017/S0260210517000614

Bondar, K. (2025). *Ukraine's future vision and current capabilities for waging AI-enabled autonomous warfare*. Center for Strategic and International Studies. https://www.csis.org/analysis/ukraines-future-vision-and-current-capabilities-waging-ai-enabled-autonomous-warfare

Bondar, K. (2026). *How Russia is building a sovereign drone ecosystem for AI-driven autonomy*. Center for Strategic and International Studies. https://www.csis.org/topics/ukraine-war

Boulanin, V., Saalman, L., Topychkanov, P., Su, F., & Peldán Carlsson, M. (2020). *Artificial intelligence, strategic stability and nuclear risk*. Stockholm International Peace Research Institute. https://www.sipri.org/publications/2020/other-publications/artificial-intelligence-strategic-stability-and-nuclear-risk

Brazil, Belgium, Ireland, Norway, & Switzerland. (2025). *Update by the Chair of the Group of Governmental Experts related to emerging technologies in the area of lethal autonomous weapons systems: Working paper on behalf of 42 High Contracting Parties* (CCW/MSP/2025/WP.5). https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Meeting_of_High_Contracting_Parties_(2025)/CCW-MSP-2025-WP.5.pdf

Brazil. (2025). *Working paper submitted by Brazil* (CCW/MSP/2025/WP.11). Meeting of the High Contracting Parties to the Convention on Certain Conventional Weapons. https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Meeting_of_High_Contracting_Parties_(2025)/CCW-MSP-2025-WP.11.pdf

Buchanan, B. (2016). *The cybersecurity dilemma: Hacking, trust and fear between nations*. Oxford University Press.

Bureau of Industry and Security. (2022). Implementation of additional export controls: Certain advanced computing and semiconductor manufacturing items; supercomputer and semiconductor end use; entity list modification. *Federal Register, 87*(197), 62186–62215. https://www.federalregister.gov/d/2022-21658

Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J

Council of Europe. (2024). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225). https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Ding, J., & Dafoe, A. (2023). Engines of power: Electricity, AI, and general-purpose military transformations. *European Journal of International Security, 8*(3), 377–394. https://doi.org/10.1017/eis.2023.1

Egan, J., & Sanders, J. (2026, January 16). *CNAS insights: Unpacking the H200 export policy*. Center for a New American Security. https://www.cnas.org/publications/commentary/cnas-insights-unpacking-the-h200-export-policy

European Union. (2026, September 7). *Convention on Certain Conventional Weapons (CCW): EU statement, Preparatory Committee of the Seventh Review Conference*. European External Action Service. https://www.eeas.europa.eu/delegations/un-geneva/convention-certain-conventional-weapons-ccw-eu-statement-preparatory-committee-seventh-review_en

Farrell, H., & Newman, A. L. (2019). Weaponized interdependence: How global economic networks shape state coercion. *International Security, 44*(1), 42–79. https://doi.org/10.1162/isec_a_00351

Garcia, D. (2018). Lethal artificial intelligence and change: The future of international peace and security. *International Studies Review, 20*(2), 334–341. https://doi.org/10.1093/isr/viy029

Goldfarb, A., & Lindsay, J. R. (2022). Prediction and judgment: Why artificial intelligence increases the importance of humans in war. *International Security, 46*(3), 7–50. https://doi.org/10.1162/isec_a_00425

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). *Generative language models and automated influence operations: Emerging threats and potential mitigations* (arXiv:2301.04246). arXiv. https://doi.org/10.48550/arXiv.2301.04246

Harjani, M., & Liu, M. C. (2026, February 26). *The uncertain future of the REAIM summit* (IDSS Paper IP26029). S. Rajaratnam School of International Studies. https://rsis.edu.sg/rsis-publication/idss/ip26029-the-uncertain-future-of-the-reaim-summit/

Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power. *Texas National Security Review, 1*(3), 36–57. https://doi.org/10.15781/T2639KP49

Horowitz, M. C., & Kahn, L. (2024). Bending the automation bias curve: A study of human and AI-based decision making in national security contexts. *International Studies Quarterly, 68*(2), Article sqae020. https://doi.org/10.1093/isq/sqae020

Horowitz, M. C., Kahn, L., & Mahoney, C. (2020). The future of military applications of artificial intelligence: A role for confidence-building measures? *Orbis, 64*(4), 528–543. https://doi.org/10.1016/j.orbis.2020.08.003

Jensen, B. M., Whyte, C., & Cuomo, S. (2020). Algorithms at war: The promise, peril, and limits of artificial intelligence. *International Studies Review, 22*(3), 526–550. https://doi.org/10.1093/isr/viz025

Jervis, R. (1978). Cooperation under the security dilemma. *World Politics, 30*(2), 167–214. https://doi.org/10.2307/2009958

Johnson, J. (2019). Artificial intelligence & future warfare: Implications for international security. *Defense & Security Analysis, 35*(2), 147–169. https://doi.org/10.1080/14751798.2019.1600800

Johnson, J. (2020). Delegating strategic decision-making to machines: Dr. Strangelove Redux? *Journal of Strategic Studies, 45*(3), 439–477. https://doi.org/10.1080/01402390.2020.1759038

Keohane, R. O. (1984). *After hegemony: Cooperation and discord in the world political economy*. Princeton University Press.

Keohane, R. O., & Martin, L. L. (1995). The promise of institutionalist theory. *International Security, 20*(1), 39–51. https://doi.org/10.2307/2539214

Liu, M. C. (2025, December 3). *International regulation of lethal autonomous weapon systems: A realistic shift in diplomatic efforts* (IDSS Paper IP25112). S. Rajaratnam School of International Studies. https://rsis.edu.sg/rsis-publication/idss/ip25112-international-regulation-of-lethal-autonomous-weapon-systems-a-realistic-shift-in-diplomatic-efforts/

Maas, M. M. (2019). How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons. *Contemporary Security Policy, 40*(3), 285–311. https://doi.org/10.1080/13523260.2019.1576464

Miller, C. (2022). *Chip war: The fight for the world's most critical technology*. Scribner.

Ministry of Foreign Affairs, European Union and Cooperation of Spain. (2026). *A Coruña Declaration: "REAIM 2026 Pathways to Action"* [Endorsement list]. https://exteriores.gob.es/en/REAIM2026/Paginas/default.aspx

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Payne, K. (2018). Artificial intelligence: A revolution in strategic affairs? *Survival, 60*(5), 7–32. https://doi.org/10.1080/00396338.2018.1518374

Perrin, B. (2025, January 24). Lethal autonomous weapons systems & international law: Growing momentum towards a new international treaty. *ASIL Insights, 29*(1). American Society of International Law. https://www.asil.org/insights

Pilz, K. F., Sanders, J., Rahman, R., & Heim, L. (2025). *Trends in AI supercomputers* (arXiv:2504.16026). arXiv. https://doi.org/10.48550/arXiv.2504.16026

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). *Official Journal of the European Union*, L 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Rivera, J.-P., Mukobi, G., Reuel, A., Lamparth, M., Smith, C., & Schneider, J. (2024). Escalation risks from language models in military and diplomatic decision-making. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 836–898). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658942

Rosert, E., & Sauer, F. (2021). How (not) to stop the killer robots: A comparative analysis of humanitarian disarmament campaign strategies. *Contemporary Security Policy, 42*(1), 4–29. https://doi.org/10.1080/13523260.2020.1771508

Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. W. W. Norton.

Schneider, J., Schechter, B., & Shaffer, R. (2023). Hacking nuclear stability: Wargaming technology, uncertainty, and escalation. *International Organization, 77*(3), 633–667. https://doi.org/10.1017/S0020818323000115

Stockwell, S. (2024). *AI-enabled influence operations: Safeguarding future elections* (CETaS Research Report). Centre for Emerging Technology and Security, The Alan Turing Institute. https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-safeguarding-future-elections

The White House. (2024, November 16). *Readout of President Joe Biden's meeting with President Xi Jinping of the People's Republic of China*. The American Presidency Project. https://www.presidency.ucsb.edu/node/375219

U.S. Department of State. (2023). *Political declaration on responsible military use of artificial intelligence and autonomy*. https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/

United Nations General Assembly. (2023). *Lethal autonomous weapons systems* (A/RES/78/241). https://undocs.org/A/RES/78/241

United Nations General Assembly. (2025). *Lethal autonomous weapons systems* (A/RES/80/57). https://digitallibrary.un.org/record/4095989

United Nations High-level Advisory Body on Artificial Intelligence. (2024). *Governing AI for humanity: Final report*. United Nations. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

United Nations Office for Disarmament Affairs. (2025). *Briefing by the Chair of the CCW GGE on LAWS on the margins of the First Committee*. https://disarmament.unoda.org/en/updates/briefing-chair-ccw-gge-laws-margins-firstcommittee