ARMS CONTROL
NEGOTIATION ACADEMY

Negotiation Pathways and Proposed Language for Governing the AI–NC3 Nexus

by Alice Saltini, Alina Hrytsenko, Fei Su

May 2026

Note: *The views presented in this memorandum are those of the authors and do not represent the views of the organizations or institutions with which they are affiliated.*

Policy Memo

TO: The UN

FROM: Alice SALTINI[1], Alina HRYTSENKO[1], Fei SU[1]

DATE: May, 2026

SUBJECT: Negotiation Pathways and Proposed Language for Governing the AI–NC3 Nexus.

---

1. Introduction

Artificial intelligence (AI) is spreading across every sector of society and the economy, and the military domain is no exception. Discussions about integrating AI into nuclear decision-making systems (including nuclear command, control, and communications (NC3)) are increasingly reaching senior diplomatic and political levels. This growing attention is reflected not only in national policy statements, but also in bilateral and minilateral declarations among nuclear-armed states, as well as in deliberations and resolutions at the United Nations and other multilateral forums.

As engagement on the AI–nuclear nexus expands at the multilateral level, identifying areas of shared understanding among states regarding the implications of AI for nuclear risk and strategic stability has become increasingly important. This is particularly relevant given the potential for AI-enabled systems to shape decision-making processes, perceptions, and escalation dynamics in ways that remain only partially understood.

The first-ever UN General Assembly resolution on this issue, *“Possible risks of the integration of artificial intelligence into command, control and communications systems of nuclear weapons,”* (Resolution 80/23) adopted on 1 December 2025, represents an initial effort to translate a fast-evolving technical debate into diplomatic language.[2] The resolution marks an important normative step by acknowledging that the risks associated with AI integration extend beyond the preservation of human authority over nuclear decisions. At the same time, the resolution’s voting pattern highlights the extent to which states diverge in interpretations of what their commitments should be in practice as AI becomes more deeply embedded in the nuclear domain.[3]

---

[1] This memo was written as a requirement of the 2025-2026 Arms Control Negotiation Academy program. Views expressed in the memo are the opinions of the authors and do not reflect the views of their employers.

[2] https://docs.un.org/en/A/RES/80/23

[3] https://digitallibrary.un.org/record/4094665?ln=en

Building on lessons learned from ongoing efforts to advance AI–nuclear discussions at the multilateral level,[4] this paper seeks to identify areas of convergence across states and to propose negotiation-ready language on AI's implications for the nuclear weapons domain that states can use in multilateral and bilateral settings. Its objective is to facilitate near-term, politically feasible commitments that reduce nuclear-related risks from AI while preserving space for national security requirements and further technical work.

This paper applies the Harvard Negotiation Method to structure practical pathways toward agreement. Section 2 surveys the current debate on AI-nuclear integration, framing the context and preparing the ground for the analysis that follows. Section 3 maps key stakeholder positions among nuclear-armed states and non-nuclear-armed states (including nuclear-hosting and Allied States, the Global South and industry). Section 4 identifies BATNAs (Best Alternatives to a Negotiated Agreement) and ZOPAs (Zones of Possible Agreement) and assesses ongoing efforts to build common ground and generate options for mutual gain. Section 5 presents adoptable text suitable for (1) outcome documents in nonproliferation, disarmament, and risk-reduction forums (e.g., UNGA First Committee and, potentially, NPT processes) and in broader AI-military governance venues (e.g., the Summits on Responsible AI in the Military Domain (REAIM)); and (2) unilateral, bilateral, or multilateral political declarations. Section 6 evaluates the practical feasibility of the proposed language and identifies potential avenues through which it could be advanced.

2. AI-Nuclear Interface

2.1 AI-NC3 (potential) use cases: present and future

NC3 systems vary significantly across nuclear-armed states, as each architecture is tailored to specific doctrines and postures. Broadly, however, NC3 denotes the integrated system that enables national leadership to receive early warning, conduct assessment and decision-making, direct force operations, and maintain assured, secure communications for the employment (or termination) of nuclear operations.[5]

Within NC3, and in adjacent intelligence and communications systems, different AI model classes already fill or could fill distinct functions. Although it is difficult to assess which specific models decision-makers are currently considering, and public information on algorithms already fielded is limited (and mostly available for the US case only), earlier-generation machine-learning tools have long been embedded for data fusion and situational awareness in air- and missile-warning architectures.[6] For example,

---

[4] https://thebulletin.org/2025/12/lessons-from-the-uns-first-resolution-on-ai-in-nuclear-command-and-control/?utm_source=ActiveCampaign&utm_medium=email&utm_content=Trump%20administration%20seeks%20to%20extinguish%20a%20beacon%20of%20climate%20science&utm_campaign=20251218%20Thursday%20Newsletter%20%28Copy%29

[5] https://www.congress.gov/crs-product/IF11697

[6] https://media.nti.org/documents/NTI_Paper_AI_r4.pdf

multi-sensor fusion systems combine data from military and civilian sensors to generate a common operational picture and assist operators in identifying anomalous tracks more rapidly. Computer-vision tools are also used to detect changes in satellite imagery and monitor activity patterns relevant to nuclear force developments. In addition, analysts and policy studies further suggest that predictive-analytics approaches may help identify indicators of adversary force movements or escalation dynamics.[7]

Contemporary AI capabilities could extend these functions. Modern models are increasingly able to process high-volume streaming data, multimodal integration across time and space, automated anomaly detection, and structured analytical summarization.[8] These advances represent a qualitative shift in scale and adaptability compared to earlier decision-support tools. Although nuclear-armed states are likely examining such applications, available evidence suggests many remain exploratory or experimental. At the same time, experts caution that today's most advanced AI systems have limitations that make them risky to use in nuclear settings. These models can produce confident but incorrect outputs ("hallucinations"), and because they operate as black boxes, it is often unclear why they reach certain conclusions. They can also be vulnerable to cyberattacks or manipulated inputs, and we cannot yet reliably ensure that they will consistently act in ways that match human intentions, goals, and values. These limitations make them ill-suited to high-stakes nuclear contexts absent strict constraints, verification, and governance. Whether, and under what conditions, states are (or should) integrate today's cutting-edge AI remains a subject of debate.[9]

2.2 Reasons for and implications of integration

Despite these concerns, states are exploring the integration of AI in nuclear-relevant decision chains because of the strategic advantages it promises: greater speed, efficiency, resilience, and lower manpower burdens. The pressure to gain (or maintain) advantage comes at the backdrop of three factors: (1) rapid technological evolution that continually expands potential applications and reach; (2) the need for a faster decision-making command, given the limited time available to decision-makers during a crisis, and hence the need for systems that support faster and better-informed decisions; and

---

7 https://securityandtechnology.org/wp-content/uploads/2025/09/Artificial-Intelligence-in-Nuclear-Command-Control-Communications.pdf

8 https://securityandtechnology.org/virtual-library/report/nuclear-command-control-and-communications-nc3-a-primer-on-strategic-warning-decision-support-and-adaptive-targeting-subsystems/

9 https://europeanleadershipnetwork.org/policy-brief/assessing-the-implications-of-integrating-ai-in-nuclear-decision-making-systems/ ; https://securityandtechnology.org/virtual-library/report/ai-nc3-integration-in-an-adversarial-context-strategic-stability-risks-and-confidence-building-measures/

(3) a competitive environment of high-tempo operations and fast diffusion of dual-use technologies, where the perceived cost of not adopting AI is ceding advantage to adversaries.[10]

Several nuclear-armed states, including China, France, Pakistan, the United Kingdom and the United States, have affirmed that ultimate nuclear-use decisions remain in human hands. Although such assurances represent an initial effort to articulate normative red lines, they are unlikely to be sufficient in the absence of upstream guardrails on how AI outputs are generated, validated, presented, and gated before reaching senior decision-makers. Moreover, states have not converged on operational definitions of "AI" or "human control". High-level assurances alone therefore do not address the technical, operational, and strategic risks that arise when AI interacts with NC3 systems, increasing the potential for misperception, miscalculation, and unintended escalation.[11]

These risks extend well beyond technical failure modes. Automation bias may lead human operators to over-rely on AI-generated recommendations. In crisis scenarios, this bias could lead to hasty or misinformed decisions, particularly under conditions of high stress or compressed timelines. Shortened decision windows increase the risk that leaders act on partial or ambiguous information, potentially escalating conflicts unintentionally.[12] Operational complexity is further heightened by the entanglement of conventional and nuclear platforms. Modern militaries increasingly integrate AI into conventional systems for logistics, surveillance, and targeting. These systems can generate signals (such as missile launches, troop movements, or high-frequency sensor activity) that automated detection systems might misinterpret and hence shape nuclear decisions.[13]

Finally, AI can alter strategic stability dynamics by intensifying competitive pressures among nuclear-armed states. Nations may feel compelled to adopt AI-enhanced nuclear systems to maintain strategic parity or gain perceived advantages, potentially fueling an AI-driven arms race. When one state pursues these capabilities, others may fear falling behind, even if the actual performance of AI systems is uncertain. Strategic risks are then compounded by the difficulty of multilateral verification and transparency. Unlike conventional arms systems, whose capabilities can often be observed through

---

[10] https://www.sipri.org/sites/default/files/2025-06/2025_6_ai_and_nuclear_risk.pdf ; https://www.war.gov/News/Releases/Release/Article/4376420/war-department-launches-ai-acceleration-strategy-to-secure-american-military-ai/ ; https://www.gov.uk/government/publications/the-strategic-defence-review-2025-making-britain-safer-secure-at-home-strong-abroad ; https://www.defense.gouv.fr/sites/default/files/aid/Report%20of%20the%20AI%20Task%20Force%20September%202019.pdf ; https://www.iar-gwu.org/print-archive/blog-post-title-four-xgtap ;

[11] https://www.sipri.org/publications/2025/other-publications/pragmatic-approaches-governance-artificial-intelligence-nuclear-nexus ; https://warontherocks.com/2024/12/beyond-human-in-the-loop-managing-ai-risks-in-nuclear-command-and-control/

[12] https://fas.org/wp-content/uploads/2025/07/June2025_AIxNC3_FAS.pdf

[13] https://www.armscontrol.org/act/2025-09/features/artificial-intelligence-and-nuclear-command-and-control-its-even-more

physical deployments or testing, AI systems are largely software-based, can be updated rapidly, and may operate in ways that are not externally visible. This makes it difficult for other states to assess capabilities or intentions accurately, particularly during crises.[14]

3. Stakeholders Map

3.1 Nuclear-armed states

Nuclear-armed states have a paramount interest in maintaining credible deterrence, and most are reportedly exploring AI to enhance decision support, early warning, and other nuclear-relevant military functions. Advances in AI are often associated with expectations of faster information processing, improved situational awareness, and enhanced operational resilience.[15]

Several nuclear-armed states have publicly affirmed that ultimate authority over nuclear-use decisions should remain in human hands. Such assurances suggest a shared recognition that fully automated nuclear decision-making would be highly destabilizing. However, these commitments remain politically voluntary and are not accompanied by common operational definitions of key concepts such as "human control," nor by verification mechanisms.

In practice, the positions of nuclear-armed states reveal a need to balance caution with competitive pressures, alongside a growing emphasis on the accelerated adoption of AI within military apparatuses.[16] At the same time, there is no P5-wide political declaration that jointly commits states to retaining human control over nuclear decision-making. Some progress has instead occurred through other channels, including bilateral statements and national or joint submissions such as the working paper submitted by France, the United Kingdom, and the United States to the 10th NPT Review Conference in 2022, as well as the 2024 REAIM Blueprint for Action, which was endorsed by several nuclear-armed states, including France, Pakistan, the United Kingdom, and the United States.[17]

---

[14] https://www.cnas.org/publications/reports/averting-ai-armageddon?utm_medium=email&utm_campaign=AI%20Nexus%20Press%20Release%20Feb%2013%202025&utm_content=AI%20Nexus%20Press%20Release%20Feb%2013%202025+Preview+CID_cb30482774269923ed5b2f65b7cbe537&utm_source=Campaign%20Monitor&utm_term=Averting%20AI%20Armageddon%20US-China-Russia%20Rivalry%20at%20the%20Nexus%20of%20Nuclear%20Weapons%20and%20Artificial%20Intelligence

[15] https://www.sipri.org/yearbook/2025/06

[16] https://media.defense.gov/2026/Jan/12/2003855671/-1/-1/0/ARTIFICIAL-INTELLIGENCE-STRATEGY-FOR-THE-DEPARTMENT-OF-WAR.PDF ; https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2025/06/25/summary-of-natos-rapid-adoption-action-plan

[17] https://docs.un.org/en/NPT/CONF.2020/WP.70 ; https://www.mofa.go.kr/www/brd/m_4080/down.do?brd_id=235&seq=375378&data_tp=A&file_seq=9

Moreover, China and the United States have jointly endorsed the principle that humans should retain control over decisions regarding the use of nuclear weapons.[18] However, differences in formulation are noteworthy. The joint France-UK-US submission and the REAIM Blueprint for Action refer to maintaining "human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment," whereas the China-US statement refers more narrowly to "the need to maintain human control over the decision to use nuclear weapons." While the former formulation appears to encompass a broader set of decision-support and operational processes surrounding nuclear use, the latter focuses more specifically on the final launch decision itself. These differences in language suggest that nuclear-armed states may have not yet converged on how far commitments on human control over nuclear use should extend across the wider decision chain.

Russia, by contrast, has not issued an equivalent formal commitment, but it remains engaged on the question of what constitutes "meaningful human control". In its explanation of vote on the UN General Assembly Resolution 80/23, Russia argued that the absence of agreed definitions of "AI" and "human control" impedes meaningful exchange, and that NC3 discussions should be limited to the P5 given sensitivities.[19]

3.2 Non-Nuclear Armed Allies

Non-nuclear-armed states that rely on the security guarantees of nuclear-armed allies, or that share similar strategic outlooks, often approach the role of AI in the nuclear domain from positions shaped by dependence and vulnerability. Heightened geopolitical tensions in recent years have reinforced the perceived importance of extended deterrence arrangements. For example, NATO allies in Europe, as well as U.S. partners such as Japan and the Republic of Korea, continue to rely on U.S. nuclear protection as a central component of their security strategies. Officials in Japan and South Korea have underscored the importance of credible and stable allied decision-making in crisis situations, including through alliance consultations on emerging technologies and strategic modernization.[20]

These states may see potential advantages in AI applications that could enhance the effectiveness and survivability of allied nuclear forces, for example through improvements in early warning and situational awareness capabilities. At the same time, they are also well aware of the possibility that AI integration could introduce new failure modes or reduce human deliberation in crises that could directly affect their security. European allies have similarly emphasized that technologies which accelerate decision-making must not undermine political control or crisis stability.[21]

---

[18] https://china.usembassy-china.org.cn/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-3/

[19] https://reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com25/eov/C1_Russia.pdf

[20] https://platform.opennuclear.org/thoughtroom/external-contributions/the-credibility-challenge-us-extended-deterrence-in-the-indo-pacific

[21] https://www.nato.int/en/what-we-do/deterrence-and-defence/emerging-and-disruptive-technologies

Accordingly, many allied countries have supported the development of international norms and voluntary frameworks promoting the responsible military use of AI. Numerous U.S. allies and partners have endorsed the 2023 Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy,[22] as well as outcome documents from the 2023, 2024 and 2026 REAIM Summits.[23] These initiatives promote best practices, including aimed at maintaining meaningful human involvement in military decision-making more broadly and improve transparency and communication channels.[24] Their endorsement reflects a broad consensus that uncontrolled AI in warfare would pose unacceptable risks, especially for states that could be caught in the middle of a great-power nuclear miscalculation.

At the same time, most U.S. allies chose to abstain in the vote on the UN General Assembly Resolution 80/23.[25] Although national positions were not always publicly elaborated, the overall pattern indicates a degree of caution among allied states regarding the scope, pace, and practical implications of emerging multilateral discussions on this issue.

3.3 Other Non-Nuclear Armed States

This group of "other non-nuclear armed states" is not monolithic. This diverse cohort of countries generally views the advent of AI in nuclear military systems with concern, rooted in both normative commitments and security considerations. A number of these states have historically supported nuclear disarmament initiatives and advocated for stronger controls on emerging weapons technologies. Concerns are often expressed that advances in military AI could exacerbate global power asymmetries, accelerate arms competition, and marginalize the security interests of states that do not possess nuclear capabilities.[26]

Within this category are states that play active roles in humanitarian arms-control debates. In multilateral forums, many have promoted precautionary approaches to military AI, emphasizing the importance of human involvement in critical decisions, enhanced transparency, and, in some cases, legally binding constraints on autonomous weapons systems. These positions are frequently framed in terms of humanitarian risks, governance gaps, and concerns about technological inequality or a widening "digital divide".

At the same time, positions within this group are not uniform. Some states emphasize their own security requirements and have been cautious about endorsing sweeping prohibitions.[27] Despite such differences, a common thread is the call for greater transparency from nuclear-armed states regarding

[22] https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/

[23] The 2026 REAIM outcome document *"Pathways to Action"* was supported by most U.S. allies, although the United States itself did not endorse the declaration.

[24] https://www.war.gov/News/News-Stories/Article/Article/3597093/us-endorses-responsible-ai-measures-for-global-militaries/

[25] https://digitallibrary.un.org/record/4094665?ln=en

[26] https://www.armscontrol.org/act/2024-11/features/global-south-and-autonomous-weapons-controls

[27] https://www.armscontrol.org/act/2024-11/features/global-south-and-autonomous-weapons-controls

AI-related developments, stronger assurances that meaningful human involvement will be maintained across nuclear decision processes, and recognition of the potential escalation risks associated with AI integration. Many states in this category were also among the proponents and supporters of Resolution 80/23. More broadly, they tend to frame global stability as a collective concern and shared responsibility, and advocate inclusive dialogue to ensure that advances in military AI do not deepen existing inequalities or become an unregulated domain of great-power competition.[28]

3.4 Industry

Industry actors play an increasingly important role in the military-AI-nuclear nexus, operating at the nexus of technological innovation, commercial incentives, and national security priorities. From major defense contractors to AI startups, firms are developing tools that can support intelligence analysis, surveillance, targeting, and aspects of command-and-control decision support, often driven by expanding dual-use markets and growing government demand.

However, the participation of industry in governance discussions on military AI presents challenges. Private firms are not traditional actors in arms control diplomacy and frequently operate outside the institutional and normative frameworks familiar to states. At the same time, commercial priorities and competitive pressures may not always align with sustained engagement in multilateral governance processes, particularly where the benefits of participation are less immediate or clearly defined. Companies often promote adoption on the basis of faster threat detection, streamlined logistics, and more effective missile defense. Many firms also recognize that failures or unintended consequences in nuclear-relevant contexts could carry severe risks.

Engagement in governance processes can therefore serve multiple purposes for industry actors, including shaping emerging standards, demonstrating responsible innovation, and maintaining access to government procurement opportunities. As a result, leading firms tend to balance efforts to advance technological development with assurances to governments regarding products' safety and reliability by. emphasizing responsible-AI principles and human-in-the-loop safeguards. At the same time, they have often shown a preference for voluntary guidelines and public-private cooperation rather than binding regulatory constraints that could affect innovation cycles or competitiveness.[29]

4. BATNAs and ZOPAs

Drawing on the mapped interests and positions, as well as the risks associated with AI integration discussed above, it is possible to assess each key stakeholder's Best Alternative to a Negotiated Agreement (BATNA) and Zones of Possible Agreement (ZOPAs) to ground the proposed language.

[28] https://www.stopkillerrobots.org/news/156-states-support-unga-resolution ; https://www.armscontrol.org/act/2024-11/features/global-south-and-autonomous-weapons-controls ; https://automatedresearch.org/news/state_position/south-africa
[29] https://www.csis.org/analysis/artificial-intelligence-and-national-security-importance-ai-ecosystem

4.1 BATNAs

For nuclear-armed states, the most credible alternative to negotiated arrangements lies in continuing the unilateral or nationally driven integration of AI into military and nuclear-related capabilities. Perceived strategic advantages associated with AI may reduce incentives to accept externally defined constraints.[30] Therefore, in the absence of negotiated frameworks, states may instead rely on unilateral confidence-building measures (CBMs) aimed at improving the safety, security, and reliability of their own systems, as well as bilateral or crisis-management mechanisms intended to reduce misperception. In some regional contexts, such as South Asia, such measures could include targeted communication channels or transparency initiatives focused on emerging technologies.

For non-nuclear-armed states, the principal BATNA involves sustained participation in multilateral processes aimed at shaping the normative and political environment surrounding AI and nuclear risk. Although their influence on operational policies is often indirect, many seek to contribute through coalition-building, advocacy for transparency and risk-reduction measures, and support for political commitments developed in international disarmament and emerging-technology forums. These efforts are intended to raise awareness of the potential risks associated with the AI–nuclear nexus and to increase the political pressure on nuclear-armed states for responsible adoptions.

For nuclear-hosting and nuclear-allied states benefiting from extended deterrence arrangements, the BATNA is shaped by the constraints of alliance politics. Their security dependence on nuclear-armed partners limits their ability to pursue independent policy initiatives that could affect allied nuclear posture. In the absence of broader agreements, these states may prioritise internal alliance consultations, transparency mechanisms, and CBMs designed to reduce escalation risks and improve crisis communication related to AI-enabled systems.

For industry actors, BATNAs is shaped by the tension between commercial incentives to advance AI innovation and growing political scrutiny in strategically sensitive domains. In the absence of negotiated governance frameworks, firms are likely to continue shaping norms through market practices, technical standards, and engagement with national regulatory processes. Some companies may also seek to influence emerging governance approaches through participation in multistakeholder initiatives or industry coalitions. At the same time, a fragmented or slow-moving regulatory environment may be viewed as an acceptable outcome for certain actors, particularly where it preserves flexibility for technological development and market competition.

30

https://usun.usmission.gov/remarks-at-the-security-councils-open-debate-on-artificial-intelligence-and-international-peace-and-security/#:~:text=Two%20months%20ago%2C%20President%20Trump,standards%20of%20our%20commercial%20sector.

4.2 ZOPAs

Despite divergent incentives, there is meaningful scope for convergence on effects-based, non-intrusive measures that improve assurance without exposing sensitive systems. A soft-law framework is particularly suited to this objective, as non-binding political commitments can enable states to acknowledge shared risks, develop practical CBMs, and foster incremental cooperation without imposing intrusive verification or formal legal obligations that could deter participation in such a sensitive domain. Over time, such measures may also help lay the groundwork for more formal and legally binding arrangements.

ZOPAs can therefore be organized along two mutually reinforcing tracks that proceed in parallel: a nuclear-armed-states track (Track A) and a broader mixed track that includes both nuclear-armed and non-nuclear-armed states (Track B). Because stakeholders hold fundamentally different incentives and levels of control over nuclear systems, any realistic negotiation approach must distinguish between: (1) issues that only nuclear-armed states can address operationally, and (2) areas where all states and stakeholders can contribute to reducing AI-related nuclear risks. Nuclear-armed states have unique operational responsibilities and sensitivities surrounding NC3, while non-nuclear-armed states and industry actors have stronger roles in shaping principles, expectations, and pressures for responsible adoption. A two-track structure reflects these different roles and creates space for progress on multiple fronts.

Track A envisions a confidential, effects-focused format among nuclear-armed states for sustained dialogue on the AI–nuclear nexus and its implications for strategic stability. The objective is to examine how AI may alter warning timelines, assessment confidence, error profiles, and escalation pathways without disclosing classified information or operational data. States would meet under a protected modality, share methodologies rather than capabilities, and confine exchanges to testing approaches, validation practices, and acceptable confidence thresholds. If geopolitical conditions and political will allow, states could go further by affirming that AI will not serve as the sole source for launch assessment or release decisions and commit to practical steps, such as deferring or abstaining on low-confidence or integrity-at-risk outputs and requiring independent corroboration before treating alerts as nuclear-relevant. In parallel, this dialogue could also cover the development of joint pilot projects exploring positive applications of AI (such as imagery change detection, structured anomaly triage, or carefully governed natural-language processing of open-source reporting) subject to independent verification and validation.

Track B focuses on broad but implementable commitments in a mixed setting. States acknowledge that, even with humans in the loop and final authority retained, the integration of AI may nonetheless have implications for decision-making. Such adoption has the potential to compress decision time, increase signaling ambiguity, and raise the risk of inadvertent escalation in ways that remain insufficiently understood and therefore warrant further study and dialogue particularly among nuclear-armed states. At the same time, states can also recognize that, when developed and deployed responsibly and within appropriate guardrails, AI integration may contribute to strategic stability and disarmament verification.

Pilot projects could explore AI's potential to reduce escalation risks, beginning with low-stakes applications that still make a meaningful contribution. These may include improving the consistency of information used in diplomatic exchanges, identifying unusual patterns in open-source data on missile tests or exercises, or supporting shared analysis of past crises to clarify escalation pathways and recurrent points of misunderstanding, before moving on to higher-stakes areas.[31] Industry here participates as an implementing partner by providing method-level transparency and joining standards work to harmonize assurance practices, allowing states to draw on private-sector capability while reinforcing a shared floor for responsible use.

This two-track approach, together with a shared understanding of both the risks and the potential opportunities as described above, aims to reduce AI-related nuclear risks through incremental and politically feasible steps that promote shared learning and confidence while avoiding intrusive verification requirements that could discourage engagement. These initial areas of convergence could provide a basis for agreed language suitable for inclusion in political declarations, voluntary codes of conduct, or outcome documents in relevant diplomatic forums, while also helping to create conditions conducive to more formal cooperation in the future.

5. Proposed language for a soft law approach

Building on the two-track framework outlined above, this section proposes a set of practical principles intended to reduce AI-related nuclear risks while remaining politically and operationally feasible. Three principles are directed at nuclear-armed states (reflecting Track A), while a further three are designed for broader adoption by all states and relevant stakeholders (reflecting Track B).

These formulations are conceived as flexible, negotiation-ready language that can be adapted to different diplomatic settings and levels of ambition. Depending on the forum and political context, selected elements could inform bilateral or minilateral political declarations among nuclear-armed states, contribute to outcome documents in processes such as the REAIM Summits, or be reflected in multilateral instruments, including future UN General Assembly resolutions or NPT discussions.

The proposed principles are intended as a starting point for incremental norm development and practical risk-reduction efforts. Their purpose is not to prescribe uniform solutions, but to offer a common basis for dialogue, confidence-building, and the gradual articulation of shared expectations regarding responsible integration.

5.1 Three principles for nuclear-armed states

31 https://thebulletin.org/2025/12/lessons-from-the-uns-first-resolution-on-ai-in-nuclear-command-and-control/?utm_source=ActiveCampaign&utm_medium=email&utm_content=Trump%20administration%20seeks%20to%20extinguish%20a%20beacon%20of%20climate%20science&utm_campaign=20251218%20Thursday%20Newsletter%20%28Copy%29

a. Nuclear armed-states recognize that, even where humans retain full authority over launch decisions, the integration of AI systems can materially affect decision-making processes. In particular, such systems (especially if integrated without appropriate safeguards, validation, and sufficient understanding of system performance) may compress decision time, increase ambiguity in signaling and perception, and raise the risk of inadvertent escalation in ways that remain insufficiently understood. These potential effects are viewed as serious and warrant sustained scrutiny, further study, and prudent risk-mitigation efforts. At the same time, nuclear-armed states acknowledge that AI may also present potential opportunities to support strategic stability and therefore underline the importance of its responsible and prudent exploration. In that respect, nuclear-armed states reiterate that, given the grave consequences associated with any potential use of nuclear weapons, avoiding miscalculation and inadvertent escalation remains a matter of overriding priority. Consequently, they commit to taking appropriate measures to deepen their understanding of AI-related risks and, as conditions permit, to share relevant validation approaches, testing methodologies, and best practices with other nuclear-armed states.
b. In recognizing these risks, nuclear-armed states affirm that AI systems shall not serve as the sole basis for launch assessment or release decisions. Decisions should not rely on AI-generated outputs that lack sufficient confidence or integrity. In that respect, they commit to taking appropriate measures to ensure the safety, security, and robustness of AI systems used in nuclear-related systems, including through rigorous testing, validation, and internal oversight practices.
c. Nuclear-armed states should consider how best to cooperate to exchange information on AI-related military incidents. States may consider establishing or expanding existing hotlines or develop other new measures in this respect.

5.2 Three principles for all states and key stakeholders

a. States recognize that, even when humans retain final authority, the integration of AI systems (especially if integrated without appropriate safeguards, validation, and sufficient understanding of system performance) can compress decision time, increase ambiguity in signaling and perception, and heighten the risk of inadvertent escalation. These potential effects are viewed as serious and warrant sustained scrutiny, further study, and prudent risk-mitigation efforts. States therefore commit to supporting further study, dialogue, and transparency measures that improve shared understanding of these risks, including through multilateral discussions and expert exchange with industry. At the same time, recognizing that AI, when used within appropriate guardrails, can enhance transparency and strengthen disarmament verification, states commit to exploring its positive applications, particularly those that have the potential to reduce nuclear risks.
b. States commit to take appropriate measures to establish national measures to enhance the safety, security, and robustness of AI systems used in nuclear-related systems, including through rigorous testing, validation, and internal oversight practices.

c. States should encourage and support the participation of industry actors and civil society in discussion at international standards bodies and technical working groups to promote harmonized approaches to assurance, robustness and responsible innovation in AI systems relevant to strategic stability.

6. Assessment of language implementability

The political feasibility of the proposed soft-law language must be evaluated in a strategic environment characterized by intensified geopolitical rivalry, weakening arms-control architectures, and declining levels of trust among nuclear-armed states. These conditions place particularly strong constraints on initiatives that touch directly or indirectly on NC3, which remain among the most sensitive components of national security policy.

Divergent strategic incentives surrounding military AI further complicate prospects for agreement. While many states acknowledge the potential risks associated with AI integration in nuclear-relevant contexts, they also perceive competitive pressures to accelerate technological adoption and preserve operational flexibility. In this environment, legally binding rules or prescriptive constraints are unlikely to gain traction in the near term. By contrast, non-binding political commitments offer a lower-cost pathway for engagement by enabling states to signal responsible intent and build shared understandings of emerging risks.

Verification challenges reinforce the appeal of such an approach. Commitments related to AI are inherently difficult to assess externally, particularly where they intersect with classified nuclear systems. Measures implying intrusive inspections, algorithmic disclosure, or detailed transparency regarding operational architectures are therefore likely to encounter resistance. An effects-based and procedural framing focusing on risk mitigation practices rather than capabilities provides a more viable basis for cooperation by allowing states to address shared concerns while preserving necessary secrecy.

Recent diplomatic developments illustrate both the opportunities and the limits of multilateral engagement. The adoption of the UN General Assembly resolution 80/23 demonstrated growing international concern and helped establish the issue as a legitimate subject of global discussion. At the same time, divergent voting behaviour highlighted sensitivities related to definitions, transparency expectations, and the potential implications for national deterrence postures.[32] These dynamics underscore the value of a differentiated engagement strategy. A two-track structure combining confidential dialogue among nuclear-armed states with broader, inclusive processes focused on risk awareness offers a pragmatic way to accommodate varying levels of political comfort and operational responsibility.

---

32 https://thebulletin.org/2025/12/lessons-from-the-uns-first-resolution-on-ai-in-nuclear-command-and-control/?utm_source=ActiveCampaign&utm_medium=email&utm_content=Trump%20administration%20seeks%20to%20extinguish%20a%20beacon%20of%20climate%20science&utm_campaign=20251218%20Thursday%20Newsletter%20%28Copy%29

Forum selection will also shape the prospects for implementation. Nuclear-armed-state engagement is likely to be most productive in formats that allow for protected exchanges, such as dedicated strategic-stability dialogues, P5-style consultations, or targeted bilateral and minilateral initiatives. In parallel, more inclusive venues, including UN General Assembly processes, REAIM-type initiatives, and other emerging-technology governance forums, can help socialize risk-reduction principles, encourage transparency practices, and build wider normative support. The NPT process could also offer an additional avenue for engagement, although discussion of AI-related issues in this forum is likely to remain politically sensitive given broader challenges facing the Treaty. Carefully framed language that focuses on how AI may affect the implementation of the NPT's three pillars (namely nonproliferation, disarmament, and peaceful uses) rather than on operational aspects of NC3 systems could nevertheless provide a constructive basis for dialogue. Such an approach may help create shared stakes in the discussion and support confidence-building without exacerbating existing tensions. The modular nature of the proposed language is intended to facilitate cross-forum diffusion, enabling states to adopt elements incrementally according to political readiness and institutional context.

Overall, while geopolitical competition and verification constraints limit the scope for ambitious regulatory outcomes, they do not preclude incremental progress. Soft-law approaches grounded in shared risk perceptions, flexible implementation pathways, and non-intrusive confidence-building measures provide a politically realistic means of reducing AI-related nuclear risks. By fostering gradual convergence and creating opportunities for practical cooperation, such measures can help stabilize expectations in the near term while laying the groundwork for more formal arrangements as trust and technical understanding evolve.

Conclusion

AI is already beginning to reshape the informational and temporal environment in which nuclear decisions are prepared and executed. Even where states affirm that authority over nuclear use remains firmly in human hands, the growing integration of AI into NC3-adjacent functions can still compress decision timelines, increase signaling ambiguity, and introduce new pathways for technical failure, misinterpretation, or manipulation. The implications are therefore not only technological but strategic, with greater potential for misperception and unintended escalation in crises.

In the current geopolitical climate characterized by intensified rivalry, weakened arms-control structures, and high sensitivity surrounding NC3 systems, ambitious regulatory approaches or intrusive transparency measures are unlikely to gain traction. This paper argues instead for a pragmatic pathway centred on incremental, effects-based cooperation through modular and non-binding commitments. Building on areas of overlap identified across stakeholders, it proposes a two-track approach that combines confidential dialogue among nuclear-armed states with broader, inclusive efforts to develop shared norms and CBMs.

Such an approach offers a politically realistic means of moving beyond general assurances of "humans in the loop" towards more substantive engagement with the implications of AI integration. While modest in scope, these steps can help stabilize expectations, foster shared learning, and create conditions for deeper cooperation over time. In this sense, the governance of the AI–nuclear nexus is likely to evolve not through sweeping agreements, but through gradual convergence on practical risk-reduction measures.

Annex

Stakeholder mapping

Stakeholder	Core Interests	Stated / Operational Positions
Nuclear-Armed States	• Maintain credible deterrence • Preserve survivability of NC3 systems • Avoid accidental/unauthorized nuclear use • Preserve strategic advantage • Modernize arsenals with emerging tech	• Commit to human control over launch decisions • Support voluntary norms; resist binding rules • Western states promote oversight and transparency • Pursue AI for intel, early warning, C2
Non-Nuclear Armed Allies	• Ensure reliability of extended deterrence • Reduce escalation risks • Maintain human authority in crises • Strengthen alliance cohesion	• Endorse voluntary norms • Advocate human-in-the-loop controls • Push nuclear allies for consultation • Support confidence-building and transparency
Other Non-Nuclear Armed States	• Prevent destabilizing AI arms races • Ensure inclusive global governance • Preference for legally-binding measures • Strengthen disarmament norms	• Advocate bans/limits on autonomous weapons • Emphasize humanitarian principles • Demand transparency from nuclear states • Push for UN-centered governance
Industry (Defense + Tech)	• Commercialize military/dual-use AI • Maintain market leadership • Avoid catastrophic AI failures • Shape regulation and standards	• Promote testing/certification over bans • Prefer flexible, innovation-friendly regulation • Increasing openness to military AI use

About the Authors:

Alina Hrytsenko is an expert in international relations and security and currently serves as a political analyst at the NATO Representation to Ukraine. She previously worked as a senior consultant at the National Institute for Strategic Studies under the President of Ukraine and has additional experience in civil service and working freelance in the fields of foreign policy and security. She is a 2024 Think Visegrad Fellow at PISM, researching Russia's strategic culture, and a graduate of the George C. Marshall European Center for Security Studies. Alina holds a PhD in Political Science two master's degrees in National Security Strategic Management and in International Relations. She is a Visiting Lecturer at the Diplomatic Academy of Ukraine. She regularly publishes in media outlets, including the Atlantic Council.

Alice Saltini is a Senior Research Lead at the Rhode Island School of Design (RISD). She also serves as a Non-Resident Expert on AI at the James Martin Center for Nonproliferation Studies (CNS) and a Research Fellow for the Centro Interdisciplinare Scienze per la Pace at the University of Pisa. She specializes in the impact of AI on nuclear decision-making. She advises governments and international organizations on managing AI-related nuclear risks, with a focus on mitigating the challenges of integrating AI into military systems and nuclear weapons, translating complex technical concepts into actionable policy solutions. She has published extensively on military applications of AI and has developed a comprehensive risk assessment framework for analyzing risks at the AI–nuclear nexus. Previously, she worked with the Institute for Security and Technology, the European Leadership Network, the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), and CNS. She holds a Master's degree in Russian Studies and a Postgraduate Certificate (PgCert) in Nonproliferation Studies from the Middlebury Institute of International Studies.

Fei Su is a Researcher with SIPRI's China and Asia Security Programme. Prior to joining SIPRI, she lived and studied in Seoul for three years, where she strengthened her fluency in Korean. She holds an MA in Public Administration from the Graduate School of Public Administration at Seoul National University, focusing on governance. Her research interests focus on regional security issues in East Asia with a special interest in North Korea, China's foreign and security policy, maritime security, cybersecurity and AI-nuclear nexus.