



ARMS CONTROL NEGOTIATION ACADEMY

AI in nuclear command and control:

**Prospects for an international agreement to keep a
human in the loop**

by Tynchtykbek Israilov, Ekaterina Mikhalevich, Haleema

Saadia, Jonas Sandbrink

March, 2025

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: Governments of France, the United Kingdom, and United States
FROM: Tynchtykbek Israilov, Ekaterina Mikhalevich, Haleema Saadia,
Jonas Sandbrink
DATE: March 27, 2025
SUBJECT: AI in nuclear command and control: prospects for an international agreement to keep a human in the loop

Introduction

The integration of artificial intelligence and automation into Nuclear Command, Control and Communication (NC3) systems could have important implications for nuclear strategic stability and escalation risks - especially if this leads to “human-out-of-the-loop” systems.¹ In this policy brief we explore the stances of different nuclear weapons states on AI-powered automation of nuclear systems and analyze the potential for arms control mechanisms for achieving a universal commitment to keep humans “in-the-loop”.

When referring to artificial intelligence (AI), we mean machine learning systems of various possible architectures that are trained on relevant data. Such AI systems could be integrated in different parts of NC3. First, AI trained to integrate complex data streams into a functional understanding of the real world could be used in early warning and situational awareness systems. Second, AI systems trained on various simulations of wargames and escalatory dynamics could be used to suggest or directly make a decision to launch. Third, AI could be used to automatically execute and target launches with reduced human involvement, including through autonomous deployment systems such as unmanned underwater vehicles (UUV). When discussing keeping “a human in the loop”, we mainly refer to critical launch decision and deployment protocols, and less to early warning and situational awareness.

Benefits of AI Integration

The AI integration with nuclear deployment and launch through automation can potentially reduce reaction time, enabling faster decision-making in the face of increasing combat speed.² AI integration can assist human decisionmakers by processing information and suggesting a range of well-calibrated

¹ Nuclear Command and Control (NC2) refers to the systems, procedures, and processes that enable a nation's leaders to exercise control over its nuclear weapons arsenal. This includes the command structure responsible for decision-making regarding the potential deployment, authorization, and use of nuclear weapons. Nuclear Command and Control and Communications (NC3) extends the concept of NC2 by incorporating the communication systems and networks that facilitate the transmission of information, orders, and instructions within the NC2 apparatus. This policy paper uses NC2 and NC3 interchangeably largely for brevity and for their interconnectedness as NC2 systems inherently rely on communication networks to function effectively.

² Fei Su and Jingdong Yuan, “Chinese Thinking on AI Integration and Interaction with Nuclear Command and Control, Force Structure, and Decision-Making,” November 13, 2023, <https://www.europeanleadershipnetwork.org/report/chinese-thinking-on-ai-integration-and-interaction-with-nuclear-command-and-control-force-structure-and-decision-making/>.

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

responses, providing them with more time to evaluate their options.³⁴ When AI is used to take humans out of the loop, it can lead to decisions at superhuman speed, potentially strengthening second strike capabilities and enhancing strategic stability. Strategic stability is achieved when the likelihood of nuclear conflict or escalation is reduced, and nuclear-armed states are confident that using nuclear weapons would not serve their strategic interests.⁵ This involves a careful balance of military capabilities, intentions, and perceptions that promote peace and prevent nuclear war.⁶ However, as we will explore in the risks section, this strategic stability may be misleading.

Using AI to automate processes can help reduce human errors, as AI systems are capable of making quick and objective decisions based on predefined strategic and ethical criteria.⁷ Compared to pre-AI automation of launches, AI systems tailored for nuclear command and control have the potential to surpass humans in decision-making and execution. When properly integrated, AI systems can efficiently analyze information and make decisions without the limitations of human information processing abilities or unpredictable emotional factors.⁸

Lastly, AI systems for nuclear command and control can potentially outperform human decision makers in identifying optimal responses, especially at extremely compressed timescales. AI systems can surpass human cognitive limitations at information processing and analysis, allowing more optimal response selection and execution.⁹ These benefits can also be leveraged when humans are in the loop as AI systems can present a tailored and broad range of retaliation or deployment options to decision makers.

Risks of AI Integration

The potential for an accidental launch due to AI integration in nuclear command and control is a significant concern. This could occur from limitations of AI systems, such as hallucinations and algorithmic biases.¹⁰ Such errors could also arise in AI systems deployed in nuclear command and control, for instance systems specialized to process and interpret a wide range of data from various sensors. Algorithmic biases can also cause miscalculations and errors due to biased training data or suboptimal design choices.¹¹ The potential for hallucinations and other errors increases significantly in the absence of a human in the loop, which serves as an additional layer of safety.

One factor that may drive errors in AI systems for nuclear command and control is a lack of good training data.¹² Real world training data for nuclear weapons deployment and relevant retaliation

³⁴ Parasuraman, Raja, et al. Complacency and bias in human use of automation: an attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 3, 2010, p. 381-410. <https://doi.org/10.1177/0018720810376055>

⁴ Huang, Y., Xue, X., & Jiang, C. (2020). Semantic integration of sensor knowledge on artificial internet of things. *Wireless Communications and Mobile Computing*, 2020, 1-8. <https://doi.org/10.1155/2020/8815001>.

⁵ Vincent Boulanin et al., "The Positive and Negative Impacts of AI on Strategic Stability and Nuclear Risk," *Artificial Intelligence, Strategic Stability and Nuclear Risk* (Stockholm International Peace Research Institute, 2020), <https://www.jstor.org/stable/resrep25355.10>.

⁶ Alexey Arbatov, "Nuclear Deterrence: A Guarantee for or Threat to Strategic Stability?," in *NL ARMS Netherlands Annual Review of Military Studies 2020: Deterrence in the 21st Century—Insights from Theory and Practice*, ed. Frans Osinga and Tim Sweijts, NL ARMS (The Hague: T.M.C. Asser Press, 2021), 65–86, https://doi.org/10.1007/978-94-6265-419-8_5.

⁷ Lowther, A. and McGiffin, C. (2019). America Needs a "Dead Hand". <https://warontherocks.com/2019/08/america-needs-a-dead-hand/>

⁸ Ivanov, S. (2022). Automated decision-making. *Foresight*, 25(1), 4-19. <https://doi.org/10.1108/fs-09-2021-0183>

⁹ Hendrycks, D. and Mazeika, M. (2022). X-risk analysis for AI research. <https://doi.org/10.48550/arxiv.2206.05862>

¹⁰ Mitrou, L., Janssen, M., & Loukis, E. N. (2021). Human control and discretion in ai-driven decision-making in government. 14th International Conference on Theory and Practice of Electronic Governance. <https://doi.org/10.1145/3494193.3494195>

¹¹ Anthony Aguirre, Emilia Javorsky, and Max Tegmark, "Artificial Escalation: Imagining the Future of Nuclear Risk," *Bulletin of the Atomic Scientists* (blog), July 17, 2023, <https://thebulletin.org/2023/07/artificial-escalation-imagining-the-future-of-nuclear-risk/>.

¹² Anthony Aguirre, Emilia Javorsky, and Max Tegmark, "Artificial Escalation: Imagining the Future of Nuclear Risk," *Bulletin of the Atomic Scientists* (blog), July 17, 2023, <https://thebulletin.org/2023/07/artificial-escalation-imagining-the-future-of-nuclear-risk/>.

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

dynamics are sparse. This means that any AI system trained for nuclear command and control would likely be trained on synthetic training data, such as that generated by computer simulations. This exacerbates the risk of algorithmic biases which could be introduced by either overfitting to a small set of real-world training data or the inherent biases in synthetic training data.

Use of AI systems for nuclear command and control increases potential vulnerabilities to cyberattacks. Cyber threats are an emerging challenge across many domains, but the stakes are uniquely high in the nuclear context.¹³ Adversaries unable to compete in traditional nuclear deterrence may seek asymmetric pathways to disrupt command and control systems through targeted hacking. Using AI in nuclear command and control can lead to additional vulnerabilities. AI systems are prone to malicious attacks called prompt injection, where an input is given in a way that exploits the AI systems properties to lead to an unintended result.¹⁴ For instance, an adversary could manipulate the sensors of an AI system connected to detection systems to mask or fake an incoming attack. This could threaten strategic stability by either causing inadvertent launches or compromising the ability to retaliate. However, keeping a human in the loop can help reduce vulnerabilities, albeit not completely eliminating them.

Another challenge of AI integration is the black box nature of these systems, which makes it difficult to assess the appropriateness of their decisions or recommendations. Unlike non-AI algorithms, AI systems are trained on a variety of data and use this training to interpret data and make decisions. However, these interpretations and decisions occur in a black box manner, and the current explainability of AI system decisions is limited. This can lead to challenges where humans cannot follow the AI's decision-making process but still presume it to be correct, even if it is erroneous.

Lastly, there is a potentially dangerous link between AI integration and Launch on Warning (LoW) posture. LoW nuclear posture is a strategic approach that entails having a country's nuclear forces ready to launch in response to early warning signals of an imminent nuclear attack.¹⁵ Thus, LoW posture depends on fast decision making. AI integration may appear to expedite decision-making by swiftly analyzing and interpreting extensive data and thus appear attractive under LoW posture. However, even where humans are in the loop, under time pressure it may be very difficult for humans to validate an AI's warning of an incoming attack and its decision to launch a preemptive strike. Thus, the combination of LoW posture and AI integration may be particularly fraught with risk.

Country Positions on AI Integration

Nuclear-armed countries generally acknowledge AI's potential benefits in various applications, but they also express shared concerns over its responsible use, particularly in relation to weapon systems, including nuclear weapons. These countries recognize that AI can enhance the efficiency and effectiveness of their weapon systems, but they emphasize the need for caution and responsibility in its implementation. The use of AI in weapon systems, especially those with devastating capabilities like nuclear weapons, demands careful consideration to prevent unintended consequences, ensure accountability, and maintain human control.

¹³ Susan Y. Pickering and Peter B. Davies, "Cyber Security of Nuclear Power Plants: US and Global Perspectives," *Georgetown Journal of International Affairs* (blog), January 22, 2021, <https://gjia.georgetown.edu/2021/01/22/cyber-security-of-nuclear-power-plants-us-and-global-perspectives/>.

¹⁴ Eugene Bagdasaryan et al., "Abusing Images and Sounds for Indirect Instruction Injection in Multi-Modal LLMs" (arXiv, October 3, 2023), <https://doi.org/10.48550/arXiv.2307.10490>.

¹⁵ Appendix 2 gives an overview of the Launch on Warning nuclear posture and includes a comparison of the LoW postures of nuclear weapons possessor states.

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

While some nuclear weapons states such as France, UK and US have proposed an internationally agreed principle to ensure human control over nuclear weapons launch systems,¹⁶ other countries, including Pakistan, India, Israel, Russia, and China, have varying stances on the integration of AI to nuclear capabilities. Pakistan is concerned about the potential destabilization of deterrence due to unregulated AI, while India is actively investing in AI for defense purposes. Israel maintains nuclear ambiguity but heavily utilizes AI in its defense forces. Russia has explored AI in missile systems, while China emphasizes achieving AI superiority in future and is apprehensive about cyber vulnerabilities in nuclear command and control systems. In absence of information about DPRK's exploration of AI, its continued development of its nuclear and missile programs remains a serious concern. It is crucial to engage in diplomatic efforts that include all nuclear-armed states in discussions on AI integration into NC3 systems to prevent escalation and ensure global stability.

The following table provides a visual representation of nuclear power nations' stances on AI integration into their respective nuclear capabilities, reflecting a spectrum of concerns and their stances.

Country	Position
China	China believes that AI and other emerging technologies will transform the nature of future combat, ¹⁷ and so, to achieve victory, the People's Liberation Army must seek superiority in this realm. Chinese analysts view the vulnerability of China's Nuclear Command, Control, and Communication systems to cyber infiltrations even if an attacker's objective was limited to cyber espionage as highly escalatory. ¹⁸ According to official Chinese sources, the CCP emphasizes the importance of people's participation in the decision-making process in weapons systems. ¹⁹
France	France (along with US and UK) have shared a paper at the United Nations warning of "the risks of relying on AI to detect nuclear threats, and using the technology in launch protocols without human involvement". ²⁰ It states "Consistent with long-standing policy, we will maintain human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment". ²¹

¹⁶ "Principles and Responsible Practices for Nuclear Weapon States: Working Paper Submitted by France, the United Kingdom of Great Britain and Northern Ireland and the United States of America" (2020 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, July 29, 2022), https://www.un.org/sites/un2.un.org/files/npt_conf.2020_e_wp.70.pdf.

¹⁷ Elsa B. Kania, "Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Powe" (Washington D.C.: Center for a New American Security, November 2017), <https://s3.amazonaws.com/files.cnas.org/documents/Battlefield-Singularity-November-2017.pdf?mtime=20171129235805>.

¹⁸ Johnson, *AI and the Bomb*.

¹⁹ Ministry of Foreign Affairs of the People's Republic of China, 'Position Paper of the People's Republic of China on Regulating Military Applications of Artificial Intelligence (AI)', 14 Dec. 2021, https://www.fmprc.gov.cn/mfa_eng/wjdt_665385/wjzcs/202112/t20211214_10469512.html;

The State Council of the People's Republic of China, 'New Generation Artificial Intelligence Development Plan', 20 Jul. 2017, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm;

Information Office of the State Council of the People's Republic of China, 'China's Military Strategy (full text)', 27 May 2015, http://english.www.gov.cn/archive/white_paper/2015/05/27/content_281475115610833.htm;

Information Office of the State Council of the People's Republic of China, 'China's National Defense in the New Era', 24 Jul. 2019, https://english.www.gov.cn/archive/whitepaper/201907/24/content_WS5d3941ddc6d08408f502283d.html.

²⁰ David Parsley, "Nuclear Powers Race for Deal to Stop AI Accidentally Launching World War III," *inews.co.uk*, March 24, 2023, <https://inews.co.uk/news/ai-nuclear-powers-deal-stop-artificial-intelligence-launch-world-war-iii-2228654>.

²¹ "Principles and Responsible Practices for Nuclear Weapon States: Working Paper Submitted by France, the United Kingdom of Great Britain and Northern Ireland and the United States of America."

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

India	India's Ministry of Defence (MoD) established a Defence AI Council (DAIC) to provide strategic direction for AI adoption in defense, promote government-industry partnerships, and review technology acquisition recommendations. The ministry emphasized capacity-building within the defense sector, with each Service Headquarter receiving a budget allocation of Rs 100 crores for AI-specific application development, recognizing AI as a 'force multiplier' for defense strategies. ²²
Israel	Israel follows a policy of nuclear ambiguity, neither confirming nor denying the existence of its nuclear arsenal. The policy is rooted in national security concerns in a volatile Middle East region. Israel has actively embraced AI technologies and has a growing AI sector. Israel supports AI research and development. Moreover, Israel Defence Forces have started using AI to select targets for air strikes and logistics. ²³ Nevertheless, PM Netanyahu's central message at the UN GA meeting in September 2023: "Like with any weapon, there is a need to handle the future development of AI with due caution". ²⁴
Pakistan	Pakistan believes that the danger of AI integration with nuclear force posture and employment policies can undermine traditional concepts of deterrence escalation and risk reduction, endangering regional and global security. Threats to peace, security and stability at the regional and global levels resulting from the use of unregulated AI in critical military decision making processes and conventional weapon systems are comparable to any category of WMDs. ²⁵
Russia	Russia has the largest nuclear arsenal globally with 2,400 strategic nuclear weapon and an estimated 1,600 deployed tactical nuclear weapons. Russia's strategic nuclear weapons are tied into Perimeter ²⁶ , an automatic system of signal rockets used to beam radio messages to launch these weapons if other means of communication are knocked out. Russia continues research and testing in this area: for instance, in 2015, a model of an automated logistics management system (including in wartime) was proposed for units of the mobile group of the Strategic Missile Forces armed with ICBMs on mobile ground-based missile systems. In 2018, it became known about the development and testing of a model based on a neural network ²⁷ as a solution for predicting the remaining operating time of elements of missile systems of the strategic missile forces. Along with Russian military's growing interest in AI, Russian researchers explore how auxiliary systems could be applied to improve logistics, security, and other functions of RVSN. As evidenced by remarks of RVSN leadership, some innovations have been introduced into practice, for example mathematical model of

²² Ambuj Sahu, "Artificial Intelligence in Military Operations: Where Does India Stand?," orfonline.org, August 2, 2019, <https://www.orfonline.org/expert-speak/artificial-intelligence-military-operations-where-does-india-stand-54030>.

²³ Marissa Newman, "Israel Quietly Embeds AI Systems in Deadly Military Operations," *Bloomberg.Com*, July 16, 2023, <https://www.bloomberg.com/news/articles/2023-07-16/israel-using-ai-systems-to-plan-deadly-military-operations>.

²⁴ Maayan Jaffe-Hoffman, "Did Benjamin Netanyahu Turn AI into a Nuclear Weapon? - Analysis," *The Jerusalem Post | JPost.com*, September 23, 2023, <https://www.jpost.com/business-and-innovation/tech-and-start-ups/article-760280>.

²⁵ "Addressing the Security and Stability Implications of Military Applications of Artificial Intelligence (AI), and Autonomy in Weapon Systems, Working Paper by Pakistan" (Conference on Disarmament, July 26, 2023), [https://docs-library.unoda.org/Conference_on_Disarmament_-__\(2023\)/CD.2334.pdf](https://docs-library.unoda.org/Conference_on_Disarmament_-__(2023)/CD.2334.pdf).

²⁶ Stilwell, B. (2022). Russia's 'Dead Hand' Is a Soviet-Built Nuclear Doomsday Device.

<https://www.military.com/history/russias-dead-hand-soviet-built-nuclear-doomsday-device.html>

²⁷ Stefanovich, D. (2019). Artificial Intelligence and Nuclear Weapons.

<https://russiancouncil.ru/analytics-and-comments/analytics/iskusstvennyy-intellekt-i-yadernoe-oruzhie/>

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

	an automated logistics management system for RVSN mobile group units armed with ICBMs on mobile ground-based missile systems ²⁸ .
United Kingdom	According to the Defence Artificial Intelligence Strategy 2022, ²⁹ UK “will ensure that – regardless of any use of AI in our strategic systems – human political control of our nuclear weapons is maintained at all times. We strongly encourage other nuclear states to make a similar commitment.” The UK supports AI systems as decision-support tools and its policy documents identify a strategic need for staying ahead of competitors through technological advances in AI within military contexts. ³⁰
United States	According to the 2022 DoD Nuclear posture review, the US would “maintain a human in the loop for all actions critical to informing and executing decisions by the President to initiate and terminate nuclear weapon employment”. ³¹ US Congress is considering legislation to prevent AI being used for launch of nuclear weapons, including that “no federal funds can be used for any launch of any nuclear weapon by an automated system without meaningful human control”. ³²

Structured Analysis of Country Positions

To evaluate the zone of possible agreement (ZOPA) for an arms control measure to retain a “human in the loop” for launch decision making and execution, we perform a structured analysis of country positions. A summary is presented in Table 2, while the full analysis can be found in Appendix 1. The structured analysis of country positions focuses on interests related to AI integration into the NC3, covering strategic considerations, responses to potential first strikes, influence from adversaries, technological prowess showcase, and the role of civil society. Additionally, interests relating to relevant arms control measures are explored, including openness to arms control agreements, verifiability concerns, willingness for unilateral commitments, concession expectations, and preferences for the participation of key actors. This analysis aims to reveal divergences and convergences in country positions, providing the foundation for defining a ZOPA and formulating potential win-win solutions for international cooperation on AI integration in the NC3 systems.

Table 2: Characterizing country positions related to AI integration and relevant arms control measures.

Legend
Strategic Incentive: Indicates the level of strategic incentive for AI integration.
Faster Strike: Indicates the emphasis on the need for a faster nuclear strike.
Adversarial Dynamics: Indicates whether a country faces particular adversarial dynamics/incentives

²⁸ Isaev, Alexander and Filatov, Vladimir, Fedorov, V., Grevtsov, A., ‘Model of automated control system of material support of military units and formations of the Strategic Missile Troops in the development of materiel and technical support system of the Armed Forces of the Russian Federation’, National Priorities of Russia. Series 1: Science and military security, 2015, No. 3 (3), P. 59.

²⁹ United Kingdom, “Defence Artificial Intelligence Strategy, Policy Paper” (Ministry of Defence, June 15, 2022), <https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy>.

³⁰ “The Risks of AI Integration with NC3, Written Evidence” (London, June 8, 2023), <https://committees.parliament.uk/writtenevidence/120293/pdf/>.

³¹ United States, “National Defence Strategy of The United States of America” (Department of Defense, October 27, 2022), <https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF>.

³² Edward John Markey, “Block Nuclear Launch by Autonomous Artificial Intelligence Act of 2023,” Pub. L. No. BUR23348 (n.d.), https://www.markey.senate.gov/imo/media/doc/block_nuclear_launch_by_autonomous_ai_act_-_042623pdf.pdf.

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

to integrate AI, for instance, if a particular adversary might be integrating AI.
Showcasing Technological Prowess: Indicates interest in showcasing technological prowess.
Lack of Moderating Voices: Indicates a lack of influence of moderating voices or civil society on nuclear policies.
Openness to Arms Control: Indicates general willingness to engage in arms control agreements.
Lack of verifiability is not an issue: Indicates whether lack of verifiability of compliance with an agreement around AI integration is not an issue for a particular country.
Amenability for a Unilateral Statement: Indicates the willingness to make a unilateral statement.
Concessions: Indicates the requirement of particular concessions to join an agreement.
Parties - P5 / N9 / NNWS: Indicates support of a country for an agreement on the level of P5, N9, or non-nuclear weapon states.

Country	Interests relating to integration of AI into NC3?	Interests relating to a minimal joint statement on human in the loop?
United States	Medium-High	High
	Strategic incentive: Medium Faster strike: High Adversarial dynamics: High Showcasing Technological prowess: High Lack of moderating voices: Medium	Openness to arms control: High Lack of verifiability is not an issue: Medium Amenability for a unilateral statement: High Concessions: Medium Parties - P5 / N9 / NNWS: P5 or N9 and NNWS
UK	Low	High
	Strategic incentive: Low Faster strike: Low Adversarial dynamics: Low Showcasing Technological prowess: Medium Lack of moderating voices: Low	Openness to arms control: High Lack of verifiability is not an issue: High Amenability for a unilateral statement: High Concessions: No concessions Parties - P5 / N9 / NNWS: P5 or N9 and NNWS
France	Low	High
	Strategic incentive: Low Faster strike: Low Adversarial dynamics: Low Showcasing Technological prowess: Medium Lack of moderating voices: Low	Openness to arms control: High Lack of verifiability is not an issue: High Amenability for a unilateral statement: High Concessions: No concessions Parties - P5 / N9 / NNWS: P5, or NNWS (likely opposed to N9)
Russia	Medium-High	Medium

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

	Strategic incentive: Medium Faster strike: High Adversarial dynamics: Medium Showcasing Technological Prowess: High Lack of moderating voices: Medium-High	Openness to arms control: Medium Lack of verifiability is not an issue: Low Amenability for a unilateral statement: High Concessions: Openness and transparency Parties - P5 / N9 / NNWS: P5
China	Medium-High	Medium
	Strategic incentive: Medium Faster strike: Medium Adversarial dynamics: Medium Showcasing Technological prowess: Medium Lack of moderating voices: High	Openness to arms control: Medium Lack of verifiability is not an issue: Medium Amenability for a unilateral statement: High Concessions: Openness and transparency Parties - P5 / N9 / NNWS: P5 (necessarily) + NNWS
India	Medium-High	Low
	Strategic incentive: Medium Faster strike: Medium-High Adversarial dynamics: High Showcasing Technological prowess: High Lack of moderating voices: Medium-High	Openness to arms control: Low Lack of verifiability is not an issue: Medium Amenability for a unilateral statement: Low Concessions: Technology transfer Parties - P5 / N9 / NNWS: N9 and NNWS
Pakistan	Medium-High	High
	Strategic incentive: High Faster strike: Medium-High Adversarial dynamics: High Showcasing Technological prowess: Medium Lack of moderating voices: High	Openness to arms control: High Lack of verifiability is not an issue: High Amenability for a unilateral statement: Low Concessions: Openness, transparency, and sharing of technology. Parties - P5 / N9 / NNWS: N9 and NNWS
Israel	High	Low
	Strategic incentive: High Faster strike: High Adversarial dynamics: High Showcasing Technological prowess: Low Lack of moderating voices: Medium	Openness to arms control: Low Lack of verifiability is not an issue: Low Amenability for a unilateral statement: Low Concessions: Low Parties - P5 / N9 / NNWS: Unlikely to join any joint statement

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

DPRK	High	Low
	Strategic incentive: High Faster strike: High Adversarial dynamics: High Showcasing Technological prowess: High Lack of moderating voices: High	Openness to arms control: Medium Lack of verifiability is not an issue: High Amenability for a unilateral statement: Low Concessions: Economic aid, recognition of nuclear status Parties - P5 / N9 / NNWS: N9, and NNWS

Considering the BATNA

The alternative to a potential agreement on limiting the AI integration into NC3 systems are individual statements made by individual countries or groups of allies. For instance, the US and UK have already made statements on maintaining a human in the loop for critical actions.³³ Such unilateral statements are a critical and potentially quite effective step in creating awareness of the risks and defusing perceived arms race dynamics related to AI integration into NC3 systems. However, if not reciprocated, then such unilateral statements still leave the possibility for adversarial dynamics leading to preemptive AI integration. Thus, unilateral statements will be less effective than a joint agreement to build trust between states.

Identifying the ZOPA

Our analysis of country positions informs the ZOPA for limiting AI integration into NC3. Generally, given current geopolitical tensions, it is very difficult to achieve any form of arms control agreement, especially any more formal or legally binding agreement. Another important and specific challenge for an agreement to limit AI integration into NC3 is the difficulty of verifying compliance. It is impossible to tell from the outside whether a human is in the loop regarding nuclear decision-making and deployment. This means that trust between potential adversaries is critical, but it also means that a formal agreement and legally binding agreement is less feasible. Therefore, a joint statement or declaration as a more loose and non-legally binding document might be the most feasible arms control tool. Such a statement could build trust around not removing humans from the loop among nuclear weapons states, and it would reduce the pressure for AI integration motivated by the fear of adversaries doing so. This would potentially present a win-win solution, as it enables nuclear weapons states to focus their resources on other areas and mitigates risks relating to malfunctioning AI and reducing human control.

Who to get to the table

One crucial question is what group of countries should agree on such a statement. There are significant risks of India and Pakistan integrating AI into NC3, thus it is not sufficient to only get P5 nuclear states on the table. One option would be to focus on small pairings, such as India and Pakistan making a joint statement. However, in this scenario India would be unlikely to make such a statement without China also making it; and China might refrain from making such a statement without the US agreeing to the same. Thus, it is important to capture nuclear weapons states

³³ United States, "National Defence Strategy of The United States of America"; United Kingdom, "Defence Artificial Intelligence Strategy, Policy Paper:"

including and beyond the P5.

One option to do so that might be acceptable to the P5 is by opening the statement to a larger set of nations, including nations without nuclear weapons. As a precedent serves the Nuclear Security Summit Series, which aimed at addressing the threat of nuclear terrorism and enhance nuclear security worldwide, included states with civilian nuclear technology.³⁴ In the case of an agreement on AI integration into NC3, it could make sense to not only include states with nuclear weapons but also states that develop advanced AI systems and could thus export such systems to other countries for integration into NC3 systems. Such countries include Canada, Germany, Japan, the Republic of Korea, and the United Arab Emirates.

Another question is who would be best placed to host the process of such an agreement. Given current tensions between Russia and the UK, France, the US, and other European countries, it is likely best if none of these countries would host this process. One option might be China; another option might be Pakistan. It is unclear how India's potential for participation would be affected by either option. The most promising option would be for a non-Western non-nuclear weapons state to host this process, such as the United Arab Emirates. Another option would be for the Declaration to be concluded under the auspices of the UN.

It may be difficult to get Israel to the table, given its general rejection of arms control agreements and continued denial of possessing nuclear weapons. However, a declaration that involves both nuclear weapons states and AI developing nations would preserve the ambiguity around Israel's nuclear weapons possession and thus serve as the best shot to get Israel to the table.

Content of a possible statement

The content of a possible declaration and joint statement should likely be simple and minimalist. AI integration is in many ways attractive to different nuclear weapons states. AI can help integrate data from many different sources, and can thus help decision-makers to understand the strategic situation. It is infeasible to assume that no AI-touched information will inform nuclear decision-making. Furthermore, Russia's Perimeter system cannot be captured by the statement. After leadership activates the system, if sensors indicate signs of nuclear explosions and communication with the General Staff has been disrupted, then launch authority is transferred to soldiers in a protected bunker. Thus, it is still a human that makes the final launch decision. A possible statement's content needs to focus on what is generally uncontroversial: having decision-making and execution of launches rest in a meaningful way with humans. Additionally, it should establish joint definitions and a shared understanding of risks.

Concretely, the joint statement could include the following points:

1. States have a shared interest in avoiding catastrophic nuclear war and escalation risks from integration of artificial intelligence.

³⁴ "FACT SHEET: The Nuclear Security Summits: Securing the World from Nuclear Terrorism," [whitehouse.gov](https://obamawhitehouse.archives.gov/the-press-office/2016/03/29/fact-sheet-nuclear-security-summits-securing-world-nuclear-terrorism), March 29, 2016, <https://obamawhitehouse.archives.gov/the-press-office/2016/03/29/fact-sheet-nuclear-security-summits-securing-world-nuclear-terrorism>.

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

2. States shall affirm the importance of maintaining human control over nuclear command and control systems and emphasize that ultimate decision-making authority should remain with human decision-makers.
3. States express willingness to engage in transparency measures and confidence-building measures to enhance trust and reduce the risk of misperception or miscalculations regarding AI integration in nuclear operations.
4. States that possess nuclear weapons acknowledge the interest and concerns of non-nuclear states regarding AI integration in nuclear operations and commit to inclusive dialogue and engagement to address shared security challenges and promote global stability.
5. States endorse a temporary moratorium on the deployment of AI technologies in critical nuclear command and control functions.
6. The joint statement should contain shared definitions for meaningful human oversight and AI integration with nuclear command and control.
7. States would establish three working groups on
 - Development of vocabulary for human oversight and AI integration with nuclear command and control.
 - Transparency and confidence building measures regarding AI integration in nuclear operations.
 - Tangible measures, including proposals for arms control agreements regarding AI integration in nuclear operations.

Conclusion

International recognition associated with the careless integration of AI into NC3 systems, and removing human oversight, is critical for mitigating risks. We conclude that in the context of geopolitical tensions, the best workable arms control agreement option would be a Declaration or Joint Statement as a means of intent or guiding principles that express the objectives of the parties involved without creating binding legal commitments.

We believe such a Declaration would be most effective if a large set of nations signed it. We consider that not only the P5 countries should conclude a Declaration, but also other states that have nuclear weapons (India, Pakistan), as well as potentially possessing nuclear weapons (Israel) and states that do not have nuclear weapons, but are actively conducting scientific research in AI sphere (Canada, Germany, Japan, Republic of Korea and United Arab Emirates).

In our opinion, the Declaration should contain points that are simple and obvious to all actors, such as:

- Universal acknowledgement of the potential/unknown dangers of AI integration into NC3;

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

- Having decision-making and execution over nuclear launches rest with humans without fail.

We recommend a non-nuclear party that is palatable to as many nuclear weapons states as possible to host the process, for instance, the United Arab Emirates.

AI in military affairs as a means of decision support is a reality of today, which reflects the promise of AI for scientific and technological progress in general. To avoid mistakes that affect the perception of threat, nuclear force modernization and nuclear doctrines must be an ongoing topic of bilateral and multilateral discussions and consultations.

Appendix 1: ZOPA Assessment

We asked for each of the different countries the following questions and rated these along low-high:

a. Interests relating to integration of AI into NC3?

1. Does AI integration/reducing the amount of humans in the loop strengthen strategic position?
2. Incentive to respond faster to a first strike of an adversary/serving as a force multiplier? (especially applicable to countries with less secure retaliation)
3. Interest in integrating if an adversary is potentially integrating, e.g. because of potential strategic advantage that the adversary would have from integration?
4. Interest in showcasing technological prowess?
5. Lack of influence of civil society voices that advocate for moderation regarding the integration of AI into NC3?

b. Interests relating to a minimal joint statement on keeping a human in the loop?

1. General openness to international agreements relating to arms control?
2. Lack of verifiability is not an issue to this country?
3. How happy to make a unilateral commitment to keep a human in the loop?
4. What concessions would it want to see from other countries as part of an agreement?
5. If there was an agreement, who should be on the table? P5, all nuclear weapons states, also non-nuclear weapons states

Country	Interests relating to integration of AI into NC3?	Interests relating a minimal joint statement on human in the loop?
China	1. Medium - judging by the agreement with Russia (2010) and based on the statements of top officials and official sources, such developments are underway, but control remains with human decision makers³⁵. 2. Medium - threat from the Korean Peninsula, conflict in the South China Sea, economic confrontation with the United States as a threat to China's national security. 3. Medium - Not a direct dynamic to pull even with US or Russia. However,	1. Medium - agree to sign a multilateral agreement with nuclear powers. The main principle is general openness (for instance, agreement between Russia and China on notifications of launches of ballistic missiles and space launch vehicles, 2010). 2. Medium - verifiability as one of the main principles of the potential agreement. At the same time, there is a need to hide their achievements in this direction. 3. High - for instance, agreement between Russia and China on notifications of launches of ballistic missiles and space

³⁵ Ministry of Foreign Affairs of the People's Republic of China, 'Position Paper of the People's Republic of China on Regulating Military Applications of Artificial Intelligence (AI)', 14 Dec. 2021, https://www.fmprc.gov.cn/mfa_eng/wjdt_665385/wjzcs/202112/t20211214_10469512.html; The State Council of the People's Republic of China, 'New Generation Artificial Intelligence Development Plan', 20 Jul. 2017, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm; Information Office of the State Council of the People's Republic of China, 'China's Military Strategy (full text)', 27 May 2015, http://english.www.gov.cn/archive/white_paper/2015/05/27/content_281475115610833.htm.

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

	general interest in having cutting-edge technological weapons systems. 4. Medium - while China generally interested in showcasing sophisticated AI capabilities³⁶, military-relevant developments are classified and information space is closed. 5. High - High lack of influence of civil society, in particular on security topics like relating to nuclear.	launch vehicles, 2010. Reuters news report that US and China are not yet ready to make joint statement.³⁷ 4. It would ask for more openness and transparency. 5. P5 (necessarily) + non-nuclear states.
France	1. Low - France's strategic position won't be significantly changed by taking human out of the loop for faster nuclear retaliation abilities. This could change if nuclear submarines become less effective deterrent. 2. Low - France has nuclear submarines and thus a small but potent retaliatory capacity. This would not be significantly changed by faster response times; for deterrence against adversaries with large arsenals, France would nevertheless likely rely on the US. France's alliance with the US means that traditional small state dynamics related to other countries having larger arsenals affect it less. No launch on warning policy 3. Low - no direct adversary that the France would feel compelled to have to keep up with. 4. Medium - France is likely generally interested in showcasing its technological prowess, however there are many ways to do this much more effectively than around automation of NC3. 5. Low - influential civil society	1. High - general openness to arms control; general alignment with UK and US; general hesitation for arms control agreement with Russia 2. High - lack of verifiability is not that great of an issue as no strong strategic dynamics with particular adversaries. 3. High - happy to make unilateral agreement (has already happened together with US and UK in soft form)³⁸ 4. No particular concessions, beyond willingness to cooperate 5. P5, or otherwise larger group of states including non-nuclear states (likely opposed to N9)

³⁶ Experts believe China is pursuing even more sophisticated technical developments called Multi-Domain Precision Warfare (MDPW). According to the Pentagon's 2022 China Military Power Report, China's People's Liberation Army is being trained and equipped to use artificial intelligence-enabled sensors and computer networks to quickly identify key vulnerabilities in the U.S. operating system, then bring together joint forces from different domains to attack targeted strikes against these vulnerabilities.

³⁷ Trevor Hunnicutt and Jeff Mason, "TAKEAWAYS - Biden and Xi Meeting: Taiwan, Iran, Fentanyl and AI," *Reuters*, November 16, 2023, sec. World, <https://www.reuters.com/world/takeaways-biden-xi-meeting-taiwan-iran-fentanyl-ai-2023-11-16/>.

³⁸ "Principles and Responsible Practices for Nuclear Weapon States: Working Paper Submitted by France, the United Kingdom of Great Britain and Northern Ireland and the United States of America."

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

India	1. Medium: It would strengthen its position against China but will complicate its strategic position against Pakistan. 2. Medium-High: India has a no first use policy (no launch on warning policy). India moving towards “cannisterisation” of nuclear weapons, which enables launch within hours. AI may be considered a force multiplier against China. 3. High: India has traditionally responded to Pakistan’s military developments. If Pakistan goes for AI integration, then India would follow suit. Indicators, however, suggest that it will be India which will do the AI integration first. 4. High: Yes, it has used its missile tests as technology demonstrators and then decided to pursue their production and employment by its strategic forces. It could do this in case of AI integration too. 5. Medium-high: Relatively vibrant civil society, but unlikely to have significant impact on nuclear issues	1. Low: India is generally not amenable to taking additional arms control commitments. India voted no on the UNGA resolution on AI in military applications. 2. Medium: India has a rudimentary capability for verification which it is strengthening with new surveillance and defence systems. 3. Low: It is very unlikely to make a unilateral statement. 4. Medium: It will ask for technology transfers in military use of AI from US and European countries. 5. High: N9, including a larger group of states that includes all nuclear states and non nuclear states
Israel	1. High: Israel has actively embraced AI technologies and has a growing AI sector. Israel Defence Forces have started using AI to select targets for air strikes and logistics.³⁹ Given this, introduction of AI to nuclear weapons will increase its position in the region both regarding first nuclear strike and retaliation capabilities. 2. High. Especially, given the latest developments in Israel and the region, integration of AI will serve as a guarantee to respond quickly to any potential attack. It will be force-multiplier against Iran and other potential conflicts in the region. 3. High: Given the complex security	1-5. Low: Although Israel has possessed nuclear weapons since the 1960s, it maintains a policy of nuclear opacity, never officially confirming the existence of its nuclear program. Accordingly, Israel has never signed the NPT,⁴⁰ and most probably won't join any other international agreement related to nuclear weapons. Would also be relatively unlikely to join a joint statement including many non-nuclear states.

³⁹ Marissa Newman, “Israel Quietly Embeds AI Systems in Deadly Military Operations.”

⁴⁰ “Israel,” *The Nuclear Threat Initiative* (blog), October 20, 2021, <https://www.nti.org/countries/israel/>.

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

	situation in the region, Israel will employ various technologies, including AI, to ensure it maintains a strategic advantage. 4. Low - Israel adheres to a policy of nuclear ambiguity, refraining from officially acknowledging the existence of its nuclear program. Consequently, it will not reveal whether it has integrated AI or not. 5. Medium - civil society active, but input on security-related topics very limited - especially on nuclear weapons	
Pakistan	1. High: Yes. Based on Johnson's argument that states less confident of their second-strike capabilities will resort to automation,⁴¹ AI integration will help strengthen Pakistan's deterrence as it will assure a retaliatory strike. 2. Medium-High: Pakistan does not have a launch on warning strategy. However, Pakistan will have an incentive to respond faster because the unpredictability and uncertainty of India conducting warfare at machine speed will compel Pakistan to automate its nuclear response. It will be a force multiplier for Pakistan. It will enhance early warning systems and increase the response time. At the same time, 3. High: Pakistan has traditionally responded to India's military developments. India's foray in AI integration would be an incentive for Pakistan to do so as well. 4. Medium: Generally, Pakistan regards mastery of technology for security purposes. So, the primary aim of AI integration would be strengthening security rather than showcasing technological prowess. 5. High - high lack of influence of civil	1. High: It would be open to an agreement on AI integration. It has already floated a working paper regarding this in the CD, Geneva.⁴² 2. High: Yes, Pakistan does not have the NTM for verification. 3. Low: It would prefer a joint statement or declaration (necessary that India is part of this) but will not be amenable for a unilateral statement. 4. High: It would ask for more openness and transparency. It might also call for sharing of technologies. 5. High: N9, including a larger group of states that includes all nuclear states and non-nuclear states

⁴¹ Johnson, *AI and the Bomb*.

⁴² "Addressing the Security and Stability Implications of Military Applications of Artificial Intelligence (AI), and Autonomy in Weapon Systems, Working Paper by Pakistan."

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

	society actors on nuclear policies of state	
Russia	1. Medium - Russia already has the Dead Hand which showcases interest in automation and ensuring retaliation but may also reduce the need to remove humans further from the loop, additionally the Petrov incident has shown importance of human in the loop. 2. High - Ukrainian crisis and NATO expansion to the east as threats to Russia's national security. 3. Medium - Russia has large nuclear potential and forces for retaliation (Dead Hand). NDCC (using AI) is designed to gather information about enemy movements from multiple sources and advise senior officers on possible responses. 4. High - Interest in showcasing technological prowess (Russian military leaders claim that the NDCC has far more computing power than the Pentagon, three times more, they say, and 15 times more storage capacity). 5. Medium-High - civil society actors moderately influential, but support state positions especially in current times,	1. Medium - agree to sign a multilateral agreement with nuclear powers. The main principle is general openness (for instance, agreement between Russia and China on notifications of launches of ballistic missiles and space launch vehicles, 2010). 2. Low - verifiability as one of the main principles of the potential agreement. Creation of an international commission based on the UN. 3. High - for instance, agreement between Russia and China on notifications of launches of ballistic missiles and space launch vehicles, 2010. 4. It would ask for more openness and transparency. 5. P5. There is no great need to sign an agreement with non-nuclear countries.
United Kingdom	1. Low - While the UK has an interest in leveraging its advanced technology base into strategic advantage,⁴³ its strategic position won't be significantly changed by taking human out of the loop for faster nuclear retaliation abilities. This could change if nuclear submarines become less effective deterrent. 2. Low - The UK has nuclear submarines and thus a small but potent retaliatory capacity. This would not be significantly changed by faster response times; for deterrence against adversaries with large arsenals, the UK would nevertheless	1. High - general openness to arms control. General hesitation around arms control with Russia. 2. High - lack of verifiability is not that great of an issue as no strong strategic dynamics with particular adversaries. 3. High - happy to make unilateral agreement (has already happened together with US and France in soft form)⁴⁵ 4. No particular concessions, beyond willingness to cooperate 5. P5, or otherwise larger group of states including non-nuclear states (likely opposed to N9)

⁴³ United Kingdom, "Defence Artificial Intelligence Strategy, Policy Paper."

⁴⁵ "Principles and Responsible Practices for Nuclear Weapon States: Working Paper Submitted by France, the United Kingdom of Great Britain and Northern Ireland and the United States of America."

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

	likely rely on the US. The UK's alliance with the US means that traditional small state dynamics related to other countries having larger arsenals affect it less. 3. Low - no direct adversary that the UK would feel compelled to have to keep up with. 4. Medium - the UK is interested in showcasing its technological prowess,⁴⁴ however there are many ways to do this much more effectively than around automation of NC3. 5. Low - civil society actors very influential	
United States	1. Medium - The US possesses a large number of nuclear weapons and has sufficient retaliatory forces to respond. However, the 2022 DoD Nuclear Posture Review emphasizes the importance of human involvement in nuclear decisions,⁴⁶ and the US Congress is considering a ban on AI-driven nuclear launches using federal funds.⁴⁷ Consequently, advocating for other countries to maintain human control in nuclear decision-making may serve as leverage to influence and control them. 2. High: The US has a significant incentive to respond quickly in the event of a nuclear attack due to the imperative to minimize damage, protect its population, and potentially prevent further escalation.⁴⁸ Swift response is crucial in order to enact effective countermeasures, mitigate the impact of the attack, and, if possible, deter	1.High: General openness to arms control; general alignment with France and UK. 2. Medium: Lack of verifiability of potential adversary countries is an issue for the US. At the same time, potentially willing to have non-verifiable agreement over no agreement (at least Biden administration). 3. High: the US has already initiated together with France and UK in soft form.⁵⁰ Reuters news report that US and China are not yet ready to make joint statement.⁵¹ 4. Medium: Concessions might relate to the limited usage of AI in NC3. 5. P5, or otherwise larger group of states including non-nuclear states (Northern Ireland was already a part of a working paper submitted by France, UK and US on Principles and responsible practices for Nuclear Weapons states in the framework of 2020 NPT Review Conference).⁵² The US and the Russian Federation agreed on a five-year extension of New START to keep it

⁴⁴ United Kingdom.

⁴⁶ United States, "National Defence Strategy of The United States of America."

⁴⁷ Edward John Markey, "Block Nuclear Launch by Autonomous Artificial Intelligence Act of 2023," Pub. L. No. BUR23348 (n.d.), https://www.markey.senate.gov/imo/media/doc/block_nuclear_launch_by_autonomous_ai_act_-_042623pdf.pdf.

⁴⁸ United States, "National Security Strategy" (The White House, October 12, 2022), <https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf>.

⁵⁰ "Principles and Responsible Practices for Nuclear Weapon States: Working Paper Submitted by France, the United Kingdom of Great Britain and Northern Ireland and the United States of America."

⁵¹ Hunnicutt and Mason, "TAKEAWAYS - Biden and Xi Meeting."

⁵² "Principles and Responsible Practices for Nuclear Weapon States: Working Paper Submitted by France, the United Kingdom of Great Britain and Northern Ireland and the United States of America."

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

	adversaries from further hostile actions. 3. High: The US has a significant interest in discouraging acts of aggression by China, Russia, and other nations. Given the increased capabilities of potential adversaries and the development of new strategies involving threats both below and above the conventional conflict threshold, there is a recognition that reliance solely on conventional forces and nuclear deterrence is insufficient. It is imperative for the U.S. defense strategy to not only maintain but also strengthen deterrence, with a specific emphasis on addressing China as the primary pacing challenge.⁴⁹ 4. High: The US actively showcases its military technological capabilities to deter potential adversaries, assure allies, and uphold its global influence. This demonstration not only reinforces strategic alliances but also underscores the US's commitment to nuclear deterrence. Ongoing modernization efforts. 5. Medium - civil society actors highly influential and vocal. Great awareness of the risks of integration, active expert community. But limited influence on actual decisions on nuclear in the past.	in force through February 4, 2026. The treaty includes a withdrawal clause that is standard in arms control agreements.
North Korea	1-5. High interest.	1. Medium - given potential to be recognized. 2. High - verification is a limited issue. 3. Low unilateral willingness. 4. Would want to have acceptance, aid, dollars. 5. N9, non-nuclear states.

Appendix 2: Launch on Warning Posture

Country	Launch on Warning (LoW)	No First Use
United	The U.S. reserves the right to use nuclear	No.

⁴⁹ United States, "National Defence Strategy of The United States of America."

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

States	weapons in response to a nuclear attack.	The U.S. maintains a policy of deliberate ambiguity regarding the first use of nuclear weapons, meaning it does not make an explicit commitment either way.
UK	The UK reserves the right to use nuclear weapons to deter aggression and to protect its national interests.	No. The UK's nuclear posture is based on a policy of deliberate ambiguity, meaning it does not make a specific commitment on whether it would use nuclear weapons first or only in response to a nuclear attack.
France	France has declared a policy of deterrence, emphasizing that its nuclear weapons serve as a deterrent against aggression and are not intended for use in conflicts of lower intensity.	No. France maintains a policy of deliberate ambiguity regarding the first use of nuclear weapons.
Russia	Russia's nuclear doctrine emphasizes the role of nuclear weapons as a deterrent against aggression. Russia's military doctrine includes the concept of "de-escalation" or "escalation control," which envisions the potential limited use of nuclear weapons in a conflict to control and de-escalate a situation.	No. Russia maintains a stance of strategic ambiguity regarding the first use of nuclear weapons.
China	China's nuclear doctrine emphasizes a defensive posture. The primary purpose of China's nuclear weapons is deterrence against aggression and to ensure the security of the state. China has provided explicit assurances that it will not use nuclear weapons against non-nuclear-weapon states or in nuclear-free zones.	Yes.
India	India's doctrine allows for the possibility of preemptive nuclear strikes in response to intelligence indicating an imminent nuclear threat.	Yes
Pakistan	Pakistan's nuclear posture is primarily based on the principle of deterrence	No.
Israel	No	No

MEMORANDUM: AI IN NUCLEAR COMMAND AND CONTROL: PROSPECTS FOR AN INTERNATIONAL AGREEMENT TO KEEP A HUMAN IN THE LOOP

DPRK	No. North Korea views nuclear weapons as a key element of its national security strategy.	No
------	--	----

The Launch on Warning (LOW) nuclear posture is a strategic approach in which a country's military forces are prepared to launch nuclear weapons in response to the detection of an imminent nuclear attack, based on early warning signals. This posture is rooted in the concept of minimizing the time between the detection of an incoming nuclear threat and the launch of a retaliatory nuclear strike.

During the Cold War, both the US and the Soviet Union maintained a form of LOW posture as part of their nuclear doctrines. The idea was that, in the event of an incoming nuclear attack, the side that detected the threat first would have an advantage in launching a counterattack. Since the end of the Cold War, the focus on LOW postures has decreased, and countries have adopted more flexible and diverse nuclear postures. For example, the US still maintains the LOW capability, its posture is not dependent on either rapid deployment or swift retaliation. Instead, US nuclear posture is designed to withstand an enemy's counterforce strike and launch a devastating retaliation, aimed at strengthening deterrence and reducing the risk of nuclear accidents caused by false alarms⁵³.

LOW is associated with significant risks of accidental nuclear war, particularly due to the potential misinterpretation of false alarms in early warning systems⁵⁴. The risks of inadvertent nuclear war from misinterpretation of false alarms have been analyzed using mathematical modeling frameworks, emphasizing the need to assess potential options to reduce these risks⁵⁵. Furthermore, the deployment of nuclear weapons abroad represents a posture decision with significant consequences for international security, highlighting the complexities associated with nuclear postures. Additionally, the type of nuclear posture a state chooses can affect the international community's willingness to intervene in emerging nuclear powers, indicating the broader implications of nuclear postures on global dynamics⁵⁶.

⁵³ Keir Lieber and Daryl G. Press, "US Strategy and Force Posture for an Era of Nuclear Tripolarity," *Atlantic Council* (blog), May 1, 2023, <https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/us-strategy-and-force-posture-for-an-era-of-nuclear-tripolarity/>.

⁵⁴ Mian, Z., Rajaraman, R., & Ramana, M. (2003). Early Warning in South Asia—Constraints and Implications. *Science and Global Security*, 11(2-3), 109-150. <https://doi.org/10.1080/714041033>

⁵⁵ Barrett, A., Baum, S., & Hostetler, K. (2013). Analyzing and Reducing the Risks of Inadvertent Nuclear War Between the United States and Russia. *Science and Global Security*, 21(2), 106-133. <https://doi.org/10.1080/08929882.2013.798984>

⁵⁶ Narang, V. (2015). Nuclear Strategies of Emerging Nuclear Powers: North Korea and Iran. *The Washington Quarterly*, 38(1), 73-91. <https://doi.org/10.1080/0163660x.2015.1038175>

The Authors

Tynchtykbek Israiilov is an experienced professional in the field of multilateral diplomacy, arms control, Small Arms and Light Weapons (SALW), non-proliferation of Weapons of Mass Destruction (WMD), and border security and management. He has successfully managed numerous projects in these areas, demonstrating his expertise and commitment to international peace and security. Tynchtykbek's contributions extend beyond project management, as he has also shared his knowledge and insights through presentations at various international, regional, and national events. Tynchtykbek holds a Master's degree in Politics and Security from the OSCE Academy in Bishkek. His research interests include international security, regional cooperation, and arms control.

Ekaterina Mikhalevich is a PhD Candidate at the International Relations Department, Saint Petersburg State University. She is a participant of visits to NATO Headquarters (Brussels, Belgium) and Headquarters of the Supreme Headquarters Allied Powers Europe SHAPE (Mons, Belgium). She is a participant of the grant project 'Malicious use of Artificial Intelligence and Challenges for Psychological Security in Northeast Asia', funded by the Russian Foundation for Basic Research and the Vietnam Academy of Social Sciences.

Haleema Saadia is a doctoral candidate at the National University of Science and Technology (NUST), Pakistan, where she is focusing on Pakistan's nuclear discourse and identity construction. She has worked at Pakistan's Strategic Plans Division (SPD) where she conducted arms control policy analysis. She is currently a South Asia Visiting Fellow at The Stimson Center and a Research Fellow at the ROADS Initiative, Pakistan. She was a Visiting Fellow at the Sandia National Laboratories and James Martin Center for Nonproliferation Studies, US. She has taught at the National University of Modern Languages (NUML). She was awarded Chancellor's Medal by the President of Pakistan and Vice-Chancellor's Medal by Quaid-i-Azam University in recognition of her academic credentials.

Jonas Sandbrink is a biosecurity researcher at the University of Oxford and a Biological Security Advisor at the UK Cabinet Office. His current work focuses on the implications of synthetic biology and artificial intelligence for biological security risks. Previously, Jonas has worked with the Future of Humanity Institute, Google DeepMind, the Nuclear Threat Initiative, and the German Parliament. He was an Emerging Leaders in Biosecurity Fellow at the Johns Hopkins Center for Health Security and an Ending Bioweapons Fellow at the Council on Strategic Risks. He holds a Bachelor in Medical Sciences and is pursuing a DPhil (PhD) at Nuffield Department of Medicine, University of Oxford.