PRAGMATIC APPROACHES TO GOVERNANCE AT THE ARTIFICIAL INTELLIGENCE–NUCLEAR NEXUS

FEI SU, WILFRED WAN, LORA SAALMAN AND VLADISLAV CHERNAVSKIKH

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

SIPRI is an independent international institute dedicated to research into conflict, armaments, arms control and disarmament. Established in 1966, SIPRI provides data, analysis and recommendations, based on open sources, to policymakers, researchers, media and the interested public.

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

Signalistgatan 9
SE-169 70 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org

PRAGMATIC APPROACHES TO GOVERNANCE AT THE ARTIFICIAL INTELLIGENCE–NUCLEAR NEXUS

FEI SU, WILFRED WAN, LORA SAALMAN AND
VLADISLAV CHERNAVSKIKH

October 2025

STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

DOI: https://doi.org/10.55163/BCHM4674

Contents

Acknowledgements

SIPRI and the authors would like to express their sincere gratitude to the German Federal Foreign Office for supporting this project with generous funding for research, workshops and publications related to the AI–nuclear nexus.

Special thanks also go to the external reviewers and SIPRI colleagues, including Dr Sibylle Bauer, Dr Vincent Boulanin, Dr Jingdong Yuan and Jules Palayer, for their constructive feedback on earlier drafts of the report, and to the many experts and officials who engaged throughout the course of the project.

Finally, the authors also wish to express their sincere appreciation to SIPRI's Editorial Department for its significant contributions to finalizing this publication.

Abbreviations

AI	Artificial intelligence
CBM	Confidence-building measure
ICBM	Intercontinental ballistic missile
LAWS	Lethal autonomous weapon system
LLM	Large language model
LOW	Launch-on-warning (posture)
LUA	Launch-under-attack (posture)
NC3	Nuclear command, control and communications
NPT	Treaty on the Non-Proliferation of Nuclear Weapons (Non-Proliferation Treaty)
P5	The five nuclear weapon states and permanent members of the United Nations Security Council
REAIM	Responsible Artificial Intelligence in the Military Domain (Summit)
SRF	Strategic Rocket Forces (Russian Federation)
SSBN	Nuclear-powered ballistic missile submarine
UAV	Uncrewed aerial vehicle
UN	United Nations
UUV	Uncrewed underwater vehicle

Summary

The rapid integration of artificial intelligence (AI) into nuclear and nuclear-related weapon systems is reshaping global security. While it accelerates the decision-making process, enhances strategic warning and improves targeting, it also introduces new forms of vulnerability and uncertainty that could destabilize deterrence and increase the risk of inadvertent nuclear escalation. These developments, unfolding amid intensifying geopolitical competition, are reshaping how states perceive threats. Although discussions on maintaining human control have gained prominence among both nuclear-armed and non-nuclear-armed states in multilateral settings, there remains limited clarity on national positions and on definitions of key concepts that include human-in-the-loop, human-on-the-loop, human control and meaningful human control. The absence of shared understanding continues to hinder the development of norms and standards for governing the AI–nuclear nexus. This report seeks to contribute to these ongoing debates.

There are three key dimensions to how AI integration may affect nuclear deterrence and strategic stability. First, the integration of AI into nuclear command, control and communications (NC3) systems raises serious concerns over reliability, transparency and human oversight. Greater reliance on AI may also drive risks associated with false positives and false negatives. Retaining human involvement in nuclear decision making has thus been a key focus for discussions among states. Nevertheless, the potential for achieving strategic advantage could mean states will remain inclined to deepen AI integration. Second, AI integration can feed into perceptions of technological imbalance that can intensify security dilemmas. Given potential impacts on the nuclear balance and second-strike credibility, fear of being left behind may push a state to pursue greater integration of automation and AI into its nuclear forces. This may prompt a nuclear-armed state to diversify its nuclear forces or to adopt a destabilizing posture such as launch on warning (LOW) or launch under attack (LUA) to preserve deterrence. Third, AI-enabled non-nuclear weapons pose additional challenges to nuclear stability by, for example, targeting NC3 infrastructure. The integration of AI into these systems magnifies their potential strategic impact, giving technologically limited states—including non-nuclear-armed states—new asymmetric tools that could disrupt or neutralize adversary assets. Nuclear-armed states have also begun to incorporate the threat posed by advanced conventional capabilities—and of non-nuclear-armed states—in their nuclear doctrines.

Approaches to human control in the nuclear domain are evolving and there are divergent perspectives among states and experts. While there is broad agreement on the need to retain human control in nuclear decision making, states have yet to clearly define which specific actions must remain under human oversight. The ways in which nuclear-armed states define the process leading up to and comprising decisions on nuclear use are thus pertinent. The debate over human control has been shaped by discussions in non-nuclear domains, particularly within multilateral processes on lethal autonomous weapon systems (LAWS). These discussions emphasize accountability, judgement and intervention capacity, although interpretations differ widely among states. Applying these principles to nuclear and nuclear-related weapon systems remains contentious.

There are pathways via which governance principles at the AI–nuclear nexus can be translated into practical action. Nuclear-armed states could work in bilateral or multilateral settings to identify specific attributes of AI systems that they find destabilizing, including from the perspective of deterrence. Greater clarity is needed on how the broad agreement on the need for human oversight in nuclear decision making can be operationalized across different nuclear-related functions. Establishing clear

principles that emphasize accountability, safety and reliability testing would need to be supported by concrete regulatory and technical measures, including safety standards and oversight mechanisms. Moreover, setting concrete and easily achievable goals for the near term can have an immediate impact and signal continued interest in advancing discussions at the AI–nuclear nexus. Multilateral forums such as Responsible AI in the Military Domain (REAIM) Summit and the working groups of the Creating an Environment for Nuclear Disarmament (CEND) initiative can be used for exchanging views, aligning principles and advancing practical cooperation among nuclear-armed and non-nuclear-armed states alike.

The importance of sustaining momentum in discussions on human control must be emphasized. Continued efforts to clarify key concepts and translate shared concerns into operational principles and standards are essential to advancing governance at the AI–nuclear nexus.

1. Introduction

Military applications of artificial intelligence (AI) are developing at a rapid pace, including the integration of AI in nuclear and nuclear-related weapon systems.[1] Coupled with deepening geopolitical tensions, these technological advancements and their military use threaten to further unsettle the global balance of power. The potential volatility is compounded by the application of AI in conventional weapons and cyber operations. By providing both nuclear-armed and non-nuclear-armed states with new avenues for gaining asymmetric advantages, such non-nuclear operations have further complicated deterrence calculations. These developments threaten strategic stability, reshape how states perceive adversarial intentions and threats in crises, alter escalation dynamics and drive military investments and deployments that fuel arms racing.

Increased awareness of the impact of AI on nuclear deterrence and strategic stability has sparked interest in reducing associated risks.[2] This has been manifested in discussions about the nexus between AI and nuclear forces in multilateral frameworks that govern the military use of AI—although these have been limited, centring on commitments to maintain human control over decisions to use nuclear weapons.[3] However, each country has its own understanding of what 'human control' means, perhaps equating it with human-in-the-loop, with human-on-the-loop or with another configuration.[4] A pragmatic approach is thus needed to analyse each country's definition of what human control means and which systems are relevant in this context. More nuanced understandings of state perspectives on human control, risk and other relevant concepts will help to identify the potential for governance at the AI–nuclear nexus, including in terms of the approach, the form and the function that related measures might take.[5]

This report seeks to contribute to these understandings, drawing from official statements, policy documents and scholarly writings that reflect domestic debates across key states, with a particular focus on nuclear-armed states, both the five nuclear weapon states recognized under the 1968 Non-Proliferation Treaty (NPT)—China, France, the Russian Federation, the United Kingdom and the United States (the P5)—and the nuclear-armed states outside the NPT—Israel, India, the Democratic People's Republic of Korea (DPRK, or North Korea) and Pakistan.[6] It then aims to chart a path forward for governance at the AI–nuclear nexus.

The report continues in chapter 2 by identifying the extent to which a shared sense of purpose exists in regulating the nexus. It does so by examining commonalities in risk assessments among key stakeholders, with a focus on the effects that integrating AI into nuclear and nuclear-related weapon systems will be perceived to have on nuclear

1 Chernavskikh, V., 'Nuclear weapons and artificial intelligence: Technological promises and practical realities', SIPRI Background Paper, Sep. 2024; and Boulanin, V. et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020).

2 Boulanin et al. (note 1).

3 Su, F., Chernavskikh, V. and Wan, W., 'Advancing governance at the nexus of artificial intelligence and nuclear weapons', SIPRI Insights on Peace and Security no. 2025/03, Mar. 2025.

4 Horowitz, M., 'Autonomous weapon systems: No human-in-the-loop required, and other myths dispelled', War on the Rocks, 22 May 2025; Saltini, A. and Pan, Y., 'Beyond human-in-the-loop: Managing AI risks in nuclear command-and-control', War on the Rocks, 6 Dec. 2024; Rautenbach, P., 'Keeping humans in the loop is not enough to make AI safe for nuclear weapons', *Bulletin of the Atomic Scientists*, 16 Feb. 2023; and Leins, K. and Kaspersen, A., 'Seven myths of using the term "human on the loop": "Just what do you think you are doing, Dave?"', Carnegie Council for Ethics in International Affairs, 9 Nov. 2021.

5 E.g. Wallander, C. A. and Keohane, R. O., 'Risk, threat, and security institutions', ed. R. O. Keohane, *Power and Governance in a Partially Globalized World* (Routledge: New York, 2002).

6 Treaty on the Non-Proliferation of Nuclear Weapons (Non-Proliferation Treaty, NPT), opened for signature 1 July 1968, entered into force 5 Mar. 1970, IAEA INFCIRC/140, 22 April 1970.

deterrence and broader strategic stability—defined here as 'the absence of incentive to use nuclear weapons first (crisis stability) and the absence of incentive to build up a nuclear force (arms race stability)'.[7] Chapter 3 then analyses the debate around human control in order to identify its utility and limitations from a governance perspective. It examines how states, or experts within those states, approach human control in the context of nuclear decision making and analyses the potential for principles to be translated into technical parameters. Chapter 4 draws on common interests identified in the preceding chapters to chart a future for governance at the AI–nuclear nexus. Chapter 5 concludes the report with key observations and reflections.

[7] Edward Warner cited in Acton, J. M., 'Reclaiming strategic stability', eds E. A. Colby and M. S. Gerson, *Strategic Stability: Contending Interpretations* (US Army War College Press: Carlisle Barracks, PA, Feb. 2013), p. 117.

2. Risk assessments

Understanding the evolving perceptions of risk that are associated with the AI–nuclear nexus is essential for designing responsible policies, fostering and maintaining strategic stability, and preventing inadvertent nuclear escalation. This chapter outlines three key dimensions of these perceptions regarding nuclear deterrence and broader strategic stability: (*a*) nuclear command, control and communications (NC3); (*b*) the nuclear balance and second-strike credibility; and (*c*) AI-enabled non-nuclear capabilities with strategic effect. The following sections also highlight where nuclear-armed and non-nuclear-armed states converge or diverge in their assessments.

Nuclear command, control and communications

While the term 'nuclear command, control and communications' is widely used in discussions of nuclear deterrence and features prominently in debates about the AI–nuclear nexus, the structure of NC3 systems remains opaque and varies significantly across nuclear-armed states. For instance, in China NC3 refers primarily to software and applications used for command-and-control technical systems.[8] In contrast, the United States' definition includes a set of infrastructures that are responsible for situational awareness, planning, decision making, force direction and force management.[9] Despite these differences, NC3 architectures are united by being designed to provide decision makers with timely warning of adversarial attacks and to allow them to effectively implement nuclear plans.

Integration of AI in NC3

Discussions on the integration of AI into NC3 have largely focused on its existing and potential use in three key areas: strategic warning (e.g. early-warning systems and intelligence, surveillance and reconnaissance), adaptive targeting (e.g. precision guidance and targeting) and decision support (e.g. data collection, data analysis and communications).[10] The AI integrated into these systems sifts through vast amounts of data, thereby enhancing the detection, tracking, targeting and strike capabilities of nuclear forces, potentially upending strategic stability. AI also facilitates the timely communication of information and analysis alongside actionable recommendations for decision makers. For example, AI could process large volumes of real-time data much faster in order to detect adversarial attacks or unusual activities, potentially giving decision makers more time to respond.[11]

However, greater reliance on AI may also increase risks from false positives (i.e. mistakenly identifying a launch or threat when none exists) and false negatives (i.e. fears over missing an incoming high-precision, prompt, stealth conventional or nuclear strike), driving even wider and faster integration of AI and autonomy.[12] This

[8] Wood, P., Stone, A. and Corbett, T., *Chinese Nuclear Command, Control, and Communications* (China Aerospace Studies Institute: Montgomery, AL, Mar. 2024).

[9] US Office of the Deputy Assistant Secretary of Defense for Nuclear Matters, *Nuclear Matters Handbook 2020 [Revised]* (US Department of Defense: Washington, DC, Aug. 2024), p 15.

[10] Saltini, A., *AI and Nuclear Command, Control and Communications: P5 Perspectives* (European Leadership Network: London, Nov. 2023; Saltini, A., Mishra, S. and Reiner, P., *Nuclear Command, Control, and Communications (NC3): A Primer on Strategic Warning, Decision Support, and Adaptive Targeting Subsystems* (Institute for Security and Technology: July 2025); and Boulanin et al. (note 1).

[11] Asghar, S., 'AI at the nexus of nuclear deterrence: Enhancing left of launch operations', Next Generation Nuclear Network, Center for Strategic and International Studies (CSIS), 23 May 2025.

[12] Hayes, P., 'Nuclear command, control and communications in the Asia-Pacific', *Global Asia*, vol. 16, no. 2 (June 2021); and Saalman, L., 'Fear of false negatives: AI and China's nuclear posture', *Bulletin of the Atomic Scientists*, 24 Apr. 2018.

is particularly dangerous when the systems are relatively immature or insufficiently tested, as illustrated by the 1983 Petrov incident caused by a malfunction in the Soviet Union's partially automated Oko early-warning system, which falsely indicated an incoming US intercontinental ballistic missile (ICBM) attack.[13] The application of automation—and its increased extent—in nuclear early-warning and command-and-control systems could further increase the risk of inadvertent nuclear use.[14] These concerns are accentuated by the uncertainty and opacity that are intrinsic to the integration of AI into weapon systems, particularly nuclear and nuclear-related weapon systems. As an example, Lu Chuanying (former head of the Research Centre for the International Governance of Cyberspace at the Shanghai Institutes for International Studies) cites the lack of transparency in AI-enabled 'critical national security facilities such as nuclear control systems' as complicating the decision-making process and contributing to doubts about the system's decisions during emergencies.[15] This is a permutation of the 'black box' problem often referred to in AI models regarding the difficulty of understanding how a decision is made and how input data influences final conclusions.[16]

Furthermore, there is common concern about a critical lack of the historical data on nuclear crises that would be needed to train AI systems effectively.[17] Unlike other domains, in which large data sets can be accumulated, the relatively few nuclear crises in history provide limited input upon which to train machine learning AI models. This lack of available data for training on the consequences of nuclear launch compels reliance on large language models (LLMs).[18] But there are doubts about the ability of systems enhanced with LLM-based AI to provide sound decision support in rapid, high-pressure—and high-stakes—scenarios.[19] Multiple studies and war games have observed the potential for LLMs to not only speed up decision making, but to contribute directly to escalation or nuclear use.[20]

State responses

Both nuclear-armed and non-nuclear-armed states have expressed concerns about the accuracy and reliability of AI systems, particularly regarding whether AI integration could introduce new vulnerabilities into NC3 and increase the risk of misinterpretation or unintended escalation. For example, corruption of the training data sets used by AI systems in their decision-support functions can potentially skew outputs and, because of the systems' opacity, lead to flawed or dangerous decisions.[21] A cyber operation

[13] Saalman, L., Saveleva, L. and Su, F., 'Mapping cyber-related missile and satellite incidents and confidence-building measures', SIPRI Insights on Peace and Security no. 2023/10, Nov. 2023.

[14] Borrie, J., 'Cold war lessons for automation in nuclear weapon systems', ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019).

[15] Lu, C. (鲁传颖), '人工智能重塑国家安全的范式和逻辑' [Artificial intelligence reshapes the paradigm and logic of national security], 人民论坛 [People's Tribune], 24 Jan. 2025 (author translation).

[16] Chernavskikh (note 1).

[17] Rautenbach, P., 'On integrating artificial intelligence with nuclear control', *Arms Control Today*, vol. 52, no. 7 (Sep. 2022).

[18] Rivera, J.-P. et al., 'Escalation risks from language models in military and diplomatic decision-making', *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery: New York, June 2024); and Chew, A. and Hemrajani, A., 'Will AI enhance decision-making in the use of nuclear weapons?', S. Rajaratnam School of International Studies (RSIS) Commentary no. 68, 2 Apr. 2025.

[19] Kallenborn, Z., 'Giving an AI control of nuclear weapons: What could possibly go wrong?', *Bulletin of the Atomic Scientists*, 1 Feb. 2022.

[20] Rivera et al. (note 18).

[21] Krantz, T. and Jonker, A., 'What is data poisoning?', IBM, 10 Dec. 2024; and Saalman, L., Su, F. and Saveleva Dovgal, L., 'Mapping cyber-related missile and satellite incidents and confidence-building measures', SIPRI Insights on Peace and Security no. 2023/10, Nov. 2023.

could manipulate this data deliberately.[22] Moreover, counterspace technologies are highlighted as potentially posing additional threats to the integrity of NC3 in a working paper submitted by the member states of the Stockholm Initiative for Nuclear Disarmament in advance of the 2026 NPT Review Conference.[23] Notably, non-nuclear-armed states have played a key role in persistently raising concerns related to emerging technologies and advocating for robust risk-reduction measures.

Retaining human involvement in nuclear decision making has thus been a key focus for discussions among states, both nuclear armed and non-nuclear armed.[24] Some progress has been achieved among the former: France, the United Kingdom and the United States submitted a joint working paper on principles and responsible practices for nuclear weapon states to the NPT Review Conference in 2022, and Chinese President Xi Jinping and the then US President Joe Biden stressed human control over nuclear decision making at a bilateral meeting in 2024.[25] Meanwhile, France, Pakistan, the UK and the USA were among the states that endorsed the Blueprint for Action unveiled at the 2024 Summit on Responsible Artificial Intelligence in the Military Domain (REAIM), which contains a point on 'maintaining human control and involvement' in nuclear decision making.[26] Studies from Western research institutes also indicate a broad consensus within the Russian expert community on the importance of maintaining a human-centred approach to nuclear weapon use; an understanding that appears to align with near-term Russian military thinking.[27] A similar discussion—albeit at a non-official level—is also ongoing in India, where retired military analysts have emphasized the importance of human oversight, transparency and confidence-building measures (CBMs) and where academic studies have examined how integrating AI into nuclear systems and nuclear power facilities could influence escalation dynamics.[28]

However, the position of nuclear-armed states may change as strategic priorities evolve and broader security dynamics develop. Already, progress towards all such states adopting high-level commitments on maintaining human control is constrained by strategic consideration, political feasibility as well as the uncertainties surrounding the future direction of technology development.[29] In pursuit of a strategic advantage—in particular the capability to operate and respond at machine speed—nuclear-armed states may be inclined to deepen AI integration. This may present an obstacle to operationalizing the principle of human control and involvement (as discussed further in chapter 3). But even if human operators retain authority over final decisions, a critical

[22] Saltini, A., 'Navigating cyber vulnerabilities in AI-enabled military systems', European Leadership Network, 19 Mar. 2024.

[23] 2026 NPT Review Conference, Preparatory committee, 'Navigating the potential impact of emerging technologies on nuclear disarmament, arms control, non-proliferation and peaceful uses of nuclear energy and technology', Working paper submitted by the Stockholm Initiative for Nuclear Disarmament (Argentina et al.), NPT/CONF.2026/PC.III/WP.35, 25 Apr. 2025.

[24] Geneva Centre for Security Policy, 'P5 experts roundtable—Online meeting on the AI nuclear nexus on 24 June 2024', Co-convenors' summary, 15 July 2024.

[25] 10th NPT Review Conference, 'Principles and responsible practices for nuclear weapon states', Working paper submitted by France, the UK and the USA, NPT/CONF.2020/WP.70, 29 July 2022; and Chinese Ministry of Foreign Affairs, 'President Xi Jinping meets with US President Joe Biden in Lima', 17 Nov. 2024. See also Su et al. (note 3).

[26] Responsible AI in the Military Domain (REAIM) Summit, 'REAIM Blueprint for Action', 9–10 Sep. 2024, para. 5.

[27] Shakirov, O., *Russian Thinking on AI Integration and Interaction with Nuclear Command and Control, Force Structure, and Decision-making* (European Leadership Network: London, Nov. 2023); and Bendett, S., *The Role of AI in Russia's Confrontation with the West* (Center for New American Security: Washington, DC, Apr. 2024).

[28] Ali, M. T., 'The role of artificial intelligence in nuclear command and control systems', United Service Institution of India, 17 Feb. 2025; Menon, P., 'India should declare that AI will not be used to autonomously launch nuclear weapons', ThePrint, 16 May 2023; Patil, S. and Rawat, R., 'Integration of artificial intelligence in nuclear systems and escalation risks', Policy brief, Asia-Pacific Leadership Network, May 2025; and Nuclear Business Platform, 'Exploring the potential of artificial intelligence in the Indian nuclear energy sector', 9 Oct. 2023.

[29] Future of Life Institute, 'Framework for responsible use of AI in the nuclear domain', Policy brief, 5 Feb. 2025.

concern remains: how can states ensure that decision makers do not become overly dependent on AI-generated recommendations and options?[30] The risk is that, over time, human operators may begin to uncritically accept machine outputs, mistaking algorithmically derived suggestions for their own independent judgement—a phenomenon known as automation bias.

The nuclear balance and second-strike credibility

Technological advancements can intensify the security dilemma: developments that enhance one state's sense of security can simultaneously heighten another state's sense of insecurity. This dynamic is evident in the context of AI integration into nuclear and nuclear-related weapon systems. In such cases, perceived technological superiority of one nuclear-armed state may affect nuclear deterrence calculations and increase the risk of escalation.[31]

Threats to second-strike forces

Fear of being left behind may push a state to pursue greater integration of automation and AI into its nuclear forces. During 2025 the USA withdrew newly introduced restrictions on its use of AI, reflecting its ongoing concerns over AI threats from and competitiveness with China.[32] While not necessarily affecting US nuclear forces, it suggests the potential for others to follow suit. Moreover, technological imbalances between nuclear-armed states could raise doubts about the survivability of the second-strike forces of a state with a smaller arsenal or less advanced capabilities. This, in turn, could increase that state's temptation to launch a first strike, undermining strategic stability.[33] Moreover, AI-enhanced capabilities—precision-guidance and target selection; navigation to ensure accurate and precise flight paths; and decision making to automate certain processes to determine the best course of action when faced with missile defences—could improve the detection and targeting of an adversary's nuclear forces.[34] Together, this could threaten an adversary's ability to execute a retaliatory strike.

For instance, a May 2024 report issued by the USA's Lawrence Livermore National Laboratory highlights threats to the sea-launched leg of the US nuclear triad due to breakthroughs in anti-submarine warfare technology, enhanced by AI, quantum computing, advanced sensing and autonomous systems.[35] It makes the case for a mobile ICBM system as a means to address vulnerabilities exposed by AI and advanced sensors that threaten the survivability of the USA's fleet of nuclear-powered ballistic missile submarines (SSBNs). Work by Tsinghua University's Center for International Security and Strategy (CISS) also suggests that AI technology has the potential to enhance US reconnaissance capabilities, with China and Russia increasingly worried that improved 'technology for tracking and homing in on mobile missile launchers will, once it

[30] Ford, C. A., 'Nuclear doctrines and prospects for P5 Dialogue', New Paradigms Forum, 18 Feb. 2025.

[31] Boulanin et al. (note 1), p. 3.

[32] US Department of Commerce, Bureau of Industry and Security, 'Framework for artificial intelligence diffusion', *Federal Register*, 15 Jan. 2025; and US Department of Commerce, Bureau of Industry and Security, 'Department of Commerce rescinds Biden-era artificial intelligence diffusion rule, strengthens chip-related export controls', 12 May 2025.

[33] Boulanin et al. (note 1), p. 3; and Su, F. and Yuan, J., *Chinese Thinking on AI Integration and Interaction with Nuclear Command and Control, Force Structure, and Decision-making* (European Leadership Network: London, Nov. 2023).

[34] Kristensen, H. M. et al., 'Nuclear notebook: Chinese nuclear weapons, 2025', *Bulletin of the Atomic Scientists*, vol. 81, no. 2 (Mar. 2025).

[35] Christenson, R., *Supplemental Second-strike: Road-mobile ICBMs in the Two-peer Environment* (Lawrence Livermore National Laboratory, Center for Global Security Research: Livermore, CA, May 2024).

reaches maturity, threaten their retaliatory capabilities'.[36] Other Chinese experts have raised many concerns that AI could weaken nuclear deterrence, intensify escalation and arms racing, challenge strategic stability, complicate accountability, increase collateral damage, lower the threshold of proliferation, and lead to misuse and abuse.[37] In a similar vein, a 2023 report by a special inquiry committee of the upper house of the British Parliament acknowledges the potential of AI to pose a threat to second-strike capabilities and to increase the likelihood of a first nuclear use by a nuclear-armed adversary in crisis.[38]

State responses

To preserve its second-strike credibility, a state could also respond by diversifying its nuclear forces (e.g. increasing the number of silos or the mobility of ICBMs) or by adopting a launch-on-warning (LOW) or launch-under-attack (LUA) posture. In fact, nearly all nuclear-armed states continue to modernize their arsenals by upgrading existing systems, developing new ones or both—driven in large part by emerging technologies such as AI.[39]

Pakistan, for instance, has focused on developing strategic countermeasures such as multiple independently targetable re-entry vehicles (MIRVs) for its medium-range ballistic missiles, in part to address India's enhanced missile defence systems incorporating AI technologies.[40] During the May 2025 conflict between these two nuclear-armed neighbours, India launched BrahMos supersonic cruise missiles integrated with AI-based guidance and target-recognition systems.[41] One of the strikes targeted the Nur Khan Pakistan Air Force Base in Rawalpindi, located near Pakistan's nuclear command-and-control centres. This raised concerns that the attack could escalate to the nuclear level.[42] This operation was particularly escalatory as an advisor to Pakistan's prime minister has said that Pakistan had only 30–45 seconds to assess whether the incoming BrahMos missile carried a nuclear warhead.[43] This also suggests a significant incentive for Pakistan to include greater AI enhancements to speed up its nuclear-related rapid response.

In contrast to the geographic proximity of India and Pakistan, the greater distances between the USA and China or Russia tend to allow for more reaction time—although

[36] Chen, Q. (陈琪) and Zhu, R. (朱荣生), '不确定性：为何担心人工智能冲击国际安全?' [Uncertainty: Why are we worried about the impact of artificial intelligence on international security?], International Security and Strategy Studies Report, vol. 8, no. 8, Tsinghua University, Center for International Security and Strategy (CISS), Aug. 2018, p. 4 (author translation).

[37] Zhu, Q. (朱启超), '人工智能武器化的发展趋势及治理策略' [The development trend and governance strategy of weaponization of artificial intelligence], 人民论坛 [People's Tribune], 5 Feb. 2025; and Zhang, D. (张东冬), '人工智能军事化与全球战略稳定' [Militarization of artificial intelligence and global strategic stability], International Cooperation Center, 1 Nov. 2022. The first of these authors is from the National University of Defence Technology and the second from Dalian Maritime University. For more examples see Su and Yuan (note 33).

[38] British Parliament, House of Lords, AI in Weapon Systems Committee, *Proceed with Caution: Artificial Intelligence in Weapon Systems*, Report of Session 2023–24, HL Paper 16 (House of Lords: London, 1 Dec. 2023), p. 51.

[39] Kristensen, H. M. and Korda, M., 'World nuclear forces', *SIPRI Yearbook 2025: Armaments, Disarmament and International Security* (Oxford University Press: Oxford, 2025).

[40] Altaf, Z. and Javed, N., 'India's MIRV expansion along with satellites, missile defense, AI and precision weapons: Destabilizing South Asia', Strategic Forecast, Center for International Strategic Studies AJK, 29 Mar. 2024; and Khan, A. A., 'Pakistan is not building an ICBM to attack America', Global Security Review, 10 July 2025.

[41] Kazmi, Z., 'Crisis by confusion: BrahMos, entanglement, and the next Indo-Pak war', Center for International Strategic Studies (CISS), 14 Aug. 2025.

[42] Chauhan, B., 'How BrahMos missile strikes forced Pakistan to agree to a ceasefire', *New Indian Express*, 11 May 2025; Kristensen, H. et al., 'How nuclear war could start', *Washington Post*, 13 June 2025; Kugelman, M., 'Reflections on the India–Pakistan crisis: Takeaways and implications for Canada–India ties', Asia Pacific Foundation of Canada, 6 June 2025; and 'India hit Rawalpindi and other airbases with BrahMos missiles before Pakistan could act, admits PM Shehbaz Sharif', *Swarajya*, 30 May 2025.

[43] '30 seconds of terror: Pakistan PM Shehbaz Sharif aide reveals what happened after India fired deadly BrahMos', *Economic Times* (New Delhi), 4 July 2025.

these differences are less pronounced when considering US military bases in the Asia-Pacific and in the High North. The USA has what can be described as a LOW or LUA posture, although the terms are often conflated. While the US option of launching ICBMs on warning of an incoming nuclear attack is often called LOW, the US departments of Defense and State tend to use the term LUA, with the attack being determined from warning information from multiple independent sensors. Nevertheless, since this assumes that US missiles would be launched before the detonation of attacking warheads, the term LUA is misleading in this case. Similarly, both LOW and LUA have been applied in external descriptions of Russia's force posture.[44] The USA has also claimed that China may be pursuing a LOW posture, citing the expansion of China's missile silos among other advances as evidence for this shift.[45] Furthermore, Yang Chengjun, a retired senior colonel of the People's Liberation Army Rocket Force, has stated that China's current nuclear reaction time is 'within a few minutes' and that it is able to 'carry out a nuclear retaliation before the enemy's nuclear weapons land'.[46]

Credible and highly effective early-warning systems are essential to making a LOW or LUA posture operationally viable. AI has not yet been directly associated with these efforts. Notably, however, China is thought to be fielding an increasingly capable layered network of sensors, with development of sophisticated strategic early-warning systems that include over-the-horizon radars, counter-low-observable radars (i.e. radars that can detect stealth aircraft and other low-observable targets) and satellites, as well as radar systems designed to detect low-flying cruise missiles.[47] The data from all these systems is thought to be collected, analysed and disseminated at joint theatre command posts.[48] Moreover, even nuclear-armed states in close geographic proximity are not immune to pursuing LOW or LUA capabilities.[49] While India—like China—upholds a no-first-use policy, some argue that it is already building up its LOW capabilities, including through the canisterization of its Agni ballistic missiles, which allows nuclear warheads to be pre-mated with mobile launchers.[50] More recently, India has reportedly been in negotiations with Russia to procure advanced long-range early-warning radar systems.[51]

While, as noted, AI has not yet been directly associated with these efforts, numerous nuclear-armed states have explored its potential integration into early-warning

[44] Harrison, T., *Options for the Ground-based Leg of the Nuclear Triad* (Center for Strategic and International Studies: Washington, DC, Sep. 2017); Wright, D., Hartung, W. D. and Gronlund, L., *Rethinking Land-based Nuclear Missiles: Sensible Risk-reduction Practices for US ICBMs* (Union of Concerned Scientists: Cambridge, MA, June 2020), pp. 8–10; Montoya, N. and Kemp, R. S., 'Launch under attack: A sword of Damocles', War on the Rocks, 17 Mar. 2023; and Sharma, R., 'Will artificial intelligence undermine nuclear deterrence? History tells us that's unlikely', *Countering WMD Journal*, no 29 (spring/summer 2025).

[45] US Defense Intelligence Agency (DIA), *Nuclear Challenges: The Growing Capabilities of Strategic Competitors and Regional Rivals* (DIA: Washington, DC, 2024), p. 17; Lowther, A. and Williams, D., 'Why America has a launch on attack option', War on the Rocks, 10 July 2023; Stockes, J. et al., *Averting AI Armageddon: US–China–Russia Rivalry at the Nexus of Nuclear Weapons and Artificial Intelligence* (Center for a New American Security: Washington, DC, Feb. 2025); and US Department of Defense (DOD), *Military and Security Developments Involving the People's Republic of China 2023*, Annual Report to Congress (DOD: Washington, DC, Oct. 2022).

[46] Yang, C. (杨承军), '核战略专家杨承军: 不宜在网络上炒作涉核问题' [Nuclear strategist Yang Chengjun: It's not appropriate to speculate on nuclear-related issues on the internet], 祖国 [Motherland], 12 May 2020 (author translation).

[47] US Department of Defense (note 45), p. 26 and passim.

[48] Wood, Stone and Corbett (note 8), p. 3.

[49] Karnad, B., 'Nuclear false alarm', Security Wise, 31 Mar. 2017; and Kazmi, A. A., 'Accidental, inadvertent or deliberate launch: The case of Indian Air Force's nuclear supersonic cruise missile', Center for International Strategic Studies (CISS), 6 Apr. 2024.

[50] Khan, A. A., 'Indian missile crisis: One step from nuclear war?', Asia-Pacific Leadership Network, 28 June 2022.

[51] 'India plans to buy this massive $4 billion wall-like radar to neutralise Chinese threats from 6,000 km away', *Economic Times* (New Delhi), 11 Dec. 2024.

systems (as discussed in the previous section). However, any technical glitch, false alarm or algorithmic error based on AI inputs could trigger escalation with potentially catastrophic consequences. While there have been debates within the expert community about the pros and cons of maintaining LOW or LUA postures, states have not abandoned these postures—on the contrary, it would appear that they adhere to them more firmly.[52] Accordingly, the recognition of the risks inherent in maintaining or adopting LOW or LUA postures, particularly if decisions rely on AI-generated inputs, could provide an opening for governance at the AI–nuclear nexus.

AI-enabled non-nuclear weapons

Integration of AI in non-nuclear weapon systems also threatens the survivability of nuclear assets through kinetic or non-kinetic means. For example, counterspace weapons such as anti-satellite (ASAT) systems can target the space-based infrastructure upon which many NC3 systems rely for early warning, surveillance and secure communications.[53] AI can also act as a force multiplier for cyber operations that target NC3 in left-of-launch (i.e. pre-launch) operations.[54] Similarly, expert-level dialogue between China and the United States has identified a number of different cyber risks that could undermine strategic stability. These include cyber espionage that targets NC3 systems and their connected networks for intelligence gathering; cyberattacks on dual-use systems (i.e. systems for both military and civilian use), even if not intended to affect nuclear operations; and the added complexity of cyber operations potentially being carried out by third parties aiming to provoke conflict.[55] Related, AI has the potential to amplify the counterforce effect of conventional capabilities (i.e. their effectiveness in targeting military assets).[56]

State responses

In recent years, nuclear-armed states have also begun to incorporate the threat posed by advanced conventional capabilities—and of non-nuclear-armed states—in their nuclear doctrines. In 2017 France acknowledged that cyber and outer space incidents could lead to 'potentially crossing the nuclear threshold'.[57] The US Nuclear Posture Review of 2018 states that the USA would consider a nuclear response against 'significant non-nuclear strategic attacks' (implied to include chemical, biological, cyber and large-scale conventional aggression).[58] The nuclear-use law adopted by North Korea in 2022 underlines the potential for pre-emptive nuclear use in the event of an attack, including by a non-nuclear-armed state.[59] Russia also revised its nuclear doctrine in November 2024 in relation to its right to use nuclear weapons in response to aggression by a non-nuclear-armed state that is supported by a nuclear-armed state.[60]

[52] Montoya and Kemp (note 44).

[53] 2026 NPT Review Conference, NPT/CONF.2026/PC.III/WP.35 (note 23).

[54] Chernavskikh, V. and Palayer, J., 'Impact of military artificial intelligence on nuclear escalation risk', SIPRI Insights on Peace and Security no. 2025/06, June 2025.

[55] Levite, A. E. et al., *China–US Cyber-Nuclear C3 Stability* (Carnegie Endowment for International Peace: Washington, DC, Apr. 2021).

[56] Chernavskikh and Palayer (note 54).

[57] Government of France, *Defence and National Security Strategic Review 2017* (Délégation à l'information et à la communication de la Défense: Paris, Oct. 2017), p. 49.

[58] US Department of Defense (DOD), *Nuclear Posture Review 2018* (DOD: Washington, DC, Feb. 2018), p. 21.

[59] Korean Central News Agency (KCNA), 'Law on DPRK's policy on nuclear forces promulgated', KCNA Watch, 9 Sep. 2022.

[60] 'Fundamentals of state policy of the Russian Federation on nuclear deterrence', approved by Russian Presidential Order no. 991, 19 Nov. 2024.

The development of AI-enabled non-nuclear weapons with strategic effect may be particularly appealing to states—both nuclear armed and non-nuclear armed—that are technologically less advanced or face resource constraints. AI presents an opportunity to develop asymmetric capabilities that could potentially alter the nuclear balance without requiring a full-scale nuclear arsenal.[61] Amid growing doubts about the credibility of US security guarantees and its shifting domestic and international priorities, for instance, US-allied states in East Asia such as Japan may seek to enhance their counter-NC3 capabilities. These include cyber capabilities and electronic attack capabilities that could extend to efforts to disrupt or neutralize key elements of an adversary's NC3 systems (e.g. radar and communications systems).[62] However, such moves to fill the vacuum left by weaker US guarantees could heighten strategic instability in the region, being perceived by other nuclear-armed states (e.g. China or North Korea) as escalatory.[63] Notably, AI-guided uncrewed aerial vehicles (UAVs)—with enhanced capabilities in target recognition and navigation in contested environments—that were deployed on a large scale by Ukrainian forces in 2024 and 2025 have targeted and damaged Russian early-warning radar sites, early-warning aircraft and strategic bombers.[64]

[61] Allen, G. and Chan. T., *Artificial Intelligence and National Security* (Harvard Kennedy School, Belfer Center for Science and International Affairs: Cambridge, MA, July 2017).

[62] Jones, C. A., 'Counter nuclear command, control, and communications', Institute for Security and Technology, 7 Nov. 2019; and Japanese Ministry of Defense (MOD), *2025 Defense of Japan 2025 Pamphlet* (MOD: Tokyo, 2025).

[63] Huang, J. (黄嘉瑜) and Ge, Y. (葛亚维), '警惕日本迈向"能战之国"' [Beware of Japan's move towards becoming a 'warlike nation'], *PLA Daily*, 24 Mar. 2025.

[64] Brown, S., 'Ukraine trained AI for its "Spiderweb" airfield drone attacks at aviation museum', *Kyiv Post*, 2 June 2025.

3. Human control in the current debate

As indicated in chapter 2, there is increasing agreement among both nuclear-armed and non-nuclear-armed states that there are risks to strategic stability posed by the integration of AI in nuclear and nuclear-related weapon systems. This helps to explain policy attention to the AI–nuclear nexus in recent years, with efforts to develop common understandings. To date, existing state approaches to governance have converged on the principle of retaining human control in nuclear decision making. At bilateral and multilateral levels, however, neither nuclear-armed nor non-nuclear-armed states have yet to fully elaborate which actions 'critical to informing and executing sovereign decisions concerning nuclear weapons employment' are to be kept in human hands.[65] This chapter disentangles views among officials and experts.

The decision-making process

While nuclear-use decisions remain the prism through which states discuss governance at the AI–nuclear nexus, states' views suggest the acknowledged impacts of a much wider range of events on those decisions. Language in official documents suggest the need for further exploration of AI's role and implications. In October 2024 the US White House issued its first-ever national security memorandum on AI, which suggested thorough safety testing of AI models and called for the administration to 'develop the capability to perform rapid systematic testing of AI models' capacity to generate or exacerbate nuclear and radiological risks'.[66] The UK cites the potential for AI to 'disrupt strategic paradigms', including through 'machine-speed escalation'.[67] While official Indian documents largely refrain from discussion of the AI–nuclear nexus, retired Indian military analysts have discussed the escalatory impact of integration of AI into nuclear systems and nuclear power facilities.[68]

These are prevalent themes, with numerous non-nuclear-armed states also expressing concerns about the increased likelihood of nuclear use specifically because of AI's role in sparking 'error and fast-tracked escalation'.[69] The states in the New Agenda Coalition observe the 'additional complexity these [technologies] introduce into deterrence calculations'.[70] The Stockholm Initiative similarly highlights multifaceted risks, with nuclear use stemming from potential misinterpretation linked to 'design flaws, unintended consequences, including from data, algorithmic and other biases, potential misuse or malicious use of the technology and the interaction of AI applications with the complex dynamics of global and regional conflicts and stability'.[71] In discussing risks linked to potential AI–nuclear integration, states often look beyond the nuclear-use decision itself to take into account the broader process that could drive nuclear confrontation.

[65] 10th NPT Review Conference, NPT/CONF.2020/WP.70 (note 25), para. 5(vii).

[66] The White House, 'Advancing the United States' leadership in artificial intelligence; harnessing artificial intelligence to fulfill national security objectives; and fostering the safety, security, and trustworthiness of artificial intelligence', US National Security Memorandum no. 25, 24 Oct. 2024, para. 3.3(f)(iii).

[67] British Ministry of Defence (MOD), *Defence Artificial Intelligence Strategy* (MOD: London, June 2022), p. 58.

[68] Ali (note 28); and Menon (note 28).

[69] 2026 NPT Review Conference, Preparatory committee, 'Measures to reduce the breadth of risks associated with nuclear weapons and measures to avoid increasing this risk', Working paper submitted by Austria et al., NPT/CONF.2026/PC.I/WP.24, 25 July 2023, para. 1.

[70] 2026 NPT Review Conference, Preparatory committee, 'Failure by the nuclear-weapon states to progress their nuclear disarmament obligations and commitments is contributing to unprecedented levels of nuclear risk', Working paper submitted on behalf of the New Agenda Coalition, NPT/CONF.2026/PC.II/WP.2, 16 May 2024, para. 8(i).

[71] 2026 NPT Review Conference, NPT/CONF.2026/PC.III/WP.35 (note 23), para. 8.

The ways in which nuclear-armed states define the process leading up to nuclear use is thus pertinent. As mentioned in chapter 2, this is often presumed to entail strategic warning, targeting and decision-support functions. At the unofficial level, experts in the policy research community have identified 'enduring features' of nuclear operations across all nuclear-armed states, including an affirmative decision to use nuclear weapons, high-level deliberations, varied data inputs from both human operators and computer systems, pre-planned military operations, and the presence of survivable assets.[72] At the official level, the USA has explicitly outlined intrinsic elements of decision making, for instance citing the need for approval of nuclear plans by the US Secretary of Defense, alignment with international law and the advisory role of senior officials.[73] Within their roles, US Strategic Command (STRATCOM) and combatant commands nominate, vet and select adversaries' strategic facilities and capabilities that are to be targeted by nuclear weapons.[74] India too has publicly detailed its nuclear command structure, with a Nuclear Command Authority (NCA) that includes a political council—the sole body that can authorize use—and an executive council—which provides inputs for decision making. Notably, India's cabinet committee on security reviews existing command and control structures, the state of readiness, the targeting strategy for a retaliatory attack and operating procedures for various stages of alert and launch.[75] These deliberative stages clearly appear as part of India's decision-making process.

Further orientation can be provided by specific conditions established in the declaratory policies of some nuclear-armed states. Russia's nuclear doctrine, updated in November 2024, presents a variety of key actions: the receipt of data on ballistic missile launches; judgement as to the criticality of threat posed by conventional weapons; and assessment as to adverse effects on critically important state or military infrastructure.[76] While Russia and three nuclear-armed states outside the NPT—India, Israel and North Korea—have not yet expressed a clear official stance on the AI–nuclear nexus, understanding these countries' nuclear decision-making processes can create a foundation for more concrete governance conversations.

Human control and judgement

The non-nuclear debate

The notion of human control, which plays a central role at the AI–nuclear nexus, has been discussed in great depth in a non-nuclear context: since 2014 there have been multilateral processes on lethal autonomous weapon systems (LAWS) in the United Nations General Assembly and under the framework of the 1980 Convention on Certain Conventional Weapons (CCW Convention).[77] In the LAWS context, the International Committee of the Red Cross (ICRC) set the standard to be that humans retain significant control over the critical functions of a weapon system: specifically, the selection and attacking of targets.[78] As summarized in a 2024 UN report on state deliberations,

[72] Durkalec, J., Péczeli, A. and Radzinsky, B., *Nuclear Decision-making, Complexity and Emerging and Disruptive Technologies: A Comprehensive Assessment* (European Leadership Network: London, Feb. 2022).

[73] US Department of Defense (DOD), *2022 National Defense Strategy of the United States of America* (DOD: Washington, DC, Oct. 2022).

[74] US Office of the Deputy Assistant Secretary of Defense for Nuclear Matters (note 9), chapter 2.

[75] Indian Ministry of External Affairs, 'The cabinet committee on security reviews [o]perationalization of India's nuclear doctrine', 4 Jan. 2003.

[76] 'Fundamentals of state policy of the Russian Federation on nuclear deterrence' (note 60).

[77] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects (CCW Convention, or 'Inhumane Weapons' Convention), opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983.

[78] Convention on Certain Conventional Weapons, Meeting of experts on lethal autonomous weapons systems, Statement by the International Committee of the Red Cross, 13 Apr. 2015.

states consider it necessary that humans retain 'sufficient information, including on the capabilities of the weapons system and the operational context, to ensure compliance with international law'. The report also lists necessary abilities for the human operators: to exercise judgement; to place limitations on types of tasks and targets as well as the duration, geographic scope and scale of use; to redefine or modify objectives or missions; and to interrupt or deactivate systems.[79]

Some nuclear-armed states have weighed in on human control in these settings. Notably, their stances do not relate to the broader integration of AI technologies, and there is a clear distinction in how some states view the use of conventional versus nuclear capabilities. There are, nonetheless, ways in which guiding principles in conventional weaponry have potential applicability to nuclear forces. France has underlined the importance of the 'human chain of command and control' and the maintenance of human responsibility and accountability.[80] Further, an opinion of the defence ethics committee of the French Ministry of Armed Forces on the use of AI in the military underlines as one of the main guiding principles defining 'clear chains of responsibility for command, control and execution regardless of the functions performed by a system incorporating artificial intelligence technologies'.[81] Pakistan's suggested provisions for a regulatory framework on LAWS echo the points on supervision and oversight in operations, listing a series of restrictions. It also cites the need for an 'effective oversight, investigative and redressal mechanism' at the national level.[82]

Within the LAWS debate, while there is general agreement that the ability for a human to intervene is the measure of control, there is also acknowledgment that modalities will largely be contextual or left to the discretion of states. Russia argues that any notion of 'meaningful' control or discussion of 'forms and degrees of human involvement' is inappropriate.[83] The UK notes that human involvement will vary depending on the 'nature of the capability, operational environment, and context of use', and the USA largely eschews control altogether and instead frames the issue as 'ensuring that machines help effectuate the intention of commanders and the operators of weapons systems'.[84] Indeed, the US Department of Defense, in a directive on autonomy in weapon systems, focuses on allowing 'commanders and operators to exercise appropriate levels of human judgment over the use of force'.[85] This emphasis on human judgement maintains that the term 'human in the loop' is misleading as it suggests continuous human interaction with a weapon platform from launch to impact. In practice, technical and operational constraints mean that such constant control is near impossible.[86]

The nuclear debate

Concerns about uncertainty and opacity linked to LAWS and general AI inclusion may be contrasted with differing and, at times, conflicting approaches to nuclear weapon systems. In the USA, for example, there is a distinction between the roles of human

[79] United Nations, 'Lethal autonomous weapons systems', Report of the secretary-general, A/79/88, 1 July 2024, para. 13.

[80] United Nations, General Assembly, First Committee, Reply received from France to Note Verbale no. ODA-2024-00019/LAWS on UN General Assembly Resolution 78/241, 24 May 2024.

[81] French Ministry of Armed Forces, Defence Ethics Committee, 'Opinion on the use of artificial intelligence technologies by the French armed forces', 14 Jan. 2025, p. 5.

[82] CCW Convention, Group of Governmental Experts on Emerging Technologies in the Area of LAWS, 'Elements of an international legal instrument on lethal autonomous weapons systems (LAWS)', Working paper submitted by Pakistan, CCW/GGE.1/2024/WP.7, 14 May 2024.

[83] United Nations, A/79/88 (note 79), p. 96.

[84] United Nations, A/79/88 (note 79), pp. 112, 115.

[85] US Department of Defense (DOD), 'Autonomy in weapon systems', DOD Directive no. 3000.09, 25 Jan. 2023, para. 4.1(c)(1).

[86] Horowitz (note 4).

control in non-nuclear and nuclear systems. While the 2022 US Nuclear Posture Review emphasizes that 'in all cases, the United States will 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', as noted above, a lower bar of human judgement is applied to advanced conventional systems.[87] Furthermore, Indian military analysts stress the need for human oversight, transparency and CBMs when arguing against the use of AI in the autonomous launch of nuclear weapons.[88] Similarly, the leader of Russia's Strategic Rocket Forces (SRF) has expressed the view that, when it comes to 'making a decision to use nuclear weapons, replacing a human is currently impractical' as AI technologies being introduced for automating even some routine operations in the SRF have not yet been 'fully studied and worked out'.[89] He stressed that nuclear weapons require 'constant human attention' due to their 'colossal power'.

Despite these indications that nuclear applications of human control must be treated differently from non-nuclear applications, there seems to be a lack of consensus on how it should be applied.[90] In part, the aversion to maintaining human control may stem from a desire to acquire strategic advantage. Some have argued this is the case with Russia, which—despite its joint calls with China for multilateral engagement on military AI applications within LAWS—has also argued that any conversation on AI governance frameworks is premature given the lack of common understanding.[91] For others, it may be a matter of focus. India and Pakistan, for instance, have largely explored military AI without explicit ties to the nuclear domain—with Pakistan's signing on to the REAIM Blueprint for Action being a notable exception. Their status as non-parties to the NPT presents additional hindrance to multilateral efforts to engage them on these issues, which in the case of Israel is exacerbated by that country's nuclear opacity.

Additionally, officials of non-nuclear-armed states have questioned the principle of 'human control and involvement' in relation to the use of nuclear weapons, leading to its omission in the updated version of the US-led 2023 Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy when the USA sought backing by other states.[92] In closed-door settings, these officials have cited concern that the focus on the human control principle may inadvertently legitimize human decisions to use nuclear weapons. Similarly, some experts argue that these commitments provide 'little more than an illusion of safety' without accompanying benchmarks and governance frameworks.[93] Moreover, without greater clarity on national positions on and definitions of human-in-the-loop, human-on-the-loop, human control and meaningful human control among both nuclear-armed and non-nuclear-armed states, such debates will continue to stymie progress on developing norms and standards that would underpin governance of the AI–nuclear nexus.

[87] US Department of Defense (DOD), '2022 Nuclear Posture Review', *2022 National Defense Strategy of the United States of America* (DOD: Washington, DC, Oct. 2022), p. 13.

[88] Ali (note 28); Menon (note 28); Patil and Rawat (note 28); and Nuclear Business Platform (note 28).

[89] Shakirov (note 27); and Biryulin, R., 'Надёжная опора безопасности и суверенитета России' [A reliable pillar of Russia's security and sovereignty], Interview of Col. Gen. Sergei Karakaev, SRF Commander, *Krasnaya Zvezda*, 17 Dec. 2024 (author translation).

[90] Su et al. (note 3).

[91] Joint statement by the Russian Federation and the People's Republic of China on Global Strategic Stability, President of Russia, 8 May 2025; Bendett (note 27); and United Nations, General Assembly, First Committee, Statement by Russia, 6 Nov. 2024.

[92] Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, US Department of State, 9 Nov. 2023; and Jenkins, B. D., US under secretary of state for arms control and international security, Remarks at the launch event for the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, 13 Nov. 2023.

[93] Saltini and Pan (note 4).

4. Pathways forward

Despite the potential for strategic advantage to be gained from the integration of AI into nuclear weapons, there is a consensus among both nuclear-armed and non-nuclear-armed states about AI's impact on nuclear deterrence and strategic stability. However, when it comes to human control in nuclear and nuclear-related weapon systems, views and approaches remain divergent. This chapter considers how governance at the AI–nuclear nexus could be further operationalized and how the debate on human control can be advanced.

Increasing understanding and awareness of risks

Despite their differences in understanding, there tends to be general recognition among state actors that AI integration into nuclear systems entails risks. In addition to their affirmation of the need for human control over the decision to use nuclear weapons, at their 2024 meeting presidents Biden and Xi stressed 'the need to consider carefully the potential risks'.[94] Similarly, the need for further exploration 'especially when applied to high-stakes areas such as nuclear technology and life-critical decisions' has been argued at meetings of the Global Partnership on Artificial Intelligence (GPAI)—an international initiative that aims to guide the responsible development and use of AI, which includes five nuclear-armed states (France, India, Israel, the UK and the USA) as members.[95] When asked about the use of AI in NC3, the British Ministry of Defence stated in 2025 that research to 'identify, understand, and mitigate against risks of AI in sensitive applications is underway'.[96]

Carrying these efforts a step further, nuclear-armed states could work in bilateral or multilateral settings to identify specific attributes of AI systems that they find destabilizing, including from the perspective of deterrence. They could then consider how such AI-related risks might manifest across different nuclear command-and-control functions and, correspondingly, means to mitigate these. Specific discussion of 'human control' or 'human judgement' in nuclear-related contexts, separate from the LAWS process, could be helpful. This could overcome cited barriers by focusing on specific functions and systems, such as NC3. A study initiated under the auspices of the UN disarmament machinery—such as the Conference on Disarmament or the First Committee of the General Assembly—could serve as a way of enabling this exchange, including by soliciting state views on priority risks stemming from the AI–nuclear nexus.

Any efforts to conduct an independent nuclear fail-safe review, such as that conducted by the USA in 2022, should consider risks stemming from technical failures of AI models and threats to information security.[97] Nuclear-armed states that engage in strategic cooperation could seek to develop common frameworks for assessing AI risks across the nuclear enterprise. This could fall under the umbrella of stockpile safety and security and could resemble past cooperation among nuclear-armed states, such

[94] The White House, 'Readout of President Joe Biden's meeting with President Xi Jinping of the People's Republic of China', 17 Nov. 2024.

[95] Indian Ministry of Electronics and Information Technology (MEITy), *Global IndiaAI Summit July 3–4, 2024: Report* (MEITy: New Delhi, 2024), p. 28.

[96] British Parliament, House of Lords, 'Nuclear weapons: Artificial intelligence—Question for Ministry of Defence', Written answer HL4966, 20 Feb. 2025.

[97] US Department of Defense (DOD), *2022 National Defense Strategy of the United States of America* (note 87), p. 13; and Dooling, M., 'Develop a risk assessment framework for AI integration into nuclear weapons command, control, and communications systems', Federation of American Scientists, 11 June 2025.

as with the development in the 1980s of permissive action link (PAL) devices to prevent arming or launching of nuclear weapons without prescribed codes.[98]

Formulating principles and standards

The focus on maintaining human control over nuclear-use decisions and related actions reflects the principle-based approach of states in governing the AI–nuclear nexus. Continuing in this manner could entail further efforts to universalize a declaration among nuclear-armed states, and beyond—given extended deterrence arrangements and potential allied consultations in nuclear decision making.[99] This could include revisiting the sentence on maintaining human control in nuclear decision making that was discussed but ultimately not included in the 2023 US-led Political Declaration.[100] States could consider elaborating principles specifically for nuclear-relevant contexts, including outside the realm of NC3. This would also be an opportunity to make the so-far Western-dominated conversation around responsible AI in the military domain more inclusive. Additional principles could centre on prudent and responsible development of AI technologies; thorough testing, evaluation, validation and verification standards; and regularity of review and updates. These would address the common concerns that nuclear-armed and non-nuclear-armed states have about AI reliability and vulnerability. Formulations bespoke to the AI–nuclear nexus can invoke potential catastrophic consequences in creating a clearer picture of rules of behaviour—the dos and don'ts—when it comes to AI–nuclear integration.

However, as suggested above, a principle-based approach cannot stand on its own. Effective governance requires accompanying practical action—for instance, through legal frameworks, regulatory standards, technical codes of conduct and export controls—to ensure legitimacy and efficacy. Given the sensitivities of the systems involved, much of this is likely to take place at the domestic level or in relations with contracted private companies. This could entail the development of AI-specific standards for safety and security across the life cycle of nuclear weapons, comparable in manner to policies and practices linked to the management of cyber risks.[101] Additionally, even within classified settings, state regulators can impose higher standards for AI systems that may be integrated in nuclear and nuclear-related weapon systems—this would essentially be modelled on the European Union (EU) framework established in the civilian space by the 2024 AI Act, which differentiates levels of risk and elaborates corresponding responsibilities on providers.[102] Along the same lines, major technology companies at the 2024 AI Seoul Summit on AI safety made a broad commitment to develop and implement what has been characterized as a kill switch should severe risks posed by a model or system be 'deemed intolerable'.[103] Such a shutdown mechanism applied in the nuclear domain could serve as a fail-safe to prevent unintended or escalatory use of force.

[98] Caldwell, D., 'Permissive action links: A description and proposal', *Survival*, vol. 20, no. 3 (May 1987).

[99] 'Sea, land, air, and NC3: Modernizing the whole nuclear enterprise', Transcript, Center for Strategic and International Studies (CSIS), 19 Nov. 2024.

[100] Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy (note 92).

[101] Dovgal, L., Su, F. and Saalman, L., 'Enhancing cyber risk reduction and the role of the European Union', SIPRI Research Policy Paper, Dec. 2024; and Saalman, L., Su, F. and Dovgal, L., *Cyber Risk Reduction in China, Russia, the United States and the European Union* (SIPRI: Stockholm, June 2024).

[102] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), *Official Journal of the European Union* L, 12 July 2024.

[103] British Department for Science, Innovation and Technology, 'Frontier AI safety commitments, AI Seoul Summit 2024', 7 Feb. 2025.

Addressing capabilities and actions

The focus of attention on NC3 to date reflects the convergence of risk priorities at the AI–nuclear nexus among at least some nuclear-armed and non-nuclear-armed states. Nevertheless, the complexity and variability of AI applications make it difficult to develop AI regulations that apply uniformly across all sectors; they therefore 'need to be sector-specific, enforcing strict rules in some sectors while allowing deviations in others'.[104]

A relevant approach could centre on renouncing or limiting the use of AI technologies in certain weapon systems.[105] For example, in 2023 members of the US Congress introduced a bill to prohibit use of government funds 'to launch a nuclear weapon using an autonomous weapons system', effectively linking the LAWS debate directly to the conversation on nuclear weapons.[106] Concern over nuclear LAWS has also been voiced by some non-nuclear-armed states in discussions in the UN General Assembly.[107] The possibility of autonomous nuclear weapon-delivery systems is not new but the widespread use of UAVs and uncrewed underwater vehicles (UUVs) and loitering munitions in the Russia–Ukraine War suggests the need to take steps to ensure that these systems are never nuclear armed, and also that they are not used to target nuclear infrastructure.[108] Relatedly, concerns have been raised about precision-guided munitions that are launched by humans but lack direct human supervision to the moment of impact.[109] Addressing all these capabilities can take different forms, including self- or mutual restraint or through export controls; for instance, in 2025 China announced a ban on the export of all civilian uncrewed vehicles for military purposes.[110]

Assuming that blanket prohibitions on capabilities are unlikely since they would have to overcome states' incentives to acquire strategic advantage, another approach would be arms control that focuses on behaviours and activities.[111] Controlled activities would centre on how AI integrated into military—and potentially nuclear—systems would operate as a means to prevent escalation. This could take inspiration from two cold war-era Soviet–US agreements: the 1972 Agreement on the Prevention of Incidents On and Over the High Seas (INCSEA) and the 1989 Agreement on the Prevention of Dangerous Military Activities (DMA).[112] Such an approach could outline procedures in the context of provocative and potentially destabilizing activities. Examples of such activities include the December 2023 Australian–British–US test use of AI algorithms

[104] Indian Ministry of Electronics and Information Technology (note 95), p. 28; and Carjueiro, D. and Celestino, V., 'A comprehensive review of artificial intelligence regulation: Weighing ethical principles and innovation', *Journal of Economy and Technology*, vol. 4 (2026).

[105] Wehsener, E. et al., *AI–NC3 Integration in an Adversarial Context: Strategic Stability Risks and Confidence Building Measures* (Institute for Security and Technology: Feb. 2023).

[106] US Congress, House of Representatives, 'Block Nuclear Launch by Autonomous Artificial Intelligence Act of 2023', Bill no. HR2894, 26 Apr. 2023; and US Congress, Senate, 'Block Nuclear Launch by Autonomous Artificial Intelligence Act of 2023', Bill no. S1394, 1 May 2023.

[107] United Nations, A/79/88 (note 79), p. 10.

[108] President of Russia, 'Presidential address to the Federal Assembly', 1 Mar. 2018; and de Troullioud de Lanversin, J., 'Ukrainian attack on Russian bombers shows how cheap drones could upset global security', *Bulletin of the Atomic Scientists*, 5 June 2025.

[109] Horowitz, M., 'The ethics & morality of robotic warfare: Assessing the debate over autonomous weapons', *Dædalus*, vol. 145, no. 4 (fall 2016).

[110] Wang, C. and Xiong, X., 'China expands export controls on drones, reflecting sense of responsibility amid conflicts', *Global Times*, 31 July 2024.

[111] Kühn, U. and Williams, H., 'Behavioral arms control and East Asia', *Journal for Peace and Nuclear Disarmament*, vol. 7, no, 1 (2024).

[112] Soviet–United States Agreement on the Prevention of Incidents On and Over the High Seas, signed and entered into force 25 May 1972, *United Nations Treaty Series*, vol. 852 (1972); and Soviet–United States Agreement on the Prevention of Dangerous Military Activities, signed 12 June 1989, entered into force 1 Jan. 1990, *United Nations Treaty Series*, vol. 1566 (1990).

to process sonar data as a means to track Chinese submarines; or China's use in May 2024 of the Taijing-4 03 radar imaging satellite (which is believed to be equipped with synthetic aperture radar and AI processors for rapid identification of targets) to capture images of a US Navy base in Norfolk, Virginia.[113]

Related, the central role of SSBNs in deterrence relationships among nuclear-armed states suggests a need to minimize undersea encounters. Yet AI and automation technologies are being developed and deployed in this domain with potentially significant implications for deterrence stability. Russia's alleged development of the Poseidon nuclear-armed autonomous UUV raises clear concerns.[114] Other nuclear-armed states are working on UUVs that can engage in anti-submarine warfare: China is alleged to be pursuing an Underwater Great Wall that can be stationed at chokepoints to work with crewed vessels to draw fire, expose adversaries' positions and even ram targets.[115] AI threatens to exacerbate the escalation risk linked to submerged operations.[116] A conversation on practical measures on safe conduct centred on AI-enabled conventional capabilities and autonomous systems can build on the voluntary 2014 Code for Unplanned Encounters in the Western Pacific, which China, France, Russia and the USA have joined.[117]

Signalling and transparency

Although there is no formal high-level political declaration on retaining human control in nuclear decision making among the P5, they have broadly accepted this principle as an implicit norm as technological development progresses. It remains important to make continued efforts to seek explicit confirmation and reaffirmation from the P5—and potentially all nuclear-armed states—on this front. At the same time, it may be valuable to identify a concrete and easily achievable goal for the near term that can have an immediate impact and signal continued interest in advancing discussions at the AI–nuclear nexus.

One such goal could involve an international ban on autonomous nuclear weapon systems. This might need to apply to all dual-capable platforms (i.e. conventional arms that can carry nuclear weapons) in order to avoid states circumventing the spirit of the commitment by simply classifying their systems as conventional. Although it is thought that no autonomous nuclear system is yet in operation, such as system would represent a significant and foreseeable common risk to strategic stability. Addressing this now would be a pragmatic and forward-looking approach, recognizing that technological development often outpaces governance efforts.

Another area of focus in the short-to-medium term could be the development of CBMs aimed at reducing the risk of nuclear escalation in the context of emerging AI capabilities. This could begin with enhancing and adapting existing crisis communication hotlines. The Moscow–Washington hotline established in 1963 after the Cuban missile crisis is often cited as a means to facilitate direct communication

[113] Capaccio, A., 'US Navy, UK, Australia will test AI system to help crews track Chinese submarines in the Pacific', Bloomberg, 2 Dec. 2023; and Honrada, G., 'China's AI makes its satellites spies in the sky', Asia Times, 11 Apr. 2022.

[114] Kaur, S., 'One nuclear-armed Poseidon torpedo could decimate a coastal city. Russia wants 30 of them', *Bulletin of the Atomic Scientists*, 14 June 2023.

[115] Wong, C., '"Underwater Great Wall": Chinese firm proposes building network of submarine detectors to boost nation's defence', *South China Morning Post*, 19 May 2016; and Sutton, H. I., 'China's new extra-large submarine drones revealed', Naval News, 16 Sep. 2022.

[116] Erästö, T., Su, F. and Wan, W., *Navigating Security Dilemmas in Indo-Pacific Waters: Undersea Capabilities and Armament Dynamics* (SIPRI: Stockholm, June 2024); and Anthony, I., Su, F. and Saalman, L., 'Naval incident management in Europe, East Asia and South East Asia', SIPRI Insights on Peace and Security no. 2023/03, Mar. 2023.

[117] Code for Unplanned Encounters at Sea, version 1.0, Western Pacific Naval Symposium, 22 Apr. 2014.

between the leaders of nuclear-armed states; and a 2013 agreement resulted in the creation of a communications link for cyber activities.[118] States could consider bilateral hotlines dedicated to addressing AI-related operations, helping to reduce ambiguity and prevent misperceptions—which could be particularly relevant in the case of India and Pakistan.

Building on these foundations, nuclear-armed states could move towards greater transparency at the AI–nuclear nexus by issuing national declarations that outline their strategies on AI integration and affirm their commitment to the safe and responsible use of AI in the nuclear weapon domain. To strengthen and legitimize these commitments, multilateral reinforcement—such as a declaration endorsed by the UN Security Council—could play a role in promoting further dialogue and transparency. Ultimately, these efforts could facilitate multilateralization of the other steps discussed, including for instance shared best-practices for safety protocols and approaches among nuclear-armed states, establishing common standards for safety and performance in the integration of AI into nuclear systems.

Considering the choice of forum

Movement to strengthen governance at the AI–nuclear nexus entails a host of decisions linked to forums, as well as the complications that this might entail. The poor condition of geopolitical relations, especially among nuclear-armed states, and the general lack of trust creates fundamental obstacles to any multilateral efforts. Additional difficulties are linked to the lack of a common technological baseline, given significant asymmetries in capabilities and levels of integration of military AI across states, even before the inclusion of the nuclear issue. Creating a dedicated forum for AI–nuclear governance is not realistic given these factors, while existing forums—on both sides of the nexus—have structural constraints and encounter questions of inclusivity and representativeness. For instance, the P5 format and the NPT exclude some nuclear-armed states, and the LAWS debate is bound by the CCW Convention.

In the short term, discussing risk reduction linked to the AI–nuclear nexus in bilateral and multilateral formats may hold the most promise, given constructive dialogue around the risk-reduction framing, interest in maintaining strategic stability and modest success in the formulation of concrete measures to take forward, including on crisis communication and information notification. Forums such as the REAIM summits and the working groups of the Creating an Environment for Nuclear Disarmament (CEND) initiative can serve as incubators for ideas. They can also facilitate further exchange of views on these topics, involving different configurations of nuclear-armed and non-nuclear-armed states, while expanding the conversation beyond NC3 and its sensitivities. These forums may also serve as stepping stones to more focused conversations, for instance among the P5 or eventually through the establishment of a UN body, whether at the General Assembly or in the context of a group of governmental experts or open-ended working group. All of these can help formalize understandings of risk linked to the AI–nuclear nexus and channel these into restraint.

Harnessing AI to reduce nuclear risk

While standing slightly apart from the above, joint action by nuclear-armed states in utilizing AI as a means to reducing nuclear risk can provide another pragmatic way

[118] Soviet–US Memorandum of Understanding Regarding the Establishment of a Direct Communications Link, signed and entered into force 20 June 1963, *United Nations Treaty Series*, vol. 472 (1963); and Nakashima, E., 'US and Russia sign pact to create communication link on cyber security', *Washington Post*, 17 June 2013.

forward. In the USA, AI already augments key aspects of maintaining the safety, security and reliability of the nuclear stockpile through modelling and simulations; transforming the nuclear weapon life cycle from the discovery of new materials to design optimization, manufacturing and ongoing surveillance; and enhancing non-proliferation through detecting and characterizing illicit nuclear activities worldwide with the analysis of vast quantities of data from diverse sources.[119] Similarly, in India there are indications that AI has aided in missile production; in life cycle management and adaptability upgrades from inertial navigation to laser-guided systems; and to circumvent such countermeasures as electronic jamming and uncrewed systems.[120] Ensuring the reliability of AI in these and other functions (e.g. safeguarding critical facilities) will bolster strategic stability.

Moreover, private enterprises can contribute significantly to reducing nuclear risk. For example, some work by OpenAI and Microsoft at the USA's Los Alamos National Laboratory is 'focused on reducing the risk of nuclear war and securing nuclear materials and weapons worldwide'.[121] Anthropic, an AI software company, has partnered with the US National Nuclear Security Administration (NNSA) to develop an AI classifier that identifies potentially harmful nuclear-related conversations.[122] In China, AI-enhanced inspection robots are used for predictive maintenance and safety monitoring at Hongyanhe Nuclear Power Plant in Liaoning province.[123]

This joint action could also include stakeholders exploring ways in which AI can have an impact on the larger information ecosystem in which nuclear decision makers operate. For example, a 2023 report sponsored by the US Department of State recommends further education and training for policymakers, decision makers and diplomats on AI and best practices in both the public and private sectors.[124] This kind of education and training, including on how to detect AI-generated content, can be especially pertinent at the AI–nuclear nexus for diplomatic and military actors. Furthermore, the proliferation of AI-generated deepfakes and disinformation during the May 2025 armed conflict between India and Pakistan underscores potential nuclear consequences even without direct AI integration in nuclear weapon systems.[125] Systematic study of direct and indirect effects on nuclear escalation of such events can expand understanding of the spectrum of risks linked to the AI–nuclear nexus.

[119] US Department of Energy, 'Artificial intelligence for national security', [n.d.].

[120] AFI, 'DRDO deploys AI for ballistic missile manufacturing', Indian Defence Research Wing, 6 Mar. 2025.

[121] Field, H., 'OpenAI partners with US National Laboratories on scientific research, nuclear weapons security', NBC News, 30 Jan. 2025.

[122] Anthropic, 'Developing nuclear safeguards for AI through public–private partnership', 21 Aug. 2025.

[123] Hongyanhe Nuclear Power Plant, '红沿河首个智能巡检机器人上线' [Hongyanhe's first intelligent inspection robot goes live], China National Nuclear Power Network, 27 Mar. 2025.

[124] Wehsener et al. (note 105).

[125] Chaudhuri, P. and Higgins, E., 'India–Pakistan conflict: How a deepfake video made it mainstream', Bellingcat, 9 May 2025.

5. Conclusions

The accelerating integration of AI into nuclear and nuclear-related weapon systems presents both opportunities for improved decision making and grave risks to global stability. While AI can enhance data processing and situational awareness, its potential to introduce errors, amplify automation bias and shorten decision timelines raises the risk of inadvertent nuclear escalation. These risks are compounded by the opacity of NC3 systems, the dual-use nature of AI and other emerging technologies and growing asymmetries between states.

Existing state approaches to the nexus between AI and nuclear forces broadly converge on the principle of retaining human control in nuclear decision making. However, key differences persist regarding which specific actions must remain under human authority and how human judgement should be exercised across complex nuclear decision chains. While some states have begun to articulate the elements of the decision-making process—including targeting, strategic assessment and authorization—many have yet to clarify their positions on how AI may support or constrain these functions. Variations in doctrine, transparency and institutional structure further complicate efforts to establish shared standards. Importantly, without concrete operational definitions, accountability mechanisms and technical benchmarks, any commitment to 'human control' may not offer meaningful safeguards.

To translate concern into action, states should develop operational principles and standards at the national level—such as on domestic oversight, life cycle safety protocols and regulation of high-risk AI use. Industry partnerships and built-in safety mechanisms can also promote responsible integration. States might also consider limits or bans on AI use in nuclear-delivery systems and other high-risk platforms, through self-restraint, export controls or arms control initiatives. Targeted CBMs—such as transparency on national policies on the application of AI in nuclear and nuclear-related weapon systems and crisis-communication hotlines focused on AI-enabled systems incidents—can reduce ambiguity and misperception.

Given current geopolitical tensions and asymmetries in AI capabilities and risk perceptions, governance at a multilateral level will be challenging. In the near term, leveraging existing forums and inviting states' submissions constitute pragmatic action that could help build momentum towards a more formal regulatory framework.

About the authors

Fei Su is a Researcher with the SIPRI China and Asia Security Programme. 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 emerging technologies.

Wilfred Wan is the Director of the SIPRI Weapons of Mass Destruction Programme. His recent research has focused on nuclear weapon risk reduction, nuclear disarmament verification and other issues related to arms control and disarmament. He also serves as the co-chair of the Alva Myrdal Centre working group on nuclear disarmament in policy and international law.

Lora Saalman is an Associate Senior Fellow at SIPRI. She was formerly vice-president of the Asia-Pacific Program at the EastWest Institute and director of SIPRI's China and Global Security Programme. She was the first US citizen to earn a doctorate from the Tsinghua University Department of International Relations, completing her coursework entirely in Chinese.

Vladislav Chernavskikh is a Research Assistant in the SIPRI Weapons of Mass Destruction Programme, focusing on nuclear disarmament and non-proliferation issues. He is also a contributor to SIPRI-led activities under the Alva Myrdal Centre working group on nuclear disarmament in policy and international law.

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE
Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org