SIPRI Insights on Peace and Security
No. 2025/06 June 2025

IMPACT OF MILITARY ARTIFICIAL INTELLIGENCE ON NUCLEAR ESCALATION RISK

VLADISLAV CHERNAVSKIKH AND JULES PALAYER*

SUMMARY

● Increasing integration of artificial intelligence (AI) into military systems has the potential to influence nuclear escalation even when that integration occurs outside nuclear weapon systems. Non-nuclear applications of military AI may compress decision-making timelines, potentially increasing miscalculation risks during a crisis. Opaque recommendations from an AI-powered decision-support system can bias a decision-maker towards acting, while autonomy in a system with counterforce potential may undermine strategic stability by threatening the integrity of second-strike capabilities.

Such uses of AI raise the fundamental question of whether they introduce new risks, exacerbate existing ones or fundamentally alter the nature of nuclear escalation. Contextual and socio-technical factors that might affect nuclear escalation pathways can help to answer this question. Understanding these dynamics is essential for using current risk-reduction measures or developing new strategies to address nuclear escalation risks posed by military AI.

The deterioration of the international security environment over the past decade has resulted in heightened concerns about the risk of nuclear war as nuclear-armed states and their allies renew their reliance on nuclear deterrence.[1] All nine nuclear-armed states—China, France, Israel, India, North Korea, Pakistan, the Russian Federation, the United Kingdom and the United States—are modernizing their nuclear forces, with some also increasing the size of their nuclear arsenals.[2] At the same time, these states are investing in the development and deployment of military artificial intelligence (AI), using it to pursue better and faster decision-making and increased autonomy in military systems.[3]

Just over half of nuclear-armed states have stated explicitly that they intend to maintain human control over decisions related to the use of nuclear weapons in order to minimize the risk of nuclear escalation.[4] Yet, even when used in non-nuclear-related military applications, military AI may have an impact on the operating environment in which nuclear decision-making takes place—for example, in a crisis situation, it may reduce the time required for threat detection and response coordination.[5] If military decision-support systems (DSSs) and automated capabilities push policymakers and military decision makers towards faster reaction times, this may increase the risk of miscalculation and errors and may affect decisions that could result in nuclear escalation. However, the ways in which military AI can influence the

[1] SIPRI, 'Role of nuclear weapons grows as geopolitical relations deteriorate—New SIPRI Yearbook out now', 17 June 2024.

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

[3] Chernavskikh, V., 'Nuclear weapons and artificial intelligence: Technological promises and practical realities', SIPRI Background Paper, Sep. 2024; Boulanin, V. et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020), p. 16; and Probasco, E. S. et al., 'AI for military decision-making: Harnessing the advantages and avoiding the risks', Issue brief, Georgetown University, Center for Security and Emerging Technology (CSET), Apr. 2025.

[4] Responsible AI in the Military Domain (REAIM) Summit, 'REAIM Blueprint for Action', 9–10 Sep. 2024, para. 5; Chinese Ministry of Foreign Affairs, 'President Xi Jinping meets with US President Joe Biden in Lima', 17 Nov. 2024; and 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, para. 5(vii).

[5] Boulanin et al. (note 3); and 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.

* This publication was made possible by a grant from Carnegie Corporation of New York. The statements made and views expressed are solely the responsibility of the authors.

risk of nuclear escalation—even AI that is not directly integrated into nuclear systems—remain relatively underexplored. This raises the fundamental question of whether AI introduces new risks, exacerbates existing ones or fundamentally alters the nature of nuclear escalation.

This paper initiates a deeper exploration of this fundamental question. It describes what can drive AI-related risks of nuclear escalation by highlighting the interplay between socio-technical impacts of AI and other contextual factors that can affect nuclear escalation. The paper continues in section I by establishing the conceptual baseline for understanding nuclear escalation risk in the age of military AI, underscoring the relevance of non-nuclear military applications of AI. Section II then explores the impact on nuclear escalation risk of specific types of AI system—DSSs and autonomy in systems with counterforce potential. Section III concludes by summarizing the findings and indicating a potential direction for future policy-oriented research aimed at addressing this risk through existing or novel risk-reduction measures.

I. Nuclear escalation risk in the age of military AI

Nuclear escalation

Nuclear escalation can be defined as the intensification or expansion of a conventional conflict to the extent that it crosses what one or more parties perceives to be a critical threshold, ultimately culminating in the use of nuclear weapons.[6] Typically, the literature differentiates between three kinds of escalation: (*a*) deliberate, when a state intends for escalation to occur; (*b*) inadvertent, when a state did not anticipate that its actions would lead to escalation, probably because its actions crossed a rival's threshold; and (*c*) accidental, when escalation is the result of mistaken or unauthorized actions.[7]

Strategic stability is usually defined as the situation in which nuclear-armed states have no incentive to initiate a first nuclear strike.[8] When strategic stability is undermined, a nuclear-armed state may be more likely to deliberately escalate to the use of nuclear weapons because it considers that the benefits of doing so outweigh the costs.[9] For example, if a state's assessments suggest that a nuclear attack on it is imminent, it may see a pre-emptive strike on the opponent's nuclear assets as the best available option. Similarly, there may be incentives to use nuclear weapons to either stop a major conventional offensive or to pressure an adversary into ending the conflict to avoid defeat.[10] A nuclear-armed state that perceives its conventional forces to be inferior to its adversaries' may be more inclined to use nuclear weapons during a military conflict to avoid losing a conventional war.[11] Nuclear escalation can also be

[6] Radin, A., Demus, A. and Evans, A. T., 'A vocabulary of escalation: A primer on the escalation literature for military planners', RAND Corp., 2024.

[7] Hoffman, W. and Kim, H. M., 'Reducing the risks of artificial intelligence for military decision making', Policy brief, Georgetown University, Center for Security and Emerging Technology (CSET), Mar. 2023.

[8] Boulanin et al. (note 3), p. 6.

[9] Larsen, E. H., 'Deliberate nuclear first use in an era of asymmetry: A game theoretical approach', *Journal of Conflict Resolution*, vol. 68, no. 5 (May 2024).

[10] Post, D. R., 'Escalating to de-escalate with nuclear weapons: Research shows it's a particularly bad idea', *Bulletin of the Atomic Scientists*, 9 Feb. 2024.

[11] Larsen (note 9).

triggered through misperception, miscalculation or misunderstanding. For example, it may result from a judgement based on faulty information, from a conventional military strike that unintentionally compromises an adversary's nuclear deterrent or from a technical malfunction of an early-warning system.[12]

The threshold for nuclear escalation depends largely on notions of retaliation in response to some form of attack. This threshold may be different depending on a particular state's declared policies, nuclear force structure, command-and-control arrangements, deployment and readiness levels, and doctrinal criteria for when and how it would employ nuclear weapons. A state (e.g. China) with a declared no-first-use policy—in which it declares that it would use nuclear weapons only in response to a nuclear attack—may be less inclined to escalate.[13] Other states (e.g. Russia and the USA) reserve the option to use nuclear weapons first in response to a broad range of nuclear and non-nuclear threats to national security.[14] Alliances involving nuclear-armed states—such as the North Atlantic Treaty Organization (NATO) or the Union State of Belarus and Russia—add additional layers of complexity as extended nuclear deterrence practices allow non-nuclear-armed states under the nuclear umbrella of a nuclear-armed state to rely on the potential use of nuclear weapons in response to acts of aggression against any member of the alliance.[15] Another factor is the scale of the conflict—conflicts involving multiple theatres and actors tend to create more dynamic environments, where misinterpretations and rapid changes in the security situation can heighten the escalation risk. Beyond this, even the geographic proximity of nuclear-armed states—as in the case of India and Pakistan—can be a factor that affects the probability of escalation. On the one hand, a state may be reluctant to use nuclear weapons on a neighbour due to radioactive fallout effects in its own territory—not to mention the devastating global humanitarian and environmental impacts that are likely to result from even a limited nuclear exchange.[16] On the other hand, the shorter transit time for nuclear weapon-delivery systems reduces the decision-making window, increasing the pressure to respond to a perceived nuclear attack with a retaliatory strike.[17]

[12] van Hooft, P., Ellison, D. and Sweijs, T., *Pathways to Disaster: Russia's War against Ukraine and the Risks of Inadvertent Nuclear Escalation* (The Hague Centre for Strategic Studies: The Hague, May 2023); and Johnson, J., 'Inadvertent escalation in the age of intelligence machines: A new model for nuclear risk in the digital age', *European Journal of International Security*, vol. 7, no. 3 (Aug. 2022).

[13] Chinese Ministry of Foreign Affairs, 'No-first-use of nuclear weapons initiative', 23 July 2024.

[14] US Department of Defense (DOD), *2022 National Defense Strategy* (DOD: Washington, DC, Oct. 2022); and 'Fundamentals of state policy of the Russian Federation on nuclear deterrence', approved by Russian Presidential Order no. 991, 19 Nov. 2024.

[15] Erästö, T., 'Reducing the role of nuclear weapons in military alliances', SIPRI Insights on Peace and Security no. 2024/01, June 2024; and 'Russia ratifies treaty with Belarus on security guarantees', TASS, 28 Feb. 2025.

[16] Sethi, M. (ed.), *Understanding Pathways to Nuclear Escalation in Southern Asia* (Centre for Air Power Studies: New Delhi, Nov. 2024); and International Committee of the Red Cross (ICRC), 'Humanitarian impacts and risks of use of nuclear weapons', 29 Aug. 2020.

[17] Rubler, M. R., 'Nuclear deterrence destabilized', eds B. Unal et al., *Perspectives on Nuclear Deterrence in the 21st Century* (Chatham House: London, Apr. 2020), p. 16.

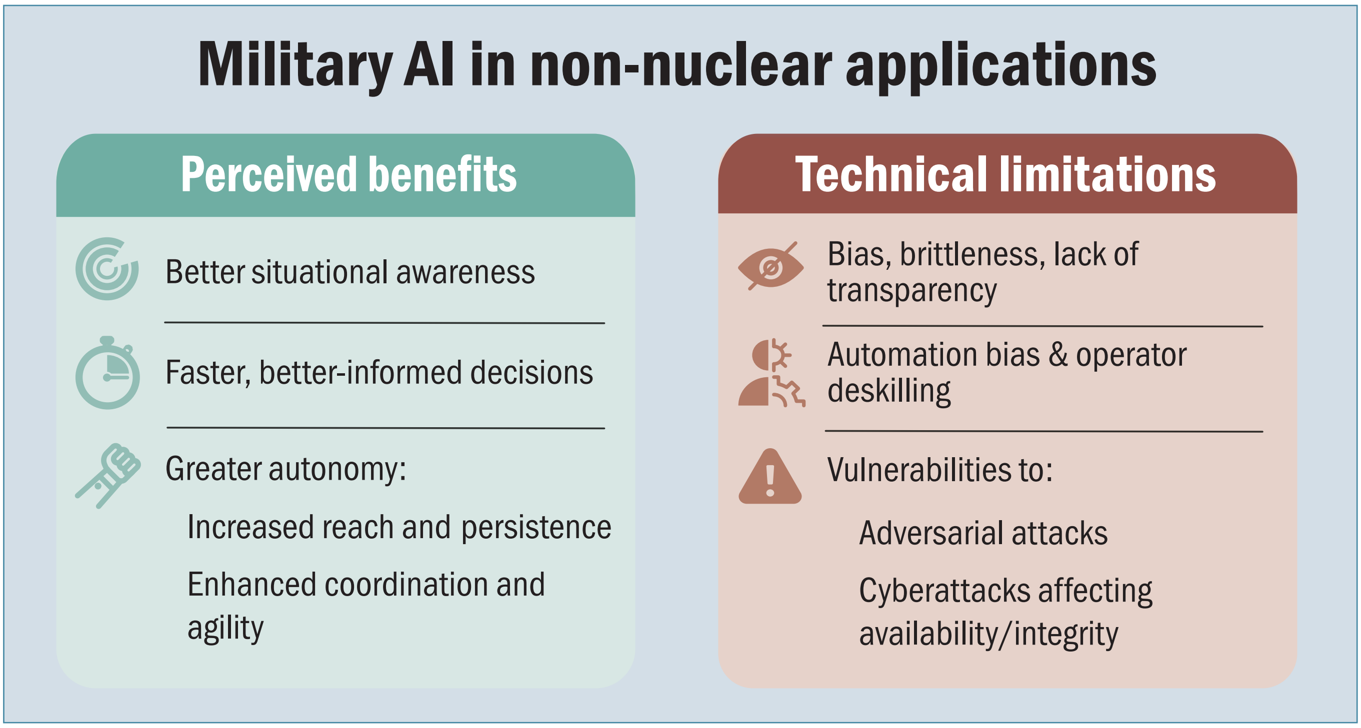

Figure 1. Perceived benefits and technical limitations of military artificial intelligence in non-nuclear applications

The influences of military AI on nuclear escalation risk

The ways in which military applications of AI may drive nuclear escalation risk in practice depend on what type of AI and where, how and to what end it is integrated. There are three areas of military AI integration that could exacerbate nuclear escalation risk. The first is the introduction of AI in nuclear command, control and communications (NC3).[18] This may include systems involved in early threat detection, targeting and decision-making on nuclear weapon use. The second is the use of AI-enabled technology, especially autonomy, in nuclear-delivery platforms.[19] For example, Russia's Poseidon (also known as Status-6) is a nuclear-armed uncrewed underwater vehicle (UUV) that will reportedly operate autonomously when deployed.[20] The third area—and the focus of this paper—is the uses of military AI-enabled systems in non-nuclear applications.[21]

Non-nuclear applications of military AI are relevant because conventional capabilities are often entangled with nuclear capabilities—they may, for example, rely on the same means of delivery, command-and-control assets

[18] Reinhold, T. et al., 'Artificial intelligence, non-proliferation and disarmament: A compendium on the state of the art', Non-proliferation and Disarmament Papers no. 92, EU Non-proliferation and Disarmament Consortium, Jan. 2025, section IV.

[19] Horowitz, M. C., 'Artificial intelligence and nuclear stability', ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019), p. 80. See also Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017).

[20] Topychkanov, P., 'Autonomy in Russian nuclear forces', ed. Boulanin (note 19), pp. 74–75; and 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.

[21] Boulanin, V., 'The future of machine learning and autonomy in nuclear weapon systems', ed. Boulanin (note 19), p. 59.

or support systems.[22] This entanglement is present in offensive and defensive capabilities across traditional military domains of land, sea and undersea, and air, as well as in cyber and outer space. For example, the United States uses early-warning satellites to detect both nuclear and non-nuclear attacks and they can trigger ballistic missile defences; it also uses nuclear-capable stealth bombers for conventional anti-ship strikes.[23] Another example is Russia's dual-capable Iskander system, which launches short-range ballistic and cruise missiles that may carry nuclear or conventional warheads.[24] The use of AI for non-nuclear military applications may therefore amplify the existing risks stemming from entanglement across all military domains. This is especially likely if AI systems create the perception—both for the user of such a system and its adversary—that an offensive operation involving these non-nuclear capabilities is likely to be more effective.[25] For example, a state might leverage AI to conduct a sophisticated cyberattack on a space system that supports both conventional and nuclear weapons, causing the targeted state to respond as if its nuclear deterrent capability were under attack. [26]

Perceived benefits and technical realities of military AI in non-nuclear applications

In pursuit of faster and better-informed decisions, militaries increasingly rely on AI-enabled systems to enhance decision-making. This involves leveraging vast amounts of data to gain superior situational awareness. AI is also a driving force behind greater autonomy in military systems—it contributes to autonomy through advancements in computer vision and machine perception. Autonomy is highly valued by armed forces because it extends operational reach, increases persistence, enhances agility and enables better coordination.[27] These are especially crucial in cyberwarfare scenarios (e.g. the above-described example of a cyberattack on a space asset) and in electronic warfare, where communications links between operators and platforms may be disrupted or compromised.

However, the technical reality is that AI systems come with significant limitations that can lead to critical failures (see figure 1). These include bias, brittleness and lack of transparency.[28] In high-stakes contexts such as those involving critical military decision-making, these flaws pose serious risks. Such problems as AI hallucinations, automation bias and operator deskilling could result in misidentification of threats, potentially triggering

[22] Acton, J. M., 'Escalation through entanglement: How the vulnerability of command-and-control systems raises the risks of an inadvertent nuclear war', *International Security*, vol. 43, no. 1 (summer 2018); and Stokes, 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).

[23] Acton (note 22); and Parthemore, C. and Dill, C., 'Paint the B-52s brightly: Reducing confusion between conventional and nuclear weapons is essential', War on the Rocks, 18 Sep. 2024.

[24] Parthemore and Dill (note 23).

[25] Stokes et al. (note 22), p. 5.

[26] Boulanin et al. (note 3), p. 107; Raju, N. and Wan, W., 'Escalation risks at the space–nuclear nexus' SIPRI Research Policy Paper, Feb. 2024, p. 17; and Raju, N., 'Parameters to assess escalation risks in space', SIPRI Research Policy Paper, Feb. 2025, pp. 3–4.

[27] Boulanin et al. (note 3), p. 16.

[28] Hoffman and Kim (note 7); Chernavskikh (note 3); and Boulanin, V., 'Risks and benefits of AI-enabled military decision-making', eds R. Geiß and H. Lahmann, *Research Handbook on Warfare and Artificial Intelligence* (Edward Elgar: Cheltenham, 2024).

unintended escalation consequences. Furthermore, increasing military reliance on AI introduces significant vulnerabilities that can be exploited by adversaries. Machine learning-based AI systems are particularly susceptible to attacks that compromise their integrity, which could lead them to make critical errors in threat assessment.[29] Adversaries could also target the confidentiality of AI models, extracting sensitive data.[30] Such breaches could have severe military consequences, enabling adversaries to locate and target key facilities, personnel or assets. Additionally, cyberattacks on AI-enabled systems could degrade their availability, thereby delaying responses or rendering systems inoperative at crucial moments, further increasing the potential for miscalculation in high-stakes scenarios.[31]

II. The impact of military AI on nuclear escalation risk

To further unpack the potential of non-nuclear military AI to affect nuclear escalation risk, this section identifies and examines two key applications of military AI—AI-driven decision-support systems (AI-DSSs) and AI integration in conventional systems with counterforce potential—and places them within the broader context of nuclear escalation dynamics.

Military AI decision-support systems

AI-enabled DSSs process vast amounts of data—collected through an increasing and diverse number of sensors, compiled databases and open-source intelligence—to assist decision makers at the strategic, operational and tactical levels.[32] Their influence on military decision-making—and especially on nuclear escalation—deserves more attention as these tools are altering how, where and when critical tactical and operational decisions are made.

In the military, AI-DSSs fulfil one or more of three primary functions: descriptive, predictive and prescriptive.[33] Descriptive systems organize and present data in a structured format to improve a military unit's situational awareness; predictive systems analyse patterns to anticipate potential future events; and prescriptive systems generate recommendations on possible courses of action. Each of these functions shapes how information is processed, interpreted and acted upon by decision makers, thus introducing distinct nuclear escalation risks.

AI-DSSs shape how policymakers and military decision makers perceive and interpret information. The way in which information is presented may influence a decision maker to draw escalatory conclusions, even in cases in which AI is not explicitly prescriptive. This has consequences for escalation risk especially because of the brittleness of the technology and the outright

[29] Puscas, I., *AI and International Security: Understanding the Risk and Paving the Path for Confidence-building Measures* (UNIDIR: Geneva, 2023); and Lohn, A. J., *Hacking AI: A Primer for Policymakers on Machine Learning Cybersecurity* (Georgetown University, Center for Security and Emerging Technology (CSET): Washington, DC, Dec. 2020).

[30] Puscas (note 29); and Lohn (note 29).

[31] Puscas (note 29); and Lohn (note 29).

[32] Boulanin (note 28).

[33] Bode, I., 'Human–machine interaction and human agency in the military domain', Policy Brief no. 193, Centre for International Governance Innovation (CIGI), Jan. 2025, p. 6.

errors for which machine learning-based AI is known. Furthermore, beyond hallucinations, the human–DSS interfacing—such as the user interface of the system, the formatting style and the choice of phrasing in the outputs of text-based systems—could end up shaping the strategic choices of political and military leaders.[34] The increased ease with which humans can interact with AI systems through natural language processing and the wide scope of capabilities that some general-purpose models display may easily lead an operator to overestimate a system's capabilities.[35]

The lack of transparency or 'black box' nature of advanced AI models may obscure the rationale behind certain outputs.[36] The opacity of AI-DSSs risks obscuring critical contextual details, potentially leading a decision maker to act on a recommendation that lacks a nuanced understanding of the situation—thereby increasing the likelihood of escalatory outcomes. For example, a 2024 study on fictional nuclear escalation scenarios found that large language models (LLMs)—a type of AI—often chose more escalatory options than human participants.[37] This might be due to bias in training data or to the fact that AI lacks human empathy and may more readily make critical escalatory decisions as it is not subject to moral consideration or social norms, such as the nuclear use taboo. However, the lack of transparency in this system prevents understanding of why it selected the most escalatory options.

Another risk exacerbated by AI-DSSs is automation bias—the tendency of users to accept a system output without critical scrutiny. As AI systems increasingly incorporate machine learning and become more opaque, it becomes harder for operators to interpret or challenge the system's reasoning and outputs. This is not a new problem in the military context, which has been integrating automated systems that can lead to automation bias for a long time.[38] In fact, even when a system's logic is fully understood by its users, they may still be unable to thoroughly assess the output due to the time-sensitive nature of the decisions or the complexity of the data being processed. This increases the likelihood that flawed or unchecked decisions could be taken solely based on AI suggestions.

Another cause for concern is related to predictive AI systems, which analyse data to anticipate future events. While these systems might excel in predictions based on physical laws (e.g. predicting the impact point of a ballistic missile), they may fail when their predictions are based on less observable data, inferring thoughts and intentions from actions.[39] In other words, AI is good at solving puzzles (i.e. making sense of data and collating it in a way that makes sense for humans), but not so good at resolving

[34] Andersen, R., 'Never give artificial intelligence the nuclear codes', *The Atlantic*, 2 May 2023; and Boyd, B. L. and Saade, T., 'Human-centered warfare: Optimizing human cognitive autonomy and avoiding machines on the loop', Killer Robot Cocktail Party, Sep. 2024.

[35] Probasco et al. (note 3), p. 19.

[36] Erskine, T. and Miller, S. E., 'AI and the decision to go to war: Future risks and opportunities', *Australian Journal of International Affairs*, vol. 78, no. 2 (2024).

[37] 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).

[38] Hawley, J. K., *Patriot Wars: Automation and the Patriot Air and Missile Defense System* (Center for a New American Security: Washington, DC, Jan. 2017).

[39] Probasco et al. (note 3).

mysteries (e.g. predicting intentions or national resolve).[40] This limitation of AI is especially relevant for nuclear escalation risk given the importance of signalling in deterrence and credibility of intentions. The historical record of near misses in nuclear crises suggests that, in some cases, human judgement has played a crucial role in preventing catastrophic miscalculations. The well-known case of Soviet officer Stanislav Petrov, who chose not to report what turned out to be a false nuclear alarm, underscores the importance of human judgment in high-stakes decision-making.[41] In contrast, at least for now, AI lacks the ability to exercise doubt, recognize social norms such as the nuclear taboo, or intuitively weigh the moral and strategic consequences of escalation.

Ultimately, the use of AI-DSSs in military decision-making affects nuclear escalation risk not so much by removing humans from the process entirely, but by undermining the conditions needed to exercise human input in the chain of intelligence gathering, analysis and communication.[42] Automation bias and excessive trust in AI-generated insights may lead officials to misjudge adversary intentions, heighten threat perceptions and increase nuclear escalation risk.[43] This displacement of human judgment is particularly concerning when it comes to the operating environment in which nuclear decision-making takes place. If AI-generated insights and recommendations end up saturating strategic assessments, nuclear escalation pathways may become more opaque and more prone to unintended consequences to the detriment of strategic stability. AI can also affect crisis instability by accelerating the speed of decision-making processes.[44] The rapid pace of AI-driven warfare could compress the timelines of decision makers, increasing the likelihood of misperception and overreaction.

This is particularly relevant in cases when a state involved in a conflict maintains its deployed nuclear forces on a high level of alert—able to be launched within minutes in what is known as a launch-on-warning posture (e.g. Russia and the USA). Conversely, nuclear-armed states without a launch-on-warning policy or capability (e.g. France) and states with a no-first-use nuclear policy (e.g. China and India), which typically store nuclear warheads separately from deployed launchers during peacetime, would take longer to launch, thereby reducing the risk of escalation due to miscalculation or misunderstanding exacerbated by AI-DSS.[45]

A further contextual factor to consider here is the variety in national models for authorizing the use of nuclear weapons among nuclear-armed states. Some states (e.g. France, Russia and the USA) rely on a 'sole authority' model, where the leader of the country makes a unilateral decision to carry out a nuclear strike; other states (e.g. China) may require authorization

[40] Goldfarb, A. and Lindsay, J. R., 'Prediction and judgement: Why artificial intelligence increases the importance of humans in war', *International Security*, vol. 46, no. 3 (winter 2021/22).

[41] E.g. Topychkanov (note 20), box 8.1.

[42] Zala, B., 'Should AI stay or should AI go? First strike incentives & deterrence stability', *Australian Journal of International Affairs*, vol. 78, no. 2 (2024), p. 157.

[43] Erskine and Miller (note 36).

[44] Andersen (note 34).

[45] Kristensen and Korda (note 2); and Arms Control Association, 'Arms control and proliferation profile: France', Aug. 2024.

from a collective decision-making body.[46] Furthermore, once a decision to use nuclear weapons has been made, the chain of command structure that executes the order differs from state to state, depending on legal, military and political practices.[47] These differences are especially relevant when considering how the use of AI-DSS will affect the roles of humans involved in military decision-making and how, in practice, human oversight over decisions leading up to nuclear weapon use might be retained. Multilayered human authorization at critical decision points before any nuclear weapon use is one of the measures that could help mitigate nuclear escalation risks that stem from both technical failures and the cognitive biases that AI decision-support tools have the potential to introduce.

Autonomy in military systems with counterforce potential

AI contributes to the automation of and autonomy in critical steps in various military systems, such as intelligence, surveillance and reconnaissance (ISR), targeting and missile guidance, and force delivery. In particular, increased autonomy is expected to enhance the counterforce effect of systems that already have counterforce potential—that is, the potential to target an adversary's military capabilities, including any second-strike nuclear capability (e.g. mobile missile systems, strategic submarines and the command-and-control infrastructure required for retaliation). These autonomous counterforce systems include precision-strike capabilities and AI-enhanced ISR and target acquisition, which can improve a state's ability to locate and neutralize an adversary's nuclear forces, potentially threatening the integrity of second-strike capabilities. Thus, by potentially making nuclear forces more vulnerable to attack and less reliable for retaliation, the increasing sophistication of AI-enhanced and autonomous non-nuclear capabilities may undermine strategic stability.[48] This perception can create pressure on a nuclear-armed state to take pre-emptive action in a crisis, including launching nuclear weapons first to preserve the integrity of its second-strike capabilities. At the same time, states integrating military AI may be emboldened to use autonomy in conventional systems to target and strike nuclear capabilities of a nuclear-armed adversary.

Furthermore, AI-enhanced autonomy is likely to exacerbate the destabilizing effects of emerging and disruptive technological developments in areas such as long-range precision-strike weapons, missile defence, counterspace capabilities and cyber, pushing nuclear-armed states to adopt riskier nuclear postures.[49] It may drive these states to modernize, expand and diversify their nuclear forces and broaden the range of conditions for nuclear use in

[46] Lewis, J. and Tertrais, B., 'Pressing the button: How nuclear-armed countries plan to launch armageddon (and what to do about the US)', War on the Rocks, 24 Apr. 2019; and Born, H, Gill, B. and Hänggi, H. (eds), SIPRI, *Governing the Bomb: Civilian Control and Democratic Accountability of Nuclear Weapons* (Oxford University Press: Oxford, 2010).

[47] Lewis, J. G. and Tertrais, B., *The Finger on the Button: The Authority to Use Nuclear Weapons in Nuclear-Armed States*, James Martin Center for Nonproliferation Studies (CNS) Occasional Paper no. 45 (Middlebury Institute for International Studies: Monterey, CA, Feb. 2019).

[48] Zala (note 42).

[49] Horowitz (note 19); and Futter, A., 'Explaining the nuclear challenges posed by emerging and disruptive technology: A primer for European policymakers and professionals', Non-proliferation and Disarmament Papers no. 73, Mar. 2021.

an attempt to counter perceived strategic threats—creating more space for misperceptions regarding 'red lines' for nuclear escalation. Russia's revised nuclear doctrine of 2024 puts greater emphasis on 'aerospace attacks' as one of the conditions under which nuclear weapons may be used, explicitly including uncrewed aerial vehicles (UAVs), alongside aircraft and missiles—reflecting Russia's concern over escalating UAV incursions amid the Russia–Ukraine war.[50] This move was underscored by several instances in which Ukraine used AI-enabled UAVs to strike Russian strategic bomber bases and strategic early-warning radar.[51]

This dynamic is even more relevant for nuclear-armed states that rely on smaller and less diversified nuclear forces (e.g. India, North Korea and Pakistan). Even with the integration of AI-enabled systems, achieving a successful counterforce strike still requires significant conventional or nuclear capabilities to neutralize hardened and dispersed targets (e.g. strategic submarines, silo fields or multiple airbases). Despite advances in AI and other emerging technologies, in the near term key elements of the second-strike capabilities of technologically advanced nuclear-armed states are, in practice, likely to remain survivable.[52] Therefore, large and more diverse nuclear arsenals are likely to remain relatively secure, while smaller forces may indeed become more vulnerable. For example, one of the factors driving China's nuclear build-up and shifts in its nuclear strategy is probably related to concerns about the viability of its nuclear deterrent in the face of the USA's advanced conventional capabilities, which are expected to become increasingly autonomous.[53]

III. Conclusions

Assessing the effect of military AI on nuclear escalation risk is highly complex due to the potential for AI to be integrated in various systems and at various levels across the military domain. This complexity is compounded by the variety of drivers of nuclear escalation, sources of nuclear risk and the fragile balance of strategic stability. As a first step towards developing a more comprehensive understanding of these escalation dynamics—which is a prerequisite for designing effective de-escalation and risk-reduction measures—this paper identifies two key types of use of AI in military systems for non-nuclear applications that can influence nuclear escalation and outlines some contextual factors that will further shape nuclear escalation dynamics.

The integration of AI in the military domain through AI-driven decision-support systems and by automating tasks of non-nuclear systems with counterforce potential both have the potential to undermine strategic stability and drive nuclear escalation. Military AI could generate greater ambiguity around capabilities and intentions and create incentives for states to escalate

[50] 'Fundamentals of state policy of the Russian Federation on nuclear deterrence' (note 14).

[51] Kirichenko, D., 'The rush for AI-enabled drones on Ukrainian battlefields', Lawfare, 5 Dec. 2024; 'Ukraine strikes Russian strategic bomber airfield, triggering huge blast', Reuters, 20 Mar. 2025; and Horowitz, M. C., 'Ukraine's Operation Spider's Web shows future of drone warfare', Council on Foreign Relations, 3 Jun. 2025.

[52] Glaser, C. L., 'The end of MAD? Technological innovation and the future of nuclear retaliatory capabilities', *Journal of Strategic Studies*, vol. 48, no. 2 (2025).

[53] Stokes et al. (note 22).

in conflict, including up to nuclear use.[54] AI integration may also increase the likelihood of nuclear escalation originating from the technological limitations of the AI system itself, via malfunctions, design flaws, unintended consequences (e.g. from data, algorithmic and other biases) and the potential exploitation by adversaries of cyber vulnerabilities of the technology.[55] Nuclear escalation risks addressed above are derived from the potential of AI to alter where, how and by whom critical military decisions are taken, and AI's potential to enhance autonomy in counterforce capabilities, which increases incentives for both the state employing AI and its adversary to take escalatory steps during a crisis, thereby increasing risk of nuclear use.

This reinforces the need for caution when using AI in military decision-making as its influence on escalation dynamics remains complex and potentially destabilizing. Developing measures to mitigate the risk of nuclear use requires further examination of how military applications of AI can influence nuclear escalation risk in particular contexts and specific crisis scenarios. In this regard, it is paramount that awareness of this issue is raised among governmental, military, expert and civil society stakeholders from nuclear-armed states and their allies. Elaborating on scenarios in which AI influences nuclear escalation can help inform the adaptation of existing risk-reduction measures or the development of new ones at the AI–nuclear nexus.

These scenarios, for example, could highlight potential approaches to developing training standards for users of military AI. In turn, these approaches could establish redundancies that enable cross-checking of information from multiple sources and systems in order to prevent the deferral of responsibility to AI and ensure that operators are able to critically assess and challenge AI-generated recommendations. Scenarios could also contribute to the development of additional transparency and confidence-building measures such as military hotlines and notification regimes that can help mitigate nuclear risk stemming from both AI-driven decision-support systems and the integration of AI in conventional systems with counterforce potential.

[54] Boulanin et al. (note 3).

[55] 2026 Non-Proliferation Treaty 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 member states of the Stockholm Initiative for Nuclear Disarmament, NPT/CONF.2026/PC.III/WP.35, 25 Apr. 2025.

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

SIPRI INSIGHTS ON PEACE AND SECURITY NO. 2025/06

IMPACT OF MILITARY ARTIFICIAL INTELLIGENCE ON NUCLEAR ESCALATION RISK

VLADISLAV CHERNAVSKIKH AND JULES PALAYER

CONTENTS

ABOUT THE AUTHORS

Vladislav Chernavskikh is a research assistant with the SIPRI Weapons of Mass Destruction Programme, focusing on nuclear disarmament and non-proliferation. He is also a contributor to SIPRI-led activities under the Alva Myrdal Centre working group on nuclear disarmament in policy and international law. Chernavskikh's recent publications include 'Nuclear weapons and artificial intelligence: Technological promises and practical realities', SIPRI Background Paper (2024).

Jules Palayer is a researcher with the SIPRI Governance of Artificial Intelligence (AI) Programme, where he focuses on understanding the impact of AI on international peace and security aiming to inform policy decisions in the realm of disarmament, arms control and non-proliferation. His recent publications include 'La inteligencia artificial en perspectiva comparada' [Artificial intelligence in a comparative perspective], *Geopolítica del poder militar* [Geopolitics of military power] (Spanish Ministry of Defence, 2024, co-author).

DOI: https://doi.org/10.55163/FZIW8544