POLICY BRIEF

APLN

How Artificial Intelligence Impacts Deterrence Stability: A Realistic Assessment

JANUARY 2026

MANPREET SETHI

Please direct inquiries to:
Asia-Pacific Leadership Network
APLN Secretariat
4th floor, 116, Pirundae-ro
Jongno-gu, Seoul, ROK, 03035
Tel. +82-2-2135-2170
Fax. +82-70-4015-0708
Email. apln@apln.network

This publication can be downloaded at no cost at www.apln.network.

Cover Photo: Artificial Intelligence and nuclear decision-making. Credit: iStock- guirong hao.

HOW ARTIFICIAL INTELLIGENCE IMPACTS DETERRENCE STABILITY: A REALISTIC ASSESSMENT

Manpreet Sethi

INTRODUCTION

Artificial Intelligence or AI is an umbrella concept. It encompasses anything that uses an algorithm that gives a machine somewhat of 'a mind of its own'. One recent report defines it as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings."[1] But the intelligence of the machine is likely to be constrained by the training or data that it has received. Its decision-making capability may also be limited by the lack of emotions and an ability to see things in context. These deficiencies have caused a good deal of apprehension about AI systems being entrusted with decision making about nuclear weapons. Hence, the insistence on the 'human in/on the loop' in nuclear command, control and communications (NC3) processes and decisions.[2]

Despite a wide acceptance of this dictum for human participation in nuclear decision-making, the reality of contemporary times is that AI enabled systems are increasingly part of nuclear architectures. Even if indirectly, these are present in other peripheral systems that feed into the NC3, such as intelligence gathering, data processing, and target identification. While these constitute use of AI for decision support, with the ultimate decision being left to human discretion, nevertheless the final human-made decision would be influenced by the kind of data and the way it has been gathered and interpreted by the AI-enabled systems. It is for this reason that military applications of AI, even when they are supposedly out of nuclear command and control structures, are yet so innate to them and are expected to impact the stability of nuclear deterrence, especially in crisis situations.

Deterrence stability could be disrupted by factors such as geopolitical dynamics, personality of leadership, or domestic compulsions. But it is equally prone to disruptions by technologies that offer a first strike advantage. Since the stability of nuclear deterrence is based on the ability to do reciprocal harm or mutual vulnerability,

[1] Tim McDonnell, Mary Chesnut, Tim Ditter, Anya Fink, Larry Lewis, "Artificial Intelligence in Nuclear Operations: Challenges, opportunities, and Impact," with contributions by Annaleah Westerhaug, *CNA Research Memorandum*, Center for Naval Analyses, April 2023, https://www.cna.org/reports/2023/04/Artificial-Intelligence-in-Nuclear-Operations.pdf

[2] "Humans on the Loop vs. In the Loop: Striking the Balance in Decision-Making," *Trackmind,* 12 February 2025, https://www.trackmind.com/humans-in-the-loop-vs-on-the-loop/

any development that allows one side to cause harm without the fear of retaliation, for instance a 'splendid first strike', can give rise to crisis instability.[3]

The use of AI across a number of military applications has ignited such fears. In fact, there are four dimensions in which use of AI enabled systems can have an adverse impact on stability of nuclear deterrence. These are briefly discussed in the following section.

AI APPLICATIONS AND IMPACTS ON STRATEGIC STABILITY

Exacerbating fears around survivability of second-strike forces

The stability of nuclear deterrence rests on the assumption of being able to respond in a way that the first user achieves no advantage because the retaliation would to inflict comparable or worse damage. It is for this reason that nations put so much emphasis on survivability of nuclear arsenals as a measure of credible deterrence. Capabilities that impact the survivability of nuclear forces and disrupt the assuredness of a second-strike capability cause crisis instability. They may compel the targeted country to resort to pre-emptive nuclear postures leaning towards weapons being held at high alert in peacetime, thereby exacerbating greater nuclear risks that inevitably accompany such hair-trigger postures.

AI enabled military applications may increase risks to survivability of second-strike forces in two ways – by improving intelligence gathering for detection of the location of nuclear forces; and by enabling better counterforce targeting leading to fear of loss of retaliatory capability. AI can improve intelligence, surveillance and reconnaissance capabilities by collecting and analysing critical information from more sources, fusing data sets from different domains, identifying correlations, and making connections including between targets and weapon systems. And, it can perform all these tasks at great speed to better inform decision makers in shorter timeframes.

Since AI can process and analyse vast amounts of data from multiple sensors promptly, a military equipped with AI could be more capable of finding, tracking and targeting adversary's nuclear assets quickly. It is even envisaged that automated pattern recognition from multiple sensors could increase the transparency of mobile land-based missiles or even track nuclear submarines (SSBNs) in the deep oceans. Since mobility and sea-based deterrence are currently seen as important attributes of survivability, a fear of their loss can trigger crisis instability.

So, the employment of AI-enabled intelligence, surveillance, and reconnaissance (ISR), such as autonomous sensor systems or automated target recognition exacerbates the fear that the opponent possesses the potential to launch a devastating first strike. This could

[3] A 'splendid first strike' is one that can disarm the adversary of their nuclear capability or decapitate their command and control in a manner that disables any retaliation.

make nations lean towards riskier postures of hair-trigger readiness; or compel the development of countermeasures to out-manoeuvre or confuse ISR efforts. For instance, Russia's AI-enabled doomsday drone, known as the Status-6 Oceanic Multipurpose System, or the Poseidon, is an autonomous vehicle that could be launched from a submarine and possesses the intelligence to elude oceanic defences to threaten delivery of a nuclear payload even in face of US conventional or nuclear counterforce or anti-submarine warfare (ASW) capabilities.[4]

Increasing fears around the robustness of NC3

NC3 is a sensitive part of nuclear deterrence and any interference with this, or threat thereof, can upset deterrence stability. In this regard, AI advances in quantum computing that can break NC3 encryption methods pose the risk of compromising the positive and negative controls of the system. Encryption ensures that only authorised individuals can access sensitive information such as nuclear codes to launch nuclear weapons. This is a critical way of enforcing negative controls to prevent unauthorised access and tampering. Quantum technology could, however, compromise the security of communication channels, allowing unauthorised access and interception. This may be used to cause deliberate miscommunication to disrupt the system. Any perception that such an act may be undertaken can increase the risk of nuclear escalation.

In another use of AI enabled cyber operations the fear of loss of sensitive information on nuclear forces would be as destabilising as the use of disinformation and deepfake technologies. Deepfakes are synthetic videos or audio recordings that are manipulated to create a false reality, which can be used to deceive and mislead individuals or groups. Disinformation campaigns can manipulate public opinion and increase tensions between nuclear-armed nations by making it easier for hostile actors to spread false narratives and create fictitious events – such as an AI generated video of a nuclear threat by a national leader – which could exacerbate military escalation.

Compressing decision-making timelines

In every conflict situation, the attempt of the belligerents is to know more and know faster. The speed of decision-making and execution could be both an asset and a liability. AI systems can expedite data acquisition and analysis from multiple sources to provide a relatively comprehensive overview to the national leadership to allow more time for decision making. But this speed could also raise the tempo of conflict by putting pressure on leaders to make decisions quickly thereby, reducing time for them to craft and consider alternatives for their consequences. This in turn could raise the risk of inadvertent escalation.

[4] Aditya Kumar, "Why U.S. Has No Defense Against Russia's Poseidon Nuclear Torpedo," *Defense News,* 31 October 2025, https://www.thedefensenews.com/news-details/Why-US-Has-No-Defense-Against-Russias-Poseidon-Nuclear-Torpedo/

Autonomous systems that can execute the six-step decision-making process – find, fix, track, target, engage, and assess (F2T2EA) – more quickly than humans could also compel them to climb the rungs of escalation ladder much faster, thereby complicating the chances of climb down. For example, real-time automatic target recognition (ATR), which utilizes deep-learning techniques to identify multiple targets efficiently, can reduce lag time for execution. As cautioned by a Centre for Global Security Research report, "the speed at which AI guided ISR could direct and execute kinetic operations could limit options for de-escalation."[5] Quick decisions also shrink the time for potential political or diplomatic actions to resolve a crisis.

In a crisis involving nuclear-armed states, slowing things down would ideally be more prudent. It may be recalled that during the Cuban Missile Crisis, President Kennedy chose the option of enforcing a naval blockade because it gave time for both sides to politically resolve the issue. AI enabled systems, on the other hand, ratchet up the tempo. Moreover, they could also divorce the decision making from the political or human context. However smart or intelligent it may be, AI would still lack human experience, intuition, context analysis capability, and creativity in its assessments. Nuclear deterrence is a mind game; its practice, therefore, is innately human. Machines could play this game clinically, and using artificially fixed conditions and expectations, raising the possibilities of deterrence instability and even breakdown. This is because no crisis is ever the same, and crisis de-escalation requires the creativity, policy innovation, and human understanding and intuition that AI lacks.

Inflating perceptions of the adversary's ability to 'win'

The mere perception that a rival has made advancements in AI systems that could offer a strategic advantage through enhanced nuclear offensive and defensive capabilities can foster suspicion, driving the other side to adopt a posture of nuclear pre-emption to avoid the potential neutralisation of its deterrent force. As stated in a 2018 report published by the RAND Corporation, "AI may be strategically destabilizing not because it works too well but because it works just well enough to feed uncertainty."[6] Perception of a large gap in the AI capabilities of the adversary could risk escalation. AI have-nots may then adopt an asymmetric escalation posture. On the other hand, a false belief in the superiority of AI enabled military systems could induce greater confidence in their use. When caught in deep geopolitical divisions and high trust deficits, nations can

[5] Zachary Davis and Michael Nacht (eds), *Strategic Latency: Red, White, and Blue - Managing the National and International Security Consequences of Disruptive Technologies*, Center for Global Security Research, Lawrence Livermore National Laboratory, February 2018, https://cgsr.llnl.gov/sites/cgsr/files/2024-08/STATEGIC_LATENCY_Book-WEB.pdf

[6] Edward Geist and Andrew J Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?," *Perspective, Perspective,* RAND Corporation, 24 April 2018, https://www.rand.org/content/dam/rand/pubs/perspectives/PE200/PE296/RAND_PE296.pdf

respond to perceived fears of how AI could undercut their nuclear deterrence and rush to build countermeasures, each action triggering new security dilemmas.

MITIGATING RISKS OF AI APPLICATIONS

As is evident, AI-enabled applications can exacerbate crisis instability. However, to some extent, nuclear deterrence remains insulated to some of the risks of deterrence instability despite these advancements, since there can never be any certainty that a first strike can prove to be effectively disarming or completely incapacitating for an adversary. The inability to rule out the possibility of retaliation despite the enhanced efficacy of an AI-enabled first strike keeps the foundation of nuclear deterrence intact. It is therefore in the interest of all nations advancing AI enabled applications to recognise and address their inherent risks and limitations.

Of course, the application of AI in warfare is an irreversible development and its usage is only likely to increase. It is also inevitable that human trust in machines will increase. In such a situation, some steps must be envisaged to ensure that the employment of each AI application is intelligently assessed and managed by humans for their benefits and risks for nuclear deterrence. Humans should be guiding the pace and direction of technology instead of being led by the technological possibilities. Human oversight over excessive autonomy should be the preferred option.

To prioritise this, there has been discussion on responsible use of AI in military applications (REAIM) and in the nuclear domain. In September 2024, the inter-governmental REAIM conference in Seoul affirmed the principle of human control in the interface between AI and NC3. This statement was signed by sixty countries of whom only three – the United States, the United Kingdom and France – were nuclear weapon states.[7] In November 2024, President Xi Jinping and President Joe Biden affirmed the need to address the risks of AI systems, improve AI safety and international cooperation, and to maintain human control over the decision to use nuclear weapons.[8] Ideally, all nuclear armed countries should commit to retaining human responsibility over nuclear decision-making, including the decision support and communication systems.

None of this can, of course, be verifiable. But it could, in fact it must, arise from an understanding that it is nobody's interest to exacerbate fears that lead to crisis

[7] Joyce Lee, "Sixty countries endorse 'blueprint' for AI use in military; China opts out," *Reuters,* 10 September 2024, https://www.reuters.com/technology/artificial-intelligence/south-korea-summit-announces-blueprint-using-ai-military-2024-09-10/

[8] Jarrett Renshaw and Trevor Hunnicutt, "Biden, Xi agree that humans, not AI, should control nuclear arms," *Reuters,* 17 November 2024, https://www.reuters.com/world/biden-xi-agreed-that-humans-not-ai-should-control-nuclear-weapons-white-house-2024-11-16/

instability. Some commonsensical steps, therefore, would be in mutual interest of all. These, inter alia, can be briefly identified as:

- Air gapping the launch command from the decision support and early warning systems;
- Taking a conscious decision to use AI only as a decision support tool and not as a replacement for human judgement, potentially prohibiting AI from generating pre-determined policy options, leaving these for human decision-makers;
- Prohibiting the use of autonomous weapons systems for nuclear delivery;
- Building greater latitude for human intervention in all stages to slow down processes where necessary;
- Banning the development, deployment, or use of AI applications capable of malicious manipulation of data in NC3 systems;
- Prohibiting cyberattacks on NC3 systems and other critical crisis communication channels;
- Ensuring availability of secure and trusted electronic and human crisis communication channels to enable deescalation.

All of the above steps, even if taken unilaterally but with a certain level of transparency, can mitigate the risks generated through increasing use of AI-enabled military applications. If bilateral or multilateral agreements could be achieved, it would be even better. But given the contemporary geopolitical tensions, these appear difficult. Nevertheless, crisis stability should be in the interest of all nuclear armed states. It would be of little use to create conditions or perceptions that hasten the nuclear weapon use out of fear of loss of the ability to retaliate. And nations with first use doctrines that believe that their AI-enhanced ability can carry out a splendid first strike and obviate retaliation may end up making a huge miscalculation.

The basic guarantee of nuclear deterrence for nations that have built reasonably secure second-strike capabilities lies in the uncertainty they have created for the adversary to remove retaliation from the equation. Over confidence induced by new technologies needs to be tempered by the fundamentals of nuclear deterrence that do not essentially change much. Therefore, intelligence must first be exercised by humans to use it effectively and gainfully in its artificial form. Stability of nuclear deterrence, and of human survival, depends on this understanding.

ABOUT THE AUTHOR

Manpreet Sethi, PhD is Senior Research Adviser at APLN and a Distinguished Fellow at the Centre for Aerospace Power and Strategic Studies (CAPSS), New Delhi, where she leads the project on nuclear security. She is an expert on a range of nuclear issues, having published over 130 papers in academic journals of repute. Over the last 25 years she has been researching and writing on subjects related to nuclear energy, strategy,

non-proliferation, disarmament, arms and export controls, and BMD. Sethi has been a member of the International Group of Eminent Persons set up by Mr Fumio Kishida, former Prime Minister of Japan, to identify pathways to a nuclear weapons free world. She is member Science and Security Board of the Bulletin of Atomic Scientists that has the responsibility of time setting on the Doomsday Clock.

ABOUT APLN

The **Asia-Pacific Leadership Network (APLN)** is a Seoul-based organization and network of political, military, diplomatic leaders, and experts from across the Asia-Pacific region, working to address global security challenges, with a particular focus on reducing and eliminating nuclear weapons risks. The mission of APLN is to inform and stimulate debate, influence action, and propose policy recommendations designed to address regional security threats, with an emphasis on nuclear and other WMD (weapon of mass destruction) threats, and to do everything possible to achieve a world in which nuclear weapons and other WMDs are contained, diminished, and eventually eliminated.


@APLNofficial


@APLNofficial


apln.network


aplnofficial