AI, NC3, and Strategic Stability: Risks and Opportunities

SCENARIO EXERCISE

Workshop Design Team: Sylvia Mishra, Christian Steins, Philip Reiner and Doug Randall

How will the integration of novel artificial intelligence into global nuclear command, control, and communication systems over the next five years transform strategic stability and deterrence dynamics?

To begin to answer this question, on April 10, 2025, the *Institute for Security and Technology (IST)* hosted the workshop, **"The Risks and Opportunities of AI in NC3: Finding Common Ground"** in Washington, DC.

This day-long, scenario-driven workshop convened more than 60 senior officials, technical experts, and civil society actors for an in-depth analysis of this crucial national security challenge. Together, the participants identified risks and opportunities presented by the potential integration of cutting-edge artificial intelligence capabilities into key dimensions of the nuclear command, control, and communications (NC3) enterprise. As these AI tools rapidly evolve in their capabilities and applications, the workshop explored outstanding strategic, operational, and tactical level considerations at this critical technological intersection through carefully crafted scenarios designed to illuminate key policy and technical considerations.

Facilitated groups explored four hypothetical worlds featuring varying levels of AI integration, governance structures, and human oversight in NC3 operations, designed to illuminate risks and opportunities through various combinations:

- The AI Stability Zone
- AI Arms Race
- Human Controlled Stability
- Asymmetric Instability

Throughout the day, participants analyzed the implications of AI integration in NC3 operations within each of those worlds, focusing specifically on three key dimensions of NC3 operations:

- Strategic Warning
- Adaptive Targeting
- Decision Support systems

To facilitate an open and thoughtful discussion, the workshop drew on hypothetical scenarios to avoid classification constraints.

Instructions

Please note that while scenario descriptions have limitations due to time and classification constraints, we urge you not to "fight the scenario."

The scenarios are purposefully crafted to propose alternative futures, and we ask that you allow yourself to suspend disbelief within reason.

Therein, within these alternative future scenarios, your expertise will help our strategic community identify diplomatic and technical solutions to advance nuclear strategic stability.

Critical Uncertainties

To what extent will AI be integrated into global NC3 systems?

What will be the state of international norms and agreements governing AI, and AI in nuclear operations and NC3 specifically?

Disclaimer

About the Institute for Security and Technology

The Institute for Security and Technology (IST) is the 501(c)(3) critical action think tank that unites technology and policy leaders to create solutions to emerging security challenges.

IST stands at the forefront of convening policymakers, technology experts, and industry leaders to identify and translate discourse into impact. We take collaborative action to advance national security and global stability through technology built on trust, guiding businesses and governments with hands-on expertise, in-depth analysis, and a global network.

info@securityandtechnology.org · securityandtechnology.org

Expanded Scenario Narratives

Scenario 1: AI Stability Zone

While AI is increasingly integrated into NC3 (Nuclear Command, Control, and Communications) systems, its deployment is governed by a complex framework of international norms governed by multilateral institutions. Spurred by close calls due to AI hallucinations involving near-miss nuclear outcomes in Ukraine and Taiwan, global leaders were jolted into collaborative action. A groundbreaking declaratory stance—initiated by the United States and China and signed by all nuclear-armed states (except for the Democratic People's Republic of Korea and Israel)—has established foundational assurances and predictability for AI's role in NC3. North Korean leader Kim Jong Un separately offered unilateral assurances in a speech, largely in line with the multilateral declaratory stance.

AI is both deeply embedded and carefully regulated within NC3 systems, guided by widely accepted principles of transparency, accountability, and human-centric design. This stability, however, is contingent not only on technical safeguards within NC3 systems, but also on the broader geopolitical environment, where ongoing cooperation in other domains reinforces the consensus around AI restraint. There is adherence to the bold commitment to maintain human-in-the-loop oversight at all stages to ensure effective human-machine teaming in nuclear decision-making, and to avoid automation bias in final decisions. To support this, robust accountability mechanisms stemming from advanced industry technology standards—such as zero-knowledge proofs—have been implemented to ensure traceability and compliance.

The primary applications of AI in NC3 are comparable to those used in conventional warfare: situation monitoring, planning, decision support, and force management. Machine learning models process vast volumes of intelligence data, improving speed and accuracy in threat detection while reducing false alarms. Strategic warning systems (in the states advanced enough to have them) now integrate real-time satellite, radar, and cyber intelligence to enhance early detection of nuclear threats. AI-powered decision support tools assist commanders in analyzing large data streams and refining response options while ensuring that launch authority remains exclusively human-controlled. Yet under the surface, some strategists question whether human-in-the-loop commitments can withstand crisis-level compression, raising concerns about overreliance on automated cues under duress.

To maintain stability, states have drawn clear red lines: AI is prohibited from autonomously altering or recommending target lists in nuclear operations. Its use is permitted to enhance force readiness and logistics, optimizing deterrence (by improving readiness and logistics, but not targeting, for example) without increasing the risk of escalation. Diplomatic agreements reinforce transparency and AI-focused confidence-building measures, ensuring that AI enhances strategic stability rather than triggering unintended crises. It surprises many, but the world has successfully rallied around a shared commitment to the responsible use of AI.

Scenario 2: AI Arms Race

AI is rapidly being integrated into NC3 (Nuclear Command, Control, and Communications) systems—but without internationally agreed-upon or even discussed controls or technical safety standards. Emerging deployments of AI during recent conflicts—such as Ukraine, Gaza, and Taiwan Strait tensions—have shown how quickly states and non-state actors are experimenting with autonomy in high-stakes environments. In the race for strategic advantage, nuclear-armed states are deploying AI-driven decision-making, force direction, and adaptive targeting, often in ways that compress decision timelines and heighten the risk of preemptive strikes, all without shared norms or oversight. The volatility within NC3 is amplified by a world already marked by geopolitical fragmentation, information warfare, and rising regional conflicts, all of which elevate the stakes of even minor AI-driven miscalculations.

A key focus has become AI-enhanced adaptive targeting in nuclear operations, where systems autonomously analyze battlefield intelligence and dynamically update target lists in real time. Lawyers have figured out how to ensure target packages are adapted quickly. While this offers clear military advantages, it also introduces serious vulnerabilities, including false novel positives and negatives, and the risk of adversarial manipulation through cyber attacks. In the absence of robust norms and technical safeguards, offensive cyber mirroring and escalation between rival nuclear states is on the rise. AI vulnerabilities supercharge this risk, creating an unstable feedback loop in which algorithmic misjudgments, deception, and premature threat classifications can spiral toward catastrophic escalation before humans can intervene.

Machine learning algorithms now shape force readiness and deployment decisions at scale, reducing human involvement in critical NC3 processes for many of the most powerful nuclear weapons states. Although human oversight is still nominally in place, many states are struggling as AI accelerates detection, targeting, and response cycles, increasing the likelihood of crisis instability and miscalculation.

Without diplomatic agreements or enforceable guardrails to preserve human authority at all critical decision points, new vulnerabilities emerge and old ones are exacerbated. Cyberwarfare and AI-enabled information operations further erode trust in command systems. With China and the United States locked in a broader AI competition and Russia accelerating asymmetric capabilities, doctrinal ambiguity around AI-enabled NC3 has become a deliberate feature—not a flaw—of nuclear deterrence. As states edge closer to delegating critical thresholds to code, the real danger may not be deliberate conflict, but a world in which no one fully

understands how war might begin.

Scenario 3: Human-Controlled Stability

States have deliberately limited the integration of AI into NC3 (Nuclear Command, Control, and Communications), prioritizing human control, system reliability, and strategic stability. The choice to restrain stems from both the existential risks posed by autonomous operations in nuclear systems and a lack of confidence in AI's reliability in high-consequence environments. It also reflects hard lessons from recent near-miss incidents, including AI misclassifications in military exercises that nearly led to real-world kinetic engagements, and false-flag drone attacks during ongoing Taiwan and Ukraine crises—events that exposed the brittleness of emerging AI systems under strategic pressure.

As a result, nuclear forces continue to operate on legacy hardware and human-controlled processes, reserving AI only for non-critical support functions such as logistics, maintenance, readiness, and route planning. While AI enhances situational awareness, states have drawn clear red lines: decision-making, force direction, high-level intelligence packages, and adaptive targeting must remain fully under human control.

While states publicly commit to prohibiting AI from assuming a dominant role in NC3—including autonomous launch, target selection, or force execution—quiet hedging strategies and compartmentalized R&D suggest that absolute restraint may be more aspirational than operational. Enforcement and verification of these restrictions remain ongoing challenges, as each nation defines "critical function" differently and transparency is often limited by national security prerogatives.

There is also growing concern that AI may have crept into certain NC3 operations—particularly in areas like intelligence filtering or cyber defense—without full appreciation of its downstream implications. Even as human control remains the policy cornerstone, some analysts warn that speed and complexity pressures may lead to silent drift toward automated dependencies, especially during moments of crisis when decision time is compressed. To address these concerns, the international community has adopted diplomatic declarations, transparency measures, and reporting mechanisms to monitor compliance with responsible AI norms. These include emerging standards focused on containment, human oversight, and technical resilience.

This world remains relatively stable, underpinned by shared restraint, analog fallback systems, and parallel redundancies. But a deeper tension is building: states that maintain strict limitations on AI risk are being outpaced by rivals who embrace it more aggressively, potentially reshaping deterrence dynamics. For now, strategic stability holds, but how long can restraint endure in a world where advantage may hinge on machines that act faster than humans ever could?

Scenario 4: Asymmetric Instability

The world is marked by uneven AI integration, regulatory breakdowns, and rising nuclear volatility. Major nuclear powers—including the United States and China—have collaboratively but unilaterally adopted cautious approaches to AI in NC3 (Nuclear Command, Control, and Communications), recognizing its destabilizing potential. Meanwhile, others have taken a divergent path. In particular, Russia, Pakistan, and the DPRK have integrated AI into their nuclear control systems aggressively and with minimal restraint, seeking asymmetric strategic advantages.

These outlier powers view AI not as a risk to be managed, but as a tool to bypass conventional deterrence frameworks. With limited transparency and few internal checks, they have accelerated the deployment of AI for early warning, situational awareness, and even target selection, often without robust human oversight. These systems, built rapidly and often on brittle architectures, are highly vulnerable to cyber intrusion, spoofing, and misinterpretation. In several known incidents, AI-enabled systems in these states have triggered false alarms or misclassified incoming threats, narrowly avoiding escalation. The destabilizing impact of AI is magnified by a fractured international landscape, where political distrust, cyber escalation, and non-NC3 AI deployments fuel misperceptions and strategic uncertainty.

The absence of enforceable global norms or verification mechanisms creates a two-tiered system: one group of states commits to responsible AI restraint in nuclear operations based on calculated risk assessments, while another exploits that very restraint to test thresholds and reshape deterrence dynamics. Multilateral forums have attempted to advance transparency and verification protocols, but these efforts are consistently undermined by the outlier powers' refusal to participate meaningfully or abide by emerging norms.

This asymmetry generates a growing sense of strategic instability. Responsible powers are now faced with a dilemma: whether to maintain restraint and risk vulnerability, or to match irresponsible actors and escalate the AI arms race. In this world, AI has become a wedge, contributing to a more fractured nuclear order with fundamentally different risk tolerances and strategic doctrines.

Nuclear deterrence, once predicated on mutual annihilation and shared red lines, now hinges on whether unstable AI systems in a few states will spark an unintended conflict, controllable or otherwise. The resulting environment is one of perpetual brinkmanship, where restraint by some is constantly threatened by recklessness from others. Without coordinated governance or credible enforcement, the global nuclear order teeters on the edge of a precipice, held hostage by actors willing to bet on automation and ambiguity in pursuit of tactical gain and perceived deterrent effect.

Scenario Exercises

Exercise 1: Scenario Development

1. Describe this scenario in a sentence.
2. What are the top 8-10 key decisions or events that would need to happen between now and 2030 for this scenario to emerge?
3. What are the specific applications of AI in this scenario? What is missing that should be here?
4. What are the biggest risks and opportunities in this world?
5. What guardrails would help prevent risk?
6. How can military advantage be gained while reducing risk?
7. Where might unintended consequences of AI emerge in this scenario?
8. What feedback loops might make this scenario more likely or more dangerous?*
9. How could non-state actors influence this world?*

Exercise 2: Scenario Implications

Decision Support:

10. How are decision-making and command efforts affected by the AI and norms factors laid out in this scenario?
11. How does AI impact the ability to maintain human oversight and escalation control?

Adaptive Targeting:

12. How does AI-driven targeting influence nuclear deterrence stability?
13. What safeguards or legal frameworks are in place to prevent miscalculations?

Strategic Warning:

14. How does AI influence threat detection and response times?
15. What are the risks of false positives or overreliance on AI-generated intelligence?

Net Assessment:

16. What concerns you most about this scenario?
17. What excites you most about this scenario?

Exercise 3: Research Implications

18. Having explored the 4 scenarios and their implications of AI on the future of NC3, what are the biggest questions that need to be addressed or answered?