



Technological Complexity and Risk Reduction:

**A Guardrails and Self-Assessment Framework for
EDTs in NC3 and nuclear weapons decision-making**

Belén Bianco

Rishi Paul

July 2024

About the 'Simulating Technological Complexity & Advancing Risk Reduction' project

This project focuses on a fast-changing, yet neglected, area of nuclear risk: mitigating the impacts of emerging and disruptive technologies (EDTs) on nuclear weapons decision-making and nuclear command, control, and communications (NC3). The fundamental aim of this project is risk reduction. We aim to do this by assisting states in identifying and mitigating nuclear use pathways and potential mistakes/miscalculations generated by the aggregate effects of EDTs on nuclear weapons decision-making processes and NC3 systems.

The 'Simulating Technological Complexity & Advancing Risk Reduction' project consists of three strands of work that seek to better understand EDTs and technological complexity, to produce detailed recommendations to mitigate nuclear risks:

1. **Strand 1** focuses on developing a framework to address likely challenges from EDTs to NC3 and nuclear decision-making. This report is part of Strand 1.
2. Through **Strand 2**, the ELN seeks to create a prototype digital tool that will simulate the highest-level nuclear weapons decision-making instances, the aggregate impact of EDTs in these processes, and the way the framework developed in Strand 1 can mitigate the risks generated by the aggregate effects of EDTs.
3. **Strand 3** will develop a sustained EDTs and technological complexity risk reduction campaign to implement the recommendations among nuclear-weapon and non-nuclear-weapon states and throughout multilateral and supra-national instances, such as the Nuclear Non-Proliferation Treaty review cycle, the Creating the Environment for Nuclear Disarmament, the Stockholm Initiative, and NATO, among others.

This work was performed with the generous support of the German Federal Foreign Office. The views and opinions of authors expressed herein do not necessarily state or reflect those of the German Federal Government. The authors would like to thank Graham Stacey, Alice Saltini, Jane Kinnimont, and Esther Kersley for their comments and suggestions to earlier drafts of this report.

About the Authors



Belén Bianco

*Former Policy Fellow,
European Leadership
Network*

Belén Bianco's work at the European Leadership Network (ELN) focused on nuclear risk reduction, arms control, disarmament, and non-proliferation, with a particular emphasis on the impact of emerging and disruptive technologies on nuclear risk. Prior to her role at the ELN, Belén worked on disarmament at various United Nations organisations, on nuclear non-proliferation and security for the Argentine government, and at weapons-focused think tanks such as the Arms Control Association and the Center for Security and Emerging Technology (CSET). Belén holds an MA in Security Studies, specialising in unconventional weapons, and a certificate in Diplomatic Studies, both from Georgetown University.



Dr Rishi Paul

*Senior Policy Fellow,
European Leadership
Network*

Dr Rishi Paul leads the implementation of Nuclear Policy projects. Rishi's research interests and expertise include nuclear risk and emerging and disruptive technologies, as well as deterrence dynamics in the context of evolving adversarial relationships, especially its effect upon the formation of nuclear strategy and escalation pathways. In late 2022, the Geneva Center for Security Policy (GCSP) awarded Rishi First Prize in its global security competition for "Innovation in International Security". Rishi holds both an MA in Strategic Studies and a PhD in nuclear strategy and ballistic missile defence from the University of Leeds.

Contents

Executive summary	5
Introduction	8
What are nuclear command, control, and communications (NC3) systems?	9
Overview of emerging and disruptive technologies (EDTs)	12
EDTs operating in aggregate	12
EDTs and nuclear weapons decision-making	12
EDTs under consideration	12
Risks from EDTs to nuclear weapons decision-making and NC3	16
Technology-inherent risks	17
Operational risks	18
Strategic risks	20
Introduction to the Guardrails and Self-Assessment Framework for EDTs in NC3 and nuclear weapons decision-making	22
GSA Framework for EDTs in NC3 and nuclear weapons decision-making	23
Potential uses for the GSA Framework	45
The path forward: opportunities for nuclear and non-nuclear weapon states	47
References	50

Executive summary

The fundamental aim of this project is to reduce risk by helping states identify and mitigate nuclear use pathways and potential mistakes/miscalculations that could arise from the complex interplay of emerging and disruptive technologies (EDTs) and the decision to use nuclear weapons. Although EDTs offer potential advantages, the rush by many nations to achieve technological superiority means that associated risks and disadvantages do not always receive the attention required. This leads to a lack of common understanding of risks and opportunities of EDTs.

Most studies address risks posed by technologies through the single lens of EDTs. However, when viewed as an aggregate, dimensions of complexity emerge revealing additional types of risks that could significantly impinge on the nuclear domain. These risks underscore the urgent need for careful management and proactive policies. Such measures are essential to maximise the benefits of advanced technologies while minimising their potential harms, all without undermining the substance and practice of deterrence.

Through a series of expert workshops, the ELN has produced a Guardrails and a Self-Assessment (GSA) Framework. This Framework aims to help the P5 and other nuclear-armed states self-evaluate the likely impact of EDTs on their NC3 systems and decision-making processes, and identify risk reduction measures. To this purpose, the GSA Framework aims to raise awareness and familiarise stakeholders at various levels with the complex interplay between a multitude of technologies, NC3 systems, and nuclear weapons decision-making.

The GSA Framework can help both states with and without nuclear weapons to implement responsible behaviours, policies, and practices and increase transparency around nuclear decision-making, fostering a more informed debate on nuclear risks. This is particularly important for the Non-Proliferation Treaty (NPT) and groups of like-minded states, such as the Creating Environment for Nuclear Disarmament (CEND) group and the Stockholm Initiative (SI).

Given the limited experience that states have in managing security discussions on EDTs, the report offers preliminary recommendations for how to address this deficiency at various stakeholder levels.

National implementation of measures outlined in the GSA Framework.

- Nuclear possessing states and technologically advanced non-possessing states adopting EDTs in their militaries should work toward the domestic adoption of the measures outlined in the GSA Framework.
- They could do so by pursuing national multistakeholder dialogues that focus on identifying requirements for the implementation of the GSA Framework and potential barriers to its operation.

The GSA Framework aims to raise awareness and familiarise stakeholders at various levels with the complex interplay between a multitude of technologies, NC3 systems, and nuclear weapons decision-making.

State Parties to the NPT should adopt a joint statement recognising the risks created by the aggregate effects of EDTs on NC3 systems and nuclear weapons decision-making. They should also convene a working group on technological complexity.

- An agreed statement between a diverse group of states would lay the groundwork to establish a working group to study the issue in depth.
- State Parties should convene a working group on technological complexity to investigate how the GSA Framework can advance disarmament and non-proliferation objectives. To ensure inclusivity and a variety of perspectives, the working group should include participation from a range of stakeholders that include academia, civil society, and the private sector.

The P5 Process should include the impact of EDTs on nuclear decision-making and NC3 in its discussions and consider risk reduction measures via implementation of the GSA Framework.

- The P5 Process should consider the risks of adopting of EDTs in the military domain and potential responses to events involving these technologies.
- The GSA Framework can guide parties through the potential risks and reveal the most appropriate risk mitigation measures.

The SI should collaborate with stakeholders in identifying GSA measures for implementation and determining which risks outlined in the Framework should be incorporated in fail-safe reviews conducted by all nuclear-weapon states.

- The SI could establish two working groups. The first would focus on identifying and prioritising GSA measures for implementation by nuclear and non-nuclear-weapon states.
- The second group could conduct a study to determine which risks outlined in the GSA Framework should be incorporated into fail-safe reviews conducted by all nuclear-weapon states.

Subgroup 3 of CEND should review the GSA Framework to evaluate its implementation and feasibility for both nuclear and non-nuclear-weapon states and assess how the aggregate effects of EDTs influence perceptions of disarmament obligations and responsibilities.

- By incorporating a review of the GSA Framework into its workplan, Subgroup 3 could address the risks posed by the aggregate effects of EDTs on nuclear weapons decision-making and NC3.
- It could also consider assessing how the aggregate effects of EDTs influence perceptions of disarmament responsibilities and obligations among nuclear and non-nuclear-weapon states.

NATO should prioritise technological complexity in its innovation activities and Implementation Strategy,¹ including analysing the implications of EDTs for NC3 systems and nuclear weapons decision-making.

- Analysing the aggregate effects of EDTs on NC3 systems and nuclear weapons decision-making structures as a separate priority area would complement current efforts and help converge two strategic topics for NATO: EDTs and nuclear deterrence.
- The NATO Secretary General's Advisory Group on EDTs could lead this work, using the GSA Framework to prioritise risks and recommend guardrails for national implementation.

Introduction

The aim of this report is to present and describe the ELN's GSA Framework, a risk mitigation tool developed under the 'Simulating Technological Complexity & Advancing Risk Reduction' project. The GSA Framework evaluates how the combined effects of EDTs might impact nuclear command, control, and communication (NC3) systems and nuclear weapons decision-making processes. It offers guardrails and self-assessment measures to limit the negative effects of such technologies.

To achieve this, the GSA Framework examines the impact of six significant and potentially disruptive technological developments: artificial intelligence (AI), autonomous systems and drones, counter-space capabilities, deepfakes, computer network operations (cyber), and quantum technologies. Unlike conventional approaches that study EDTs in isolation, the GSA Framework evaluates their aggregate effects on NC3 systems and the additional pressures and complexities these may create for the nuclear weapons decision-making process. In essence, it aims to assess the combined impact of EDTs, recognising that these effects exceed the sum of their individual parts.

The GSA Framework is focused on technological developments that are expected to mature over the next five to ten years. It serves as a predictive snapshot, incorporating informed assumptions about potential interactions among EDTs, and the risks their combined effects could impinge on the nuclear weapons domain. While studies on this topic vary in their focus, most also concentrate on a similar medium-term (five to ten, or slightly longer) horizon. This approach is shared by the NATO Science & Technology Organization, the US National Intelligence Council's Global Trends Project, and other notable efforts.²

This report does not seek to describe the technical discussions surrounding EDTs and NC3, nor does it cover every aspect of the nuclear weapons decision-making process. It also does not claim that EDTs are the only source of complexity in NC3 systems. Broader strategic, military, operational, moral, and psychological factors are also likely to play important roles in shaping the strategic landscape and may continue to dominate in some cases.³ However, understanding the various pressure points created by the aggregate effects of EDTs and identifying risk mitigation measures are critical first steps to ensure decision-makers are prepared for a future in which these challenges will arise.

Understanding the various pressure points created by the aggregate effects of EDTs and identifying risk mitigation measures are critical first steps to ensure decision-makers are prepared for a future in which these challenges will arise.

What are NC3 Systems?

Nuclear Command, Control, and Communications

The definition, design, and terminologies associated with NC3 systems are nuanced and differ between the nuclear possessor states. NC3 can broadly be defined as an information system employed within a political/military organisation. It is a general phrase that incorporates strategic and tactical systems. Consequently, combat direction, tactical data, and warning and control systems may each be considered NC3 systems.⁴

NC2 (nuclear command and control) and NC3 (NC2 plus communications) encompass overlapping systems for nuclear weapons management. Understanding the distinction and intersection between NC2 and NC3 is valuable to grasp how different nuclear possessor states exercise command and control over their nuclear arsenals and convey decisions within their military and political structures.

Command involves delegating a task by the highest political authority to its military forces. Control entails overseeing the operations of military forces as directed by command, implementing constraints through doctrines such as standard operating procedures (SOPs), and utilising communication and intelligence networks.⁵

The principles of NC2 and NC3 encompass policies, procedures, and organisational structures necessary to ensure that nuclear weapons can be commanded and controlled effectively, securely, and reliably. Whether NC2 or NC3, the key question for such a system can be summarised succinctly as the process states use to guarantee that unconventional weapons are employed only in accordance with their intended plans.⁶

NC3 ensures the effective management, secure operation, and reliable communication of a nation's nuclear forces. These systems integrate various sophisticated elements to ensure that nuclear arsenals are controlled, their use is authorised only by leaders, and that communications remain intact even during a nuclear war. Communication is a core component of any deterrent strategy, but it is often overlooked because military planners underestimate its critical importance: the greatest vulnerability in a nuclear posture is communications.⁷

The preference displayed by the P5 to ensure communications survivability underscores the size and variety of their nuclear arsenals and the need for reliability and resilience to ensure that command and control remain functional and secure even under attack; the US Department of Defence (DOD), for instance, operates through its NC3 enterprise to implement the NC2 functions. Similarly, communications survivability is also a key attribute of the UK NC3 system, which is designed to operate under all foreseeable circumstances and to ensure that a correctly authenticated UK national firing control message is sent from the National Command Authority to the ballistic missile submarines (SSBN).⁸

As the backbone of a nuclear state's ability to manage its nuclear forces – including the ability to authorise and control nuclear weapon use, communicate with forces, and maintain situational awareness – disruptions to NC2 and NC3 systems are highly destabilising for several crucial reasons. The table below identifies some of the associated risks.

Table 1: Risks associated with targeting of NC3 systems⁹

Threat	Risk
Breakdown of command and control leading to loss of authoritative command	<p>Compromised decision-making: NC3 systems enable top-level leaders to make informed decisions regarding the use of nuclear weapons. Attacking these systems can sever the chain of command, potentially leading to unauthorised or accidental nuclear weapon use.</p> <p>Reduced command effectiveness: a breakdown in NC3 could mean that nuclear forces might not receive orders in a timely manner, leading to confusion and potential misuse of nuclear assets.</p>
Increased risk of miscalculation and accidental launch resulting from misinterpreted signals	<p>False alarms: attacks on NC3 systems could generate false alerts or obscure real threats, leading to misinterpretation of data and potentially causing premature or accidental nuclear launches.</p> <p>Automated retaliation risks: some states may have automated response systems that could be triggered incorrectly if NC3 systems are attacked, leading to unintentional escalation.¹⁰</p>
Threat to deterrence stability as a result of undermining rival's second-strike capability	<p>Eroding retaliation confidence: the assurance that a state can retaliate (second-strike capability) after absorbing a nuclear strike is crucial for deterrence. Attacks on NC3 could undermine this capability by causing doubts about the ability to execute a nuclear reprisal.</p> <p>Perceived weakness: if a state's NC3 systems are perceived to be vulnerable, it might embolden adversaries to take more aggressive actions, thereby destabilising the strategic balance.</p>
Crisis instability resulting from the risk of rapid escalation	<p>Pre-emptive strike incentives (first strike instability): fearing that its NC3 systems might be disabled, a state might feel compelled to launch its nuclear weapons pre-emptively during a crisis, thereby increasing the likelihood of nuclear conflict.</p> <p>Escalation to nuclear use: an attack on NC3 systems could be seen as the precursor to an attempt to decapitate the state's leadership or neutralise its nuclear forces, potentially leading to immediate escalation to nuclear use as a defensive measure.</p>
Loss of critical information resulting from impairment of situational awareness and decision-making	<p>Reduced visibility or blindness in crisis: NC3 systems provide critical situational awareness about the status of nuclear forces and the strategic environment. Attacking NC3 systems can blind decision-makers to incoming threats or the actual status of their forces.</p> <p>Delay in critical decisions: the inability to obtain or verify information quickly can delay crucial decisions, which in a fast-moving crisis could mean the difference between deterrence holding or failing.</p>
Destabilisation of international relations and threat to global security	<p>Erosion of trust: attacking NC3 systems can erode trust between nuclear-armed states, leading to increased tensions and reduced willingness to engage in arms control or disarmament talks.</p> <p>Encouraging arms races: if states perceive their NC3 systems as vulnerable, they might invest more in offensive capabilities to pre-emptively disable adversary NC3, leading to an arms race and reduced global stability.</p>
Breakdown of nuclear safety protocols resulting from compromised communication channels	<p>Impaired communication: reliable communication is essential for implementing safety protocols, including nuclear risk-reduction measures and crisis management. Attacks on NC3 systems can disrupt these channels, increasing the risk of inadvertent escalation.</p> <p>Failure in de-escalation: the inability to communicate effectively with other nuclear-armed states during a crisis can prevent timely de-escalation and conflict resolution.</p>

Operational and strategic uncertainty resulting from increased uncertainty

Ambiguous postures: without robust and reliable NC3, a state's nuclear posture becomes uncertain, as adversaries cannot be sure of the state's command and control over its nuclear forces. This uncertainty can lead to worst-case scenario planning and heightened alert levels.

Unpredictable responses: if NC3 systems are compromised, the responses from the attacked state may become unpredictable, increasing the likelihood of unintended and rapid escalation to the nuclear level.

Strategic cultures and principal functions of NC2 - NC3

Strategic culture predicts that different states adopt different NC2 and NC3 structures, based on domestic politics and interests, cultural and decision-making norms, civil-military relations, and historical experiences.¹¹ These factors also influence nuclear learning, technological development, succession procedures, and security environments.¹² Organisational behaviour and bureaucracies further shape how these states design and execute their nuclear weapons decision-making processes.¹³ This is significant as the NC3 enterprise may involve hundreds or thousands of people interacting with each other across various technical systems.¹⁴

Despite the differences in national approaches to NC2 and NC3, the fundamental principles and functions of these systems often intersect and merge with one another. For example, John Gower argues that since the withdrawal of US weapons from the UK, the UK's concentration of its deterrent in the sea launched ballistic missile system eliminated the need for additional complexities in NC3 systems, focusing solely on national command and control over its SSBN force across all scenarios.¹⁵ However, effective communications are integral to the UK's SSBN, crucial for maintaining the effectiveness of each NC2 component, underscoring the overlap in core communications functions between NC2 and NC3.¹⁶ In the US context, NC2 relies on a survivable network of communications and warning systems that ensure secure connectivity from the president to all nuclear-capable forces, while the ability to move trusted data and advice depends on NC3.¹⁷

Given the breadth, complexity, diversity, and secrecy surrounding of NC2 and NC3 architectures among nuclear possessor states, this report omits discussing the myriad of their structural and technical specifications, which is beyond the scope of the GSA Framework. For simplification, the GSA Framework refers to NC3 and not NC2 because *effective* communications are embedded, although to varying degrees, within each nuclear possessor states' systems.

Overview of emerging and disruptive technologies (EDTs)

The fundamental tenets of nuclear weapons decision-making have not significantly evolved since the Cuban Missile Crisis. Nevertheless, EDTs such as autonomous weapons and drones, counter space capabilities, cyber offensive capabilities, AI, deepfakes, and quantum technologies are increasingly interacting with the nuclear domain.

EDTs operating in aggregate

The complexity of technological innovation as they relate to security and strategy raises several important issues. One issue is that nuclear weapons cannot be considered in isolation from EDTs because states' increasing dependency on these technologies to conduct conventional operations has causal effects on escalation pathways and nuclear weapons decision-making. While most nuclear possessor states treat nuclear weapons as a separate category of warfare (for good reasons) and distinct from conventional war, this can lead stakeholders to focus on nuclear postures and doctrines without considering the wider strategic landscape, thereby overlooking the impact of EDTs on the nuclear domain.¹⁸ Yet, EDTs have the potential to fundamentally change the way in which nuclear operations are conducted and how nuclear command and control functions.¹⁹

In this context, the GSA Framework concentrates on EDTs that can potentially impact NC3 systems and nuclear weapons decision-making processes. This scope thus encompasses both new technologies and established ones that have evolved in relevance within the defence and nuclear weapons domain.²⁰ Ultimately, the GSA Framework addresses challenges posed by technological developments that could reshape the landscape of nuclear weapons decision-making, and it evaluates the *aggregate* effects of EDTs rather than *isolated* impacts of individual technologies.

EDTs and nuclear weapons decision-making

Despite advancements in EDTs, the fundamental tenets of nuclear weapons decision-making have not significantly evolved since the Cuban Missile Crisis. Nevertheless, EDTs such as autonomous weapons and drones, counter space capabilities, cyber offensive capabilities, AI, deepfakes, and quantum technologies are increasingly interacting with the nuclear domain.

EDTs introduce benefits, as well as risks; they can benefit decision-making processes through improved sensors, data analysis and management, and real time situational awareness, among other things. However, as the technologies significantly increase the volume of data, knowledge, and factors to be considered, and operate at speeds beyond human capability, nuclear weapons decision-making will become increasingly complex. Decision-makers may face a situation where multiple EDTs interact, adding layers of complexity, creating internal feedback and biases, and generating unexpected and potentially unexplainable outcomes. In this context, the rapid evolution of EDTs and their intricate interactions when combined have the potential to exert a profound influence on the global nuclear weapons decision-making landscape.

EDTs under consideration


In a first iteration of this project, the ELN and organisations that it partnered with identified six emerging and/or disruptive technologies with the potential to significantly impact future nuclear weapons decision-making.²¹ These technologies (listed in Table 2 below) were chosen based on a comprehensive review of patent and non-patent literature, focusing on factors such as

novelty, growth trajectory, and potential for disruptive applications in the military domain.

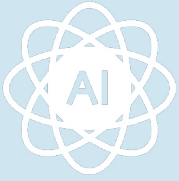
Hypersonic technologies, particularly their military applications in the form of hypersonic glide vehicles (HGVs) and cruise missiles (HCMs) were excluded for a few reasons. First, the transformative nature of HGVs and HCMs on the strategic balance remains unclear. Second, some ballistic missiles, such as the US Minuteman III or the Russian RS-28 Sarmat can already travel at hypersonic speeds.²² Thus, ballistic missiles and HGVs present advancements that fit in a continuum, as they share many similarities, rather than the latter having disruptive characteristics.²³ Last, the additional benefits of HGVs and HCMs over advanced ballistic and cruise missiles (which can already evade defences, manoeuvre, and strike targets with a high degree of accuracy) are debated.²⁴

The table below provides a concise overview of the implications of the selected EDTs on NC3 and nuclear weapons decision-making.

Table 2: Overview of the implications of selected EDTs on NC3 and nuclear weapon decision-making²⁵

<p>Autonomous weapons and drones</p> 	<p>Autonomous weapons and drones present both challenges and opportunities for nuclear weapons decision-making and NC3 systems. They can enhance situational awareness and provide new capabilities for resilience and redundancy, but they also introduce risks of rapid escalation, reduced decision-making timeframes, and vulnerabilities to cyber and electronic warfare. From a decision-making perspective, drone swarms combined with neural networks engender a range of capabilities. They could analyse big portions of terrain in detail, remain airborne for extended periods, and have the power to strike without human involvement. In addition, they can detect human presence, recognise faces and interpret human emotions. They could also neutralise weapons and specialise in specific roles, as well as communicate autonomously, mimic human visual systems for nuclear detection, and employ AI-based decision-making to overwhelm defences.</p>
<p>Counter space capabilities</p> 	<p>Counter-space capabilities in a context of increased commercialisation of space introduce several complexities into nuclear weapons decision-making and NC3.</p> <p>Counter space capabilities that target satellites used for early warning, communication, and intelligence, surveillance, and reconnaissance (ISR) can be particularly risky, as these systems constitute key nodes in states' NC3. These satellites are also entangled with non-nuclear weapons and are typically dual use, meaning that they enable both nuclear and non-nuclear operations. Kinetic attacks or potential accidents that affect these assets would compromise a core function of NC3 by reducing situational awareness.</p> <p>Yet, the risk of space congestion within five years may make it too risky to neutralise enemy satellites.</p>
<p>Cyber offensive capabilities</p> 	<p>The integration of novel cyber technologies with drones, AI, and deep fakes could threaten NC3 and nuclear weapons decision-making. Machine learning could enable daily generation of numerous malware versions. Remote access trojan malware could target critical infrastructure networks and manipulate AI in hardware to eliminate valid threat warnings and/or create fake hazards.</p> <p>A potential shift of computing power to 'internet of things' (IoT) devices will create vulnerabilities, particularly near targets like nuclear facilities. Hackers could compromise counter-measures in industrial robots and simulate cyber-attacks on infrastructure plants. Novel combinations could involve malware code generation, rapid spreading through IoT devices, and automatic triggering of cyber-attacks upon proximity to a target. The manipulation of AI in hardware could allow for server corruption, impacting infrastructure including nuclear weapons facilities.</p>

Artificial intelligence



AI will reshape NC3 and nuclear weapons decision-making, offering advantages in data validation and countering risks related to information overload and machine-driven deception. While AI promises autonomous machine speed decision-making and augmented reality systems for enhanced situational awareness, it may also be a primary source of systemic risk. 'Distributed AI' – AI distributed across multiple devices – will enable machines to operate both in tandem and autonomously. This could limit the options available to decision-makers during an attack, as they may be compelled to either endorse the machine's chartered course of action, or rely solely on machine-mediated outputs to suggest an alternative path.

The distinction between AI focused on detection and targeting, and AI as a trusted advisor in escalation and decision-making will be crucial. Machine deception at machine speed emerges as a critical risk, necessitating multiple AIs to cross-check and validate sources and reasoning within nuclear NC3. Some AI designs could improve trust between machines and humans, but outcomes are uncertain. The potential for system cascades triggered by mismatches between AI recommendations and human intentions poses a threat to well-judged conflict management.

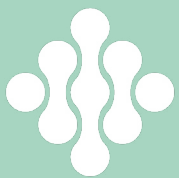
Deepfakes



Deepfake technology poses a significant and imminent risk to NC3 system, by providing state and non-state actors inexpensive and simple means to create near-indistinguishable fake images, videos, and text. This automated deception threatens to erode trust in intelligence sources, political leaders, and digital platform security. Countermeasures are emerging, including blockchain co-processors and machine generated trustworthiness scores. However, the risks persist, with generative adversarial networks (GANs) used for detection systems, leading to a potential 'machine vs machine' environment.

The convergence of technologies introduces challenges such as neuro-linguistic programming intercepting and rewriting text and synthetic audio deepfakes manipulating command orders, news, and political statements.

Quantum technologies



The unprecedented speed of quantum computers, potentially operating 100 trillion times faster than conventional supercomputers, could introduce a 'machine vs machine' dynamic in which speed will dominate and compound the complexity inherent to NC3 systems and the uncertainty faced by nuclear weapons decision-makers. This technology has the potential to have a multiplier effect, complicating decision-making across other technologies. This shift may lead to asymmetric shocks and expose critical systems like encryption and security networks to vulnerabilities.

Hybrid quantum/classical computers could compromise key NC3 elements such as submarine tracking and space command and control, accelerating real-time threat detection and introducing complexities in decision-making.

Quantum-based gaming will revolutionise strategic decision-making and the evaluation of emerging nuclear risks. Conversely, quantum technologies have the potential to magnify the effects of complexity and uncertainty, rendering interconnected networks more susceptible to deception and attacks.

The selected EDTs are interconnected, globally widespread, and inherently both dual-use (applicable for both military and civilian purposes) and dual capable (intervening in both nuclear and non-nuclear operations). Given this, the GSA Framework examines their aggregate and collective effects on nuclear weapons decision-making, rather than the impact of individual technological developments. The main objective is to develop a comprehensive understanding that will inform strategies for navigating the intricate intersection of disruptive technologies and nuclear weapons decision-making and planning.

Risks from EDTs to nuclear weapons decision-making and NC3

These technologies have the potential to accelerate information flow, compress decision-making timeframes, involve more stakeholders, and exacerbate existing ambiguities.

Navigating nuclear weapons decision-making has always been a complicated task – EDTs are poised to multiply this complexity exponentially. These technologies have the potential to accelerate information flow, compress decision-making timeframes, involve more stakeholders, and exacerbate existing ambiguities. Acting in aggregate, these factors could heighten misperception, misunderstanding, and miscalculation, thereby increasing the risk of nuclear escalation and unintended nuclear use.

To address these challenges and develop measures to mitigate the impact of EDTs on nuclear decision-making, the GSA Framework’s starting point is risk identification. The GSA Framework simplifies this process by categorising risks at three levels: technological, operational, and strategic.

- **Technology-inherent risks** stem from the characteristics, limitations, or complexities of the technology itself, including the design and development of technology-based systems.
- **Operational risks** arise from the operational environment, practices, and human factors involved in the management and use of the technology, often related to human-machine interaction.
- **Strategic risks** encompass potential threats or uncertainties that could undermine stability, security, or the balance of power, influencing the calculations of states and actors involved.

While these categories are interconnected – with technology-inherent risks, for instance, leading to operational risks, potentially escalating into strategic risks – they are delineated separately to facilitate the development of targeted risk mitigation strategies at each level. It should be noted that the list of risks described below does not intend to be an exhaustive enumeration of all potential risks generated by EDTs, nor to cover all possible technological scenarios and combinations. Furthermore, as technologies constantly evolve and develop, new risks are likely to emerge. We encourage states to also consider, assess, and mitigate against additional risks as they appear.

Table 3: Risks of EDTs to nuclear weapons decision-making and to NC3, by category

Technology-inherent risks		
Identifier	Definition	Description
Malfunction	Malfunction of EDTs (AI and quantum technologies in particular) integrated into systems critical for NC3, including decision support, situational monitoring, detection, and early warning systems.	NC3 EDT-augmented systems, especially those integrating AI and quantum technologies, are vulnerable to malfunction due to algorithmic errors and hallucinations, potentially generating incorrect outputs with high levels of confidence. In decision support, situational monitoring, detection, and early warning systems for NC3 and nuclear weapons decision-making, these kinds of malfunctions could lead to incorrect assessments of threats, misinterpretation of data, and/or erroneous recommendations, potentially resulting in inappropriate actions being taken under false pretence or beliefs.
Cyber-attacks	Susceptibility of EDTs-augmented systems – such as decision support, situational monitoring, detection, and early warning systems – to cyber-attacks.	EDTs-augmented systems that involve AI and quantum technologies may represent attractive targets for cyber-attacks. First, malicious actors may attempt to exploit vulnerabilities in these technologies to launch cyber-attacks aimed at data breaches and unauthorised disclosure of sensitive information, for espionage, blackmail, or propaganda campaigns. Second, these technologies are particularly susceptible to cyber spoofing attacks, where malicious actors manipulate the systems to deceive or trick decision-makers. For instance, AI systems could be spoofed by feeding them manipulated data, causing them to make incorrect decisions or provide misleading recommendations. Similarly, quantum technologies, if compromised, could be vulnerable to spoofing, potentially leading to erroneous outputs or compromised security protocols within NC3 systems and nuclear weapons decision-making processes. Lastly, cyber-attacks could potentially compromise the availability of these systems, leading them to cease functioning and resulting in severe consequences for situation awareness and/or ongoing operations. These breaches would pose significant risks to the integrity and security of NC3 and nuclear weapons decision-making processes. ²⁶
Computing constraints	Underperformance of AI-powered autonomous systems tasked with ISR and delivery duties, due to computing limitations on the battlefield.	The underperformance of AI-powered autonomous systems, such as AI-powered drones, satellites, or ground vehicles tasked with Intelligence, Surveillance, and Reconnaissance (ISR) and delivery duties, due to computing limitations on the battlefield may lead to incomplete situational awareness, delayed response times, impaired targeting accuracy, increased risk of strategic miscalculation, failed delivery, and erosion of confidence in NC3 systems. Persistent underperformance of ISR capabilities may undermine the reliability and effectiveness of NC3 systems, resulting in slower decision-making, reduced agility in responding to nuclear threats, and potential escalation of conflict with nuclear-armed adversaries. ²⁷

Operational risks

Identifier	Definition	Description
Automation bias and trust gaps	Automation bias and trust gaps in EDTs-augmented NC3 systems, especially those incorporating AI and quantum technologies.	Automation bias in NC3 EDTs-augmented systems, especially those incorporating AI and quantum technologies, pose significant risks to nuclear weapons decision-making processes. Automation bias may lead decision-makers and military personnel to blindly trust AI- and quantum-generated outputs, potentially without critically evaluating the underlying data, computation, or algorithms, thereby increasing the likelihood of misinterpretations and flawed decisions. Concurrently, trust gaps may emerge due to the opaque nature of AI and quantum technologies, where decision-makers struggle to verify the outputs provided, leading to scepticism or distrust and subsequent rejection of the suggested course of action or information provided.
Information uncertainty	Uncertainty over the reliability of information obtained, collected, and processed within NC3 systems.	EDTs pose a significant challenge in discerning genuine from fabricated media content, with deepfakes presenting a notable concern. The proliferation of deepfake technology has the potential to create widespread uncertainty over the reliability of information obtained, collected, and processed within NC3 systems and nuclear weapons decision-making processes. Operators may struggle to differentiate between authentic and manipulated media content, leading to hesitancy in basing operational decisions on potentially compromised information. Moreover, the susceptibility of EDT-augmented NC3 systems to cyber-hacks further exacerbates information uncertainty, as malicious actors can exploit vulnerabilities to manipulate data and deceive operators. The convergence of deepfakes and cyber threats introduces unprecedented risks to the integrity and trustworthiness of information sources, undermining the effectiveness and reliability of NC3 systems which military operators depend on.
Difficulty of attribution	Difficulty of attribution between attacks or third-party spoofing.	The synergistic utilisation of EDTs such as cyber offensive capabilities, AI, and deepfake technologies significantly heightens the difficulty of discerning whether a reported threat is authentic or if the system has fallen victim to sophisticated spoofing techniques. This difficulty in attribution can lead to uncertainty and hesitation in response efforts, potentially allowing malicious actors to exploit vulnerabilities. If military operators are unaware that the systems have been spoofed, they might fail to respond appropriately. Conversely, if operators believe the threat is authentic, malicious actors can continue their attack undeterred.

Overreliance	Overreliance on EDTs in NC3 systems.	The increasing integration of AI, quantum computing, and other EDTs in NC3 systems – such as decision support, situational monitoring, detection and early warning systems – can create a dependency that may compromise the effectiveness of military operations. This overreliance could lead to a gradual erosion of critical thinking skills and human judgment, as system operators become accustomed to operating in an environment in which EDTs-augmented systems dictate the path forward. Overreliance becomes particularly complex when the availability of these systems may be compromised, leading to a scenario where they cease functioning altogether. In such situations, military operations may lack the skills to effectively compensate for the sudden absence of these advanced systems, jeopardising operational effectiveness.
Lack of training	Military personnel may misinterpret outputs, over-rely on flawed information, neglect biases in AI algorithms, or ignore signals that reflect that a system has been hacked, compromising nuclear weapons decision-making within NC3.	If military personnel are unable to comprehend the operational mechanisms of EDTs-augmented systems, particularly those integrating AI and machine learning technologies, they may fail to recognise the inherent risks and the diverse range of errors these systems can generate. Consequently, there's a heightened risk of making critical decisions grounded in flawed or untrustworthy data. Moreover, they may inadvertently disregard biases ingrained within AI systems, such as data bias, algorithmic bias, implicit bias, interaction bias, feedback loop bias, and social bias.

Strategic risks

Identifier	Definition	Description
Erosion of trust	Erosion of trust between states generated by information uncertainty.	The proliferation of deepfake technology, coupled with the inherent risks of malfunction, cyber-hacks and underperformance of EDTs-augmented systems, exacerbates the already complex landscape of information uncertainty. These technology-inherent risks impact the operational landscape as they undermine the reliability and authenticity of information obtained, collected, and processed within NC3 systems, making it difficult for operators to differentiate between authentic and manipulated or erroneous content. At the strategic level, the proliferation of deepfakes creates fertile ground for mis- and disinformation campaigns and propaganda efforts aimed at sowing discord, undermining alliances, and destabilising geopolitical dynamics. As trust between states erodes due to heightened uncertainty and scepticism over the authenticity of information, the risk of misperception, miscalculation, and conflict escalation in nuclear contexts becomes increasingly pronounced.
Lack of understanding	Lack of understanding among decision-makers and defence planners regarding the potential effects of integrating EDTs into NC3 systems, of the utilisation of defensive and offensive EDTs capabilities on NC3 systems, and of the consequences of accidents arising from the deployment of these technologies.	Without a practical understanding of EDTs and their potential impacts on decision support, situational monitoring, detection and early warning systems within NC3, nuclear weapons decision-makers are at risk of inadvertently introducing errors or making flawed decisions, potentially leading to unintended consequences or escalation of nuclear risks and creating vulnerabilities for accidental or unauthorised use of nuclear weapons. Additionally, the lack of comprehension regarding the consequences of employing defensive and offensive EDT capabilities may result in inadequate preparedness and response strategies, leaving nuclear operations vulnerable to disruptions or unintended consequences. Furthermore, the lack of understanding of defence planners and national security strategists about the implications of EDTs, such as those that allow for autonomy in the nuclear weapons decision-making process, could undermine the expectation of rationality that underlies deterrence strategies, and in turn affect the strategic calculations of decision-makers. Last, the insufficient awareness of defence planners and national security strategists on the implications of EDTs in NC3 could result in lack of budgeting to develop safety, security, and reliability measures to mitigate technology inherent risks of EDTs.
Geopolitics	Geopolitical competition and premature technology deployment, leading to their use beyond their initial purposes, including in applications where thorough testing has not been conducted.	In an environment of heightened geopolitical competition, nations may feel compelled to swiftly adopt EDTs to gain strategic advantages over adversaries or as an asymmetric response to the perceived technological advancements of a rival. However, hastily deploying EDTs beyond their initial purposes, including in applications where thorough testing has not been conducted risks introducing malfunctions or vulnerabilities, particularly in critical domains like nuclear weapons decision-making. Uncertainty surrounding these technologies could undermine strategic stability and deterrence postures and may trigger unintended escalation.

Proliferation	Proliferation risks due to increased development and/or ownership of EDTs by private actors within the technology and arms industries.	As the defence industry landscape evolves, moving away from the traditional dominance of governments and militaries in defence technology and weapons development, innovation has become increasingly decentralised. This shift often results in EDTs and their enabling systems and technologies – such as materials, parts, components, infrastructure, and processing and computing services – not being exclusively developed or owned by governments. This decentralisation heightens the risks associated with proliferation as states cede the responsibility for the security of these technologies to the private sector, increasing the likelihood of them falling into the hands of malicious non-state actors, proxies, and those seeking to exploit vulnerabilities for nefarious purposes.
Control	Control risks due to development and/or ownership of EDTs by private actors within the technology and arms industries.	The increased involvement of the private sector in the development of EDTs deepens states' reliance on the private sector, thereby affording these actors greater access, influence, and power. The heightened involvement of the private sector broadens the spectrum of actors engaged in nuclear weapons decision-making and operating NC3 systems, potentially complicating governance structures and introducing novel vulnerabilities and challenges, such as security breaches, conflicts of interest, and decreased state control over critical infrastructure.
Autonomy and deterrence practices	Escalation and erosion of deterrence arising from the introduction of increasing autonomy in NC3, particularly concerning the potential for nuclear weapons launch decisions to be made without direct human control.	Delegating decision-making authority to autonomous systems heightens unpredictability and complexity, increasing the potential for misinterpretation or miscalculation of signals or events. The lack of direct human involvement could lead to scenarios where automated systems misinterpret incoming data, triggering unwarranted escalation or even a nuclear response based on flawed information. In this context, nuclear deterrence is a psychological phenomenon and is practiced according to a set of beliefs, which includes humans exercising rationality. However, if decision-making is delegated to autonomous systems without human judgement, this could erode deterrence.
Situational awareness and crisis stability	Escalation, misinterpretation, miscalculation, and compromised crisis stability arising from reduced situational awareness due to disruptions and malfunctions of NC3 systems and early warning systems.	NC3 early warning systems play a crucial role in providing real-time information about potential threats and inform the nuclear weapons decision-making process. Disruption to these systems, whether due to deliberate targeting by adversaries, technical malfunctions, or unforeseen accidents, can limit or degrade the ability of decision-makers to accurately assess a situation, leading to misunderstandings or miscalculations that could escalate tensions. The vulnerability of these systems is exacerbated by the burgeoning commercialisation of space, which has led to increased congestion and competition for orbital resources. This congested environment heightens the risk of accidental interference or collision, further jeopardising the integrity of NC3 early warning capabilities, and ultimately, the strategic calculations of policy planners and decision-makers.

Introduction to the Guardrails and Self-Assessment Framework for EDTs in NC3 and nuclear weapons decision-making

The GSA Framework assists states in steering the military application of EDTs at multiple levels, reducing risk in decision-making and managing potential escalation. It consists of measures that states can implement independently and immediately, without requiring negotiation or coordination with other nations.

The GSA Framework is comprised of two components. Firstly, a set of 'guardrails' encompassing guidelines for decision-makers and operators, including recommendations for technology assessments, awareness raising, and training measures, best practices, and unilateral pledges. Secondly, a 'self-assessment checklist' designed to facilitate risk identification and mitigation through a series of open-ended questions. This checklist aids in operationalising the guardrails and can be utilised by states to evaluate the resilience of their nuclear weapons decision-making structures and NC3 systems against the risks posed by EDTs.

The two distinct aspects of the GSA Framework target different stakeholders but are equally applicable to all. The guardrails are designed for discussions with the P3, aiming for the mechanism's adoption to quickly identify and mitigate EDTs-related risks in the nuclear field. The self-assessment checklist is intended for nations less inclined to adopt guardrail measures, enabling independent audits of nuclear powers and planners.

The proposed measures fall into the following categories:

- a. Assessment:** identification of the areas where ongoing examination of both the immediate and far-reaching impacts of EDTs in NC3 and nuclear weapons decision making is essential.
- b. Awareness raising and training:** initiatives aimed at increasing awareness among specific actors about the risks posed by EDTs to nuclear weapons decision-making and NC3.
- c. Best practices:** recommendations which nuclear weapons decision-makers, defence and national security strategists, and military operators could adopt to mitigate the risks associated with EDTs in the nuclear weapons decision-making process and NC3 systems.
- d. Pledges:** unilateral actions that a state should publicly commit to undertake or refrain from, to effectively mitigate the risks associated with EDTs in nuclear weapons decision-making and NC3 systems.

The GSA Framework, which is presented on the next page, links each identified risk with corresponding guardrails and self-assessment measures to address them. It should be noted that not all risks have associated risk reduction measures in every category. The GSA Framework does not aim to provide an exhaustive list of all possible risk mitigation measures that states can adopt in each category, but rather highlight clear examples that states can follow to mitigate the risks of EDTs on nuclear weapons decision-making processes and NC3 systems.

Introduction

A Guardrails and Self-Assessment (GSA) Framework for Emerging and Disruptive Technologies (EDTs) in Nuclear Command, Control, and Communications (NC3) and nuclear weapons decision-making

The ELN's GSA Framework seeks to assist policymakers in addressing risks arising from the aggregate effects of EDTs on NC3 systems and nuclear weapons decision-making.

EDTs:

- Autonomous weapons and drones
- Counter-space capabilities
- Cyber offensive capabilities
- Artificial Intelligence
- Deepfakes
- Quantum technologies

Categorisation of risks:

- Technology-Inherent: Characteristics and limitations of technology itself.
- Operational: Human and environmental factors.
- Strategic: Threats to stability, security, or balance of power.

Risk mitigation measures:

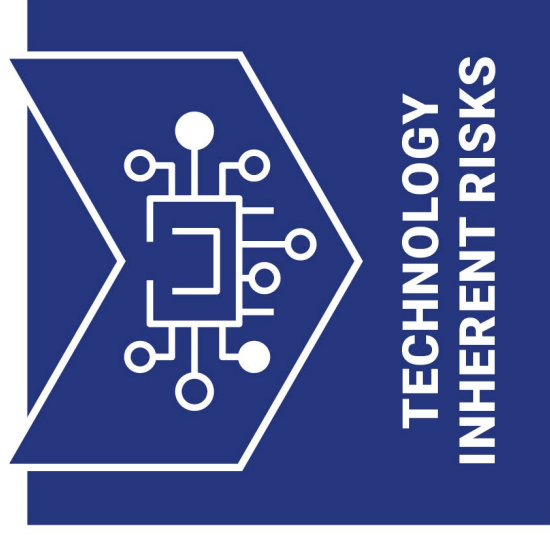
- Guardrails: Guidelines, recommendations, awareness, training, best practices, pledges.
- Self-Assessment checklist: Open-ended questions for risk identification and mitigation.

Categorisation of Guardrails and Self-Assessment measures:

- Assessment: Ongoing EDT impact examination.
- Awareness raising and training: Increasing awareness initiatives.
- Best practices: Recommendations for decision-makers and military operators.
- Pledges: Unilateral commitments by states.



Categorisation of risks of EDTs operating in aggregate to NC3 and nuclear weapons decision-making



- Malfunction
- Cyber-attacks
- Computing constraints

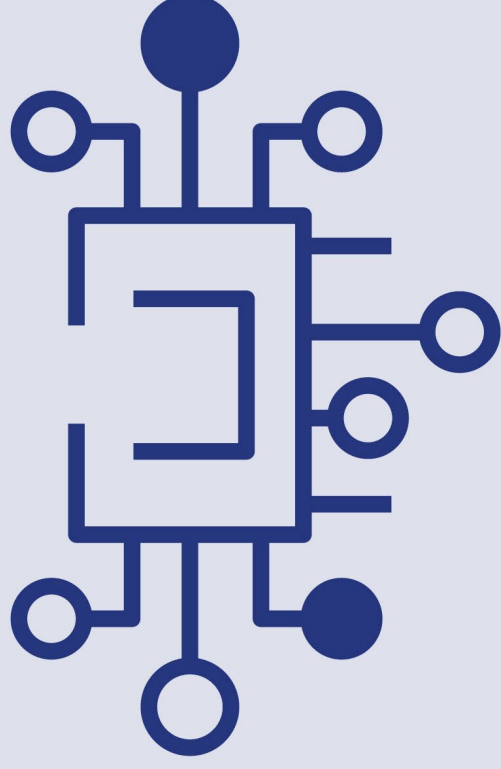


- Automation bias and trust gaps
- Information uncertainty and difficulty of attribution
- Overreliance
- Lack of training



- Erosion of trust
- Lack of understanding
- Geopolitics
- Proliferation
- Control
- Autonomy and deterrence practices
- Situational awareness and crisis stability





Technology inherent risks

- Malfunction
- Cyber-attacks
- Computing constraints





Technology inherent risks

MALFUNCTION

Malfunction of EDTs (AI and quantum technologies in particular) integrated into systems critical for NC3, including decision support, situational monitoring, detection, and early warning systems.

A. ASSESSMENT

Guardrail

States should conduct technology assessments of EDTs integrated in NC3 systems regularly, and as part of their national defence reviews.

Self-Assessment

- i. What measures are used to evaluate the effectiveness, reliability, and transparency of EDTs integrated into NC3 infrastructure?
- ii. How are technology assessments integrated into national defence reviews and strategic planning processes to ensure alignment with broader defence objectives and priorities?
- iii. What criteria is used to evaluate the performance, accuracy, and security of EDTs incorporated into NC3 systems?

B. BEST PRACTICE

Guardrail

States should ensure that EDTs-augmented systems incorporated into NC3 are only employed in applications where thorough testing has been conducted.

Self-Assessment

- i. How does the State ensure that EDTs-augmented systems incorporated into NC3 are only employed in applications where thorough testing has been conducted?
- ii. How is the performance and accuracy of EDTs during testing validated, and what metrics, criteria, and benchmarks are used to evaluate their success in meeting operational objectives and requirements?

C. BEST PRACTICE

Guardrail

States should conduct a Fail-Safe Review of the safety, security, and reliability of nuclear weapons, and of the safety, security, reliability, and resilience of NC3 and integrated tactical warning/attack assessment systems, especially in the context of potential malfunctions of relevant EDTs-augmented systems incorporated into NC3.

Self-Assessment

- i. Are Fail-Safe Reviews conducted iteratively and periodically to account for evolving technological advancements, operational requirements, emerging threats, and adversaries' changing postures and doctrines?
- ii. How are the findings and recommendations from previous Fail-Safe Reviews incorporated into subsequent iterations?

Technology inherent risks

Malfunction

Cyber-attacks

Computing constraints

MALFUNCTION

D. BEST PRACTICE

Guardrail

States should ensure relevant level of human control and implement significant redundancy and backup systems that can provide fail-safes in case of malfunction, alternative systems running in parallel, fallback procedures for manual intervention, and backup communication channels for critical alerts and notifications.

Self-Assessment

- i. Has the State identified and prioritised key functions and operations within EDTs-augmented systems incorporated in NC3 that require redundancy and backup support to ensure continuity and reliability under adverse conditions or unforeseen events?
- ii. How robust and resilient are redundancy and backup systems in providing fail-safes for critical functions in the event of malfunctions or disruptions of EDTs-augmented systems incorporated in NC3?

E. PLEDGE

Guardrail

States should publicly commit not to incorporate EDTs-augmented systems – especially AI-powered ones coupled with increasing degrees of automation – into NC3, unless these are reliable, transparent, and trustworthy.

Self-Assessment

- i. How effectively has the State communicated its commitment to ensure the reliability and trustworthiness of EDTs-augmented systems, particularly those with increasing degrees of automation, before their incorporation into NC3?



CYBER-ATTACKS

Susceptibility of EDTs- augmented systems – such as decision support, situational monitoring, detection, and early warning systems – to cyber-attacks.

A. ASSESSMENT

Guardrail

States should conduct vulnerability testing of EDTs-augmented systems in NC3. These can help identify potential weaknesses and vulnerabilities that could be exploited by cyber-attacks. This includes conducting penetration testing, vulnerability scanning, red team exercises, security audits, and scenario-based simulations.

Self-Assessment

- i. What vulnerability testing procedures for NC3 systems are conducted with the goal of identifying potential weaknesses and vulnerabilities susceptible to exploitation in cyber-attacks?
- ii. How are vulnerability testing procedures conducted?

B. BEST PRACTICE

Guardrail

States should conduct a Fail-Safe Review of the vulnerability of nuclear weapons, NC3 and integrated tactical warning/attack assessment systems, especially in the context of potential cyber-attacks to relevant EDTs-augmented systems incorporated into NC3.

Self-Assessment

- i. How comprehensively are Fail-Safe Reviews conducted to assess the vulnerability of nuclear weapons, NC3 infrastructure, and integrated tactical warning/attack assessment systems to potential cyber-attacks targeting relevant EDTs-augmented systems within NC3?
- ii. What methodologies and tools are employed in Fail-Safe Reviews to simulate, model, or analyse the impact of cyber-attacks on relevant EDTs-augmented systems, and how do these assessments inform risk management and mitigation strategies?

C. BEST PRACTICE

Guardrail

States should implement best practices in cyber security such as secure coding practices, encryption of sensitive data, network segmentation, and secure configuration management to prevent unauthorised access and data breaches.

Self-Assessment

- i. How comprehensively are best practices in cybersecurity integrated into policies, procedures, and operational practices, encompassing aspects such as secure coding, encryption, network segmentation, and configuration management to mitigate the risk of unauthorised access and data breaches?
- ii. What measures are in place to promote awareness and adherence to secure coding practices among defence contractors, and in particular developers and software engineers, including training, guidelines, and code review processes to identify and address potential vulnerabilities in software applications and systems?



Technology inherent risks

Malfunction

Cyber-attacks

Computing constraints

CYBER-ATTACKS

D. BEST PRACTICE

Guardrail

States should adopt continuous monitoring to detect and promptly mitigate anomalous activities such as cyber hacking attempts. This entails deploying security monitoring tools, intrusion detection systems, and automated response mechanisms to identify and respond to suspicious behaviour in real-time.

Self-Assessment

- i. How effectively is continuous monitoring integrated into cybersecurity strategies and operational practices, ensuring real-time detection and response to anomalous activities and potential security threats?
- ii. What security monitoring tools and technologies are deployed to collect and analyse data, including intrusion detection systems, log management solutions, and behaviour analytics platforms, to identify indicators of compromise and suspicious behaviour?
- iii. How are baseline patterns of normal behaviour for relevant EDTs-augmented systems and personnel interactions established, allowing for the timely detection of deviations or anomalies that may indicate potential cyber hack attempts?

COMPUTING CONSTRAINTS

Underperformance of AI-powered autonomous systems tasked with ISR and delivery duties, due to computing limitations on the battlefield.

A. ASSESSMENT

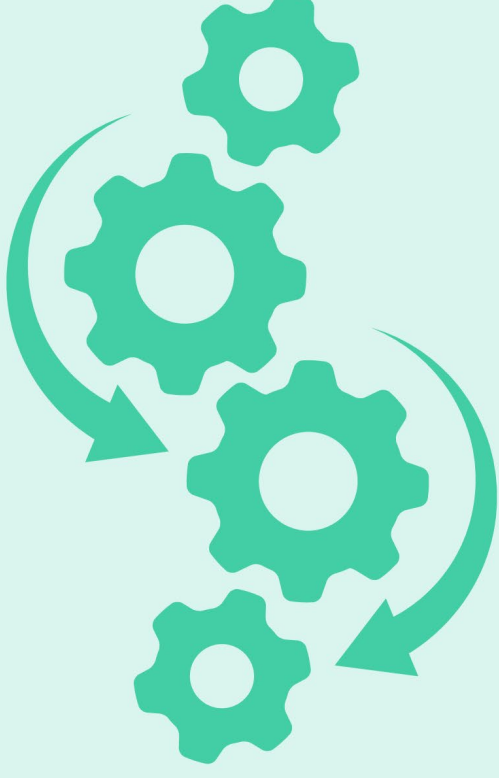
Guardrail

Conduct regular assessments of computing infrastructure to identify potential limitations that could affect the performance of AI-powered autonomous systems. This includes evaluating processing power, memory, network bandwidth, and latency to ensure sufficient resources are available for ISR and delivery tasks.

Self-Assessment

- i. How systematically are regular assessments conducted to evaluate the computing infrastructure supporting AI-powered autonomous systems, with a specific focus on identifying potential limitations that could impact performance in ISR and delivery tasks?
 - ii. What criteria and metrics are utilised in assessing processing power, memory capacity, network bandwidth, and latency of the computing infrastructure, and how do these align with the requirements and demands of AI algorithms and applications used in ISR and launch operations?





Operational risks

- Automation bias and trust gaps
- Information uncertainty and difficulty of attribution
- Overreliance
- Lack of training





Operational risks

Automation bias and trust gaps

Information uncertainty and difficulty of attribution

Overreliance

Lack of training

Operational risks

AUTOMATION BIAS AND TRUST GAPS

Automation bias and trust gaps in EDTs-augmented systems in NC3, especially those incorporating AI and quantum technologies.

A. ASSESSMENT

Guardrail

States should conduct regular evaluations of EDTs-augmented systems in NC3 to identify instances of automation bias and trust gaps. This includes analysing decision-making processes, assessing user interactions with the systems, ensuring data diversity, and evaluating the reliability and accuracy of AI and quantum-generated outputs.

Self-Assessment

- i. How systematically are regular evaluations conducted to assess NC3 systems for signs of automation bias and trust gaps with a specific emphasis on understanding how these phenomena may impact decision-making processes and outcomes?
- ii. What methodologies and approaches are utilised in evaluating NC3 systems, including analysing decision-making processes, assessing user interactions with the systems, ensuring data diversity, and scrutinising the reliability and accuracy of AI and quantum-generated outputs?

B. AWARENESS RAISING AND TRAINING

Guardrail

States should provide training and awareness programs to military system operators on the risks associated with automation bias in EDTs-augmented systems. Users should be trained to recognise potential biases, interpret system outputs critically, and make informed decisions based on a combination of automated recommendations and considered human judgment.

Self-Assessment

- i. How comprehensive and tailored are the training and awareness programs provided to military system operators regarding the risks associated with automation bias in EDTs-augmented systems, considering the unique operational context and interconnected requirements of NC3 and nuclear weapons decision-making processes?
- ii. What topics and concepts are covered in the training and awareness programs, including explanations of automation bias, transparency, and their implications for decision-making, as well as strategies for recognising, mitigating, and addressing these phenomena within EDTs-augmented systems?

Operational risks

Automation bias and trust gaps

Information uncertainty and difficulty of attribution

Overreliance

Lack of training

AUTOMATION BIAS AND TRUST GAPS

C. PLEDGE

Guardrail

States should pledge to uphold ethical principles and responsible use of AI and quantum technologies in the military domain.

Self-Assessment

- i. How deeply ingrained are ethical principles and considerations of responsible AI and quantum technology use within decision-making and special operating procedures, particularly in the context of NC3 and nuclear weapons decision-making?
- ii. What specific ethical guidelines, frameworks, or codes of conduct have been established or adopted to govern the development, deployment, and operation of AI and quantum-augmented systems in NC3 and nuclear weapons decision-making?

INFORMATION UNCERTAINTY AND DIFFICULTY OF ATTRIBUTION

Uncertainty over the reliability of information obtained, collected, and processed within NC3 systems.

Difficulty of attribution between attacks or third-party spoofing.

A. BEST PRACTICE

Guardrail

States should deploy robust information assurance practices to safeguard the integrity and authenticity of data within NC3 systems, such as encryption, digital signatures, access controls, secure communication protocols, blockchain, and hand-coding.

Self-Assessment

- i. How comprehensively are information assurance practices, such as encryption, digital signatures, access controls, and secure communication protocols, integrated into NC3 systems to protect the integrity and authenticity of data?
- ii. What mechanisms are employed to protect sensitive data and information flows within NC3 systems?



Operational risks

Automation bias and trust gaps

Information uncertainty and difficulty of attribution

Overreliance

Lack of training

INFORMATION UNCERTAINTY AND DIFFICULTY OF ATTRIBUTION

B. BEST PRACTICE

Guardrail

States should strengthen detection and forensic capabilities to identify and mitigate the impact of deepfakes, spoofing, and cyber-attacks on NC3 systems, and to facilitate attribution. Measures could include advanced threat detection technologies, anomaly detection algorithms, real-time monitoring systems, and provenance.

Self-Assessment

- i. How robust and comprehensive are detection and forensic capabilities in identifying and mitigating the impact of deepfakes, spoofing, and cyber-attacks on NC3 systems, as well as facilitating processes to determine attribution?
- ii. What threat detection technologies and methodologies are employed to detect anomalies, suspicious activities, and potential indicators of deepfake manipulation, spoofing, or cyber-attacks within NC3 systems, including intrusion detection systems, anomaly detection algorithms, and behaviour analytics?
- iii. How effectively do detection and forensic capabilities support the attribution process, including the collection, preservation, and analysis of digital evidence to determine the origin, nature, and intent of cyber-attacks against NC3 systems?

OVERRELIANCE

Overreliance on EDTs in NC3 systems.

A. ASSESSMENT

Guardrail

States should conduct regular assessments of automation bias in EDTs-augmented systems to evaluate the effectiveness of these systems in achieving their objectives and identifying any instances of overreliance on automated outputs.

Self-Assessment

- i. How are the processes for conducting assessments of automation bias in EDTs-augmented systems structured to evaluate the effectiveness of these systems in achieving their objectives in NC3 and nuclear decision-making processes?
 - ii. Are there measurable indicators in place to signal instances of overreliance on EDTs within NC3 systems? These could encompass patterns of behaviour, decision outcomes, or feedback from operators and users.

Operational risks

Automation bias and trust gaps

Information uncertainty and difficulty of attribution

Overreliance

Lack of training

OVERRELIANCE

B. AWARENESS RAISING AND TRAINING

Guardrail

States should establish training programs for system operators and relevant personnel on the capabilities and limitations of EDTs that are integrated into NC3 systems. Training should emphasise the importance of maintaining a balanced approach to decision-making, recognising the strengths and weaknesses of EDTs, and exercising human judgment in critical situations.

Self-Assessment

- i. How comprehensive and tailored are the training programs established for system operators and relevant personnel on the capabilities and limitations of EDTs integrated into NC3 systems?
- ii. How effectively do the training programs emphasise the importance of maintaining a balanced approach to decision-making, considering both the strengths and weaknesses of EDTs, and the critical role of human judgment in complex and uncertain situations?

C. PLEDGE

Guardrail

States should pledge to prioritise human oversight and accountability in NC3 and establish mechanisms for reviewing and auditing automated tasks, as well as holding individuals responsible for errors or failures attributed to overreliance on EDTs.

Self-Assessment

- i. How are mechanisms for reviewing and auditing automated decisions established and integrated into NC3 processes to ensure transparency, accountability, and compliance with ethical principles and legal frameworks?
- ii. What specific criteria or thresholds are used to trigger reviews and audits of automated decisions within NC3, including factors such as decision complexity, potential impact, and stakeholder concerns?

Operational risks

Automation bias and trust gaps

Information uncertainty and difficulty of attribution

Overreliance

Lack of training

LACK OF TRAINING

Military personnel may misinterpret outputs, over-rely on flawed information, neglect biases in AI algorithms, or ignore signals that reflect that a system has been hacked, compromising nuclear weapons decision-making within NC3.

A. AWARENESS RAISING AND TRAINING

Guardrail

States should launch awareness raising campaigns targeting military system operators to highlight the importance of recognising and addressing biases in AI algorithms. Military system operators should be trained on the various types of biases that can affect automation in EDTs-augmented systems and the potential consequences of overlooking them in nuclear decision-making.

Self-Assessment

- i. How targeted are the awareness raising campaigns launched to highlight the importance of recognising and addressing biases in AI algorithms among military system operators and relevant personnel?
- ii. How are the potential consequences of overlooking automation bias in EDTs-augmented systems communicated to military system operators, including risks related to inaccurate assessments, flawed recommendations, and compromised decision-making processes in nuclear scenarios?

B. AWARENESS RAISING AND TRAINING

Guardrail

States should initiate targeted awareness campaigns aimed at military system operators, emphasising the critical importance of remaining vigilant against potential cyber-attacks. These campaigns should include comprehensive training on how to identify signals indicative of a system breach and underscore the potential ramifications of disregarding such indicators in nuclear weapons decision-making.

Self-Assessment

- i. Are military personnel adequately trained to recognise signals of a potential system breach?
- ii. Has comprehensive education been provided regarding the potential consequences of disregarding indicators of cyber-attacks in the context of nuclear decision-making?



Operational risks

Automation bias and trust gaps

Information uncertainty and difficulty of attribution

Overreliance

Lack of training

LACK OF TRAINING

C. BEST PRACTICE

Guardrail

States should conduct hands-on simulation exercises to familiarise system operators with automation bias in EDTs-augmented systems and to recognise cyber hacks in realistic scenarios.

Self-Assessment

- i. What specific scenarios and contexts are incorporated into the exercises to simulate real-world conditions and challenges faced by system operators in NC3 and nuclear weapons decision-making environments?
- ii. How are the simulation exercises designed to provide opportunities for system operators to interact with EDT-augmented systems, explore different functionalities, recognise cyber breaches, and practice decision-making processes in dynamic and complex situations?

D. BEST PRACTICE

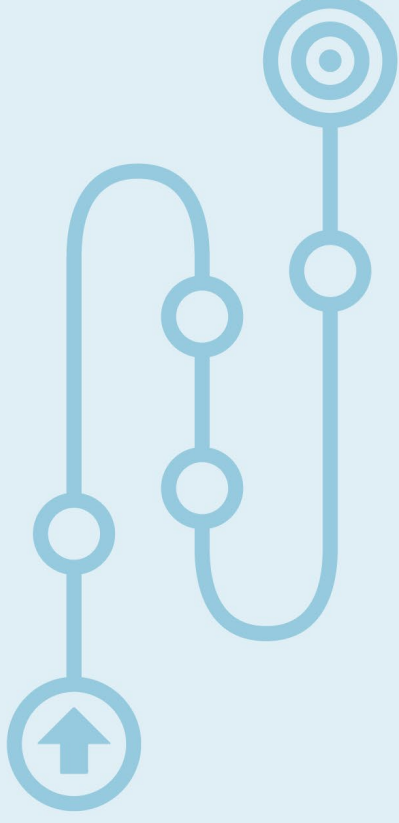
Guardrail

States should implement regular skills assessments and proficiency testing to evaluate the competence of system operators in recognising cyber breaches and preventing automation bias in the use of EDTs-augmented systems. They should also identify areas for improvement and provide targeted training and support to address knowledge gaps and enhance skills.

Self-Assessment

- i. How comprehensive and systematic are the skills assessment and proficiency testing processes implemented to evaluate the competence of system operators in recognising cyber breaches and in preventing automation bias in the use of EDTs-augmented systems?
- ii. What specific criteria, metrics, and performance indicators are used to assess the proficiency of system operators in recognising cyber breaches and in preventing automation bias in the use of EDTs-augmented systems?





Strategic risks

- Erosion of trust
- Lack of understanding
- Geopolitics
- Proliferation
- Control
- Autonomy and deterrence practices
- Situational awareness and crisis stability

Strategic risks

Erosion of trust

Lack of understanding

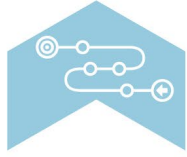
Geopolitics

Proliferation

Control

Autonomy and deterrence practices

Situational awareness and crisis stability



Strategic risks

EROSION OF TRUST

Erosion of trust between states, generated by information uncertainty.

A. AWARENESS RAISING AND TRAINING

Guardrail

States should conduct awareness-raising campaigns to educate decision-makers involved in national security decisions about the risks associated with cyber-attacks and with deepfake technology, especially within nuclear weapons decision-making contexts.

Self-Assessment

- i. How comprehensive and targeted are the awareness-raising campaigns conducted to educate decision-makers involved in national security decisions about the risks associated with cyber-attacks and deepfake technology within nuclear weapons decision-making?
- ii. How are awareness-raising campaigns tailored to address the varying levels of knowledge, expertise, and roles of decision-makers within the national security and nuclear weapons decision-making apparatus, ensuring relevance and accessibility of information across diverse audiences?

B. BEST PRACTICE

Guardrail

States should establish and test crisis management mechanisms, including widespread communication infrastructure, direct lines, bilateral consultative protocols, and channels of communication between adversaries and allies.

Self-Assessment

- i. What specific communication channels and protocols are established to enable timely and secure information exchange between relevant stakeholders, including government agencies, military entities, international partners, and adversaries in the event of a nuclear crisis?
- ii. How are crisis management mechanisms tailored to address the unique challenges and dynamics of nuclear weapons decision-making contexts, including considerations of escalation risks, strategic stability, and crisis de-escalation measures?

Strategic risks

Erosion of trust

Lack of understanding

Geopolitics

Proliferation

Control

Autonomy and deterrence practices

Situational awareness and crisis stability

EROSION OF TRUST

C. BEST PRACTICE

Guardrail

States should adopt digital media signing protocols and ensure that all media content exchanged in official and diplomatic communications is signed using this method.

Self-Assessment

- i. What specific digital media signing standards and technologies have been adopted to ensure the authenticity, non-repudiation, and tamper-evidence of media content, including documents, images, videos, and audio recordings, exchanged in official channels?
- ii. How effectively have digital media signing protocols been integrated into official and diplomatic communications processes to authenticate and verify the integrity of media content?

LACK OF UNDERSTANDING

Lack of understanding among decision-makers and defence planners regarding the potential effects of integrating EDTs into NC3 systems, of the utilisation of defensive and offensive EDTs capabilities on NC3 systems, and of the consequences of accidents arising from the deployment of these technologies.

A. AWARENESS RAISING AND TRAINING

Guardrail

Organise campaigns specifically tailored for nuclear weapons decision-makers and defence planners, to raise awareness about the potential effects of EDTs on NC3 systems and nuclear weapons decision-making among these actors.

Self-Assessment

- i. What specific channels and mediums are utilised to disseminate information about EDTs, ensuring maximum reach and engagement among the target audience of nuclear weapons decision-makers and defence planners?
- ii. How are the content and messaging of the awareness campaigns customised to address the unique concerns, knowledge gaps, and decision-making contexts of nuclear weapons decision-makers and defence planners?

B. BEST PRACTICE

Guardrail

Encourage continuous learning among nuclear weapons decision-makers and defence planners to stay updated on EDT advancements and their implications for nuclear weapons decision-making.

Self-Assessment

- i. What mechanisms are in place to facilitate nuclear weapons decision-makers, defence planners, and national security strategists to stay updated on EDT advancements and their potential effects on NC3 systems and nuclear weapons decision-making processes?
- ii. What forums or platforms exist for knowledge sharing and exchange among nuclear weapons decision-makers and defence planners regarding EDTs and their implications for nuclear weapons decision-making?

Strategic risks

Erosion of trust

Lack of understanding

Geopolitics

Proliferation

Control

Autonomy and deterrence practices

Situational awareness and crisis stability

GEOPOLITICS

Geopolitical competition and premature technology deployment, leading to their use beyond their initial purposes, including in applications where thorough testing has not been conducted.

A. BEST PRACTICE

Guardrail

States should define explicit uses for the incorporation of EDTs in the military domain. The reliability of such systems should be tested and must only be deployed within those defined uses across their entire life cycle.

Self-Assessment

- i. How robust are systems for monitoring and enforcing compliance with the defined uses of EDTs, including mechanisms for regular review and adjustment as needed?
- ii. What measures have been implemented to ensure that EDTs are only deployed within the explicitly defined uses throughout their entire life cycle?

B. PLEDGE

Guardrail

States should commit to responsible EDT adoption by pledging to prioritise thorough testing and validation processes before deployment, especially in NC3 systems and nuclear weapons decision-making.

Self-Assessment

- i. Have clear testing frameworks and protocols been established to guide the validation process of EDTs?
- ii. Have budgetary considerations been made to support comprehensive testing procedures, including personnel, equipment, and facilities?

Strategic risks

Erosion of trust

Lack of understanding

Geopolitics

Proliferation

Control

Autonomy and deterrence practices

Situational awareness and crisis stability

PROLIFERATION

Proliferation risks due to increased development and/or ownership of EDTs by private actors within the technology and arms industries.

A. AWARENESS RAISING AND TRAINING

Guardrail

Raise awareness among defence industry stakeholders about the proliferation risks associated with EDTs.

Self-Assessment

- i. What is the current level of awareness among defence industry stakeholders regarding the potential risks of proliferation associated with EDTs?

B. BEST PRACTICE

Guardrail

States should implement stringent regulatory frameworks to control the development, transfer, and export of EDTs and their enabling technologies and systems — such as materials, parts, components, infrastructure, and processing and computing systems.

Self-Assessment

- i. How effective are existing regulatory frameworks in addressing proliferation risks associated with EDTs and their enabling technologies?
- ii. Are there adequate mechanisms in place to monitor and enforce compliance with regulatory requirements for EDTs?

C. BEST PRACTICE

Guardrail

States should intensify cooperation with the private sector to ensure safeguard and safety measures are designed and implemented in the development of EDTs to be utilised in the military domain.

Self-Assessment

- i. Are there established channels of communication and collaboration between government entities and private sector partners specifically focused on limiting proliferation of EDTs?
- ii. Are there initiatives to provide training and education to private sector personnel involved in EDT development regarding safety protocols and best practices to prevent the proliferation of these technologies?

Strategic risks

Erosion of trust

Lack of understanding

Geopolitics

Proliferation

Control

Autonomy and deterrence practices

Situational awareness and crisis stability

CONTROL

Control risks due to development and/or ownership of EDTs by private actors within the technology and arms industries.

A. ASSESSMENT

Guardrail

States should evaluate the potential implications of private sector control over EDTs systems, their critical components, and their enabling technologies.

Self-Assessment

- i. Are there mechanisms in place to identify and analyse the specific risks associated with private sector control over EDTs in the context of national security and defence?
- ii. Have assessments been conducted to understand the potential vulnerabilities introduced by private sector involvement in EDT development and ownership?

B. BEST PRACTICE

Guardrail

States should establish clear guidelines and standards governing private sector involvement in EDT development for military applications. These guidelines should aim to prevent any single actor from gaining disproportionate access, influence, or power that could adversely impact nuclear weapons decision-making processes.

Self-Assessment

- i. How do these guidelines ensure equitable participation and prevent any single actor from gaining undue access, influence, or power over the nuclear weapons decision-making process?
- ii. What mechanisms are in place to regularly assess and update guidelines and standards governing private sector involvement in EDT development for military applications?

AUTONOMY AND DETERRENCE PRACTICES

Escalation and erosion of deterrence arising from the introduction of increasing autonomy in NC3, particularly concerning the potential for nuclear weapons launch decisions to be made without direct human control.

A. BEST PRACTICE

Guardrail

States should ensure that autonomous systems integrated into NC3 are designed with fail-safe mechanisms to prevent unauthorised or erroneous actions.

Self-Assessment

- i. Are there clear protocols in place for human intervention and override in the event of a failure or unexpected behaviour of autonomous systems within NC3?
- ii. How transparently have the fail-safe measures implemented in autonomous systems within NC3 been communicated to allies and adversaries?

Strategic risks

Erosion of trust

Lack of understanding

Geopolitics

Proliferation

Control

Autonomy and deterrence practices

Situational awareness and crisis stability

AUTONOMY AND DETERRENCE PRACTICES

B. BEST PRACTICE

Guardrail

States should refrain from delegating launch authority of nuclear weapons to machines. Human judgement and control over these decisions should be retained at all times.

Self-Assessment

- i. What mechanisms exist for ongoing monitoring and evaluation of the effectiveness of human control measures in NC3 systems?

C. BEST PRACTICE

Guardrail

States should mandate human oversight in the target identification process for nuclear weapons launches, thereby preventing autonomous systems from independently selecting targets.

Self-Assessment

- i. How clear are the policies and guidelines regarding human involvement in the target identification process for nuclear weapons, ensuring that autonomous selection by machines is strictly prohibited?
- ii. What safeguards exist to ensure human involvement in the identification of targets of nuclear weapons launches?

D. PLEDGE

Guardrail

States should make public commitments to uphold principles of human control and accountability in nuclear weapons decision-making processes.

Self-Assessment

- i. What steps have been taken to integrate the principles of human control and accountability into policies, procedures, and operational practices related to nuclear weapons decision-making?
- ii. How has the commitment to retaining human judgment over nuclear weapon launch authority been communicated to allies, adversaries, and the broader international community to prevent the erosion of nuclear deterrence?

Strategic risks

Erosion of trust

Lack of understanding

Geopolitics

Proliferation

Control

Autonomy and deterrence practices

Situational awareness and crisis stability

SITUATIONAL AWARENESS AND CRISIS STABILITY

Escalation, misinterpretation, miscalculation, and compromised crisis stability arising from reduced situational awareness due to disruptions and malfunctions of NC3 systems and early warning systems.

A. BEST PRACTICE

Guardrail

States should refrain from conducting direct-ascent anti-satellite missile tests. Such tests produce substantial debris, posing a significant risk to crucial space infrastructure integral to early warning systems.

Self-Assessment

- i. Have clear policies and guidelines been established within defence and space agencies regarding the prohibition of direct-ascent anti-satellite missile tests?

B. BEST PRACTICE

Guardrail

States should be transparent and provide prior notification of any authorised close encounters with space infrastructure, such as designated activities for inspection purposes.

Self-Assessment

- i. Are there contingency plans and established communication channels to address any unexpected developments or emergencies during close encounters?
- ii. Have clear protocols and special operating procedures been established for notifying and coordinating with affected parties prior to conducting any close encounters?

C. PLEDGE

Guardrail

States must pledge not to attack NC3 systems of adversaries, critical for nuclear stability, security of nuclear arsenals, and facilitating reliable communication channels for decision-makers during times of crisis.

Self-Assessment

- i. How clearly has the commitment to not target adversaries' NC3 systems been communicated?
- ii. Are there established protocols and guidelines in place to ensure adherence to the pledge not to attack NC3 systems of adversaries, across all levels of military and government decision-making?

Potential uses for the GSA Framework

Considering the rapid pace at which EDTs are evolving and the pressure to adopt these technologies in the military domain, the GSA Framework is designed for swift, independent, and unilateral implementation by states, without the need for lengthy negotiations.

Considering the rapid pace at which EDTs are evolving and the pressure to adopt these technologies in the military domain, the GSA Framework is designed for swift, independent, and unilateral implementation by states, without the need for lengthy negotiations. This is not to suggest that the GSA Framework should replace any bilateral or multilateral efforts to address the risks EDTs pose to nuclear weapons decision-making and NC3 systems. Such negotiations are necessary but can take years or even decades to materialise, and in the current geopolitical context, time is a luxury we do not have.

We hope the GSA Framework will be a step in the right direction in mitigating risks and ensuring that the adoption of EDTs in the military domain does not create undue dangers or push states towards nuclear use. The GSA Framework can and should be implemented by nuclear possessor and non-possessor states, as EDTs-related pressures towards nuclear use can arise from the actions of either. Additionally, it can guide discussions around the risks generated by EDTs in multilateral and bilateral nuclear conversations. Below are some ways decision-makers and policy planners can use this resource:

National level

- The GSA Framework can help identify actions that may exacerbate nuclear risk by increasing the likelihood of misperceptions, misunderstandings, unintended escalation, and erosion of crisis stability. It also suggests measures each state can implement to mitigate those risks.
- Nuclear possessor states:
 - The **guardrails elements** can guide the adoption of responsible behaviour and risk reduction measures to protect nuclear weapons decision-making processes and NC3 systems from EDT risks.
 - The **self-assessment questions** can facilitate conversations with a wide range of stakeholders about the risks and their perceived responsibilities, including technology developers, military planners and operators, and nuclear weapons decision-makers.
 - For nuclear possessor states that are less inclined to adopt risk mitigation or guardrails measures, the self-assessment questions can help internally evaluate the resilience of nuclear weapons decision-making structures and NC3 systems against the risks posed by EDTs.

Bilateral dialogues

- Among nuclear possessor states, the GSA Framework can serve as a primer for conversations on strategic concerns. It could be included in bilateral discussions about strategic stability, nuclear risk, and EDTs. It can also guide multi-stakeholder dialogues aimed at achieving empathic understanding between allies and adversaries.²⁸
- The GSA Framework can help the non-possessors engage with enhanced equity in conversations on nuclear risks with nuclear

possessor states. It can equip them with the pertinent questions to ask and key topics to focus on, facilitating a thorough review of the policies and practices of nuclear powers and nuclear planners.

Multilateral instances

- The GSA Framework can be used by both nuclear and non-nuclear-weapon states in multilateral fora focused on nuclear risks, such as the NPT and its related initiatives, including CEND, the SI, and the P5 Process. Although not all these initiatives explicitly address EDTs, these technologies can significantly impact the nuclear order and states' abilities to comply with treaty obligations.²⁹ In fact, at the 10th NPT Review Conference, State Parties emphasised the need for regular dialogue on the implications of EDTs.³⁰ The GSA Framework can guide these discussions, ensuring that the impact of EDTs on nuclear weapons decision-making processes and NC3 systems are thoroughly considered and addressed. By providing a structured approach, the GSA Framework helps states navigate the complexities of EDTs and their impact on nuclear risk, facilitating a more informed and collaborative effort to enhance global nuclear risk mitigation strategies.
- The GSA Framework can support NATO's efforts to adapt its deterrence posture by offering an innovative tool for concrete nuclear risk reduction related to technological complexity.
- In these settings, the GSA Framework can enable non-nuclear-weapon states to engage more equitably in discussions about the impact of EDTs on nuclear risk. It can also be used to advocate for their participation in NATO exercises.

The path forward: opportunities for nuclear and non-nuclear weapon states

Notwithstanding its core purpose of showing how risks can be mitigated, the GSA Framework serves as a primer for discussion within a plethora of forums: the NPT, the P5 Process, CEND, NATO, and the SI.

Trust amongst states has significantly deteriorated in recent years, driven by great power conflicts, the war in Ukraine, nuclear modernisation, and the unravelling of arms control agreements. The unchecked development of EDTs adds further complexity to an already unstable situation. Currently, the prospects for meaningful arms control talks are slim. In this challenging context, the GSA Framework is designed to operate at various stakeholder levels; our recommendations flow from its benefits and potential uses.

Notwithstanding its core purpose of showing how risks can be mitigated, the GSA Framework serves as a primer for discussion within a plethora of forums: the NPT, the P5 Process, CEND, NATO, and the SI. While it aims to reduce risks through a set of measures, its true test lies in its acceptance at the political and operation levels.

To foster this acceptance, the following recommendations could help the suggested risk mitigation measures gain political traction.

Recommendations

- 1. Nuclear possessor states and technologically advanced non-possessor states developing EDTs should consider implementing the measures outlined in the GSA Framework.**

The GSA Framework's strength lies in its ability to be swiftly and unilaterally implemented. Implementing its recommendations could involve a domestic multistakeholder dialogue that includes all parties involved in the development of EDTs including military and political decision-makers, military system operators, and specialised agencies. The focus should be on defining roles and responsibilities, establishing requirements and timelines for implementation, identifying implementation challenges, and devising strategies to overcome these challenges.

- 2. State Parties to the NPT are encouraged to adopt a statement recognising the risks created by the aggregate effects of EDTs on NC3 systems and nuclear weapons decision-making. Additionally, they could convene a working group to study the issue in depth, using the GSA Framework as a starting point and key frame of reference.**

State Parties to the NPT should adopt a joint statement acknowledging the challenges posed by EDTs to nuclear weapons-decision making and NC3 systems. This statement would lay the groundwork to establish a working group to study the issue in depth.

The working group should include representatives from both nuclear-weapon and non-nuclear weapon states and be co-chaired by one member from each group. Its mandate may be twofold: first, to study how the combined effects of EDTs affect nuclear weapons decision-making and present challenges to the achievement of the Treaty's objectives; second, to recommend EDTs risk mitigation measures for both nuclear and non-nuclear weapon states to advance the disarmament and non-proliferation agendas. The GSA Framework offers a foundational, though non-exhaustive, list of risks and risk

mitigation measures that States Parties can consider an expand upon. The working group may also consider involving the private sector, as many of these technologies are being developed by these actors, making their collaboration essential for advancing discussion around EDTs.³¹ To ensure inclusivity and a variety of perspectives, the working group may also include representatives from civil society and academia.

3. The five nuclear-weapon states should incorporate the impact of EDTs on nuclear decision-making processes and NC3 systems into their discussions on nuclear doctrines within the P5 Process. This could include evaluating risk reduction measures and the potential implementation of the GSA Framework.

The P5 Process, consisting of meetings and conferences among the five recognised nuclear-weapon states under the NPT, focuses on their unique responsibilities under the treaty. Among other topics, these discussions have been addressing increased transparency around the P5's nuclear doctrines to contribute to long-term risk reduction. The P5 Process should also consider the risks of adopting EDTs in the military domain and potential responses to events involving these technologies. The GSA Framework could serve as starting point for these conversations, guiding the P5 Process through the risks posed by EDTs to NC3 systems and nuclear weapons decision-making, and potential risk mitigation measures.

4. The SI should collaborate with policy and technical experts to identify and prioritise GSA measures for implementation. Additionally, it could determine which risks outlined in the GSA Framework should be incorporated in fail-safe reviews conducted by all nuclear-weapon states.

The SI has already demonstrated interest in advancing discussions on EDTs, calling for research, analysis, education, and awareness regarding the effects of emerging technologies on nuclear risks.³² It has also urged nuclear-weapon states to take steps to limit the potential for new technologies to create new nuclear risks and exacerbate existing ones.³³

To further these goals, the SI could establish two working groups. The first would focus on identifying and prioritising GSA measures for implementation by both nuclear and non-nuclear weapon states. The second group could study which risks outlined in the GSA Framework should be incorporated into fail-safe reviews conducted by nuclear-weapon states. While the United States is currently undertaking such a review, the SI could also explore strategies to incentivise other nuclear-weapon states to follow suit.

5. The CEND Subgroup 3 on 'Interim measures to reduce the risk associated with nuclear weapons' should review the GSA Framework to evaluate its implementation and feasibility for both nuclear and non-nuclear-weapon states. The subgroup should also assess how the aggregate effects of EDTs influence perceptions of disarmament obligations and responsibilities.

The CEND Subgroup 3 focuses on identifying factors that could contribute to the risk of nuclear weapons use, including EDTs, and on assessing measures to address these risks. The GSA Framework, with its emphasis on risk identification and mitigation, could thus significantly enhance the work of Subgroup 3. By incorporating a review of the GSA Framework into its workplan, Subgroup 3 could address the risks posed by the aggregate effects of EDTs on nuclear weapons decision-making and NC3. Additionally, the subgroup could consider assessing how the aggregate effects of EDTs influence perceptions of disarmament responsibilities and obligations among nuclear and non-nuclear-weapon states.

6. NATO should prioritise technological complexity in its innovation activities and Implementation Strategy,³⁴ including analysing the implications of EDTs for NC3 systems and nuclear weapons decision-making. The GSA Framework would serve as a useful roadmap for NATO's Advisory Group on Emerging and Disruptive Technologies to prioritise risks and recommend guardrails for NATO members to adopt.

NATO has shown a deep interest in EDTs and has been developing policies since 2019, producing various strategies, roadmaps, and implementation plans.³⁵ It has identified specific technologies and considered their implications for deterrence, defence, and capability development. Analysing the aggregate effects of EDTs on NC3 systems and nuclear weapons decision-making structures, as a separate priority area, would complement these efforts and help converge two strategic topics for NATO: EDTs and nuclear deterrence. The NATO Secretary General's Advisory Group on EDTs could lead this work, using the GSA Framework to prioritise risks and recommend guardrails for national implementation.

References

1. NATO, "Foster and Protect: NATO's Coherent Implementation Strategy on Emerging and Disruptive Technologies" (2021).
2. Office of the Chief Scientist, "Science & Technology Trends 2020-2040: Exploring the S&T Edge", (NATO Science & Technology Organization, March, 2020), https://www.nato.int/nato_static_fl2014/assets/pdf/2020/4/pdf/190422-ST_Tech_Trends_Report_2020-2040.pdf; National Intelligence Council Strategic Futures Group, "The Future of the Battlefield" (Global Trends, March 2021), <https://www.dni.gov/index.php/gt2040-home/gt2040-deeper-looks/future-of-the-battlefield>, (accessed 31 May, 2024); Diane DiEuliis and Peter Emanuel, "Cyborg Soldier 2050: Human-Machine Fusion and its Implications", Section 2 in: Strategic Latency Unleashed: The Role of Technology in a Revisionist Global Order and the Implications for Special Operations Forces (Center for Global Security Research Lawrence Livermore National Laboratory, 2021), <https://cgsr.llnl.gov/content/assets/docs/StratLatUnONLINE.pdf>.
3. Jacek Durkalec, Anna Péczeli, Brian Radzinsky, "Nuclear decision-making, complexity and emerging and disruptive technologies: A Comprehensive Assessment" (European Leadership Network, February 2022), p.7, <https://europeanleadershipnetwork.org/wp-content/uploads/2022/01/LLNL.pdf>.
4. United States Naval Academy, "Chapter 20: Command, Control and Communication", (Fundamentals of Navy Weapons Systems, n.d.), <https://fas.org/man/dod-101/navy/docs/fun/part20.htm>, (accessed 31 May, 2024). The US Department of Defense, for instance, defines the NC3 enterprise as: "the collection of activities, processes, and procedures performed by appropriate commanders and support personnel who, through the chain of command, allow for decisions to be made based on relevant information, and allow those decisions to be communicated for forces for execution". (Department of Defense, "2016 Nuclear Matters Handbook: New Revised Edition, Authoritative Guide to American Atomic Weapons, History, Testing, Safety, Security, Delivery Systems, Physics and Bomb Designs, Terror Threats", p. 268, (Washington, D.C.: Progressive Management, 2016).
5. Salma Shaheen, "Nuclear Command and Control Norms: A Comparative Study", p.3 (London: Routledge, 2019).
6. Scott D. Sagan, "The Origins of Military Doctrine and Command and Control Systems", p. 16 (quoted in Salama Shaheen, p. 3, op cit., 2000)
7. Paul Bracken, "Communication Disruption Attacks in a Nuclear Context", (Defense.info, 25 October, 2019), <https://defense.info/re-shaping-defense-security/2019/10/communication-disruption-attacks-in-a-nuclear-context/>.
8. John Gower, "United Kingdom: Nuclear Weapon Command, Control, and Communications", p.7, (NAPSNet Special Reports, September 12, 2019) <https://nautilus.org/napsnet/napsnet-special-reports/united-kingdom-nuclear-weapon-command-control-and-communications/?view=pdf>.
9. The risks presented were identified during an ELN workshop titled "Nuclear weapons and new tech: identifying new challenges to nuclear command, control, and communications," held in London in November 2023, and subsequently synthesised by the authors.
10. Allegedly the Russian Strategic Rocket Force can ensure the automated launch of a retaliatory nuclear strike when command chain and combat management systems are destroyed, and personnel missile units are dead. Known in Russia as 'Perimeter', and in the west as 'Dead Hand', experts remain divided on whether such a system exists.
11. Scott D. Sagan, "The Origins of Military Doctrine and Command and Control Systems", p. 16 (quoted in Salama Shaheen, p. 8, op cit., 2000)
12. Ibid
13. Salma Shaheen, 2019, op cit, pp. 7 - 9.
14. Bruce G. Blair, "The Logic of Accidental Nuclear War", p. 71., (The Brookings Institution, Washington D.C, 1993).
15. John Gower, "United Kingdom: Nuclear Weapon Command, Control, and Communications", p. 4, (NAPSNet Special Reports, September 12, 2019), <https://nautilus.org/napsnet/napsnet-special-reports/united-kingdom-nuclear-weapon-command-control-and-communications/?view=pdf>.
16. Ibid.
17. Office of the Deputy Assistant Secretary of Defense for Nuclear Matters, "Nuclear Matters Handbook 2020 (Revised)", p. 17, (Washington, D.C., 2020), <https://www.acq.osd.mil/ncbdp/nm/NMHB2020rev/docs/NMHB2020rev.pdf>.

18. One of the key debates among US and Chinese scholars involves the co-mingling of nuclear and conventional command and control infrastructure, which creates risks for inadvertent escalation. For further reading see: Lewis, J. W., & Litai, X, "Making China's nuclear war plan", 68(5), 45–65, p. 61, (Bulletin of the Atomic Scientists, 2012) <https://doi.org/10.1177/0096340212459155>.
19. Natasha E. Bajema and John Gower, "A Handbook For Nuclear Decision-Making and Risk reduction in an Era of Technological Complexity", p. 67, (Jane E. Nolan Center, Council on Strategic Risks, 2012). <https://councilonstrategicrisks.org/wp-content/uploads/2022/12/NuclearTechnologicalComplexity-Dec22.pdf>.
20. Andrew Futter, "Explaining the Nuclear Challenges Posed by Emerging and Disruptive Technology: A Primer for European Policymakers and Professionals", p. 2, (EU Non-Proliferation and Disarmament Consortium, No. 73, March, 2012) https://www.nonproliferation.eu/wp-content/uploads/2021/03/EUNPDC_no-73_FINAL-1.pdf
21. For more information on the first iteration of the project, see: Katarzyna Kubiak, Sylvia Mishra and Graham Stacey, "Nuclear weapons decision-making under technological complexity: pilot workshop report. Global Security", (European Leadership Network, March, 2021), <https://www.europeanleadershipnetwork.org/wp-content/uploads/2021/03/ELN-Pilot-Workshop-Report-1.pdf>.
22. Andrew Futter, 2021, op cit, p. 3.
23. Ibid.
24. Ibid.
25. The following threats posed by EDTs were identified by the ELN and project partners during a previous iteration of the project.
26. For a comprehensive and detailed analysis of the risks posed by cyber-attacks to AI, refer to: Andrew J. Lohn, "Hacking AI. A Primer for Policy Makers on Machine Learning Cybersecurity", (Center for Security and Emerging Technology, December, 2020), <https://cset.georgetown.edu/publication/hacking-ai/>
27. For a comprehensive overview of the computing limitations of onboard AI-powered systems, consult: Kyle A. Miller and Andrew J. Lohn, "Onboard AI: Constraints and Limitations", (Center for Security and Emerging Technology, August, 2023), <https://cset.georgetown.edu/publication/onboard-ai-constraints-and-limitations/>
28. Sebastian Brixey-Williams, Alice Spilman and Nicholas J. Wheeler describe a "Multi-Stakeholder Dialogue" as an instance that brings together "two or more parties to exchange their perceptions, explore possibilities for reaching new shared understandings, and identify practices that could enable the parties to reduce distrust, and potentially build trust, leading to the reduction of strategic risks." ("The Nuclear Responsibilities Toolkit: A Practical Guide for Thinking, Talking and Writing", p. 30, (BASIC & ICCS, January, 2022), https://basicint.org/wp-content/uploads/2022/01/BASIC_Nuclear-Responsibilities-Toolkit_2nd-Edition.pdf.)
29. Heather Williams, "Remaining relevant: Why the NPT must address emerging technologies", p. 7, (King's College London, August, 2020), <https://www.kcl.ac.uk/csss/assets/remaining-relevant-new-technologies.pdf>
30. United Nations, "2020 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons", (UN Document NPT/CONF.2020/CRP.1/Rev.2. 2020), <https://app.unidir.org/sites/default/files/2023-08/2020NPTRevConDraft.pdf>
31. Heather Williams has identified the need to include this set of stakeholders on EDTs. See Heather Williams, "The Nuclear Order and Emerging Technologies", p.8, (Centre for Science and Security Studies, King's College London, February, 2022), <https://www.kcl.ac.uk/csss/assets/the-nuclear-order-and-emerging-technologies.pdf>.
32. United Nations, "A Nuclear Risk Reduction Package", p. 4, (Working paper submitted by the Stockholm Initiative, to the 2020 Review Conference of the parties to the Treaty on the Non-Proliferation of Nuclear Weapons, UN document NPT/CONF.2020/WP.9/Rev.1, 2022) <https://www.regeringen.se/>
33. Ibid, p. 3.
34. NATO, "Foster and Protect: NATO's Coherent Implementation Strategy on Emerging and Disruptive Technologies" (2021).
35. For a timeline of NATO's evolution towards the development of its EDTs policies, refer to: NATO, "Evolution", (Emerging and disruptive technologies, 30 May, 2024). https://www.nato.int/cps/en/natohq/topics_184303.htm#evolution, (accessed 30 May 2024); For more information on NATO's approach and initiatives linked to EDTs and its digital transformation, refer to: NATO, "Digital Transformation", (n.d.), https://www.nato.int/cps/en/natohq/topics_184303.htm, (accessed 30 May 2024); and NATO, "Emerging and Disruptive Technologies", (2024), https://www.nato.int/cps/en/natohq/topics_184303.htm.

The European Leadership Network (ELN) is an independent, non-partisan, pan-European network of over 450 past, present and future European leaders working to provide practical real-world solutions to political and security challenges.

Published by the European Leadership Network, July 2024.

Published under the Creative Commons Attribution-ShareAlike 4.0

© The ELN 2024

The European Leadership Network itself as an institution holds no formal policy positions. The opinions articulated in this policy brief represent the views of the author(s) rather than the European Leadership Network or its members. The ELN aims to encourage debates that will help develop Europe's capacity to address the pressing foreign, defence, and security policy challenges of our time, to further its charitable purposes

We operate as a charity registered in England and Wales under Registered Charity Number 1208594.

European Leadership Network
8 St James's Square
London, SE1Y 4JU
United Kingdom

Email: secretariat@europeanleadershipnetwork.org

Tel: 0203 176 2555

Website: europeanleadershipnetwork.org

Follow us    