EUROPEAN LEADERSHIP NETWORK

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

TECHNOLOGY INHERENT RISKS

- Malfunction
- Cyber-attacks
- Computing constraints

OPERATIONAL RISKS

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

STRATEGIC RISKS

- 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?

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?

Technology inherent risks

Malfunction

Cyber-attacks

Computing constraints

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

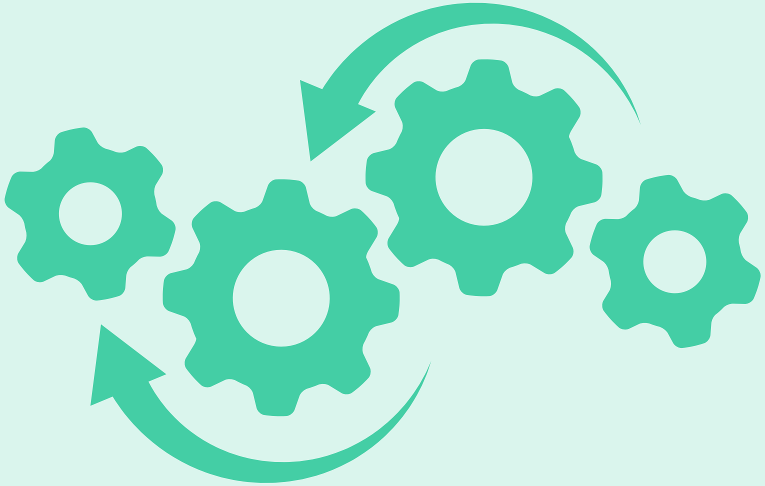
Guardrail

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

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?

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.

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?

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

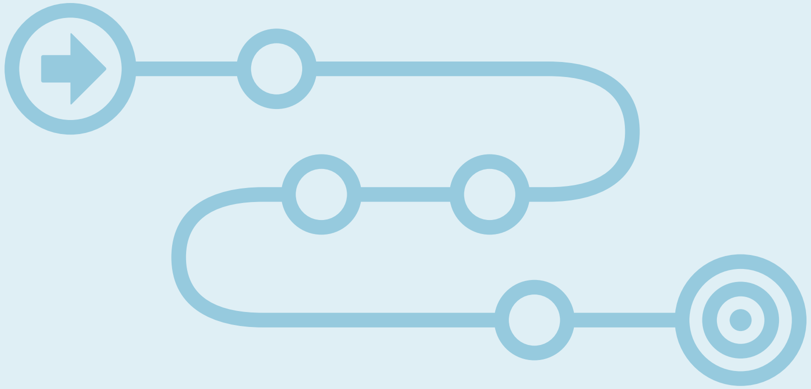
Guardrail

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?

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?

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?