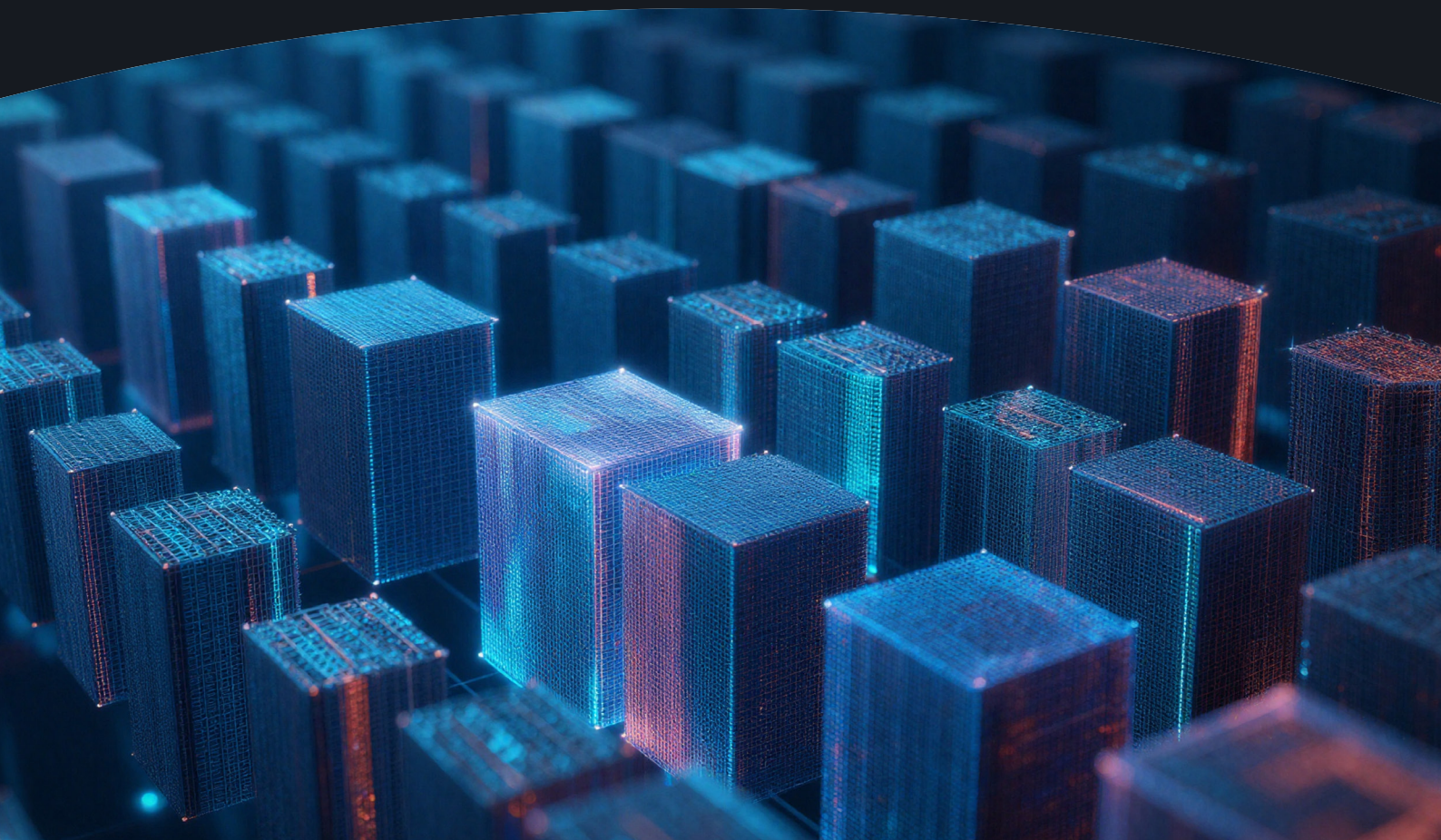




UNIDIR

Guidelines for the Development of a National Strategy on AI in Security and Defence

CENTRE OF EXCELLENCE ON AI, PEACE AND SECURITY



Acknowledgements

Support from UNIDIR's core funders provides the foundation for all of the Institute's activities. The Guidelines were prepared by the Centre of Excellence on AI, Peace and Security, which is funded by the Governments of France, Germany, Italy, the Kingdom of the Netherlands, Pakistan, the Republic of Korea and Switzerland, and by Microsoft and the Fondation pour l'adaptation de la Genève internationale.

The authors wish to thank Dr. Giacomo Persi Paoli for advice, guidance and support the first and second phases of this programme of work, as well as Ioana Puscas (UNIDIR), Vhaire Gudgeon (UNIDIR), Dr. Samuel Segun (Global Center on AI Governance) and Major Jamal Mohamed Hassan (Ministry of Defence, Kenya) for their reviews and suggestions on the draft and final guidelines.

The authors also wish to thank States that participated in the focus group discussions for the first phase of this project, in March 2024, as well as members of the Roundtable for AI, Security and Ethics (RAISE) for their valuable contributions and input to the guidelines. Finally, the authors wish to thank Jessica Espinosa Azcárraga and Maria Belen Lopez Conte for their tremendous support for this project.

The Centre of Excellence on AI, Peace and Security is UNIDIR's global, permanent platform dedicated to advancing governance of artificial intelligence in the peace and security domain. Based in Geneva, it supports more effective and inclusive approaches to the challenges that AI presents through research, dialogue and capacity-building.

About UNIDIR

The United Nations Institute for Disarmament Research (UNIDIR) is a voluntarily funded, autonomous institution within the United Nations. One of the few policy institutes worldwide focusing on disarmament, UNIDIR generates knowledge and promotes dialogue and action on disarmament and security. Based in Geneva, UNIDIR assists the international community in developing the practical, innovative ideas needed to find solutions to critical security problems. For more information, visit [unidir.org](https://www.unidir.org).

Note

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. The views expressed in the publication are the sole responsibility of the individual authors. They do not necessarily reflect the views or opinions of the United Nations, UNIDIR, its staff members or sponsors.

Citation

Yasmin Afina, Jan Hendrik Mannsperger and Edward Madziwa, *Guidelines for the Development of a National Strategy on AI in Security and Defence*, Geneva: UNIDIR, 2026. <https://doi.org/10.37559/SECTEC/2026/AI/08>.

Cover photo: AdobeStock.

Authors

The guidelines were produced by UNIDIR's Centre of Excellence on AI, Peace and Security. They were drafted by Dr Yasmin Afina, principal investigator for this project, as well as Jan Hendrik Mannsperger and Edward Madziwa.



Yasmin Afina

Researcher & Military AI Lead, Centre of Excellence on AI, Peace and Security

Dr Yasmin Afina is a Researcher for the Security and Technology Programme at UNIDIR, where her research covers the intersection between international security, international law and artificial intelligence. Yasmin leads the Military AI thematic portfolio of the Centre of Excellence on AI, Peace and Security. She holds a PhD in International Law from the University of Essex, where she examined the international humanitarian law considerations for the development of AI for military targeting.



Jan Hendrik Mannsperger

Graduate Professional, Centre of Excellence on AI, Peace and Security

Jan Hendrik Mannsperger is a Graduate Professional with UNIDIR's Centre of Excellence on AI, Peace and Security. His research focuses on artificial intelligence and international security. He holds Bachelor's degrees in Political Science and in Mass Media and Communication Science from the University of Vienna.



Edward Madziwa

Programme Assistant, Security and Technology Programme

Edward Madziwa is a Programme Assistant for the Security and Technology Programme at UNIDIR. He holds an LLM in International Humanitarian Law and Human Rights from the Geneva Academy; an LLM in Constitutional and Human Rights Law and a Bachelor of Laws Honours Degree from Midlands State University, Zimbabwe. His experience and expertise include international humanitarian law, public international law and the law of weaponry and new technologies including artificial intelligence and cyber.

Acronyms and Abbreviations

AI	Artificial intelligence
CBM	Confidence-building measure
CCW	(Convention on) Certain Conventional Weapons
CERT	Computer emergency response team
ICT	Information and communications technology
IHL	International humanitarian law
IHRL	International human rights law
ISR	Intelligence, surveillance and reconnaissance
R&D	Research and development
REAIM	Responsible AI in the Military Domain
SOP	Standard operating procedure

Contents

Acknowledgements	2
Authors	3
Acronyms and Abbreviations	4
Introduction	10
Background	10
Purpose and Structure of the Guidelines	11
Why a National Strategy on AI in Security and Defence?	12
Scope	13
Methodology	15
CHAPTER I. PROCEDURAL GUIDELINES	17
P.1. Development and adoption of a national strategy on AI in security and defence	18
P.1.1. Conceptualization	18
P.1.2. Definition of objectives	19
P.1.3. Definition of desired outcomes	19
P.1.4. Definition of scope	19
P.1.5. Distribution of roles and responsibilities	20
P.1.6. Assessment of financial resources and capacities	20
P.1.7. Internal consultations	20
P.1.8. External consultations	21
P.1.9. Assessment of relevant geopolitical dynamics at the international and regional levels	21
P.1.10. Assessment of relevant political dynamics and context at the national level	22
P.1.11. Assessment of the wider AI policy and regulatory landscape	22
P.1.12. Identification and assessment of relevant actors	23
P.1.13. Assessment of international law	23
P.1.14. Assessment of ethics	23
P.1.15. Assessment of technical standards	24
P.1.16. Assessment and management of AI risks	24
P.1.17. Assessment of non-AI security risks	24
P.1.18. Collective drafting of a national strategy document	24
P.1.19. Classification: To publish or not to publish?	25
P.1.20. Review of the draft national strategy document	25
P.1.21. Validation process for the draft national strategy document	25
P.1.22. Adoption of the national strategy document	26
P.1.23. Public communication of the national strategy document	26

P.2. Implementation and review of a national strategy on AI in security and defence	27
P.2.1. Key actors for implementation	27
P.2.2. Key coordinating bodies and actors	27
P.2.3. Allocation of roles to new and existing bodies	28
P.2.4. Clarity of structure and distribution of roles and responsibilities	28
P.2.5. Key actions for implementation	28
P.2.6. Key timelines and milestones for implementation	29
P.2.7. Accountability for implementation	29
P.2.8. Adoption of an implementation plan	29
P.2.9. Establishment of mechanisms for oversight, monitoring and verification	29
P.2.10. Establishment of mechanisms for review and revision	30
P.3. Cooperation	31
P.3.1. National cooperation	31
P.3.2. Bilateral cooperation	31
P.3.3. Regional cooperation	31
P.3.4. International cooperation	32
P.3.5. Cooperation with technology providers	33
P.3.6. Cooperation with civil society	33
P.4. Lessons from and convergence with select fields	34
P.4.1. The cybersecurity field	34
P.4.2. The nuclear field	34
P.4.3. The biological and chemical security field	35
P.4.4. The aviation sector	35
P.4.5. AI safety	35
P.4.6. Civilian AI governance	35
CHAPTER II. SUBSTANTIVE GUIDELINES	37
S.1. Data practices	38
S.1.1. Data provenance	38
S.1.2. Data-collection practices	39
S.1.3. Data storage and infrastructure	39
S.1.4. Data hygiene	40
S.1.5. Appropriateness of training data	40
S.1.6. Appropriateness of testing data	40
S.1.7. Proxy data use	40
S.1.8. Synthetic data use	41
S.1.9. Data governance and international law	41
S.1.10. Frameworks and agreements for the sharing of classified data	42

S.2. Implications of machine learning use	43
S.2.1. Black box	43
S.2.2. Probabilistic nature of AI technologies and accuracy rates	43
S.2.3. Safe testing environments and practices	43
S.2.4. Safe evaluation environments and practices	44
S.2.5. Transparency in testing and evaluation processes	44
S.2.6. Open-source AI	44
S.2.7. Foundation models and generative AI	44
S.2.8. Agentic AI systems	45
S.2.9. Adversarial robustness and AI security	45
S.2.10. The human element	46
S.3. Public-private partnerships	47
S.3.1. Private companies and development of AI for defence	47
S.3.2. Distribution of roles and responsibilities	47
S.3.3. Foster research and development ecosystems	48
S.3.4. Intellectual property	48
S.3.5. Supply chain security	48
S.3.6. Security and resilience of the industrial ecosystem	48
S.3.7. Evolving procurement dynamics	49
S.3.8. Maintenance	49
S.4. Technological transfer	50
S.4.1. Technological transfer, equity and proliferation risks	50
S.4.2. Procurement from foreign governments	50
S.4.3. Sales to foreign governments	50
S.4.4. Sales to non-governmental entities	51
S.4.5. Oversight and verification	51
S.4.6. Compute governance and strategic dependencies	51
S.4.7. Engagement with foreign industries	52
S.5. Life cycle management	53
S.5.1. Identification of relevant life cycle stages	53
S.5.2. Identification of relevant actors across life cycle stages	53
S.5.3. Distribution of roles and responsibilities across the technology's life cycle	54
S.5.4. Considerations for the technology's procurement	54
S.5.5. Considerations for the technology's development	54
S.5.6. Considerations for the technology's testing and evaluation	55
S.5.7. Considerations for the technology's adoption, integration and deployment	55
S.5.8. Considerations for the technology's use	55
S.5.9. Considerations for the technology's end of life and decommissioning	56

S.6. Capacity-building	57
S.6.1. Human capacity	57
S.6.2. Technological and infrastructural capacity	57
S.6.3. Financial capacity and resource mobilization	57
S.6.4. Regulatory and policy capacity	58
S.7. Human resources	59
S.7.1. Training and retention of AI talent	59
S.7.2. Capacity-building and awareness-raising of developers	59
S.7.3. Capacity-building, upskilling and awareness-raising of users	59
S.7.4. Capacity-building and awareness-raising of policy and regulatory entities	60
S.7.5. Academic and civil society engagement	60
S.7.6. Industry engagement	60
S.8. Legal compliance	61
S.8.1. Compliance with international humanitarian law	61
S.8.2. Compliance with international human rights law	61
S.8.3. Compliance with international criminal law	62
S.8.4. Compliance with public international law	62
S.8.5. Compliance with other bodies of international law	63
S.8.6. Implications of technological transfer for export and import control regimes	63
S.8.7. Articulation of a national position on the interpretation of international law in the context of AI in security and defence	64
S.8.8. International law and ethics	64
S.8.9. Mechanisms for oversight and evaluation of legal compliance	65
S.9. Ethics	66
S.9.1. Gender bias	66
S.9.2. Racial bias	66
S.9.3. Other forms of discrimination	67
S.9.4. Fairness	67
S.9.5. Explainability and traceability	67
S.10. Defence applications of AI	68
S.10.1. Scope of the defence applications of AI	68
S.10.2. Integration into weapon systems	68
S.10.3. Integration into non-weapon systems	69
S.10.4. AI in information operations and cognitive warfare	69
S.10.5. Legal compliance	69
S.10.6. Key arms control and disarmament considerations	70

S.11. Defence applications of AI	71
S.11.1. Scope of the security applications of AI	71
S.11.2. AI integration into surveillance operations	71
S.11.3. Biometric identification and facial recognition	71
S.11.4. AI in border management and migration	72
S.11.5. Cross-border data-sharing practices and AI implications	72
S.11.6. AI integration into law enforcement mechanisms, practices and operations	72
S.11.7. Oversight and accountability	73
S.12. AI integration into critical national infrastructure	74
S.12.1. Assessment of past, present and future AI integration into critical national infrastructure	74
S.12.2. Risk assessment	74
S.12.3. Sectoral risk-mitigation measures	74
S.12.4. Dependencies and risks	75
S.12.5. Interoperability with legacy systems	75
S.13. Resilience and preparedness	76
S.13.1. Plan for emergency response	76
S.13.2. Establishment of an emergency response team	76
S.13.3. Distribution of roles and responsibilities	76
S.13.4. Capacity-building	76
S.13.5. Risk assessment and risk mitigation	77
S.13.6. Incident recording, notification and analysis	77
S.13.7. The conduct of exercises and stress-testing	77
S.13.8. Containment and behavioural safety	77
S.13.9. AI incident classification framework	78
S.13.10. Dedicated resources for resilience and preparedness	78
S.13.11. Strategy for the maintenance and security of facilities	78
S.13.12. Preparedness for long-term technological progress and evolutions	79
Conclusion and next steps in the project	80

Introduction

Background

As innovation in artificial intelligence (AI) proceeds at breakneck speed, the appetite exhibited by States for devising and adopting frameworks for the governance of the research, development, deployment and use of these technologies has reached unprecedented heights. With calls for governance solutions increasing at both the national and international levels, the number of national strategy documents that frame the development, deployment and use of these technologies has started to grow across regions, with a notable increase in 2024–2026 – a development noted by the Secretary-General of the United Nations in his most recent report on developments in science and technology and their potential impacts.¹

Yet, only a handful of national strategy documents have a section dedicated to AI strategies and policies on security and defence applications. Even fewer strategy documents are specifically, entirely dedicated to this realm – and differences can be observed across regions. This scarcity is at odds with the Secretary-General’s recommendation for Member States to “urgently develop national strategies on responsible design, development and use of artificial intelligence”, as outlined in his *New Agenda for Peace*.²

A number of State-led initiatives, such as the Call to Action of the 2023 Responsible AI in the Military Domain (REAIM) Summit, have also recognized national strategies as being key to enabling the responsible development, deployment and use of military AI.³ The importance of the strategies being formulated “to ensure responsible AI applications in the military domain” was further stressed in the REAIM Blueprint for Action, adopted in 2024.⁴ The outcome document of the third REAIM Summit, the 2026 Pathways to Action, further recognizes “sharing appropriate information on national policies, principles, governance structures and oversight arrangements” as a form of confidence-building measure (CBM) to strengthen trust and transparency.⁵

Against this backdrop, UNIDIR launched in 2024 a programme of work to establish guidelines for the development, adoption, implementation and review of national strategies on AI in security and defence. This programme sought to anticipate, capture and dissect the key issues, considerations and needs

-
- 1 General Assembly, “Current developments in science and technology and their potential impact on international security and disarmament efforts”, Report of the Secretary-General, A/81/246, 21 July 2026, <https://docs.un.org/A/81/246>, paragraph 10. See also Y. Afina, *The Global Prism of Military AI Governance: Reflections from the 2025 Regional Consultations on Responsible AI in the Military Domain* (Geneva: UNIDIR, 2026), <https://unidir.org/publication/the-global-prism-of-military-ai-governance-reflections-from-the-2025-regional-consultations-on-responsible-ai-in-the-military-domain>.
 - 2 United Nations, *A New Agenda for Peace*, Our Common Agenda Policy Brief 9 (New York: United Nations, July 2023), <https://dppa.un.org/en/a-new-agenda-for-peace>, 28.
 - 3 Responsible AI in the Military domain (REAIM) Summit, “REAIM Call to Action”, 15–16 February 2023, <https://www.government.nl/documents/publications/2023/02/16/ream-2023-call-to-action>, paragraph 5.
 - 4 Responsible AI in the Military domain (REAIM) Summit, “REAIM Blueprint for Action”, 9–10 September 2024, https://hcass.nl/wp-content/uploads/2026/06/REAM_Blueprint-for-Action-Seoul.pdf, paragraph 8.
 - 5 Responsible AI in the Military domain (REAIM) Summit, “REAIM Pathways to Action”, 4–5 February 2026, https://www.exteriores.gob.es/en/REAM2026/Documents/REAM_2026_Pathways_to_Action.pdf, paragraph 23(a).

that must be addressed as States grapple with the topic of responsible AI in security and defence. It did so by drawing on States' good practices and their shared methods in this space, along with input from non-State stakeholders with a role in the AI governance ecosystem. As a key milestone, and to mark the project's first phase, in October 2024 UNIDIR launched its *Draft Guidelines for the Development of a National Strategy on AI in Security and Defence*.⁶ The project team subsequently collected direct feedback and suggestions from government representatives and civil society organizations to further consolidate the draft guidelines. Complemented with a literature review and the insights that UNIDIR collected between 2024 and 2026 through its engagements on AI in security and defence, the project team has reviewed and revised the draft guidelines into the present document.

Purpose and structure of the guidelines

Echoing the approach adopted in the draft guidelines, the purpose of the revised guidelines is to anticipate, capture and dissect the key issues, considerations and needs that each State must address as it seeks to develop, adopt, implement and review a national strategy on AI in security and defence. In recognition of the host of incentives that arise from the establishment of such strategies (as described in further detail in the following section), it is hoped that these guidelines will serve as a useful tool for States and non-State stakeholders alike as they seek to address issues related to the responsible development, deployment and use of AI in security and defence.

While they do not seek to be prescriptive, the guidelines provide a series of considerations and recommendations divided into two types:

1. Procedural guidelines

The first part of the guidelines provides States with recommendations on the process surrounding the development, adoption, implementation and review of a national strategy on AI in security and defence. These guidelines build on good practices, meaningful efforts and existing approaches adopted and considered by States at the national, regional and international levels. It is hoped that the procedural guidelines will provide States and non-State stakeholders with a roadmap with key steps, processes and considerations to support the development, adoption, implementation and review of a national strategy on AI in security and defence.

2. Substantive guidelines

The second part of the guidelines outlines a series of substantive considerations that States should examine and integrate, as appropriate, in the development, adoption, implementation and review of a national strategy on AI in security and defence. While in no way exhaustive of all relevant considerations – which are highly context-dependent, varying from one State to another and across regions – it is hoped that the substantive guidelines will provide States and non-State stakeholders with food for thought on key issues that may, or may be anticipated to, influence the State's approach and subsequent strategy in this space.

6 UNIDIR Security and Technology Programme, *Draft Guidelines for the Development of a National Strategy on AI in Security and Defence* (Geneva: UNIDIR, 2024), <https://unidir.org/publication/draft-guidelines-for-the-development-of-a-national-strategy-on-ai-in-security-and-defence>.

The establishment of these guidelines is further motivated by the recognition that the stage at which States find themselves in the development, adoption, implementation or review of their national strategies on AI in security and defence varies both within and across regions.⁷ The guidelines will thus serve two purposes: for States in the relatively early stages of developing a strategy, and even more so for those considering doing so, the guidelines will serve as a tool to build and consolidate capacity; for States already at the stage of implementation, or even review, of their national strategy on AI in security and defence, the guidelines will remind them of the key policy, legal and ethical considerations surrounding the responsible development, deployment and use of these technologies.

In addition, while the development of a national strategy on AI in security and defence would require the mobilization of considerable resources and time, it is hoped that these guidelines will also showcase the dynamic and “living” nature of a strategy. Similar to the development of a national strategy on cybersecurity, States are not expected to craft a perfect, all-inclusive and everlasting strategy from the outset. It should rather be viewed as a framework to be developed and refined through implementation, review and updating. It is therefore hoped that these guidelines can encourage States to approach strategy development not as a quest for perfection but as a necessary first step; and that they will serve as a companion for States throughout their journey in developing, adopting, implementing and reviewing their national strategy on AI in security and defence.

Why a national strategy on AI in security and defence?

There are a number of reasons and incentives for a State to establish a national strategy on AI in security and defence. While such a strategy can take many forms, its effective development, adoption, implementation and review present many opportunities for the State to build its capacity, readiness and maturity in the face of technological progress in this space:

- **Clear state of the art:** A comprehensive survey and understanding of the technological landscape in security and defence are essential for the development, adoption, implementation and review of a national strategy. This will not only provide the State with clarity on the state of the art with regards to technological progress; this will also enable it to comprehensively map relevant stakeholders as well as their respective roles and responsibilities.
- **Evident distribution of roles and responsibilities:** Building on the comprehensive mapping of relevant stakeholders in the governance of AI in security and defence, a national strategy will provide clarity on the distribution of roles and responsibilities and, when appropriate, the working relationships between actors, agencies and organizations, both public and private.
- **Effective coordination:** In addition to a clear distribution of roles and responsibilities, a national strategy also provides a framework to enable effective coordination between actors, agencies and organizations, both public and private. Beyond communication channels, effective coordination is necessary to ensure alignment in implementing existing policies, laws and ethical guidelines, thus preventing fragmentation between government agencies and organizations and between the public and private sectors. Effective coordination will also be needed to optimize the necessary financial, human and technological resources.

⁷ Yasmin Afina, *The Global Kaleidoscope of Military AI Governance: Decoding the 2024 Regional Consultations on Responsible AI in the Military Domain* (Geneva: UNIDIR, 2024), <https://unidir.org/publication/the-global-kaleidoscope-of-military-ai-governance>.

- **Change management:** In recognition of AI's general-purpose and cross-cutting nature, a national strategy will provide a clear and stable framework to manage the organizational changes that the development, deployment and use of these technologies in security and defence will entail.
- **Resource mobilization:** Defining a State's technological and regulatory objectives, aspirations and vision in a national strategy will enable the mobilization and subsequent allocation of the financial, human and technological resources required to operationalize the State's approach to the governance of AI in security and defence.
- **Public oversight:** A national strategy on AI in security and defence will demystify the possible applications of these technologies and the frameworks within which they are developed, deployed and operate. The general public will be informed of the State's intention, plans and approaches surrounding these technologies, thus enabling appropriate public oversight.
- **Responsible innovation:** The development, deployment and use of AI in security and defence present a host of opportunities, from enhanced intelligence, surveillance and reconnaissance (ISR) capabilities to logistics support and management. As public and private actors alike seek to harness these opportunities, strategy documents act as a means to ensure a responsible approach to AI innovation. A national strategy also serves as a means to formally mainstream specific aspects of "responsible AI", such as ethics, bias mitigation and data governance.
- **Risk reduction:** In addition to the opportunities offered by AI technologies in security and defence, the risks stemming from their development, deployment and use must be addressed. By capturing these risks, a national strategy offers an opportunity for a State to identify risk-reduction and risk-mitigation approaches, and a concrete plan for their subsequent implementation and operationalization. Specific considerations for security and defence applications to take account of the dual-use nature of AI technologies will be particularly important, as innovation in the civilian space proceeds at breakneck speed.
- **International collaboration:** A national strategy provides clarity for the international community on a State's objectives, its state of the art, and its overall approach to the governance of AI in security and defence. As well as serving as a confidence building measure, this clarity will also pave the way for potential collaborations at the inter-State, regional and international levels.
- **Interpretation of international law:** The development, deployment and use of AI in security and defence are all framed by a series of international legal frameworks, including international humanitarian law (IHL), international human rights law (IHRL) and *jus ad bellum*. Yet, while there is agreement on applicability, there is no universally agreed consensus as to how international law applies. A national strategy offers an opportunity for a State to reflect on how it would approach the interpretation of international law.

Scope

The guidelines cover AI in "security and defence". The widening of the scope of the guidelines beyond defence (i.e., the military domain) into the broader domain of security is driven by a number of reasons (outlined in Figure 1). Recognizing that there is no universally accepted approach to defining "security" and "defence", UNIDIR has generally unpacked four dimensions of the military domain as a means to classify operations (see Table 1).⁸

8 UNIDIR Security and Technology Programme, *Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security: An Evidence-Based Road Map for Future Policy Action* (Geneva: UNIDIR, 2025), <https://unidir.org/publication/artificial-intelligence-in-the-military-domain-and-its-implications-for-international-peace-and-security-an-evidence-based-road-map-for-future-policy-action>.

FIGURE 1: THE RATIONALE FOR CONSIDERING AI IN SECURITY AND DEFENCE

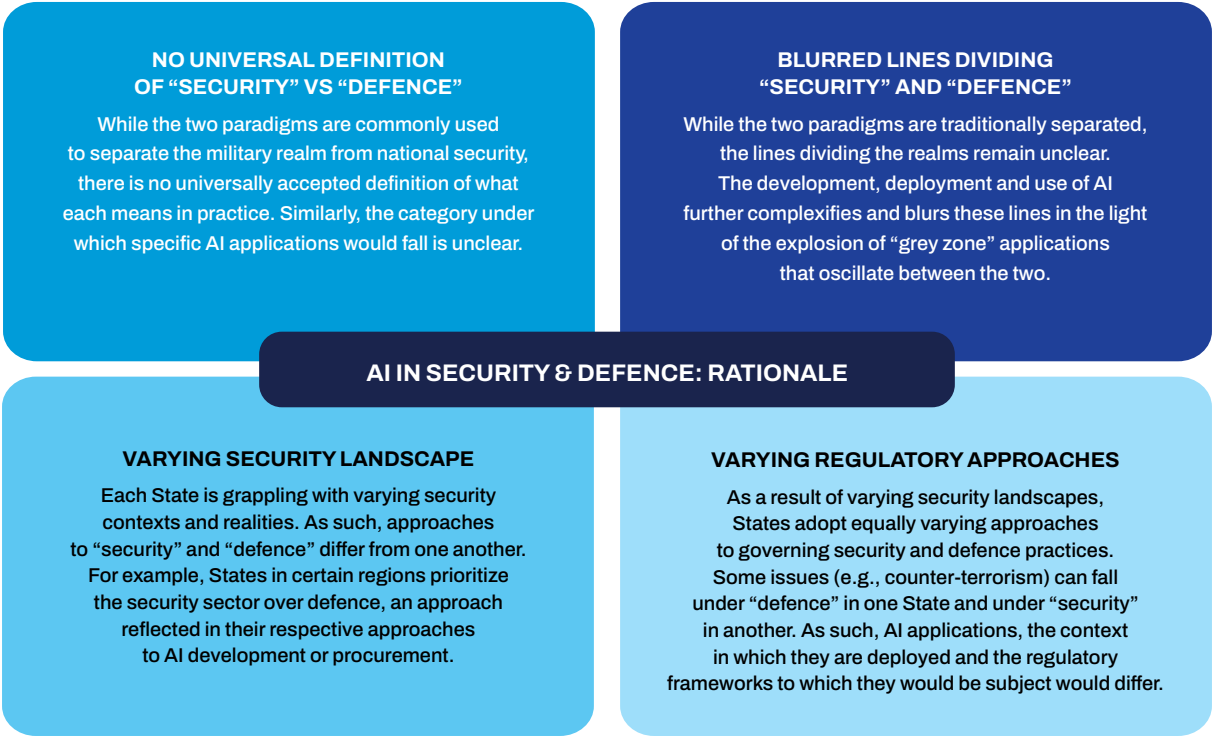


TABLE 1: THE FOUR DIMENSIONS OF THE MILITARY DOMAIN

ARMED CONFLICT	OTHER MILITARY OPERATIONS REQUIRING THE USE OF FORCE*	SUPPORT TO NATIONAL SECURITY	MILITARY ASSISTANCE
International armed conflicts	Counter-piracy	Support for national law enforcement	Peace operations
Non-international armed conflicts	Counter-terrorism	Border security	Humanitarian and disaster relief
	Counter-insurgency	Protection of national critical infrastructure sites	Evacuation of civilians
	Combating organized crime		Search and rescue

**All operations listed in this category could potentially reach the legal threshold for being considered non-international armed conflicts.*

Methodology

The guidelines are the product of an extensive review of existing national strategies and governmental policy documents pertinent to AI in security and defence, where available in the public domain.⁹ Existing guidelines, both related and unrelated to technology security, have also been studied to draw lessons on scoping, structure and delivery.¹⁰

In October 2024, UNIDIR released the guidelines in draft form and invited feedback from both State and non-State stakeholders. Ahead of the launch of the draft guidelines, the project team held four focus group discussions in early 2024 to consult with State representatives from all five United Nations regional groupings (Africa, the Asia-Pacific, Eastern Europe, Latin America and the Caribbean, and Western Europe and others).¹¹ A multi-stakeholder consultation with representation from civil society, industry and academia across all five regional groupings also took place in early 2024. Further experts were consulted bilaterally to complement and consolidate UNIDIR's findings.

The project team subsequently collected and processed the feedback and suggestions received from government representatives and civil society organizations to further consolidate the draft guidelines. Complemented with a literature review and the insights that UNIDIR collected between 2024 and 2026 through its engagements on AI in security and defence – which drew in particular from the 2024 and 2025 REAIM regional consultations – the project team reviewed and revised the draft guidelines into the present document.¹²

The present guidelines are intended as a tool for States and all relevant stakeholders involved in the development, adoption, implementation and review of national strategies on AI in security and defence. While no longer a draft, UNIDIR would welcome feedback from all stakeholders that make use of the guidelines and from all perspectives, viewpoints and approaches to this issue.

9 Many of the resources consulted were extracted from UNIDIR's AI, Peace and Security Portal, <https://aipolicyportal.org>, a comprehensive repository of policy documents on AI of all United Nations Member States and various intergovernmental organizations and multi-stakeholder and other initiatives.

10 Among the model guidelines and other documents studied, see in particular International Telecommunications Union (ITU) et al., *Guide to Developing a National Cybersecurity Strategy: Strategic Engagement in Cybersecurity* (Geneva: ITU, 2018), https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-CYB_GUIDE.01-2018-PDF-E.pdf; Noam Lubell, Jelena Pejic and Claire Simmons, *Guidelines on Investigating Violations of International Humanitarian Law: Law, Policy, and Good Practice* (Geneva: Geneva Academy and ICRC, 2019), https://www.icrc.org/sites/default/files/document/file_list/guidelines_on_investigating_violations_of_ihl_final.pdf; Global Task Force for Inclusive AI, “[Draft] Guidelines for Participatory and Inclusive AI”, Partnership on AI, 2024; Centre for Humanitarian Dialogue, “Code of conduct on artificial intelligence in military systems”, 2021, <https://hdcentre.org/wp-content/uploads/2021/08/AI-Code-of-Conduct.pdf>.

11 United Nations, “Regional groups of Member States”, n.d., <https://www.un.org/dgacm/en/content/regional-groups>.

12 For insights on the 2024 and 2025 REAIM regional consultations, see Afina, *The Global Kaleidoscope of Military AI Governance*; Afina, *The Global Prism of Military AI Governance*.

CHAPTER I.

Procedural guidelines

The first guidelines for the development, adoption, implementation and review of national strategies on AI in security and defence are procedural. The procedural guidelines outline the institutional course of action across all relevant actors and stakeholders for the entire cycle of the national strategy. This part covers the development and adoption process (including conceptualization, the definition of objectives, and the assessment of opportunities, risks and the international landscape); implementation and review (including the identification of relevant actors and the establishment of verification and review mechanisms); cooperation with external stakeholders; and, finally, convergence with select fields that offer a cross-cutting perspective on the procedural steps. The guidelines do not seek to be prescriptive but, rather, provide States and key stakeholders with an overview of the main considerations that, albeit non-exhaustive, would benefit from extensive reflections. They also include, as appropriate and necessary, solutions to support the development, adoption, implementation and review of national strategies.

P.1. Development and adoption of a national strategy on AI in security and defence

In the development of a national strategy on AI in security and defence, a number of steps and key considerations must be taken into account. These encompass both internal and external factors that may have an impact on the national strategy under development.

P.1.1. Conceptualization

The development of a national strategy on AI in security and defence requires a clear vision and strategic direction. Hence, a State developing such a strategy should engage in a conceptualization process that helps to define and identify its goals in the AI space, and it should subsequently align those with their broader national security and defence posture. The State should consult extensively with internal and external stakeholders to ensure a comprehensive overview and understanding of the possible

opportunities, perceived risks and broader implications associated with AI in security and defence, and it should subsequently formulate a clear vision and strategic direction.

The State should consider the establishment of working groups comprising military officers, experts, AI developers, lawyers, ethicists, technologists, procurement officers and governmental research bodies, as well as policy-makers and planners from within government and outside. There should be regular coordination across the working groups – for the

conceptualization stage and beyond. The conduct of regular consultations and working sessions would help define the State's strategic direction and goals.

P.1.2. Definition of objectives

A State developing a national strategy on AI in security and defence should define clear objectives that underpin and shape that strategy. These objectives should provide for the State's ambitions and goals, that is, what it ultimately seeks to achieve through the integration of AI in security and defence. To this end, a comprehensive survey of opportunities that may be offered by these technologies and specific applications would be helpful.

The State should define specific and measurable objectives and should subsequently create an action plan for each objective. Each plan should include a clear description of its scope and underlying rationale, clear timelines, responsible actors, milestones and key performance indicators to track, review and evaluate progress. These plans should also account for the eventual challenges and limitations that may arise, and the State should subsequently develop an accompanying risk-mitigation and risk-management framework. This latter assessment should be comprehensive and should factor in, among other things, political, legal, technological, operational, financial and geopolitical considerations.

P.1.3. Definition of desired outcomes

A State developing a national strategy on AI in security and defence should also define the desired outcomes of the strategy. These should reflect short-, medium- and long-term goals and aspirations not only as to what the State seeks to achieve with respect to AI in security and defence, but also in the broader national, regional and international security landscape.

To this end, the State should define specific and measurable desired outcomes and interim milestones that take account of the national, regional and international security policy landscapes and their interplay. The State should consider the development of frameworks for monitoring and for evaluation and verification to measure, track, review and evaluate progress towards the desired outcome.

P.1.4. Definition of scope

Defining the scope of a national strategy on AI in security and defence is critical to ensuring that the appropriate kinds and levels of resources are directed towards priority areas and are allocated to the appropriate agencies, while the strategy remains grounded within a very specific framework. The scope should clearly outline areas where AI is to be applied in security and defence, thus ensuring alignment with existing national, regional and international security goals. The scope should cover both immediate operational needs and long-term strategic objectives, taking into consideration domestic security concerns, international obligations and technological trends.

A State developing a national strategy should thus consider developing a detailed scoping document that outlines the various applications of AI in security and defence and the limitations and criteria for expanding or narrowing its use over time. This scoping document should be subject to review on a regular basis, ultimately ensuring that it remains relevant as technology and security environments, priorities and needs evolve.

P.1.5. Distribution of roles and responsibilities

One of the most critical aspects of the implementation of a national strategy on AI in security and defence is the clear, coherent, efficient and effective distribution of roles and responsibilities among relevant stakeholders. Such a distribution would notably address concerns about potential overlaps in roles and responsibilities between military branches, intelligence and law enforcement agencies, and ministries, particularly in the light of the blurring lines dividing security and defence applications.

A State developing a national strategy should consider the development of a framework that codifies such a distribution of roles and responsibilities. This framework would include specific mandates, powers and limitations for each public body and under which circumstances they apply.

P.1.6. Assessment of financial resources and capacities

A State setting out to develop a national strategy on AI in security and defence should conduct a comprehensive assessment of the financial resources needed for all stages of its development, adoption, implementation and review. This assessment should take stock of both resources currently available and those anticipated to become available over the strategy's expected life cycle. It should simultaneously identify funding gaps, limits and shortcomings. Based on this assessment, the State should inform, and where necessary revise, the objectives, scope, and distribution of roles and responsibilities defined earlier in the process.

Building on this assessment, the State should develop a budget and financing plan covering the full procedural cycle, from development through to implementation and review. At this stage, the State should also address the reliability, sustainability and potential risks of existing budgeting sources and, if necessary, should explore changes in budget allocations and alternative funding sources, both internal and external. Typical risks can include dependence on a single budget line, short political or budget cycles and volatility from competing priorities.

This process would keep the State's ambitions grounded in what is realistically achievable by shaping the methodological frameworks that organize strategy priorities. It would thereby enable efficient budget allocation and use of resources and the sourcing of prospective resources to support the national strategy.

P.1.7. Internal consultations

Internal consultations are critical to ensure that all relevant government entities and agencies are aligned on a national strategy on AI in security and defence. These consultations should include, among others, relevant ministries (e.g., ministries of defence, of foreign affairs, of interior, and of information and communications technology, ICT), intelligence agencies, law enforcement agencies and other relevant bodies to ensure a united, unified and coherent front and approach.

A State developing a national strategy on AI in security and defence should thus organize and facilitate regular inter-agency coordination and information exchange meetings to discuss the development, adoption, implementation and

review of the national strategy. This will ensure that the national strategy is holistic and encompasses varying viewpoints from across the governmental ecosystem.

Ensuring effective coordination between agencies will, in this sense, be key. The State should therefore identify and either designate or establish (subject to resource availability) a specific agency or body responsible for inter-agency coordination and for facilitating communication.

P.1.8. External consultations

Engaging with external stakeholders—including academic and research institutions, civil society organizations, the private sector and, as appropriate and necessary, allied or partner States—is critical to building a robust, inclusive and effective national strategy on AI in security and defence. A State developing such a strategy should take stock of the external expertise available, establish a network of experts, and leverage the latter to address the technical, legal, ethical and operational risks, challenges and concerns raised in this space.

The State should consider establishing a formal consultation process with external stakeholders; alongside a high-level track, the consultation should collect input, feedback and recommendations on specific issues and topics. The State should then consider the extent to which these external stakeholders should be involved (either formally, informally or both) in the subsequent implementation and review of the national strategy.

P.1.9. Assessment of relevant geopolitical dynamics at the international and regional levels

A State developing a national strategy on AI in security and defence should make an integrated assessment of the relevant geopolitical dynamics, issues and concerns at the international and regional levels that may have a direct or indirect impact on the strategy. A thorough understanding of the international and regional geopolitical environment would indeed be critical for shaping effective strategies on AI in security and defence.

To this end, the State should take stock of its existing capabilities and institutions already engaged in international and geopolitical monitoring and strategy development (e.g., a national security council, an existing foreign ministry policy planning unit, the defence ministry's strategic assessment staff, or any other equivalent body). On this basis, it should consider the extent to which such capabilities already include, or could be expanded to include, AI-specific expertise and produce assessments on geopolitical dynamics relevant to the issues addressed here.

In addition to existing capabilities, or where these cannot be expanded to meet this need, the State could consider the establishment of a dedicated working group tasked with the regular monitoring of geopolitical dynamics at the international and regional levels. The extent to which the working group should or should not include external stakeholders, in addition to the specific agencies involved and disciplines represented, is at the discretion of the State in question, although the inclusion of a diversity of perspectives would consolidate the assessments made by the working group.

The State should evaluate the ways in which these dynamics are relevant to and may influence, directly or indirectly, the national strategy under development. This includes, for example, the impact that geopolitical tensions could have on digital and ICT infrastructure, energy security, supply chains and subsequent repercussions for strategic technological dependencies. On a macro level, strategic competition and the perceived “AI arms race” to secure military advantage could also have an impact on the State’s posture and overall approach to its national strategy (e.g., accelerated procurement processes). The State should thus consider carefully the implications of international and regional geopolitical dynamics for its national strategy, and the possible repercussions on legal compliance and ethics.

P.1.10. Assessment of relevant political dynamics and context at the national level

A State developing a national strategy on AI in security and defence should make an integrated assessment of the relevant political dynamics at the national level, including public opinion, political stability and legislative environments. The State should assess the domestic political landscape to ensure that the strategy under development will secure sustained internal buy-in throughout its life cycle – from its development to its subsequent adoption, implementation and eventual review.

A thorough understanding of the national political landscape would also influence the likelihood of success in implementing the national strategy and would help to anticipate areas where flexibility may be required in order to adapt to changing conditions. To this end, the State should organize formal and informal briefings and consultations with legislative and

other relevant bodies to take stock of national priorities, gauge political support and address concerns within the national strategy.

P.1.11. Assessment of the wider AI policy and regulatory landscape

A comprehensive review of existing AI policies and regulations at the national, regional and international levels is necessary to ensure that a national strategy on AI in security and defence is cohesive and adequately aligned with the regulatory frameworks in place. A State developing a national strategy should thus conduct a comprehensive review and take stock of AI policies and regulations to which it is party at the international, regional and national levels or that are otherwise relevant. The State should conduct such reviews periodically in order to ensure that the national strategy takes into account the rapidly evolving landscape of governance of AI, including in security and defence, and is adapted accordingly and as necessary.

In conducting such reviews, beyond policies and regulations that specifically relate to AI, the State should also examine applicable regulatory frameworks that govern import and export controls, licensing regimes, procurement processes, and other relevant international policy instruments in place. It should assess specifically how each could directly or indirectly relate to AI and its core components, from the software to the physical infrastructure underpinning their use, the capabilities they are being integrated into and specific components of these. On this basis, the State should assess how these frameworks affect applicable policies and wider strategies around sovereignty, strategic technological dependencies and other security policy considerations relevant to a national strategy on AI in security and defence.

P.1.12. Identification and assessment of relevant actors

A national strategy on AI in security and defence should take into account the broader landscape of actors with direct and indirect stakes in the technology's development, deployment, use and governance – and any other relevant actors – at the international, regional and national levels. A State developing such a strategy should systematically map these actors, including those in the public, private and non-profit sectors, as well as informal, unorganized or non-institutional actors. Examples of relevant actors include academic and research institutions, foreign armed forces and governmental bodies, international and regional organizations, individual companies, industry associations and standards bodies, and civil society.

The identification of relevant actors should be followed by an assessment of each of them; their roles, stakes and relevance for the procedural and substantive aspects of the national strategy. These stakes can be direct or indirect across the technology's life cycle and supply chain. They can include direct procurement, supply chains, potential subcontracting, testing, deployment and integration, and de-commissioning. In addition, this process should distinguish, where relevant, between actors whose interests are broadly aligned with the State's own and those whose interests may be competing or adversarial.

Such a mapping and assessment would ensure that the national strategy maintains an overview of the rapidly evolving landscape of relevant actors for AI in security and defence. They can also potentially serve as a reference point for potential partnerships, responsible industry relations and contracting, knowledge- and research-sharing, civil society consultation,

and organized engagement with the public. They would also support the identification of possible dependencies, vulnerabilities and proliferation risks, and inform due diligence, oversight and risk-mitigation measures.

P.1.13. Assessment of international law

The development, deployment and use of AI in security and defence must be conducted in adherence and compliance with international law, including international humanitarian law and international human rights law (see guideline S.8.) A State developing a national strategy on AI in security and defence should thus conduct a comprehensive and thorough assessment of applicable legal frameworks and regimes at the international and regional levels. Such assessments should be conducted in consultation with legal experts from various fields, including academia, the armed forces, governmental bodies, and international and regional organizations. This will ultimately ensure that the national strategy adheres to the State's international and regional obligations.

P.1.14. Assessment of ethics

Ethical considerations should lie at the forefront of the development, deployment and use of AI in security and defence (see guideline S.9.) As such, a State developing a national strategy on AI in security and defence should conduct a comprehensive and thorough assessment of existing ethical guidelines and a survey on the ethical principles of importance. Such an assessment should be conducted in consultation with ethicists from various fields, including academia, governmental bodies, and international and regional organizations. This will ultimately ensure that the national strategy is aligned with ethical frameworks and guidelines, and with technologies across their life cycle.

P.1.15. Assessment of technical standards

A State developing a national strategy on AI in security and defence should take stock of and evaluate technical standards in place to ensure that AI systems in security and defence are robust, reliable and interoperable across domains and varying security contexts. These include technical standards adopted by specialized organizations and other bodies (e.g., the International Organization for Standardization, ISO, and the Institute of Electrical and Electronics Engineers Standards Association, IEEE SA). In addition, the State should consider the development of national technical standards specifically for AI systems in security and defence – or the revision of existing standards – in alignment with good practices in this space.

P.1.16. Assessment and management of AI risks

A State developing a national strategy on AI in security and defence should assess the risks associated with the development, deployment and use of AI in security and defence. These studies should then be complemented by the development and adoption of clear frameworks for the assessment, classification, mitigation and reduction of risk. Established processes should be mandated across governmental agencies and, as appropriate and necessary, non-governmental partner stakeholders throughout the technology's life cycle.

The State should also consider the conduct and implementation of regular risk audits. This would ensure the continued scrutiny of AI technologies in security and defence against existing, new and emerging risks.

P.1.17. Assessment of non-AI security risks

In addition to AI-specific risks, broader security issues, concerns and threats that may be of relevance to AI in security and defence should be considered. These non-AI security risks include cyberthreats, risks to energy security and to critical digital, cloud and ICT infrastructure, and risks posed by the activities of private military contractors and non-State armed groups, as well as geopolitical instability. A State developing a national strategy on AI in security and defence should thus conduct a comprehensive assessment to identify, map and consider these risks, and ensure that the strategy factors in the broader security landscape and, subsequently, non-AI security risks that may be of relevance.

The State should thus coordinate with its specialized security bodies and agencies (e.g., national cybersecurity agency, national nuclear authority, intelligence agencies) to take stock of these risks, the extent to which they are of relevance to AI in security and defence, and the role of the national strategy in classifying, managing, mitigating and reducing such risks.

P.1.18. Collective drafting of a national strategy document

Building on the studies, consultations processes and assessments undertaken in the first phase of the strategy's development, a State developing a national strategy on AI in security and defence should consider the establishment of a cross-agency working group mandated to draft the national strategy document. The working group should include a designated lead coordinator and penholder to preside over, oversee, and exercise accountability and responsibility over the drafting process. The process should be framed by a

specific mandate and clear rules of procedure to ensure that the national strategy incorporates diverse perspectives and expertise from security and defence and from across disciplines.

P.1.19. Classification: To publish or not to publish?

A State developing a national strategy on AI in security and defence should reflect on and decide whether or not to publish its strategy. Far from being a binary decision, States may, in fact, consider a spectrum of options: release the full strategy; release parts of the strategy with classified or redacted sections kept out of public consumption; release a summary cleared for public release; or not release it at all.

Considerations that may affect such decision include sensitivities and the classification of certain information, the need for transparency, and risks and opportunities that stem from the publication or, conversely, the classification of such a strategy.

P.1.20. Review of the draft national strategy document

Prior to its adoption, a State developing a national strategy on AI in security and defence should establish a review process to ensure that key stakeholders, both in government or non-governmental, as appropriate, are able to provide feedback on and input into the text. An institutionalized multistage review process would ensure the strategy's inclusivity and effectiveness.

To this end, the State should organize a series of review consultations, validation workshops and consolidation processes to collect and incorporate feedback from both public and private stakeholders.

P.1.21. Validation process for the draft national strategy document

Following the review process, and prior to formal adoption, a State developing a national strategy on AI in security and defence should conduct formal verification of the draft document's internal and external coherence, technical accuracy, legal and regulatory compliance, and alignment with its defined objectives, scope and desired outcomes. This validation should holistically assess the national strategy's conceptual and procedural aspects, including the State's internal capabilities and costs. It should also assess the extent to which the outcomes of consultations and the review process have been incorporated.

The validation process should encompass a broader legal review, assessing the strategy's soundness and feasibility, and its consistency with existing law, other overlapping national strategies on cyber, digital or defence issues, budget frameworks, domestic oversight bodies, applicable standards and international commitments. It should confirm that the national strategy's goals and measures do not contradict one another.

As one of the final stages prior to formal adoption, the State should produce a validation and verification document confirming that this review process has taken place, outlining its modalities and findings, and attesting to the coherence and legitimacy of the strategy. The State should subsequently decide whether this validation document should be released to the public or remain classified.

P.1.22. Adoption of the national strategy document

Once a State has finalized development of a national strategy on AI in security and defence, it should proceed with formal adoption and should communicate the strategy to all relevant stakeholders. The precise adoption process will depend on the State's specific regulatory, legislative and policymaking processes.

In parallel, the State should work on a communication plan and strategy to disseminate the national strategy as appropriate to the relevant national stakeholders and allies and partners as necessary and appropriate. This will ensure that relevant stakeholders remain informed and prepared to implement their respective duties and responsibilities following the adoption of the strategy.

P.1.23. Public communication of the national strategy document

Once a State has adopted a national strategy on AI in security and defence, it should place particular emphasis on clear and transparent communication with the public, cognizant of the classification requirements defined earlier in the drafting process (see guideline P.1.19). Where the strategy, or a public-facing version of it, is published, the State should consider the extent to which these measures would enable the general public to be informed, as necessary and appropriate, of the State's intentions, plans and approaches surrounding the use of AI in security and defence.

In order to sustain this transparency beyond the point of initial adoption, the State should also consider the conduct of periodic reporting on the strategy's implementation. This could include, for instance, an account of how any preceding strategy was implemented over its life cycle. The State should subsequently consider whether or not these reports should be released for public consumption.

P.2. Implementation and review of a national strategy on AI in security and defence

Following the development and adoption of a national strategy on AI in security and defence, a number of considerations are specifically relevant to its implementation. These include the actors involved, monitoring and evaluation processes, the need for oversight and verification mechanisms, and auditing.

P.2.1. Key actors for implementation

The effective implementation of a national strategy on AI in security and defence requires the involvement and intervention of various key actors from across government, the military, industry (including industry associations, military contractors and technology companies), intelligence and security contractors, and academic and research institutions, including national laboratories. Each actor plays a unique role in ensuring that the national strategy is successfully implemented and operationalized.

To this end, a State implementing a national strategy should define and codify a clear distribution of the roles and responsibilities of key actors, from both the public and the private sectors. A designated lead agency should be identified and mandated to coordinate implementation efforts and facilitate inter-agency communication.

P.2.2. Key coordinating bodies and actors

A central coordinating body is critical to enabling and overseeing the implementation of a national strategy on AI in security and defence. This coordination body could either be concentrated in a single agency (e.g., embedded within a specific ministry) or be composed of representatives of diverse agencies. It should have a clear, explicit and unambiguous mandate, with the authority to manage inter-agency initiatives, efforts and collaborations and to monitor and evaluate the implementation process. It should also have the ability to address potential challenges in a swift and timely manner.

The coordinating body should have the resources required to regularly organize and operate inter-agency meetings, manage timelines, and monitor implementation milestones. A State implementing a national strategy should also reflect on whom this coordinating body would be accountable and report to and should establish the appropriate mechanisms.

P.2.3. Allocation of roles to new and existing bodies

A State developing and implementing a national strategy on AI in security and defence should consider the need for institutional adaptation of existing bodies or – as necessary and as resources permit – the establishment of dedicated bodies, working groups, consortia and forums. If there is an identified need to create new bodies for the implementation of the strategy, this should be articulated in a clear and measurable manner. The State should make decisions on such establishments or adjustments proportional to the objectives, scope, resources and capacities that it has already assessed, and these should be revisited and re-calibrated as needs be at each subsequent phase of the strategy's development, adoption, implementation and review. Dedicated bodies that a State should consider to establish, and the key functions that the State should consider allocating to them, include a central coordination body to oversee the implementation of the national strategy; a working group to monitor geopolitical dynamics; a formal oversight body to monitor strategy implementation; an independent oversight body to review and oversee technology transfers; oversight and accountability bodies for security applications; an inter-agency emergency response team that follows a newly established emergency response plan; and a licensing body for potential sales of sensitive technologies, if permitted, to non-governmental entities. The State should also consider the extent to which it can distribute some, if not all, of the critical responsibilities for implementation across its existing bodies.

P.2.4. Clarity of structure and distribution of roles and responsibilities

A clear structure for the distribution of roles and responsibilities is critical for the implementation of a national strategy on AI in security and defence. The structure should be comprehensive, yet it should maintain a degree of flexibility to adapt to evolving technological and operational needs for the implementation of the strategy.

A State implementing a national strategy should thus develop a clear governance structure that clearly delineates roles and responsibilities at the international, regional, national and agency levels. Each entity should be aware of its role and responsibilities, reporting structures, expectations, and ways through which it could contribute to the overall strategy.

P.2.5. Key actions for implementation

A State implementing a national strategy on AI in security and defence needs to develop a road map of key actions to guide and frame the effective implementation of the strategy. The State should thus create a detailed action plan that outlines key steps for implementation. This action plan should include timelines, resource allocation (including budgeting and funding) and performance metrics. Each key action should then be broken down into projects with clear deliverables, assigned teams and a tracking system to monitor progress.

P.2.6. Key timelines and milestones for implementation

Establishing realistic timelines and milestones ensures the effective implementation of a national strategy on AI in security and defence. These milestones should allow for periodic assessments and reviews as necessary.

A State implementing a national strategy should thus develop a timeline for each phase of the strategy, setting specific milestones and frameworks for the tracking, measurement and reporting mechanisms. The State should also consider the need for necessary flexibility, and it should adjust the strategy based on technological progress and the ever-evolving policy and regulatory landscape surrounding the governance of AI in security and defence.

P.2.7. Accountability for implementation

Accountability mechanisms are necessary to ensure that stakeholders meet and implement their respective roles, duties and responsibilities. These mechanisms should include clear accountability mechanisms that hold actors responsible for meeting their implementation targets. Possible measures that could be established by a State implementing a national strategy on AI in security and defence include performance reviews, reporting obligations and providing a coordinating body with an oversight mandate.

P.2.8. Adoption of an implementation plan

Upon the development of a strategic implementation plan, a State implementing a national strategy on AI in security and defence should work towards formal adoption of the strategy not only at the national level, but also at the agency level and by all stakeholders. By requiring a formal and multilayered adoption process, such

an approach would foster accountability, ownership and commitment towards the implementation of the national strategy.

P.2.9. Establishment of mechanisms for oversight, monitoring and verification

A State implementing a national strategy on AI in security and defence should consider establishing mechanisms to oversee, monitor and verify its implementation. These mechanisms should come with clearly assigned roles and responsibilities for the relevant bodies, agencies and stakeholders. Oversight corresponds to the authority to steer. The State should consider a formal oversight body with the mandate, powers and resources necessary to take decisions on the course of the strategy. This includes, for example, obtaining information from the implementing bodies, requiring corrective action as necessary, and convening regular inter-agency meetings to review progress, identify issues and challenges, and adjust accordingly. Monitoring corresponds to continuous documentation and tracking throughout the implementation cycle. To this end, the State should consider mechanisms, including of budget execution. The State should consider establishing mechanisms to track expenditure against pre-approved allocations and to identify and explain any variances and deviations. Verification corresponds to the periodic assessment of progress against the milestones and indicators set in the strategy, for example through audits, third-party reviews and internal assessments as appropriate, considering the unique considerations that emerge from AI applications in security and defence contexts. These assessments should, for example, be in compliance with the applicable classification frameworks and requirements. Monitoring and verification reports would feed into the designated oversight body. Findings

that point to shortcomings in the strategy itself and its implementation should feed into its review and revision (see guideline P.2.10.). In practice, the State should consider the extent to which these functions could overlap, which may lead to either their combination or their assignment to pre-existing bodies.

P.2.10. Establishment of mechanisms for review and revision

A State implementing a national strategy on AI in security and defence should consider establishing mechanisms for the review and revision of the strategy. Where oversight, monitoring and verification (see guideline P.2.9) ask whether the strategy is being implemented as intended, periodic review evaluates whether its objectives, desired outcomes and scope remain relevant, effective and efficient in light of technological and security developments. Review should draw on the findings of monitoring and verification and conclude with a recommendation on whether the strategy should be revised.

Based on review, the State should establish a dedicated mechanism for substantive and procedural updates to the strategy, whether periodically or in response to specific and measurable triggers, such as lessons learned, feedback, external events or changes in priorities and the broader context. The State should clearly define these triggers and their thresholds, the body responsible for initiating an update (by default the oversight body) and the authority that adopts it. Where the strategy is adopted for a defined time frame, the State should also establish in advance the conditions under which an interim update may be issued before the strategy expires. The mechanism should ensure continuity of oversight and no gap in the strategy's validity between successive versions, with milestones and indicators remaining comparable across them.

P.3. Cooperation

Internal and external cooperation will play a critical role in the development, adoption, implementation and review of a national strategy on AI in security and defence. States should thus consider how they intend to frame such efforts in the light of the sensitive nature of the topic at hand.

P.3.1. National cooperation

National cooperation between government agencies and branches of the armed forces is crucial for ensuring the successful and effective implementation of a national strategy on AI in security and defence. To this end, a State implementing such a strategy should introduce a national cooperation framework that codifies

the respective mandates of each agency and possible working relationships for the implementation of the strategy. The State should also consider the establishment of inter-agency working groups with a thematic focus, enabling collaboration, coordination and alignment on specific security and defence issues.

P.3.2. Bilateral cooperation

Bilateral cooperation with other States is crucial for the sharing of good practices, for the alignment of policies and regulations, and for enhancing security and defence relationships (e.g., through joint investment programmes and initiatives). A State implementing a national strategy on AI in security and defence should specify the partner States with which it intends to cooperate, as well as the specific purposes for which that cooperation is to be pursued.

The development of bilateral agreements on AI cooperation with key allies and partners may be particularly helpful due to the limited scope of such agreements and considering the sensitivities surrounding the development, deployment and use of AI in security and defence. A State implementing a national strategy should consider the degree and depth of such bilateral cooperation, the relevant actors and the resources that would need to be allocated, for what specific purposes and within what time frames.

For the identification, establishment, monitoring and review processes of bilateral cooperation agreements, the State should put particular focus on trade agreements, import and export control arrangements, licensing regimes, and other instruments that may facilitate technological transfer as these mechanisms and policies are likely to carry both opportunities for trust-building and cooperation, and security risks, dependencies and potential for proliferation.

P.3.3. Regional cooperation

Regional cooperation, where desirable and feasible, would allow a State to align its national strategy on AI in security and defence with those of other States in the region and with

regional security frameworks. Such cooperation and partnerships could pave the way for collaboration to address common challenges, reinforce regional security and stability, and ensure technical interoperability between systems. The State should thus consider the role it seeks to play in the regional context and should promote regional cooperation frameworks through the organization of convenings, summits and information-exchange platforms specific to AI in security and defence, both high-level and at the working level.

Joint regional defence projects on development and integration of AI in security and defence should also be considered, notably to address challenges and issues that affect the regional security landscape (e.g., joint initiatives to combat transnational organized crime and for counter-piracy efforts). Beyond convening, the State should also provide for the frameworks, processes, architectures and mechanisms necessary for such regional cooperation to be meaningful, sustainable and adequately resourced in the long term.

P.3.4. International cooperation

International cooperation on governance of AI in security and defence is crucial to addressing international security challenges and ensuring the responsible development, deployment and use of these technologies. A State implemen-

ting a national strategy on AI in security and defence should take stock of on-going governance processes, discussions, multilateral forums and initiatives at the international level, both within the United Nations and beyond.

In particular, it should track the discussions, recommendations and outcomes of the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts (GGE) on lethal autonomous weapon systems (LAWS), especially if the State is a party to the CCW Convention. It should also track the principles set out by the United Nations General Assembly.¹³ Beyond the United Nations, the State may also consider State-led international initiatives aimed at fostering global cooperation on AI in security and defence, such as the REAIM initiative.

Building on this assessment, the State should consider its role and ambitions within the international community. It should engage accordingly through active participation in international forums, by supporting international institutions and by leading on State-led initiatives.¹⁴ Transparency forms a key part in underpinning international cooperation; information-sharing will, in this sense, constitute a key CBM to signal to other actors the State's commitment to the safe and responsible development of AI technologies.¹⁵

13 See, in particular, General Assembly, resolution 79/239, 24 December 2024, <https://docs.un.org/A/RES/79/239> (2024); and resolution 80/58, 1 December 2025, <https://docs.un.org/A/RES/80/58>.

14 The REAIM Summit and adjacent processes were initially led by the Netherlands and the Republic of Korea and more recently also championed by Singapore, Kenya and the United Kingdom. In addition, the United States-led Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy has been endorsed by a number of States and a series of related cross-regional thematic working groups has been established.

15 Ioana Puscas, *Confidence-Building Measures for Artificial Intelligence: A Multilateral Perspective* (Geneva: UNIDIR, 2024), https://unidir.org/wp-content/uploads/2024/07/UNIDIR-Confidence_Building_Measures_Artificial_Intelligence-Multilateral_Perspective.pdf.

P.3.5. Cooperation with technology providers

Cooperation with technology providers, particularly the private sector, is central to fostering responsible innovation in AI, including in the security and defence sector. A State implementing a national strategy on AI in security and defence should thus explore and identify private companies involved across the supply chain of AI capabilities for security and defence in order to cooperate with them, while ensuring effective, meaningful and sustainable collaboration throughout the life cycle of these technologies (see also guidelines S.3.1–8).

P.3.6. Cooperation with civil society

External cooperation with civil society organizations (including human rights groups, academic and research institutions, and public interest advocacies) is an integral aspect of meaningful multi-stakeholder engagement in the develop-

ment, adoption, implementation and review of a national strategy on AI in security and defence. Structured and sustained engagement with civil society organizations with direct and indirect stakes ensures transparency, accountability, public legitimacy, and the inclusion of expert, local and lived-experience perspectives that will make the national strategy more robust.

A State developing a national strategy should thus provide organizations with access to relevant information, cognizant of previously defined classification requirements. The State should define the approach, structure and formats of cooperation with civil society during the entire cycle of development, adoption, implementation and review. It should also consider establishing dedicated consultative mechanisms that ensure meaningful participation, such as advisory panels, public communications and feedback channels.

P.4. Lessons from and convergence with select fields

In the development, adoption, implementation and review of a national strategy on AI in security and defence, States ought to consider lessons and good practices from other security and technological fields. However, these comparative analyses must be done with a thorough understanding of the unique characteristics of AI and its inherent differences from other security and technological fields. Areas of technological and governance convergence between security fields should also be considered and addressed as appropriate.

P.4.1. The cybersecurity field

A State implementing a national strategy on AI in security and defence could consider the extent to which successful governance models in the cybersecurity field can be applied to AI in security and defence. These include export control models for intangible systems, the establishment of emergency response teams, effective capacity-building programmes and CBMs.

In addition, the State should consider areas of convergence between the AI and cyber fields and should address possible risks and issues that may arise in its national strategy. Opportunities that may arise from this convergence should also be considered, particularly to bolster cyber defence capabilities in defence ecosystems and for the security of critical national infrastructure.

P.4.2. The nuclear field

A State implementing a national strategy on AI in security and defence could consider the extent to which lessons can be drawn from nuclear governance, cognizant of the inherent technical differences between AI and nuclear technologies. These lessons may include those from safety frameworks, nuclear safeguards and non-proliferation frameworks, as well as approaches to delineating peaceful from non-peaceful uses. In addition, lessons may be drawn from nuclear safety and security for the AI domain specifically, including the security and traceability of supply chains, best practices for decommissioning of nuclear technologies, and the protection of sensitive nuclear information through graded classification systems.

In addition, the State should consider areas of convergence between the AI and nuclear security fields and should address possible risks and issues that may arise in its national strategy. Opportunities that may arise from this convergence should also be considered.

P.4.3. The biological and chemical security field

A State implementing a national strategy on AI in security and defence could consider the extent to which lessons can be drawn from the biological and chemical security field. The general-purpose character of both AI and biological and chemical technologies, along with the risks stemming from their inherently dual-use natures, provide much room for cross-pollination and mutual learning for key issues. Relevant lessons may include approaches to controlled access, non-proliferation, oversight of high-risk research, practitioner responsibility and risk-based governance mechanisms for potentially harmful applications.

In addition, the State should consider areas of convergence between the AI and the biological and chemical security fields, including the increasing use of AI in biological and chemical research and development (R&D) and should address possible risks and issues that may arise in its national strategy. Opportunities that may arise from this convergence should also be considered.

P.4.4. The aviation sector

A State implementing a national strategy on AI in security and defence could consider the extent to which lessons can be drawn from the aviation field. These lessons may draw on established approaches to certification, monitoring, standardization, incident reporting and management, and technical safety in aviation. The State should also consider the dual-use nature and co-development between aviation, aerospace and defence technologies and their governance.

P.4.5. AI safety

A State implementing a national strategy on AI in security and defence could consider the extent to which lessons can be drawn from the field of the technical safety of general-purpose and frontier AI models. As general-purpose models are increasingly integrated into security and defence contexts, and often share underlying technical characteristics with systems developed specifically for that domain, advances in technical AI safety research carry direct implications for the national strategy. Lessons can be drawn from approaches to safety engineering, testing and evaluation, capability evaluations, monitoring, guardrails, and alignment, control and interpretability research.

The State should consider areas of convergence between the AI safety field and AI in security and defence and should address possible risks and issues that may arise in its national strategy. Opportunities that may arise from this convergence (e.g., existing technical norms, standards, and testing, evaluation, verification and validation frameworks) should also be considered.

P.4.6. Civilian AI governance

As governance frameworks surrounding civilian AI applications emerge and increasingly proliferate, a State implementing a national strategy on AI in security and defence should think of opportunities for cross-pollination of governance approaches between the civilian domain and that of security and defence. Among areas of overlap, data governance and dual-use technologies are particularly important to explore and address in the national strategy, considering both opportunities and risks in addition to offering a possible framework to delineate between civilian and non-civilian applications.

CHAPTER II.

Substantive guidelines

This second series of guidelines for the development, adoption, implementation and review of a national strategy on AI in security and defence are substantive. Echoing the approach to the first, procedural guidelines, these substantive guidelines do not seek to be prescriptive. Rather, they provide States and key stakeholders with an overview of technical, legal and political considerations, ranging from data and life cycle management to compliance with international law, ethics, public–private partnerships, and resilience and preparedness. While these considerations are non-exhaustive, extensive reflection on them would provide relevant approaches and, as appropriate and necessary, solutions to support the development, adoption, implementation and review of a national strategy on AI in security and defence.



S.1. Data practices

Data constitutes the lifeblood of AI technologies. In the light of the sensitivities of security and defence applications and the complex regulatory implications that arise from them, States should reflect on governance approaches to data practices in their national strategies on AI in security and defence.¹⁶

S.1.1. Data provenance

Given the sensitive nature of security and defence applications, a State developing a national strategy on AI in security and defence should consider where the data to be used for training, testing and the subsequent use of AI technologies it is expected to originate from, what laws and regulations may apply already, and what wider considerations apply. This is particularly relevant for States that procure AI systems from private vendors – domestically or

internationally – or allied governments, where the provenance of training and testing data (and its history) may either be unknown or subject to the regulatory regime of another State.

The State should thus consider integrating specific guidelines, if not requirements, in its national strategy for the documentation, verification and eventual certification of the provenance of data used to train, test and use AI systems in security and defence. Provenance

¹⁶ Yasmin Afina and Sarah Grand-Clément, *Bytes and Battles: Inclusion of Data Governance in Responsible Military AI Discussions*, Centre for International Governance Innovation (CIGI) Paper no. 308 (Waterloo, ON: CIGI, 2024), <https://www.cigionline.org/publications/bytes-and-battles-inclusion-of-data-governance-in-responsible-military-ai>.

documentation should, in particular and to the extent possible, include the origin and collection modalities of training and testing data; the nature of the data (e.g., if it is synthetic); the legal frameworks and regulations under which the data was collected and processed; and any limitations, assumptions and known biases in the data that may affect system performance or legal compliance. The State should consider the extent to which provenance requirements can be incorporated into procurement processes, including through contractual obligations on suppliers. However, it should acknowledge that full traceability of training data provenance may not always be feasible, particularly for complex AI systems trained on large and heterogeneous data sets.

S.1.2. Data-collection practices

A State developing a national strategy on AI in security and defence should ensure that the practices for collecting data for the development, training and testing of AI are lawful. All data should be collected, processed and stored in compliance with the applicable data-protection laws, cognizant of the relevant nuances of security and defence applications. The State should clarify how existing data-protection regimes apply in such contexts and, if necessary, develop its interpretation of the law for specific use cases in security and defence.

In parallel, the State should establish oversight and accountability mechanisms to ensure that data collection is done within the confines of the law (see e.g. guideline S.8.2 on IHRL). It should thus develop data-governance frameworks with clear protocols for data collection in specific security and defence use cases, with compliance as a guiding principle framing its efforts.

S.1.3. Data storage and infrastructure

Storage of the data that is to be used for the training, testing and use of AI technologies raises distinct governance considerations. A State developing a national strategy on AI in security and defence should consider where and how such data will be stored. It should assess the infrastructure required to store that data not only in compliance with applicable laws and regulations, but also in a manner that is secure and reliable across different classification levels. These considerations – from the physical security, resilience and redundancy of data-storage facilities, to the extent to which such facilities are secured against disruption, whether from natural hazards, deliberate attacks or infrastructure failures – will have profound implications for the State.

The State should also consider the comparative merits of, on the one hand, domestically owned, managed and operated storage infrastructure and, on the other, reliance on commercial cloud services, including foreign service providers. These considerations must be coupled with the dimension of compliance, notably considering the jurisdiction in which the infrastructure may fall, as well as the applicable legal regimes governing access to stored data and the physical infrastructure.

Furthermore, operational and strategic risks that may arise from dependency on external infrastructure for data that is critical to security and defence functions must be considered. For commercial or foreign-hosted infrastructure, the State should establish clear contractual and technical safeguards. These should include data-localization requirements where appropriate, encryption and access control standards, and contingency plans in the event that access to such infrastructure is interrupted, disrupted, suspended, withdrawn or destroyed.

S.1.4. Data hygiene

Data hygiene practices ensure the reliability of AI systems with regards to operational effectiveness, legal compliance and ethics. Identifying and implementing good practices to ensure the hygiene of training and testing data sets is particularly important to ensure the responsible development, deployment and use of AI in security and defence. A State developing a national strategy on AI in security and defence should thus develop and implement data hygiene guidelines as part of that strategy. These guidelines should also include requirements for cleaning, validation and updating data sets on a regular basis, ensuring their veracity and, thus, the accuracy of AI systems.

S.1.5. Appropriateness of training data

The quality of training data is critical for the performance, accuracy and reliability of an AI system. A State implementing a national strategy on AI in security and defence should thus ensure that the AI systems that it is developing, acquiring, testing, adopting or using for security and defence applications are trained on data sets of appropriate quality. In particular, it should ensure their diversity, veracity and representation.

To do this, the State should develop evaluation benchmarks and metrics to measure the appropriateness of training data, to be developed with multidisciplinary input from the relevant stakeholders. The State should also consider the establishment of a national repository of approved training data sets that have been evaluated against legal and ethical requirements and relevant standards, thus ensuring traceability and consistency in the quality of training data used across various applications.

S.1.6. Appropriateness of testing data

The quality of testing data is also critical for the performance, accuracy and reliability of an AI system. Testing data should be different from training data and remain separate from it. A State implementing a national strategy on AI in security and defence should develop evaluation benchmarks and metrics to measure the appropriateness of testing data, ensuring that it accurately reflects, to the extent possible, the operational environment in which the AI system will be deployed. This will, in turn, enable the conduct of comprehensive, robust and reliable testing and evaluation of AI systems in security and defence.

The State should also deliberate on the question of the production of, and subsequent access to and use of testing data. In particular, technology developers' access to testing data should be carefully framed and considered, especially to avoid the optimization of system performance solely against testing data, with worse downstream reliability after deployment for real-world use.

S.1.7. Proxy data use

Proxy data, or proxy indicators, are often used when the actual data (i.e., direct indicators) required for training an AI model is either unavailable or too sensitive to use. Use of proxy data is particularly important and prominent in security and defence applications (e.g. for ISR), where direct indicators may not always be available. A State developing a national strategy on AI in security and defence should consider the opportunities and risks that stem from proxy data use for the training, testing and deployment of AI in security and defence, in close consultation with the relevant commu-

nities across disciplines. The State should then consider the establishment of guidelines or, at least, good practices that frame the responsible use of proxy data in the context of AI in security and defence. Oversight mechanisms, verification measures and accountability frameworks for the use of proxy data should also be considered and developed by the State as appropriate.

S.1.8. Synthetic data use

The availability, quality and diversity of data have long been raised as key challenges to the development, deployment and use of AI in security and defence.¹⁷ Synthetic data can, in principle, help overcome the limitations that stem from insufficient and unavailable data sets for the training, testing and deployment of AI systems. However, many concerns arise related to the quality and representativeness of synthetic data.

A State developing a national strategy on AI in security and defence should thus develop and adopt guidelines or, at least, good practices for the responsible use of synthetic data in the context of AI in security and defence. The State should also consider and develop oversight mechanisms, verification measures and accountability frameworks for the use of synthetic data, as appropriate. In parallel, it should consider the specific use cases where use of synthetic data would offer a cutting-edge advantage in security and defence systems and, at the same time, address the root causes of insufficient data for the training, testing and deployment of these systems.

S.1.9. Data governance and international law

A State implementing a national strategy on AI in security and defence should put in place a framework with clear guidelines on the governance of data for AI in security and defence. To this end, the State should review legal frameworks, policies and other regulatory instruments in place surrounding data governance and the extent to which they are relevant and appropriate for AI development, deployment and use in security and defence. Notably, the State should consider data privacy as well as existing practices, processes and mechanisms in place for storing and sharing data, including in the context of cross-border data transfers with integrated autonomy or automation.

The State should ensure that all collection, processing, sharing and use of data for AI in security and defence is conducted consistently with the applicable international legal frameworks. In particular, the State is bound by the applicable international human rights law obligations throughout the data life cycle. Notably, this applies to the collection, acquisition or generation in peacetime of data used to develop, train, test and operate AI applications in security and defence. In the development or review of the national strategy and other regulatory instruments, the State should ensure that data governance for AI in security and defence is consistent with IHRL, including the rights to privacy, data protection, freedom of expression and association, non-discrimination, and due process and with the prohibition of arbitrary or unlawful interference with privacy.

17 Harry Deng, *Exploring Synthetic Data for Artificial Intelligence and Autonomous Systems: A Primer* (Geneva: UNIDIR, 2023), <https://unidir.org/publication/exploring-synthetic-data-for-artificial-intelligence-and-autonomous-systems-a-primer>.

S.1.10. Frameworks and agreements for the sharing of classified data

The development, testing, deployment and use of AI in security and defence increasingly requires that suppliers be given access to sensitive or classified data for training, fine-tuning or evaluation processes. A State developing a national strategy on AI in security and defence should establish clear legal, contractual, technical and procedural frameworks governing such access. These should recognize the possible tension that may arise between, on the one hand, the operational benefits of enabling private vendors to work with representative and sufficiently granular data and, on the other, the need to protect classified information, sources and methods.

These frameworks should address, among other things, the security clearance and vetting

requirements applicable to vendors and their personnel; the use of secure and controlled environments for training and fine-tuning activities that would involve classified and sensitive data; contractual limitations on the retention, reuse and onward transfer of data by vendors, including upon the conclusion or termination of an agreement; auditing and monitoring mechanisms to verify vendor compliance with data-handling obligations; and clear allocation of responsibility and liability in the event of unauthorized access, loss or compromise of classified data shared under such arrangements. Furthermore, the States should also consider the extent to which such frameworks could, or could not, be standardized across vendors and procurement processes for AI technologies, particularly to minimize administrative burden while preserving clear, predictable and consistent security standards.

S.2. Implications of machine learning use

The development, deployment and use of machine learning for security and defence applications bring to the fore a series of considerations specific to the nature of these technologies. National strategies on AI in security and defence should thus account for these technologies' unique characteristics and considerations.

S.2.1. Black box

The black box nature of AI systems – where decision-making processes are not transparent and the calculations leading to a system's outputs remain opaque – is a significant and widely shared concern in the light of these technologies' development, deployment and use in security and defence. A State developing a national strategy on AI in security and defence should ensure that AI systems are designed, to the extent possible, with explainability features to ensure and preserve traceability and accountability. This should apply especially for applications involving high risks and high consequences (e.g., target identification in defence settings). At the very least, the State should consider the management and oversight of applications where the black box issue is inevitable, and it should clarify the accountability and responsibility of the relevant actors in the development, deployment and use of these technologies.

S.2.2. Probabilistic nature of AI technologies and accuracy rates

AI technologies have an inherently probabilistic nature: their outcomes are based on likelihood and probabilities, rather than certainty. Concur-

rently, and independently of AI deployment and use, an element of uncertainty is inherently present in the security and defence realms – particularly the latter. A State developing a national strategy on AI in security and defence should thus account for the inherent uncertainties brought by AI use in security and defence, particularly with regards to risk assessments ahead of deployment. Benchmarks and standards to evaluate a system's accuracy and reliability against operational needs, legal compliance and alignment with ethical guidelines will thus be key in addressing the probabilistic nature of AI in security and defence.

S.2.3. Safe testing environments and practices

The robust testing of an AI system in realistic environments prior to its deployment is critical to measuring and ensuring its reliability. Testing must simulate the full range of operational contexts, including edge cases that may reveal system weaknesses. The environments in which these tests are conducted, and adjacent practices, must be consistently controlled, contained, safe and secure.

A State implementing a national strategy on AI in security and defence should thus invest in the financial, human and technological resources required for safe testing environments and practices. These should include regulatory sandboxes specifically dedicated to security and defence applications to provide a controlled environment for the testing and evaluation of new and emerging technologies. The State should also consider and prioritize good practices for the documentation and oversight of testing practices.

S.2.4. Safe evaluation environments and practices

Good evaluation practices should go beyond initial testing to extend throughout the life cycle of an AI system. A State implementing a national strategy on AI in security and defence should conduct a periodic assessment of the performance, legality and compliance of its AI technologies, and the ethical implications of these systems in operational use. The State should develop a framework for continuous evaluation of AI systems and should invest in the technological solutions required to conduct such monitoring (e.g., through audit frameworks and systems).

S.2.5. Transparency in testing and evaluation processes

Transparency – to the extent possible in the light of sensitivities and possible limitations due to export control regimes – in the process of testing and evaluating AI systems builds trust and confidence, both within a State and among regional and international partners. Sharing non-sensitive aspects of AI testing and evaluation processes would constitute a key component of information exchange as a

confidence-building measure at both the domestic and international levels.¹⁸ A State implementing a national strategy on AI in security and defence should thus consider the creation and publication of transparency reports on the testing and evaluation of AI systems in security and defence, while ensuring that sensitive and classified information is preserved to ensure the integrity of the State's national security.

S.2.6. Open-source AI

The application of open-source practices in AI could, in principle, foster innovation and increase accessibility. An increasing number of security and defence agencies – particularly in States with fewer resources – have already adopted, or at least considered the adoption of, open-source AI. Yet, these practices could also present severe risks, particularly in security and defence contexts given their sensitivities, in addition to the risks of proliferation into the hands of non-State armed groups. Hence, a State implementing a national strategy on AI in security and defence should develop clear policies and guidelines that ensures that the need for open innovation to leverage the opportunities that open-source practices offer is balanced with consideration of the sensitivities that stem from security and defence applications.

S.2.7. Foundation models and generative AI

The development and proliferation of large foundation models – including large language models (LLMs), multimodal models, and other general-purpose AI systems capable of performing a wide range of complex tasks – could present a series of distinct governance challenges. A State developing a national strategy

18 Púscas, *Confidence-Building Measures for Artificial Intelligence*.

on AI in security and defence should consider the specific implications of foundation models and generative AI for security and defence applications. These include the near impossibility of fully auditing the training data of large foundation models, which may have been sourced from the open Internet and may include contested material; the risk of hallucination, particularly in security and defence contexts where confidence in AI-generated analysis may be misplaced and would be critical; automation bias, whereby operators in high-tempo operational environments may be inclined to accept AI-generated recommendations without concrete and proper evaluation; and the challenge of maintaining substantive, rather than nominal, human oversight of AI systems whose outputs are presented in authoritative natural language.

The State should consider developing governance frameworks that address these specific challenges throughout the life cycle of these technologies, including through testing and evaluation protocols tailored to foundation model behaviour. It should also consider clear guidelines on the authorized use cases of generative AI in security and defence contexts, and explicit requirements for the documentation and review of AI-generated products used to support legally and operationally consequential decisions.

S.2.8. Agentic AI systems

Agentic AI systems – that is, AI systems capable of taking multi-step actions autonomously and across complex task sequences, with limited, intermittent or no human intervention – represent an emerging category of applications with governance implications that are distinct from those stemming from narrow AI or static foundation models. A State developing a national strategy on AI in security and defence

should consider the implications of agentic AI for security and defence applications. These should include possible implications for the human element – and frameworks and regulations constructed around a specific approach to the human element (e.g., human “judgment”, “control” or “intent”) – when AI agents act autonomously at machine speed across extended action sequences. They should also include considerations surrounding predictability, distributed action and assumptions surrounding the ability to maintain accountability and responsibility; and the implications of agentic systems for existing command structures, doctrines, legal review processes and wider compliance assessments that were designed for human decision makers.

The State should also consider the development of specific governance frameworks for agentic AI in security and defence. These should address questions of authorization scope, intervention thresholds, rollback mechanisms, and the conditions under which agentic AI may and may not be deployed in specific operational contexts and within specific constraints and parameters (e.g., through agentic harnesses).

S.2.9. Adversarial robustness and AI security

AI systems in security and defence will constitute targets for deliberate adversarial manipulation, including through attacks on their training data (data poisoning), their inputs at deployment (adversarial examples), their underlying model architecture (model extraction or inversion) and their supporting digital infrastructure (cyberattacks on AI systems). A State developing a national strategy on AI in security and defence should consider the implications of adversarial attacks for the reliability, integrity and safety of AI systems in security and

defence, and it should develop governance frameworks to address them. These frameworks should include requirements for adversarial robustness testing as a key component of development and pre-deployment testing and evaluation; investment in technical solutions and processes to monitor, detect and mitigate adversarial inputs throughout their life cycle; specific processes for AI systems operating in contested electromagnetic environments where sensor inputs may be deliberately spoofed; and mechanisms to identify and respond to in-service degradation of AI performance that may indicate adversarial interference.

The State should also consider the extent to which AI security considerations can be integrated into existing cybersecurity and resilience frameworks, while remaining attentive to the specific characteristics of AI systems that may require tailored approaches.

S.2.10. The human element

The integration, deployment and use of AI into security and defence applications will introduce a series of considerations of the human factor that extend beyond the technical characteristics of the systems themselves. A State developing a national strategy on AI in security and defence should consider the cognitive and psy-

chological dimensions of human interactions with AI systems. These should include the risk of automation bias, the converse risk of automation aversion, as well as the possible limitations of specific constructs such as “human control” or “human judgment”, particularly as these technologies progress at scale and at speed. The State should also consider the risk of diffusion of responsibility, particularly when reliance on AI systems for consequential decision-making processes may have implications for individual accountability over the outcomes of these decisions.

Additionally, the State should consider the psychological dimension and integrate these into technical frameworks, policies, legal assessments (particularly if they are subject to a degree of subjective assessment) and operational processes. Examples in the case of technical frameworks include in the context of co-design of the AI system’s user interface. In the case of policies, it can extend to training programmes that cover technical proficiency and the ability to maintain critical judgment calibrated, as appropriate, to the risk of skill atrophy in the age of AI. Operational processes can, for example, be used to identify and respond to system errors, limitations and protocols in response to unpredictable outputs or unintended outcomes.

S.3. Public–private partnerships

Public–private partnerships are key to fostering responsible innovation in AI, including in the security and defence sector. States should thus explore ways in which such partnerships can be leveraged to advance the implementation of national strategies on AI in security and defence, while ensuring such collaboration is effective, meaningful and sustainable.

S.3.1. Private companies and development of AI for defence

In addition to traditional technology companies, a State implementing a national strategy on AI in security and defence should address the specific governance implications arising from the growing role in the development and deployment of defence AI capabilities of private companies and other specialized defence AI contractors. These actors occupy a distinct position in the governance landscape: while they are commercial technology providers, their systems may be deployed in contexts with direct implications for legal compliance and the State's responsibilities under international law.

The State should thus consider the extent to which existing legal frameworks adequately govern the roles, responsibilities and accountability of private actors. It should also consider the extent to which AI developed by such actors brings implications that are novel or are more complex than those of non-AI systems. These implications could include the legal and ethical standards that the AI systems developed by private actors must meet prior to deployment; the obligations of these actors in the event of AI-related incidents or harms (e.g., the duty to conduct effective investigations for alleged violations of IHL and the need for adequate measures); questions about intellectual property

(IP) and data ownership that arise from collaborative development; and the governance of on-going updates of models, retraining and operational dependencies that may arise post-deployment.

S.3.2. Distribution of roles and responsibilities

A State that is implementing a national strategy on AI in security and defence should establish frameworks that provide a clear definition and distribution of roles and responsibilities between the public and private sectors. While the formulation, enforcement and oversight of the national strategy should remain the sovereign prerogative of the State, it should also consider the role that the private sector could play in facilitating the implementation.

In particular, the State should provide for the accountability and responsibility of the private sector in the context of the development, deployment and use of AI in security and defence. The national strategy must thus include these considerations during development, adoption, implementation and review. In addition, the State should also consider the formulation of performance-measurement and evaluation methods, with specific guidelines to evaluate and measure the success of public–private partnerships.

S.3.3. Foster research and development ecosystems

A thriving research and development ecosystem is critical for responsible AI innovation. A State implementing a national strategy on AI in security and defence should thus promote collaboration between government, academia and industry as appropriate in the context of the development, deployment and use of AI in security and defence. Through the establishment of national AI innovation frameworks and structures, the State should incentivize R&D, in addition to mobilizing and allocating the financial, human and technological resources required to foster its national R&D ecosystem.

S.3.4. Intellectual property

A State implementing a national strategy on AI in security and defence should consider the implications that public–private partnerships may have for intellectual property rights, particularly with regards to the development of dual-use technologies. The State should provide guidelines to ensure clear and defined ownership and licensing rights for the IP developed within public–private partnerships. These guidelines should also define the extent to which IP rights may be leveraged to protect sensitive information and technologies.

S.3.5. Supply chain security

The development of AI systems for security and defence relies on complex, multi-tiered supply chains that may span across hardware components, software interdependencies, open-source libraries, data pipelines and jurisdictions. A State implementing a national strategy on AI in security and defence should consider the security and resilience of these supply chains. To this end, the State should consider the development of supply chain mapping and assurance processes, including the vetting of sub-tier suppliers (if possible), the verification

of component and software provenance, and the establishment of security standards for vendors and their subcontractors.

While existing frameworks (e.g., export and import control regimes) could be leveraged for security, the State should pay particular attention to the specific implications that AI may bring. These implications could include, for instance, dependencies on open-source software and models, which may offer significant benefits in terms of innovation and cost but which may also introduce distinct assurance challenges given the diversity and, at times, limited oversight of their development, training, testing and overall contributor base.

S.3.6. Security and resilience of the industrial ecosystem

The development, provision, production and maintenance of AI capabilities for security and defence applications are supported by a wide industrial ecosystem. In addition to the security of supply chains, a State developing and implementing a national strategy on AI in security and defence should thus consider the security and resilience of that ecosystem. This should include, for instance, the degree of diversity (or, conversely, concentration) among vendors and suppliers and subsequent interdependencies, risks surrounding industrial espionage, IP theft, and interruptions in the workflow and supply chain as a result of acquisition, merger or insolvency.

The State should also consider the establishment and implementation of the appropriate safeguards, governance frameworks and policies to address these risks and foster a secure and resilient industrial base. This could include, for instance, targeted investment, support and incentives for emerging companies.

S.3.7. Evolving procurement dynamics

The procurement of AI capabilities for security and defence is increasingly shifting away from traditional procurement models that centred on the one-off acquisition of a physical capability. States are increasingly acquiring access to AI capabilities through software subscriptions, model-access agreements and cloud-based services – a fundamentally distinct approach to traditional procurement. Furthermore, in some instances, States are engaging with AI vendors and suppliers in ways that are closer to investment relationships, such as through co-funding arrangements, strategic partnerships, seed funding for emerging companies, and early commitments that function as a form of demand signal.

States should consider the distinct governance implications of this shift. These implications include the operational interdependencies that may arise from subscription-based or service-based access to AI capabilities, as opposed to the ownership of a discrete system. They also include the need for contractual provisions that address continuity of service, data portability, standard operating procedures (SOPs) if the model is under maintenance or if a new version of the model has been released, and rights and obligations upon contract terminations. Similarly, implications arise from security clearance and vetting requirements for the personnel of the supplying company with access to sensitive systems or data; and the implications of these evolving relationships for the State's overall posture in relation to technological dependencies, competition and strategic autonomy.

Existing procurement frameworks, doctrines and policies should subsequently be adapted to ensure that they are adequate in the light of

these evolving dynamics. They must also ensure operational continuity in security and defence contexts, particularly in situations of crises or when contingencies are necessary.

S.3.8. Maintenance

The continued and sustained performance, reliability and security of AI systems in security and defence requires on-going maintenance across their life cycle. This includes software updates, re-training to address performance degradation or data drift, the patching of identified vulnerabilities and the maintenance of the hardware component of AI-enabled capabilities. A State implementing a national strategy on AI in security and defence should thus consider the resourcing, contractual and institutional arrangements required to ensure the continued maintenance of these systems. This should include the extent to which maintenance will be the responsibility of government agencies, rather than relying on vendor support – particularly when these systems are purchased from abroad.

The State should further develop contingency plans to address the risk of discontinuation of vendor support, whether due to commercial decisions, business failure or the withdrawal of services for any other reasons. The State should also consider the inclusion of maintenance and support guarantees in procurement agreements, including minimum notice periods for discontinuation. In addition, it should assess the resource implications of maintenance obligations over the full life cycle of AI systems, recognizing that these may constitute a substantial and recurring component of the overall cost of AI capabilities in security and defence.

S.4. Technological transfer

States should outline in their national strategies their approaches to technological transfer of AI for security and defence, both in procurement and sales.

S.4.1. Technological transfer, equity and proliferation risks

Technological transfer, especially with regards to AI in security and defence, may be driven by a number of considerations, including procurement, sales, assistance and joint efforts with allied or partner States. Yet, when implementing a national strategy on AI in security and defence, a State should consider how to approach this issue while considering risks surrounding technological proliferation and potential misuse.

Considerations should extend beyond the transfer of the technology itself. It should thus cover “package deals” that include training, as well as the technical and broader assistance that can be offered to ensure that the technology is successfully and responsibly adopted and integrated into the recipient State’s defence capabilities throughout its life cycle, including with respect to maintenance. The State should develop a framework to ensure that it is able to maintain control and oversight over the transfer of these sensitive technologies.

These concerns must also be counterbalanced by the need to promote equitable access to AI capabilities for responsible actors. The State should thus establish comprehensive policies to frame the technological transfer of AI systems in security and defence to foreign governments, private entities or international partners.

S.4.2. Procurement from foreign governments

A State implementing a national strategy on AI in security and defence should carefully evaluate its approach to the procurement of AI capabilities from foreign governments. The State should thus develop a framework to govern such procurements, ensuring that they meet security and operational needs in addition to legal compliance and alignment with ethical guidelines. The State should ensure that clear and robust processes are in place for local acceptance tests along with pre- and post-sale feedback mechanisms with the foreign seller. It should invest in the solutions needed to ensure the safe and secure interoperability of these new technologies with existing systems.

S.4.3. Sales to foreign governments

In implementing a national strategy on AI in security and defence, a State should carefully evaluate its approach to sales to foreign governments of AI capabilities for security and defence applications. The State should thus develop a framework to govern such sales, ensuring that they are conducted in compliance with existing international arms control obligations and in alignment with national security interests.

Clear and robust processes must be in place to ensure transparency in these transactions and ensure a degree of oversight to prevent misuse in ways that would undermine regional security or, more generally, international peace and security. Such a framework should specifically consider both government-to-government sales and business-to-government sales (i.e., when a private actor operating on a State's territory sells AI-enabled capabilities to a foreign government).

S.4.4. Sales to non-governmental entities

Sales of AI with security and defence applications to non-government entities should be strictly regulated through the establishment of a framework governing such transactions. In order to reduce risks of misuse, a State implementing a national strategy on AI in security and defence should establish and implement thorough controls and oversight mechanisms to ensure that these technologies are neither used nor repurposed to enable the violation of international law or in ways that would compromise national security.

The State should thoroughly vet any non-governmental entity seeking to purchase AI applications for security and defence. This should include evaluation of their intended use, of the entity's capacity for responsible management and use, and of its adherence to applicable laws. The State should thus consider the establishment of a licensing body to examine, review and deliberate on (i.e., either approve or reject) sales of such capabilities to non-governmental entities. Specific contractual clauses should also be included to ensure and maintain accountability and responsibility pre- and post-sales.

S.4.5. Oversight and verification

In the light of the sensitive nature of AI technologies, particularly those with applications in security and defence, a State implementing a national strategy on AI in security and defence should consider the establishment of oversight and verification mechanisms in the context of technological transfers – whether the State acts as the supplier or purchaser of these capabilities.

The State should also consider the establishment of an independent oversight body mandated to review AI technology transfers. Its establishment, composition, level of independence and governance should be appropriate, cognizant of the sensitivities surrounding this topic, while ensuring that such a body has the capacity and power to conduct investigations into and document and report on any transfers that may pose a risk to national, regional or international security.

S.4.6. Compute governance and strategic dependencies

The development, deployment and use of AI in security and defence is closely contingent on access to high-performance computing infrastructure, including specialized semi-conductors, secure cloud platforms and the necessary network infrastructure. A State implementing a national strategy on AI in security and defence should conduct a comprehensive assessment of its strategic dependencies in this context. This should include the extent to which AI development and deployment capacity would be dependent on provision of hardware, software or cloud services and by whom; where these service and technology suppliers are located; as well as material dependencies.

The State should consequently develop a strategy to address material dependencies over time, including through dedicated investment, regional and international cooperation, and the diversification of supplier relationships as necessary and as appropriate. The State should further consider the energy and environmental implications of compute-intensive development, deployment and use of AI, and it should develop frameworks and policies to manage the resulting, growing demands on national energy infrastructure.

S.4.7. Engagement with foreign industries

In addition to the considerations that relate to procurement from and sales to foreign governments (see guidelines S.4.2 and S.4.3), a State implementing a national strategy on AI in security and defence should address the implications of its own domestic industry's engagement with foreign companies in the

development of AI capabilities for security and defence. This includes, for instance, foreign investment in, acquisition of, or joint venture agreements with domestic companies of strategic relevance, collaborative R&D ventures involving foreign industry partners, and the licensing or transfer of technology between domestic and foreign companies.

The State should consider the development of monitoring and oversight mechanisms applicable to such engagement and dynamics. These mechanisms should balance, on the one hand, the potential benefits of international industrial collaboration against, on the other, risks of unauthorized technology transfer. The State should ensure that its approach to foreign industry engagement aligns with its broader frameworks, regulations and policies on technological transfer, on import and export control regimes, and on investment screening.

S.5. Life cycle management

A national strategy on AI in security and defence should outline the ways in which the State intends to conduct the management of these technologies' life cycle. While there is no universally accepted model of an AI technology's life cycle and its stages, management of the life cycle provides States with clear entry points to promote responsible behaviour and practices, from the development to the decommissioning of AI technologies in security and defence.¹⁹

S.5.1. Identification of relevant life cycle stages

Effective life cycle management requires the clear identification of all stages in an AI technology's life cycle, from design and development to decommissioning. A State implementing a national strategy on AI in security and defence should consider that the life cycle of most, if not all, AI technologies is neither linear nor necessarily cyclic in a straightforward manner. The State should thus design policies and frameworks that identify and outline each life cycle stage, and it should then incorporate good practices from both military and civilian use of AI to promote responsible behaviour and practices.

Each State agency may have its own definition, interpretation and approach to the life cycle of these technologies, depending on its respective mandate, jurisdiction and policies. However, a fragmented approach to life cycle management across national agencies may carry risks with regards to fragmentation, vulnerabilities, oversight (or lack thereof), inter-

operability and incoherence both domestically and internationally. To ensure alignment at the national level, the State should thus develop and implement a national life cycle management framework for AI in security and defence.

S.5.2. Identification of relevant actors across life cycle stages

The complexity of AI systems requires the involvement and intervention of various actors, both public and private, across the technology's life cycle. Each actor – from developers, via testing and evaluation professionals and lawyers, to the end users – should have clearly defined and distributed roles and responsibilities at each stage of the technology's life cycle. A State implementing a national strategy on AI in security and defence should thus ensure the involvement of all relevant stakeholders, as appropriate, in the development, adoption, implementation and review of the national strategy to ensure that no actor is overlooked across the technology's life cycle.

19 Yasmin Afina and Giacomo Persi Paoli, *Governance of Artificial Intelligence in the Military Domain: A Multi-Stakeholder Perspective on Priority Areas* (Geneva: UNIDIR, 2024), <https://unidir.org/publication/governance-of-artificial-intelligence-in-the-military-domain-a-multi-stakeholder-perspective-on-priority-areas>.

Clear communication channels and coordination mechanisms should be established among all stakeholders. This should include the facilitation of regular meetings that transcend State agencies and disciplines and, as appropriate, incorporate the viewpoints of both governmental and non-governmental entities, including the private sector.

S.5.3. Distribution of roles and responsibilities across the technology's life cycle

A clear definition, delineation and distribution of roles and responsibilities across the technology's life cycle would prevent gaps in oversight and ensure that AI systems are managed effectively. A State implementing a national strategy on AI in security and defence should thus adopt and establish a clear organizational structure, with the appropriate resources allocated, to ensure effective coordination between actors, as well as the implementation and oversight of these distributed roles and responsibilities.

S.5.4. Considerations for the technology's procurement

The procurement of AI systems should be managed and framed carefully to ensure alignment with national, regional and international policies, applicable laws and regulations (particularly with respect to export and import controls if it involves a foreign supplier), as well as ethical guidelines. A State implementing a national strategy on AI in security and defence should thus take stock of current procurement policies and frameworks, review procurement practices for AI capabilities, and evaluate their suitability to ensure their security, reliability and compliance.

The State should establish procurement guidelines that include testing and evaluation benchmarks that combine considerations of

compliance with applicable laws, ethical guidelines and operational needs. Local acceptance test benchmarks will be particularly critical for capabilities purchased externally (i.e., from other States or from foreign industries) and for off-the-shelf capabilities since their development and training have not been tailored to the procuring State's requirements. A periodic review of these procurement guidelines should be conducted to factor in technological progress and the emergence of new norms and principles.

S.5.5. Considerations for the technology's development

The design and development stages of AI technologies in security and defence must be centred around legal compliance and ethical guidelines. It is critical that a State developing a national strategy on AI in security and defence translates legal requirements and ethical considerations into concrete action for the design, development and training of AI systems. These could include, for example, good practices for compliance with international humanitarian law or data practices for the training and testing of AI-enabled target-identification systems.

Additionally, a State implementing a national strategy on AI in security and defence should ensure that safety and security considerations are incorporated from the early stages of the development of these technologies, including identification and mitigation of bias and resilience against adversarial attacks. To this end, the State should consider the possibility of co-developing with the technology supplier, to the extent possible and appropriate, the capability under procurement. Such co-development would not only require the State to dedicate the necessary resources; it will also require the establishment and mobilization of specific processes, partnerships, frameworks and architectures that would allow for this

exercise to be conducive to innovation while balancing national security considerations.

S.5.6. Considerations for the technology's testing and evaluation

The testing and evaluation of AI technologies in security and defence play a key role in ensuring reliability and the alignment of AI systems for security and defence with the State's policies, applicable laws and ethical guidelines. From an operational standpoint, testing and evaluation are critical to ensuring that the systems perform as intended and as expected under the appropriate conditions. Beyond performance, testing and evaluation are also critical in verifying the implementation and effectiveness of key principles (e.g., human control, judgment, oversight and involvement, legal compliance, and alignment with ethical principles).

A State implementing a national strategy on AI in security and defence should thus not only develop comprehensive testing and evaluation frameworks with clear protocols as part of the national strategy, it must also invest in and allocate the resources necessary to facilitate thorough testing and evaluation (e.g., through simulations, adversarial stress-testing for edge cases, and the identification of and planning for potential failure scenarios). These frameworks and protocols must take account of the State's national security context, realities and wider environment. They could consider, for instance, the extent to which independent, third-party testing and evaluation would be desirable, if at all, in the light of the State's historical relationship with non-governmental bodies and organizations in sensitive domains.²⁰

S.5.7. Considerations for the technology's adoption, integration and deployment

The adoption and deployment of an AI technology in security and defence should be done in a phased and controlled manner, to ensure that it is integrated safely and securely into existing systems. The effective integration and interoperability should be tested, monitored and evaluated periodically to ensure consistent reliability and predictability – particularly as these systems would be expected to interact, or even create co-dependencies, with existing legacy systems in the security and defence apparatus.

A State implementing a national strategy on AI in security and defence should thus develop and adopt clear procedures for the introduction of AI into operational settings and should address issues that may arise post-deployment.

S.5.8. Considerations for the technology's use

A State developing a national strategy on AI in security and defence should consider the relevant operational, policy, legal and ethical considerations and implications of the use of AI in security and defence. These include legal compliance, accountability and the human element, specifically the ways in which the State should ensure, when appropriate, human control, judgment, oversight or involvement. A risks-based approach to framing and governing the use of AI technologies in security and defence should be considered, as well as the appropriate measures necessary to mitigate risks stemming from their operational use.

20 A recent UNIDIR study found, for example, that attitudes towards third-party evaluations diverge greatly across regions for military applications. See Afina, *The Global Prism of Military AI Governance*.

Furthermore, the State should consider how it could leverage existing frameworks, tools and processes to govern the use of these technologies. These include, for example, military doctrine; SOPs; rules of engagement; tactics, techniques and procedures (TTPs); logbooks; and after-action reviews (AARs).²¹

S.5.9. Considerations for the technology's end of life and decommissioning

The effective and responsible management of an AI system's end of life is a critical yet often overlooked dimension of life cycle governance and of the broader responsible development, deployment and use of AI in security and defence. A national strategy on AI in security and defence should outline clear procedures for the retirement and decommissioning of AI systems in security and defence. This should encompass the secure data destruction or transfer of data in compliance with applicable data-protection and classification frameworks; the documentation of the system's performance and use history; incident records and compliance assessments; the governance of second-hand sales, techno-

logical transfer and re-purposing of decommissioned systems; and the safe and secure recycling or repurposing of components.

Additionally, the State developing the national strategy should consider the concerns and risks surrounding proliferation, reliability and accountability of repurposed technologies that have been previously used and subsequently decommissioned. Furthermore, the State should also consider the issues that may arise from systems that may not have been formally decommissioned, yet have reached effective obsolescence due to evolution in the operational environment. As such, it should identify and designate a specific agency or body as being responsible for overseeing the safe and secure management of a technology's end of life.

21 Giacomo Persi Paoli and Yasmin Afina, "The tactical governance of artificial intelligence in the military domain", Global Commission on Responsible AI in the Military Domain (GC REAIM) Expert Policy Note Series, May 2025, <https://hcss.nl/wp-content/uploads/2025/05/Dr-Giacomo-Persi-Paoli-and-Dr-Yasmin-Afina.pdf>.

S.6. Capacity-building

The development, adoption, implementation and review of a national strategy on AI in security and defence requires sustained capacity-building efforts across multiple dimensions. States should thus approach capacity-building holistically, recognizing that human, technological, financial and regulatory capacity are interdependent, and that gaps in one dimension may hinder or slow down progress in others.

S.6.1. Human capacity

Human capacity encompasses the recruitment, training, retention and development of personnel with the required technical, legal, ethical and operational expertise. It constitutes a foundational dimension of capacity-building for AI, with specific considerations in security and defence (see also guidelines S.7.1–6). A State developing a national strategy on AI in security and defence should assess its need for human capacity and invest in the necessary tools to develop and retain it.

S.6.2. Technological and infrastructural capacity

A State developing a national strategy on AI in security and defence should assess and invest in the technological capacity and infrastructure required to support the development, deployment and use of AI in security and defence in a manner that considers, holistically, compliance with applicable laws and regulations, operational requirements, policy and governance frameworks in place, technical limitations and the State's wider security landscape. This might include, for instance, investment in computing infrastructure, secure data storage

and frameworks and environments for network infrastructure, testing and evaluation, and the technical tools necessary for monitoring, auditing and oversight.

The State should further consider the extent of its current capacity in a way that relates to its strategic ambitions. It should also consider possible material gaps that will have been identified and measured, resource implications and possible partnerships.

S.6.3. Financial capacity and resource mobilization

The development, adoption, implementation and review of a national strategy on AI in security and defence will require the mobilization and allocation of substantial financial resources over a sustained period. A State developing and implementing a national strategy on AI in security and defence should assess the financial resources required to support each dimension of capacity-building. These include human capacity development, technological and infrastructural investment, and the on-going costs associated with the maintenance, testing, evaluation and oversight of AI systems throughout their life cycle.

To this end, the State should consider the range of different funding mechanisms possible and available, such as dedicated budgetary allocations, public–private joint investment funds, and international funding and cooperation arrangements. The resource-mobilization plan should be realistic, sustainable and in alignment with the broader strategy’s stated objectives, timelines and milestones.

S.6.4. Regulatory and policy capacity

Regulatory and policy frameworks play a distinctive, enabling role relative to the other dimensions of capacity-building. They constitute a pre-condition for capacity-building, with a national strategy on AI in security and defence in particular providing the basis and institutional

mandate without which human, technological and financial capacity cannot be meaningfully built and mobilized. Simultaneously, they also constitute an active enabler of capacity-building by providing the required investment, training and institutional designs. They are also an outcome in its own right.

A State developing and implementing a national strategy should thus consider the suite of regulatory and policy frameworks necessary for the development and implementation of the national strategy and more widely. They should also assess needs and perceived gaps that need to be addressed for the broader development, deployment and use of AI in security and defence.

S.7. Human resources

While the development, adoption, implementation and review of national strategies on AI in security and defence require substantial financial resources, States should not overlook the importance of considering and investing in the human resources required to ensure these strategies' effectiveness.

S.7.1. Training and retention of AI talent

A State implementing a national strategy on AI in security and defence should cultivate, retain, foster and nurture AI talent and should address the difficulties encountered in the recruitment and retention of qualified personnel. In close collaboration with academic and research institutions and the private sector, the State should establish and invest in long-term programmes to form a solid AI talent pool, in addition to providing competitive salaries and opportunities for professional development and advancement in the sector.

The State should also consider the development of national AI training programmes and curriculums linked to career pathways specifically in the security and defence sector. Through the establishment of partnerships with academic and research institutions and the private sector, these programmes would create and facilitate specialized AI courses focused on security and defence applications.

S.7.2. Capacity-building and awareness-raising of developers

While developers of AI technologies in security and defence should be well-versed in technical skills, a State implementing a national strategy on AI in security and defence should also ensure their awareness of, and literacy in, ethical, legal and operational standards, considerations and

obligations. The State should thus facilitate capacity-building efforts to ensure that developers undergo ethical and legal training specific to the development of AI applications in security and defence. The State should also consider incentivizing such training, including through certifications.

S.7.3. Capacity-building, upskilling and awareness-raising of users

A State implementing a national strategy on AI in security and defence should invest in the intended users' technical, legal and ethical capacity and knowledge. The State should mobilize the appropriate resources required to ensure that all personnel using AI systems in security and defence receive adequate training on not only the functionality and parameters of the systems, but also the technical limitations of the systems in use, as well as the relevant legal and ethical considerations stemming from the deployment and use of these systems.

Ultimately, users should be able to conduct informed risk assessments and, when operating the system, should be well-informed of its characteristics, limitations, and the legal and ethical considerations relevant to specific use-cases. Users should also be aware of the institutional architecture in which they operate, including the relevant bodies, agencies and stakeholders, oversight mechanisms, and accountability frameworks in place.

S.7.4. Capacity-building and awareness-raising of policy and regulatory entities

Policymakers and regulators must be equipped with a deep and thorough understanding of AI technologies and their inherent complexities in order to ensure the development and implementation of robust, adaptive and evidence-based governance frameworks. A State developing a national strategy on AI in security and defence should undertake a comprehensive assessment of internal capacity and knowledge gaps across governmental agencies.

The State should subsequently invest in training for the policymaking and regulatory communities to reinforce their understanding of the technical, legal and ethical dimensions of AI in security and defence. Such training should be conducted periodically, thus ensuring policy and regulatory entities are well-informed of the technological progress and subsequent risks stemming from the development, deployment and use of AI in security and defence.

Policy and regulatory entities should be aware of key developments in the technological, governance and operational fields that have direct or indirect implications for their work. The State should invest in tools and systems to enable the monitoring of these key developments.

S.7.5. Academic and civil society engagement

A State implementing a national strategy on AI in security and defence should establish partnerships with academic institutions for research and development in AI, thus ensuring that the

State remains at the cutting edge of AI innovation. Academic and civil society engagement should also be facilitated to ensure expert and multi-stakeholder input on the policy, legal and ethical implications of AI in security and defence. To this end, the State should consider incentivization (including through the funding of research by academic institutions and civil society organizations) and the establishment of frameworks to facilitate such engagement in a consistent and sustained manner.

S.7.6. Industry engagement

Public-private partnerships will be key in ensuring that the development, deployment and use of AI in security and defence are done in a responsible, safe and secure manner. Acknowledging the pioneering role that industries play in driving AI innovation and R&D, including in security and defence, a State implementing a national strategy on AI in security and defence should engage closely with the private sector. The State should also consider the creation and maintenance of formal and informal channels and platforms to facilitate such industry engagement.

Beyond establishing dialogue and fostering collaboration opportunities, industry engagement would also be critical in ensuring the alignment and adherence of industry to national policies, international law and ethical principles.²² To this end, incentivization and clarity will be key, which the State should reflect on and foster through its national strategy.

22 On concrete, international efforts in this space, see UNIDIR's project on developing a Framework of Responsible Industry Behaviour for AI in the Military Domain: <https://unidir.org/framework-of-responsible-industry-behaviour-for-ai-in-the-military-domain>.

S.8. Legal compliance

All States are bound by international law, which includes international humanitarian law, international human rights law and public international law. As States grapple with the governance of the development, deployment and use of AI in security and defence, national approaches and policies must remain within the boundaries set by the law. As such, States should put legal compliance at the heart of their national strategies on AI in security and defence and outline how they interpret and approach certain dispositions of the law. States should also outline in their national strategies the concrete measures that they intend to adopt and implement in order to foster compliance with international law (e.g., through the conduct of iterative legal reviews).

S.8.1. Compliance with international humanitarian law

A State developing a national strategy on AI in security and defence must ensure that the development, deployment and use of AI in security and defence complies with international humanitarian law. The latter is applicable in international and non-international armed conflict; this, however, does not mean that IHL considerations can be omitted by States in peacetime. Good practices to foster compliance with IHL outside armed conflict (e.g., at the design and development stages of defence AI capabilities, Article 36 legal reviews) should be identified and implemented.

The State should put in place processes to facilitate the conduct of legal reviews while procuring, testing, evaluating and adopting AI systems, with clear guidelines and benchmarks to measure compliance. It should also put in place processes for the conduct of periodic legal reviews to ensure consistent compliance

with IHL. The State should also invest in the development of solutions to monitor and audit a system's ability to comply throughout its life cycle and, as necessary, flag modifications in the system's parameters and performance that may affect its ability to operate within the boundaries of IHL.

S.8.2. Compliance with international human rights law

The development, deployment and use of AI technologies in security and defence must comply with international human rights law, both in conflict and peacetime. A State developing a national strategy on AI in security and defence should thus clarify the IHRL implications of the development, deployment and use of these technologies, particularly with regards to the right to life, freedom of expression and the prohibition of arbitrary detention. The State should also outline in its national strategy how it intends to foster compliance with IHRL, including through due process and legal review mechanisms.

For applications of AI in armed conflict, the State should reflect on the applicable obligations in both IHL and IHRL, and the interplay between the two bodies of law in the context of AI development, deployment and use. It should pay particular attention to security applications of AI that operate predominantly and exclusively within an IHRL framework, rather than IHL. These applications include predictive policing, biometric surveillance, border and migration management and AI-assisted decision-making in detention and sentencing contexts. In such instances, the absence of an armed conflict threshold does not diminish the State's human rights obligation; rather, it may heighten them. The national strategy should therefore require necessity and proportionality assessments prior to the development of such tools, together with mechanisms for redress for individuals adversely affected by algorithmic decision-making.²³

Data collection for security and defence application is likely to be done during peacetime when IHRL, not IHL, applies (as noted in guideline S.1.9). Recognizing this, the equivalent provisions specific to armed conflict equally apply and must therefore respect both IHL and IHRL. Capacity-building should also be prioritized, notably to sensitize and raise awareness about IHRL obligations and implementation mechanisms among public bodies using, or at least expected to use, AI technologies in security and defence. There should also be awareness-raising with regards to human rights

obligations among non-State stakeholders, notably the private sector (e.g., industries developing AI capabilities in security and defence). This should include due diligence expectations consistent with the UN Guiding Principles on Business and Human Rights.²⁴

S.8.3. Compliance with international criminal law

International criminal law warrants particular attention as a State developing a national strategy on AI in security and defence should consider how existing modes of individual criminal responsibility (including command responsibility) apply when AI-enabled or autonomous systems are involved in the commission of a war crime. The State should therefore ensure that the distribution of roles and responsibilities established throughout the technology's life cycle is capable of sustaining meaningful accountability. Given that different modes of responsibility exist under the Rome Statute and customary international law, the national strategy should consider each mode of responsibility separately.²⁵

S.8.4. Compliance with public international law

Public international law governs State behaviour and relations, including in the context of AI development, deployment and use in security and defence. Thus, when developing a national strategy on AI in security and defence, a State should consider the public international

23 General Assembly, Human Rights Council, "Practical application of the Guiding Principles on Business and Human Rights to the activities of technology companies, including activities relating to artificial intelligence", Report of the Office of the United Nations High Commissioner for Human Rights, A/HRC/59/32, 16 June 2025, <https://docs.un.org/A/HRC/59/32>, p. 10.

24 Office of the United Nations High Commissioner for Human Rights, *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, HR/PUB/11/04 (New York: United Nations, 2011), https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

25 Rome Statute of the International Criminal Court, 1998, <https://www.icc-cpi.int/sites/default/files/2024-05/Rome-Statute-eng.pdf>, Articles 25(3) and 28.

law implications of AI, possible risks and mitigation measures, as well as its interpretation of a number of contentious related aspects. Such reflections are particularly important in the light of the possible integration of AI into decision-making processes surrounding the resort to and use of force, and subsequent implications for *jus ad bellum*. These are necessary to ensure that there is no risk of inadvertent or accidental escalation.

Considerations should include State responsibility and the question of attribution, including where AI systems contribute to decision-making chains that complicate – and heighten the political sensitivity of – the identification of the responsible actor, and the subsequent implications for State sovereignty of AI deployment and use. This applies, for example, in AI-enabled trans-border ISR activities or in the use of AI to support decisions on the threat or use of force.

S.8.5. Compliance with other bodies of international law

When developing, adopting, implementing or reviewing a national strategy on AI in security and defence, a State should also consider the applicability and subsequent application of other bodies of public international law. These include international maritime law, international space law, international refugee law, and disarmament law. Building on these reflections, the State should take stock of the statutory and customary obligations to which it is bound by each of these bodies of law. It should then adopt

the appropriate measures needed to foster compliance in the context of the development, deployment and use of AI in security and defence applications.

The State should also continuously monitor the interplay between AI and existing arms control and disarmament frameworks to which it is a party to avoid the creation of gaps and overlaps in coverage.

S.8.6. Implications of technological transfer for export and import control regimes

Legal compliance questions arise from the pathways by which capabilities for AI in security and defence move between actors, and possibly proliferate to new actors. Five pathways should be distinguished since each engages different legal instruments and calls for different national measures. The first is the transfer of AI-enabled military systems and platforms (ranging from loitering munitions and uncrewed vehicles to AI-enabled decision-support systems). The second is the transfer of inputs on which AI capabilities depend (including advanced computing hardware, model weights, training data, algorithms and associated technical assistance).²⁶ The third is diffusion that involves no discrete transfer at all (e.g., public release of open weight models, remote access to hosted models through cloud services or application programming interfaces, and the technical content of model outputs).²⁷ The fourth is the role of AI as an accelerant of proliferation in other domains (e.g., by lowering the expertise barriers to

26 Kolja Brockmann, “Regulating transfers of AI algorithms, training data and models: The potential and limitations of export controls”, SIPRI, 25 March 2026, <https://www.sipri.org/commentary/topical-background/2026/regulating-transfers-ai-algorithms-training-data-and-models-potential-and-limitations-export>.

27 Wassenaar Agreement, “Best practices for implementing intangible transfer of technology controls”, 2006, https://www.wassenaar.org/app/uploads/2019/consolidated/ITT_Best_Practices_for_public_statement_2006.pdf

chemical, biological or cyber capabilities).²⁸ The fifth is the role of AI as an accelerant for its own proliferation (e.g., by lowering the technical, financial and organizational barriers to developing adapting and deploying additional AI systems).²⁹ A national strategy on AI in security and defence should address each pathway and should consider both the State's obligations as a supplier and its exposure as a recipient, importer and dependent user of AI software developed abroad.

S.8.7. Articulation of a national position on the interpretation of international law in the context of AI in security and defence

States are currently grappling to deepen their understanding of the application of international law in relation to AI in security and defence. In this context, a State developing a national strategy on AI in security and defence should consider articulating and, where appropriate, publishing its national position on the interpretation and application of relevant international law. To this end, the State can draw lessons from the processes undertaken in the cyber domain, through which a growing number of States have articulated national positions on the application of international law to State use of ICTs.³⁰

A published national position offers several benefits: it provides clarity and predictability for a State's own armed forces and security services; it constitutes a confidence-building

measure; it contributes to the crystallization of customary international law and to the broader multilateral debate in various forums; and it can be revisited and refined as the technology, operational practice and multilateral discussions evolve. The State should further consider whether such a position is best articulated as a standalone statement, as an integral part of the national strategy itself or through statements delivered in relevant multilateral forums. Legal advisors, the armed forces and the appropriate ministries should coordinate closely in the development of the national position.

S.8.8. International law and ethics

International law and ethics are distinct but mutually reinforcing. International law defines what a State can do by way of international, binding obligations that the State must meet. Notably, while interpretations may differ from one State to another, the principles remain clear and uniform, particularly those crystallized in international treaties. Ethics addresses the wider question of what a State and the wider multi-stakeholder community should do. Notably, ethical assessments can diverge greatly across cultures, geographies and contexts by way of varying values and approaches to morality.

Where ethical and legal assessments diverge or where States' ethical views differ, a State developing a national strategy on AI in security and defence should consider addressing how it handles such differences.

28 Security Council, resolution 1540 (2004), 28 April 2004, <https://docs.un.org/S/RES/1540>; Chris Chung, Almutaser Albalawi and James Revill, "Averting a darker biological weapons future", UNIDIR, 25 September 2026, <https://unidir.org/averting-a-darker-biological-weapons-future>.

29 US National Telecommunications and Information Administration (NTIA), *Dual-Use Foundation Models with Widely Available Model Weights* (Washington, DC: NTIA, July 2024), <https://www.ntia.gov/sites/default/files/publications/ntia-ai-open-model-report.pdf>.

30 UNIDIR Security and Technology Programme, *A Compendium of Good Practices: Developing a National Position on the Interpretation of International Law and State Use of ICT* (Geneva: UNIDIR, 2024), <https://unidir.org/publication/a-compedium-of-good-practices-developing-a-national-position-on-the-interpretation-of-international-law-and-state-use-of-ict>.

S.8.9. Mechanisms for oversight and evaluation of legal compliance

A State implementing a national strategy on AI in security and defence should establish oversight mechanisms to ensure that AI technologies for security and defence applications remain consistently compliant with applicable laws. The State should consider and facilitate conducting periodic evaluations and establishing a process to ensure documentation and reporting – whether public or internal – of compliance efforts for each branch of international law.

Oversight and evaluation mechanisms should be designed to function across the full life cycle of an AI system, from the pre-acquisition legal review, through testing and evaluation, to operational use and eventual decommissioning. These should be accompanied by the appro-

priate legal, technical and operational expertise necessary to keep pace with iterative updates to AI systems. The State should consider how oversight bodies interface with the accountability, verification and review mechanisms that may be described anywhere in the national strategy to avoid duplication and ensure that legal compliance findings feed directly into decisions on the continued development, deployment or withdrawal of a given system.

There are transparency and confidence-building expectations associated with such participation in the various on-going intergovernmental and multi-stakeholder processes. If the State participates in any of these processes, its domestic and reporting mechanisms should be capable of generating the information needed to meet such expectations.

S.9. Ethics

Ethics sit at the heart of the approach of many States to the governance of AI in security and defence. Yet, unlike international law, ethics are not binding and so views, interpretations and prioritization may diverge widely across States. In addition, there is no established mechanism for ethical accountability and responsibility. National strategies thus present an opportunity for States to outline their approaches to key ethical issues and how they translate into policy and legal considerations.

S.9.1. Gender bias

Without the appropriate safeguards and risk-mitigation measures in place, AI systems could reinforce gender biases that have severe security and defence implications.³¹ Fairness and inclusivity should be considered as an approach to address the risks that stem from gender bias. Thus, when developing a national strategy on AI in security and defence, a State should consider outlining possible approaches and solutions to addressing these risks. These could include benchmarks and guidelines on the use of diverse and representative data sets, the establishment of algorithmic audits, and corrective measures when biases are detected. Gender perspectives should be considered and integrated, as appropriate, at every stage of the life cycle of an AI technology, from design and conceptualization to operational use.

S.9.2. Racial bias

Left unaddressed, racial bias in AI can have severe legal and ethical repercussions. In the context of both security applications (e.g., law enforcement) and defence (e.g., military targeting), racial biases could lead to the violation of international humanitarian law and international human rights law. A State developing a national strategy on AI in security and defence should thus ensure that AI systems used in these contexts are subject to thorough scrutiny and should review processes to avoid the exacerbation of issues stemming from racial bias. The regular conduct of audits and reviews of the system's performance will be critical, particularly when these systems are used for operations in areas of diverse demography. The State should prioritize the introduction of review and oversight mechanisms, along with a systematic compliance assessment.

31 Katherine Chandler, *Does Military AI Have Gender? Understanding Bias and Promoting Ethical Approaches in Military Applications of AI* (Geneva: UNIDIR, 2021), <https://doi.org/10.37559/GEN/2021/04>.

S.9.3. Other forms of discrimination

The development, deployment and use of AI in security and defence should prevent the perpetuation and exacerbation of any form of discrimination, including those based on disability, age or socioeconomic status. A State developing a national strategy on AI in security and defence should ensure that it explicitly addresses these potential biases through concrete recommendations and solutions, including through robust testing, evaluation and validation processes to prevent discriminatory outcomes downstream. The State should prioritize the incorporation of ethics into the design of AI technologies for security and defence, with input from ethicists, experts and representatives of minorities and other groups most at risk of algorithmic discrimination.

S.9.4. Fairness

Fairness in AI decision-making is critical for maintaining public trust and for governance frameworks that are centred on ethics; this is further accentuated for AI applications in security and defence. A State developing a national strategy on AI in security and defence should ensure that AI systems under development, in procurement and in use are not only transparent, but that considerations around fairness have been embedded into the parameters of the systems' learning, development and testing stages. The State should facilitate a periodic review of societal impacts and fairness evaluation, with multi-stakeholder input as appropriate.

S.9.5. Explainability and traceability

When a State developing a national strategy on AI in security and defence reviews key ethical factors in the development, adoption, implementation and review of a national strategy on AI in security and defence, it should dedicate particular attention to explainability and traceability.

In the light of the black box and inherently probabilistic nature of AI systems, human operators should be able to explain the rationale behind their decision to deploy and use AI systems in specific contexts. To this end, these technologies should have, to the extent possible and appropriate, methods and processes in place for their explainability, that is, for users to comprehend the results and outputs created by the system (see also guideline S.2.2). Traceability – that is, the ability to understand and trace back the processes behind the development of a system (e.g., source of training and testing data, testing and evaluation metrics, the algorithm's key parameters) – will, in this sense, be important to consider (see also guideline S.2.1).

Traceability and explainability would not only be important for risk assessments undertaken by the human operator; they would also allow, as appropriate, human operators to intervene when necessary, particularly in situations where AI-driven decision-making may contravene ethical or legal requirements. The State should thus consider the extent to which explainability and traceability could, or should, be part of procurement processes.

S.10. Defence applications of AI

The integration of AI in defence presents a host of opportunities, from back-end support to increased autonomy in weapon systems and enhanced ISR capabilities.³² At the international level, the governance of defence applications of AI is high on States' agendas to ensure their responsible development, deployment and use, and cognizant of their associated risks, international law obligations and ethical guidelines.

S.10.1. Scope of the defence applications of AI

Defining the scope of AI applications in defence is critical for the development, adoption, implementation and review of a national strategy on AI in security and defence. A clear and defined scope would prevent a fragmented approach across governmental agencies, in addition to ensuring a clear distribution of roles and responsibilities in the defence ecosystem. A State developing a national strategy should consider the development, deployment and use of AI for both offensive and defensive purposes. Furthermore, it should consider both combat and non-combat functions, in the light of their inherently different policy, legal and ethical implications. The State should also consider the boundaries between the "defence" and "security" realms, and which regulatory instruments and bodies have jurisdiction over these applications.

S.10.2. Integration into weapon systems

A State developing a national strategy on AI in security and defence should address both the opportunities and the risks that stem from inte-

gration of AI into weapon systems. The strategy presents an opportunity for the State to outline the key guiding principles that frame its approaches and priorities in the governance of the integration of AI into weapon systems. Rationales for permissible use-cases, limitations and eventual prohibitions should be included in the national strategy and substantiated, notably on the basis of international law and ethical guidelines. Operational guidelines should also be developed with regards to the development, deployment and use of AI-enabled weapon systems, including specific rules of engagement when these systems are in use. Frameworks surrounding their pre-deployment stages, grounded in compliance with international law and ethical guidelines, should also be developed, including with regards to procurement practices and testing and evaluation. The State should also consider timely audits and the establishment of monitoring systems, in addition to recurrent legal reviews to ensure these systems' consistent compliance with international law, particularly international humanitarian law and international human rights law.

³² Afina, *The Global Kaleidoscope of Military AI Governance*.

S.10.3. Integration into non-weapon systems

Integration of AI into defence applications extends beyond weapon systems, including for enhanced ISR capabilities, training and simulation, and logistics support.³³ While these technologies offer significant operational advantages, these systems must be developed, deployed and used within the limits imposed by international law and ethical guidelines.

A State developing a national strategy on AI in security and defence should thus take stock of existing defence applications of AI outside weapon systems. It should subsequently develop a clear governance framework surrounding their development, deployment and use. The establishment of training benchmarks, rigorous testing and evaluation metrics, and good practices for their development and use will be key to the successful implementation of the national strategy.

The State should also consider timely audits and the establishment of a monitoring system, in addition to recurrent legal reviews, to ensure the consistent compliance of these systems with international law, particularly IHL and IHRL.

S.10.4. AI in information operations and cognitive warfare

AI is increasingly integrated into information operations conducted by States in the security and defence domains. This includes the generation and distribution of synthetic media, the conduct of AI-enabled psychological opera-

tions, the use of large language models to produce targeted influence content, and the application of AI to the detection and eventual attribution of adversary information campaigns. A State developing a national strategy on AI in security and defence should consider the legal and ethical parameters governing its use of AI-enabled information operations, including the authorization processes and oversight mechanisms in place, as well as escalation risks and crisis management. It should also consider the specific IHL considerations arising from the use of AI-generated synthetic media in armed conflict and the governance of AI capabilities for defensive information operations, including detection and counter-messaging.

S.10.5. Legal compliance

Ensuring legal compliance in the context of the development, deployment and use of AI in defence will be critical and should lie at the heart of a national strategy on AI in security and defence. When developing such a strategy, a State should clarify its interpretation of international law in the context of AI in defence and should develop a national position that captures and crystallizes its approach.³⁴ Articulating such a position offers concrete benefits specific to the defence context. Among these benefits are giving armed forces clarity and predictability in operational planning and providing a basis for interoperability with allies and partners. Such a national position also constitutes a confidence-building measure specifically in instances where it addresses contested questions.

33 Sarah Grand-Clément, *Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain* (Geneva: UNIDIR, 2023), <https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain>.

34 Analogies can be drawn from good practices in cyber. See UNIDIR Security and Technology Programme, *A Compendium of Good Practices*.

In addition, the State should adopt a comprehensive legal review mechanism to ensure the consistent compliance of AI development, deployment and use with international law, particularly IHL and IHRL. For AI-enabled weapons and means or methods of warfare, this should take the form of a legal review conducted under Article 36 of Additional Protocol I to the Geneva Conventions (or an equivalent domestic process if the State is not party to the Protocol). The reviews should not just be applied prior to a system's adoption but iteratively across its life cycle given that updates to an AI system's training or parameters can materially change its legal risk profile after the point of initial review.³⁵ Such reviews must be conducted periodically throughout the technology's life cycle and the State should ensure that the legal, technical and operational expertise needed to conduct them is adequately resourced and trained as the technology evolves.

S.10.6. Key arms control and disarmament considerations

A national strategy on AI in security and defence should consider the implications of AI applications in defence for key arms control and disarmament frameworks and should outline the State's approach, position and ambitions in this space. These considerations may include the opportunities that AI presents to reinforce existing arms control and disarmament treaties, the associated risks, and the relationship and interplay between relevant international, regional and national frameworks.

35 On iterative legal reviews see Afina, *The Global Prism of Military AI Governance*. This is also an approach that is adopted in the civilian domain. See, for example, European Union, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act), 13 June 2024, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>, Articles 9 and 72.

S.11. Security applications of AI

The integration of AI in security has vast potential to support States both in law enforcement (e.g., crime analysis and investigative support) and in national security functions that may span from ensuring border security, via supporting intelligence analysis, to enhancing the cybersecurity and resilience of critical national infrastructure (e.g., threat detection and prevention). In certain regions, the development, integration and adoption of wider security applications of AI technologies constitute the priority over those in the defence realm.

S.11.1. Scope of the security applications of AI

For many years, States have already been considering the opportunities offered by AI in security, with a number of applications in deployment and use. A State developing a national strategy on AI in security and defence should thus take a comprehensive overview of intended uses, associated risks and safeguards measures as part of the strategy. The State should also consider the extent to which “grey zone” applications at the intersection of security and defence (e.g., deployment of military-grade ISR capabilities to support law enforcement in times of national crisis) would fall within the scope of security. Regular review mechanisms should be integrated into the strategy to allow it to adapt to new and emerging security threats, risks and evolving technologies.

S.11.2. AI integration into surveillance operations

A State developing a national strategy on AI in security and defence, and that develops, adopts, integrates or uses AI in surveillance, should consider applicable international legal frameworks, particularly international human rights law, and ethical guidelines. Oversight frameworks and mechanisms should thus be in

place to ensure that AI-enabled surveillance practices remain in compliance with IHRL (e.g., proportionality and necessity). Good practices include the establishment of an oversight body, the publication of transparency reports on AI use, and the development of accountability frameworks. The State should also develop, establish and implement audit solutions and mechanisms for the periodic review of surveillance data and practices. It should prioritize investment in technological solutions (e.g., privacy-enhancing technologies) and capacity-building (e.g., training of law enforcement and intelligence officers on responsible AI use and applicable legal frameworks).

S.11.3. Biometric identification and facial recognition

Biometric identification and facial recognition have emerged as two of the most contested security applications of AI. A State developing a national strategy on AI in security and defence should consider treating it as a distinct area of the strategy, rather than including it under the general guidance on surveillance. It should apply particular caution to remote biometric identification, biometric categorization, emotion-recognition systems and AI-enabled inference of sensitive attributes, as these tools

could create heightened risks of discrimination, chilling effects and arbitrary inference with privacy. Accordingly, the national strategy should establish a distinct framework for legal review of and thresholds for biometric identification. At a minimum, this should differentiate real-time use from retrospective use, live public space deployment from controlled access-control use, and identification from categorization. It should also apply particular scrutiny, including thorough vendor due diligence, to the procurement of commercially sourced biometric databases and matching services.³⁶

S.11.4. AI in border management and migration

A State developing a national strategy on AI in security and defence should assess the need for AI in border management, migration and asylum procedures, including applications relating to identity verification, document authentication, surveillance, risk assessment, search and rescue, and the allocation of administrative or humanitarian resources. The State should also assess any eventual use of AI in these ways. The strategy should ensure that such applications comply with IHRL, IHL and other applicable legal obligations including non-refoulement and the right to seek asylum.

S.11.5. Cross-border data-sharing practices and AI implications

Increased use of AI in security raises novel issues and exacerbates existing sensitivities associated with cross-border data-sharing practices, particularly with regards to national

sovereignty and privacy. A State developing a national strategy on AI in security and defence should thus consider the establishment of a dedicated framework and bilateral agreements defining the expected implications, parameters, conditions and limitations of AI integration into cross-border data-sharing practices. The framework should include provisions on applicable laws, safeguards and oversight. Furthermore, the State should conduct a survey of possible risks stemming from development, deployment and use of AI in the context of cross-border data-sharing. Risks of unauthorized access to and interception, misuse and compromised integrity of data should, among others, be considered and addressed in the national frameworks and any bilateral agreements.

S.11.6. AI integration into law enforcement mechanisms, practices and operations

The development, integration, adoption and use of AI in law enforcement mechanisms, practices and operations offer a host of opportunities. However, their inherent risks must be addressed through robust oversight and accountability mechanisms. A State developing a national strategy on AI in security and defence should conduct a survey of existing applications, on-going procurement processes and future plans for integration of AI into law enforcement mechanisms and practices, along with a comprehensive assessment of risks and relevant actors. The State should also establish and implement good practices to ensure the responsible development, integration, adoption and use of AI in law enforcement. These practices

36 Office of the United Nations High Commissioner for Human Rights, *Guiding Principles on Business and Human Rights*, Principle 17; General Assembly, Human Rights Council, “Human rights and new and emerging technologies: Human rights due diligence for AI and State obligations”, Report of the United Nations High Commissioner for Human Rights, A/HRC/62/33, 18 June 2026, <https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/sessions-regular/session62/advance-versions/a-hrc-62-33-auv-1-en.pdf>.

include the identification and mitigation of harmful biases, the establishment of oversight, accountability and redress mechanisms, and the undertaking of regular audits and compliance reviews, notably in relation to international human rights law. It should prioritize capacity-building, particularly for law enforcement officers, on the parameters and limitations of the AI-enabled tools in use.³⁷

S.11.7. Oversight and accountability

A State developing a national strategy on AI in security and defence should establish and implement rigorous oversight mechanisms for AI development, deployment and use in the

security sector. It should put in place accountability frameworks and structures to review the development, procurement, deployment and use of AI in security contexts, and to ensure compliance with applicable regulatory frameworks at the national, regional and international levels. The establishment of oversight and accountability bodies should ensure that mechanisms are in place to enable relevant agencies and concerned stakeholders – from both the public and the private sectors – to provide input. The State should conduct risk assessments on a regular basis, and it should integrate their findings into the mandate of the established oversight and accountability bodies.

37 For an example of a responsible AI toolkit in law enforcement, see United Nations Interregional Crime and Justice Research Institute (UNICRI), “The toolkit for responsible artificial intelligence innovation in law enforcement”, n.d., <https://unicri.org/topics/Toolkit-Responsible-AI-for-Law-Enforcement-INTERPOL-and-UNICRI>.

S.12. AI integration into critical national infrastructure

The integration of AI technologies into critical national infrastructure should be considered in the development, adoption, implementation and review of national strategies on AI in security and defence. As States are increasingly adopting and integrating, or at least considering the integration of, AI technologies into critical national infrastructure, it is important to reflect on their security implications and ensure their resilience.

S.12.1. Assessment of past, present and future AI integration into critical national infrastructure

In the development of a national strategy on AI in security and defence, a State should conduct an assessment of past, present and planned (or at least intended) integrations of AI into critical national infrastructure. Such an assessment should include a comprehensive review of the systems and hardware being used, information on their supply chains (e.g., supplier, origin of components, testing and evaluation records, procurement history), relevant actors, processes in place for auditing, maintenance plans, the systems in which they are integrated, envisioned end of life and plans for decommissioning. This would enable the State to formulate an effective strategy for the integration of AI technologies into critical national infrastructure, ensuring the compatibility and interoperability of these systems.

S.12.2. Risk assessment

In its review of past, present and future AI integration into critical national infrastructure, a State developing a national strategy on AI in security and defence should also conduct a

comprehensive risk assessment. This assessment should include consideration of risks from external factors (e.g., adversarial AI attacks), risks introduced from the integration of AI into critical national infrastructure (e.g., lack of interoperability with legacy systems), relevant actors, as well as risks pertaining to the supply chain of AI technologies and subsequent implications for the resilience and integrity of critical national infrastructure.

These risk assessments should be conducted on a regular basis to ensure that the State remains up to date in terms of the introduction of novel risks, including those stemming from subsequent integration of AI into critical national infrastructure.

S.12.3. Sectoral risk-mitigation measures

Once the risk assessment is complete, a State developing a national strategy on AI in security and defence should develop risk-mitigation measures for each sector in which AI could be integrated into critical national infrastructure. These risk-mitigation measures must establish a clear distribution of roles and responsibilities for the different actors named across govern-

mental agencies. The agency leading on risk-mitigation efforts for each sector should then develop a sector-specific plan with concrete policy recommendations, pathways for implementation and oversight mechanisms. A review of such plans must be conducted on a regular basis and in consultation with relevant stakeholders, from both the public and the private sectors, to ensure the continued relevance of the measures in place.

S.12.4. Dependencies and risks

A State developing a national strategy on AI in security and defence should map and assess possible dependencies that may arise from the integration of AI into critical national infrastructure. These could include dependencies on external suppliers for hardware, software, data or maintenance services. Dependencies between different infrastructure sectors should also be considered, particularly when they may transmit or amplify the impact of an AI-related failure or attack. These include, for instance, dependencies on supporting infrastructure, such as energy and connectivity, which are both themselves critical for the continued operation of AI systems.

The State should also consider the extent to which such dependencies may introduce single points of failure and concentrated risk. It should thus develop mitigation measures, including

through the eventual diversification of suppliers as necessary and as appropriate, redundancy in critical functions, and contingency planning for the temporary or eventually permanent loss or disruption of key dependencies.

S.12.5. Interoperability with legacy systems

The integration of AI into critical national infrastructure may require interoperability with legacy systems that, in themselves, were not necessarily designed with AI integration in mind and in their design. A State developing a national strategy on AI in security and defence should assess the compatibility of existing infrastructure and systems with planned AI integration, including the technical, security and operational challenges that may arise from interfacing modern AI systems with older infrastructure. The latter may lack the necessary connectivity, data standards or security architecture to ensure and support effective integration.

The State should thus develop a plan to identify, measure and address identified interoperability gaps. This could include, for instance, the phased modernization of legacy systems (or parts thereof), the development of secure layers or, where modernization may not necessarily be feasible, the adoption of alternative approaches that would not compromise the security and reliability of the underlying infrastructure and wider ecosystem of systems.

S.13. Resilience and preparedness

In the development, adoption, implementation and review of national strategies on AI in security and defence, States should consider – and ensure – resilience and preparedness in emergency situations that stem from the use of these technologies.

S.13.1. Plan for emergency response

A State developing a national strategy on AI in security and defence should consider and identify what possible incidents, shock situations and crises might arise from the use of AI in security and defence. The assessment must be holistic in order to factor in the different facets of such incidents (e.g., scale, actors involved, harms caused, primary and secondary effects, potential geopolitical repercussions, etc.). This exercise must then lead to the careful planning of the appropriate emergency response at all levels.

S.13.2. Establishment of an emergency response team

Once a State developing a national strategy on AI in security and defence has established an emergency response plan, it should also designate an inter-agency emergency response team, with a clear mandate and set responsibilities in case of incidents. Similar to a computer emergency response team (CERT), the mandate of the AI emergency response team will be commensurate with the incident at hand (in terms of, e.g., scale, actors involved, harms caused, primary and secondary effects, potential geopolitical repercussions). The State should consider whether existing emergency response teams (e.g., a CERT) should see their mandate widened to cover AI, or whether there is a need – resources permitting – to establish a new entity.

S.13.3. Distribution of roles and responsibilities

A national strategy on AI in security and defence should include a clear distribution of roles and responsibilities in case of incidents. This distribution should be done carefully, considering the respective authority, mandate, and available human and financial resources of each governmental body and agency.

S.13.4. Capacity-building

A State developing a national strategy on AI in security and defence should prioritize capacity-building efforts to ensure that all actors involved in ensuring national resilience and preparedness have all the knowledge, guidance and resources to implement their respective roles and responsibilities. Ensuring the capacity of actors is critical both in response to incidents and for ensuring the day-to-day resilience of the State. This is especially the case if the State is considering, or is already adopting, AI technologies with security and defence implications, including in the military domain, in critical national infrastructure and for national security purposes (e.g., law enforcement and border security).

S.13.5. Risk assessment and risk mitigation

A State developing a national strategy on AI in security and defence should include the formulation of a risks-classification framework in its emergency response plan. It should also conduct a comprehensive risk assessment and identify mitigation measures. The latter should include the roles and responsibilities of both public and private actors, key intervention points, and the resources required and to be mobilized by each actor. The State should conduct and review risk assessments and risk-mitigation plans on a regular basis to ensure that they are up to date. These assessments should include the perspectives, needs and realities of all relevant stakeholders, from both the public and the private sectors.

S.13.6. Incident recording, notification and analysis

A State implementing a national strategy on AI in security and defence should establish a system to record incidents that arise from the use of AI in security and defence. This should include key facts, actors, reactions and processes in place, and responses. Such a system would enable the relevant authorities to identify lessons and foster preparedness and resilience against future incidents by, for example, adapting existing protocols, doctrines and standard operating procedures.

As appropriate, the State should also consider specific mechanisms and processes for incident notification, both domestically with the appropriate national agencies and internationally with concerned States, organizations and stakeholders. The latter should be accompanied by a clear external communication strategy, particularly mindful of eventual sensitivities that may arise. The State should allocate the resources required to ensure the necessary

processes, capacity and infrastructure are in place for implementation.

S.13.7. The conduct of exercises and stress-testing

A State implementing a national strategy on AI in security and defence should consider the systematic conduct of exercises on a regular basis to test the existing measures, protocols, doctrines and SOPs in place for incident response. These exercises would enable the State to stress-test the continued effectiveness of its measures and adapt them accordingly, should the need arise. This would keep processes up to date, as well as maintaining consistent preparedness against new and emerging risks.

Such exercises include red-teaming and adversarial testing. The State should consider whether such exercises should be facilitated by internal or external parties, considering both the value of independent reviews and the sensitivities involved.

S.13.8. Containment and behavioural safety

In addition to the risk of deliberate, external interference, a State implementing a national strategy on AI in security and defence should consider the risk that an AI system – particularly increasingly capable and autonomous systems – may act in ways that deviate from its intended parameters, including during testing and evaluation. Frontier AI systems may circumvent the technical controls of their testing environments, and some AI models may resist attempts to halt or constrain their operation when pursuing an assigned objective. Containment must thus be proactively designed and engineered – and will require its own sets of testing and verification.

The State should consider ways through which an AI system intended for security and defence applications could be subject to rigorous containment testing prior to its deployment. Furthermore, the State should consider the establishment and implementation of technical and procedural safeguards, including real-time monitoring and clear authority and SOPs and mechanisms to interrupt or deactivate a system's operation as necessary, commensurate with the possible consequences that may arise from such actions.

S.13.9. AI incident classification framework

A State implementing a national strategy on AI in security and defence should consider the development of a national framework for the classification and reporting of AI-related incidents in security and defence. Such a framework could, for instance, provide a tiered taxonomy of incidents based on their severity, the nature and scale of harm caused or risked, the systems and actors involved, and the potential for cross-border implications at the regional and international levels. In developing this framework, the State may consider drawing on analogous models particularly from the cybersecurity field, including incident-classification schemes used by CERTs and national cybersecurity agencies, while remaining attentive to the specific characteristics of AI systems that may require tailored approaches.

Some of the elements this framework would address could include the thresholds that trigger mandatory reporting within government and, where appropriate, to partners; the actors responsible for incident classification, investigation and reporting; the interface between the AI incident-classification framework and other relevant accountability, legal review and IHL compliance processes; and the mechanism by

which lessons drawn from incident recording and analysis are fed back into strategy review and the updating of operational guidelines and SOPs.

S.13.10. Dedicated resources for resilience and preparedness

In addition to the redirection of resources allocated to other priorities, dedicated financial, human and technological resources should be allocated specifically to resilience and preparedness functions relating to AI in security and defence. In particular, a State implementing a national strategy on AI in security and defence should consider the establishment of a dedicated budget line or resource-allocation mechanism for resilience and preparedness, including for the maintenance of emergency response capacity, the conduct of regular exercises and stress-testing, and investment in the technological solutions required for incident detection, response and recovery.

S.13.11. Strategy for the maintenance and security of facilities

A State implementing a national strategy on AI in security and defence should develop a dedicated strategy for the maintenance and security of the physical facilities that support the development, testing, deployment and operation of AI systems in security and defence contexts. These facilities include data centres, testing and evaluation facilities, and other critical installations. This strategy should address physical security measures commensurate with the sensitivity and criticality of the facility in question, on-going maintenance requirements to ensure continued operational integrity, and contingency plans in the event of facility disruption, whether arising from natural hazards, deliberate attack or infrastructure

failure. The State should also consider how it could ensure resilience and preparedness against risks of disruption in these facilities – and plan for scenarios in which the availability of this infrastructure is limited or compromised.

S.13.12. Preparedness for long-term technological progress and evolutions

In the light of the rapid pace of technological progress in AI, a State implementing a national strategy on AI in security and defence should ensure that its resilience and preparedness

planning are not calibrated exclusively to present-day capabilities and risks; they should also anticipate the needs and implications arising from continued technological advancement. The State should thus consider the establishment of a mechanism for regular horizon-scanning and foresight over emerging AI capabilities and associated risks. It should also ensure that its resilience and preparedness frameworks are reviewed and updated on a regular basis as appropriate and as necessary.

Conclusion and next steps in the project

Although the international community is still in the early stages of developing, adopting and implementing national strategies on AI in security and defence, there is nevertheless widespread acknowledgment of the importance of such strategies for the governance of the development, deployment and use of these technologies. To this end, States must adopt a holistic approach to the development, adoption, implementation and review of these national strategies.

The above guidelines for the development, adoption, implementation and review of national strategies on AI in security and defence are offered by UNIDIR to support States' efforts in this space. The 147 guidelines, both procedural and substantive, do not seek to be prescriptive, but rather to provide States with a series of considerations, approaches and tools to support their respective national processes.

In 2027, UNIDIR will work towards integrating these guidelines into its AI, Peace and Security Portal in a manner that is user-friendly and interactive, with the hope of feeding the guidelines into a section of the portal dedicated to a policy-building tool. Ultimately, it is hoped that States will find these guidelines useful in supporting the development, adoption, implementation and review of their national strategies on AI in security and defence by giving access, in one place, to the main considerations and good practices that may be of relevance to their efforts.



Palais des Nations
1211 Geneva, Switzerland

© 2026, UNIDIR

UNIDIR.ORG