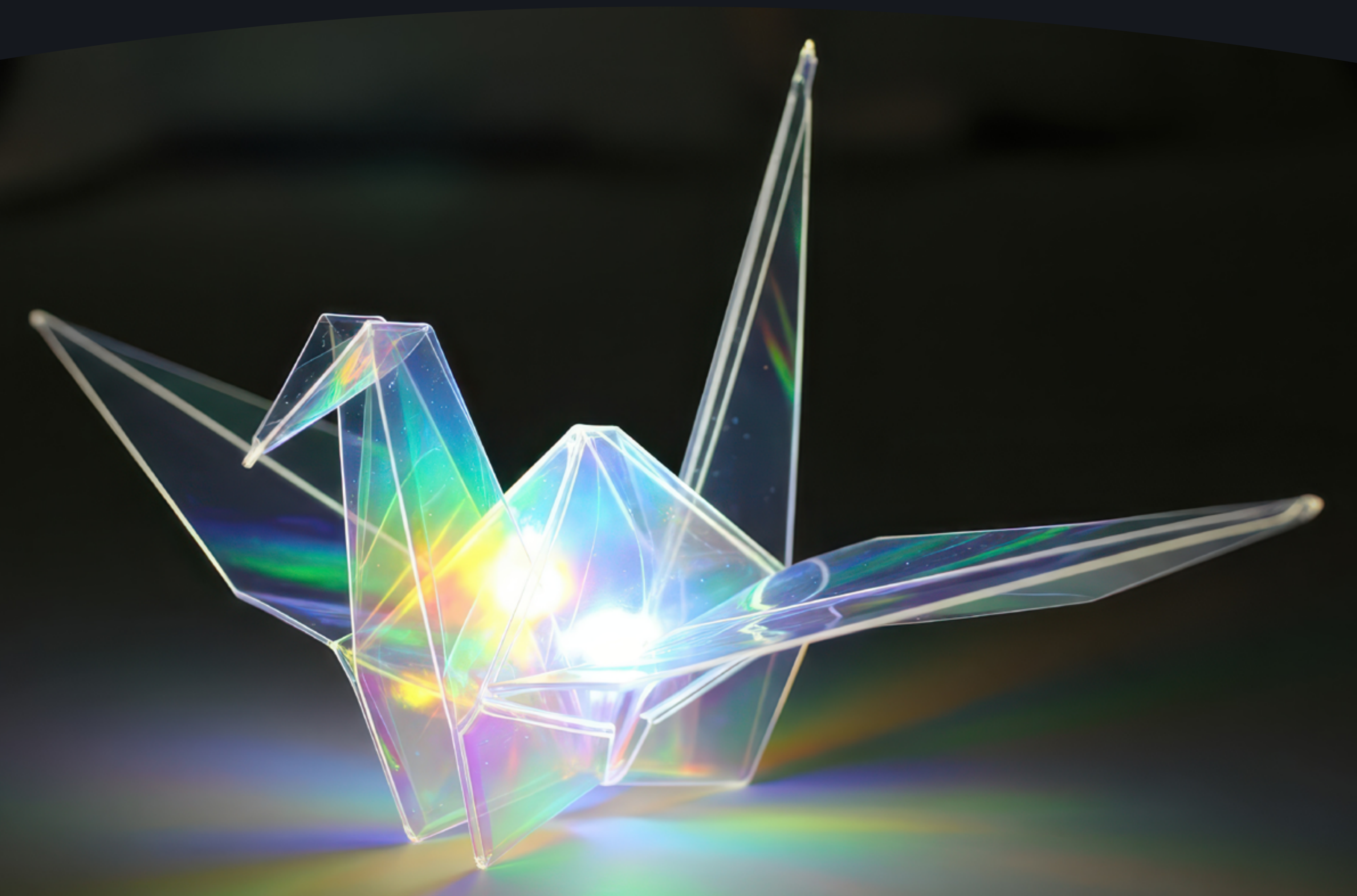
SUMMARY REPORT

The Global Prism of Military AI Governance

Reflections from the 2025 Regional Consultations on Responsible AI in the Military Domain

YASMIN AFINA



Acknowledgements

Support from UNIDIR's core funders provides the foundation for all of the Institute's activities. This study was funded by the Republic of Korea and Spain and implemented by UNIDIR's Security and Technology Programme (SecTec), which is supported by Czechia, France, Germany, Italy, the Kingdom of the Netherlands, Norway, Switzerland and Microsoft.

The author wishes to thank Dr. Giacomo Persi Paoli (UNIDIR) for advice, guidance and support for this programme of work and for his review of this report. Thanks for their support throughout the conduct of the regional consultations, their reviews and suggestions are also due to Dominique Steinbrecher, Edward Madziwa, Gosia Loj, Jessica Espinosa Azcárraga and Jessica Lee Abowitz (UNIDIR), as well as Husham Ahmed, Michael Karimian and one reviewer who chose to remain anonymous. The author also wishes to thank Spain, the Republic of Korea and the Kingdom of the Netherlands for convening the regional consultations, along with France, Kenya and Pakistan, which acted as hosts.

About UNIDIR

The United Nations Institute for Disarmament Research (UNIDIR) is a voluntarily funded, autonomous institute 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 to develop the practical, innovative ideas needed to find solutions to critical security problems.

About the Security and Technology Programme

Contemporary developments in science and technology present new opportunities as well as challenges to international security and disarmament. UNIDIR's Security and Technology Programme seeks to build knowledge and awareness on the international security implications and risks of specific technological innovations and convenes stakeholders to explore ideas and develop new thinking on ways to address them.

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 necessary reflect the views or opinions of the United Nations, UNIDIR, its staff members or sponsors.

Citation

Yasmin Afina, "The Global Prism of Military AI Governance: Reflections from the 2025 Regional Consultations on Responsible AI in the Military Domain", UNIDIR: Geneva, 2026.

About the author

This report was produced by UNIDIR's Security and Technology Programme. It was drafted by Yasmin Afina, who attended all five regional consultations.

Acronyms & Abbreviations

AGI	Artificial general intelligence
AI	Artificial intelligence
ASEAN	Association of Southeast Asian Nations
CBM	Confidence-building measure
DSS	Decision-support system
ICC	International Criminal Court
IHL	International humanitarian law
IHRL	International human rights law
ISR	Intelligence, surveillance and reconnaissance
LAWS	Lethal autonomous weapon systems
NC3	Nuclear command, control and communications
NSAG	Non-state armed group
R&D	Research and development
RAISE	Roundtable for AI, Security and Ethics
REAIM	Responsible AI in the Military Domain
TPNW	Treaty on the Prohibition of Nuclear Weapons
UX	User experience

Table of contents

Executive summary

In the run-up to the third Summit on Responsible Artificial Intelligence in the Military Domain (REAIM), to be held in A Coruña, Spain, on 4–5 February 2026, the Governments of Spain, the Republic of Korea and the Kingdom of the Netherlands, in partnership with France, Kenya and Pakistan, conducted a series of five regional consultations on artificial intelligence (AI) in the military domain. Facilitated by UNIDIR, the consultations sought to build on the 2024 REAIM Regional Consultations and the 2023 and 2024 Summits, in addition to capturing evolutions in national views and policies on responsible AI in the military domain, regional priorities and multi-stakeholder engagement over the year.

The present report seeks to capture the main takeaways from the five regional consultations, summarizing participants' views along with some of UNIDIR's observations. Specifically, these observations are centred around the following themes, which constituted common threads across all regional consultations (while minor adjustments were made for each regional event to factor in its respective local context and realities):

- **National policies and practices**

Participants were invited to reflect on their respective key priority areas at the national level for responsible AI in the military domain, as well as the opportunities and challenges perceived. This segment of the consultations also provided an opportunity for states to share existing national strategy documents and other policy frameworks of relevance to responsible AI in the military domain, along with national good practices.

- **Looking back – post-REAIM 2023 and 2024 reflections**

This segment of the regional consultations invited states to reflect on the REAIM initiative since the inaugural summit, in February 2023. Beyond participation and endorsement, this segment was also aimed to take stock of the possible impact that the REAIM initiative may have had within regions.

- **Looking ahead – reflections for the 2026 REAIM Summit**

Building on past experiences, existing policies and aspirations, participants were provided with an opportunity to reflect on the journey ahead for REAIM in the context of the third summit and beyond, from both substantive and procedural perspectives.

Through this overview of national views, policies and good practices, the consultations reaffirmed, as in 2024, that there is no single pathway to responsible AI governance in the military domain. Instead, regional perspectives continue to be shaped by distinct security environments, legal traditions, threat perceptions and levels of technological maturity. At the same time, the 2025 consultations revealed clearer patterns of convergence around a core set of concerns, including compliance with international law, risk reduction and management, the centrality of industry engagement, as well as the need to translate high-level principles into operational practice.

Across regions, states reported tangible progress since 2024, notably through the development, review or implementation of national AI strategies and defence-specific frameworks, as well as the identification of concrete good practices in areas such as procurement, legal review, institutional coordination and capacity-building. These developments attest to the catalytic effect of sustained regional dialogue, which the regional consultations have enabled, while also highlighting persistent asymmetries in capacity and perceived readiness.

At the same time, the consultations underscored that the risk landscape has become more complex. Participants consistently raised concerns related to the proliferation of AI-enabled capabilities to non-state armed groups; escalation dynamics, including in contexts involving nuclear-armed states; the convergence of AI with cyber and other technological fields; as well as challenges associated with the dual-use nature of these technologies and the security of supply chains.

In addition, this report provides a comprehensive overview of some of the takeaways from the discussions held with the multi-stakeholder community. One key objective of the consultations, in acknowledgment of the importance of multi-stakeholder engagement, is to take stock of the views of regional representatives from industry, civil society, academia and research institutes, as well as regional and international organizations. Altogether, their contributions enabled UNIDIR to take stock of diverse perspectives on the current state of affairs in each region and their respective contexts, in addition to thematic deep dives on both substantive and governance areas.

Complementing these discussions, a scenario-based multi-stakeholder tabletop exercise provided an opportunity to examine how responsible AI principles may be operationalized across the life cycle of AI-enabled military capabilities, from the design and development stages via procurement and use to after-action review. The exercise reinforced the importance of design choices, user interfaces, guidelines for responsible procurement and documentation practices in shaping accountability and compliance in practice, among other things. Furthermore, the exercise enabled the collection of data around specific assurance measures, which subsequently provided a visualization of priority areas, patterns and divergences for the operationalization of responsible AI principles in the military domain.

Finally, the consultations provided an opportunity to reflect on the REAIM journey, three years since the inaugural summit, in February 2023. States were provided with an opportunity to reflect and look back, identifying both the success factors and limitations that REAIM as an initiative has encountered since its inception. Participants were also provided with an opportunity to reflect and look ahead, identifying substantive areas of priority that they are keen to see further pursued within REAIM and beyond, along with formulating a series of concrete recommendations for the way ahead.

Taken together, the 2025 regional consultations indicate a gradual and natural evolution from exploration and stocktaking to operationalization and implementation. Some of the key takeaways from the consultations include:

From principles to operationalization

A central conclusion emerging from the consultations is that efforts towards responsible AI governance in the military domain must increasingly focus and rest on implementation. As such, the effectiveness and success of responsible AI governance in the military domain will increasingly depend on the availability of practical tools, guidance, frameworks and processes that support operationalization and coherent implementation across the full life cycle of military AI.

Structured engagement with industry

While states generally acknowledged the important role of the private sector in 2024, one year later, participants across regions recognized that industry actors play a decisive role in shaping AI-enabled military capabilities. This occurs through system design choices, data practices, testing and evaluation, as well as post-deployment maintenance and support. A more structured engagement with industry is increasingly seen, not as an optional add-on to international deliberations, but as a necessary condition for translating responsible AI principles and commitments as a critical pathway to action.

Capacity and trust-building

Echoing the 2024 regional consultations, discrepancies in states' capacity emerged as a structural feature of the global AI governance landscape. Participants emphasized that the digital divide – reflected through uneven access to technical expertise, data, infrastructure and institutional resources – continues to shape both risk perceptions and governance options. Additionally, trust-building was repeatedly identified as a necessary complement to capacity-building. Participants noted the criticality of structured, neutral and independent channels for engagement with states and industry, without which mistrust, uncertainty and misaligned expectations could be exacerbated.

Strategy for REAIM's next phase

Looking ahead, participants consistently emphasized the importance of continuity, coherence and complementarity across efforts, initiatives and processes. REAIM's sustainability was seen as contingent on its ability to evolve alongside the technology and the governance landscape, including by providing space to address implementation challenges that cut across the public and private sectors, but also without duplicating or competing with United Nations-based processes; many states expressed a preference for the latter as a forum for inter-state deliberations over initiatives that sit outside the organization. While there have been discussions as to how future REAIM efforts could further expand their substantive scope and depth, a significant proportion of the conversations focused on enhancing outreach to increase inclusivity and equity, in addition to deepening and sustaining impact. This includes supporting implementation through the formulation of a clear strategy for REAIM's future pathway, guidance and shared good practices; outlining and strengthening linkages with relevant United Nations, regional and sectoral initiatives; and preserving institutional memory amid an increasingly complex agenda.

1. Background

Since its inaugural edition, the Responsible Artificial Intelligence in the Military Domain (REAIM) Summit has played a significant role in promoting mutual understanding and fostering dialogue among diverse stakeholders on the responsible application of artificial intelligence (AI) in the military domain. While the first summit, held in The Hague in February 2023, successfully put the topic high on the political agenda of many states, its two co-hosts – the Kingdom of the Netherlands and the Republic of Korea – have sought to build on this momentum and deepen their engagement globally. In partnership with regional partners, five regional consultations were held in the first half of 2024, in Singapore, Istanbul, Nairobi, Santiago and online.[1]

The 2024 REAIM Regional Consultations provided a unique platform to capture regional perspectives and to identify areas of convergence and divergence in states' perceptions of the responsible development, deployment and use of AI in the military domain. The Republic of Korea, as host of the second REAIM Summit, in September 2024, crystallized these perspectives in the summit's outcome document: the Blueprint for Action. The latter was specifically aimed at identifying the key principles that underpin responsible AI in the military domain, including compliance with international law, other relevant legal frameworks and regional instruments, and with ethics, as well as the need for safeguards to reduce risks of malfunction or unintended consequences.

Against this backdrop, and in acknowledgment of the tremendous value of these regional platforms for dialogue, a second series of regional consultations was held in the run-up to the third edition of the REAIM Summit, to be held in A Coruña, Spain, on 4–5 February 2026. In partnership with regional partners, five regional consultations were held by Spain, the Republic of Korea and the Netherlands over the course of 2025:

- Asia-Pacific: Seoul, Republic of Korea, 20–21 May 2025 (hosted by the Republic of Korea)
- Europe and North America: Geneva, Switzerland and online, 5 June 2025 (co-hosted by France)[2]
- West Asia and the Middle East: Islamabad, Pakistan, 17–18 June 2025 (co-hosted by Pakistan)
- Africa: Nairobi, Kenya, 27–28 August 2025 (co-hosted by Kenya)
- Latin America and the Caribbean: New York, United States of America, and online, 8–9 December 2025 (hosted by Spain, the Republic of Korea and the Netherlands)[3]

[1] For a summary report of the 2024 REAIM Regional Consultations, see Y. 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/.

[2] While Canada and the United States of America joined the European and North American regional consultation, Mexico joined the Latin American and Caribbean event.

[3] Following practice from the consultation in 2024 in the region, the 2025 Latin American and Caribbean regional consultation extended beyond the military to include the broader security domain (including law enforcement, border security, as well as efforts to counter transnational organized crime) in recognition of the region's security landscape and realities.

Facilitated by UNIDIR in its capacity as knowledge partner, the consultations comprised a series of plenary and breakout group discussions. These sessions were designed to build on the perspectives shared in the 2024 regional consultations, in addition to the outcomes of the 2024 REAIM Summit and other initiatives. Specifically, they were organized around the following core themes and guiding questions, with minor adjustments made for each regional consultation to factor in its local context and realities:

National policies and practices

Participants were invited to reflect on their key priority areas at the national level for responsible AI in the military domain, as well as the opportunities and challenges perceived. In addition, states shared views on the ways through which AI is being integrated (if at all) into their respective military systems, building on reflections around what constitutes and defines the "military domain". This segment of the consultations also provided an opportunity for states to share existing national strategy documents and other policy frameworks of relevance to responsible AI in the military domain, along with national good practices to enable the responsible development, deployment and use of these technologies.

Looking back – post-REAIM 2023 and 2024 reflections

This segment of the regional consultations invited states to reflect on the REAIM initiative since the inaugural summit, in February 2023. In addition to taking stock of endorsements (or the lack thereof) of both summits' outcome documents, the participants reflected on the grounds for their support for endorsement or, conversely, the procedural and substantive obstacles and limitations standing in the way of such endorsement. Beyond participation and the endorsement, this segment also aimed to take stock of the possible impact that the REAIM initiative may have had within regions, factoring in not only the summits and their outcome documents but also the 2024 regional consultations.

Looking ahead – reflections for the 2026 REAIM Summit

Building on past experiences, existing policies and aspirations, participants were provided with an opportunity to reflect on the journey ahead for REAIM in the context of the third summit and beyond, from both substantive and procedural perspectives. Substantively, participants shared views on the themes and areas they would like to see reflected and prioritized in the programme of the third REAIM Summit and its outcome document, grounding these reflections on national relevance and the regional context. Additionally, participants were asked to share their hopes for the impact of the summit and its outcomes at the national, regional and international levels. Procedurally, participants were invited to reflect on good practices from the 2023 and 2024 REAIM Summits and their adjacent activities and, conversely, areas where there is room for innovation and change. At the more strategic level, participants were also invited to reflect on the direction that the REAIM initiative should take and how it should interact with other ongoing processes and efforts, taking into account in particular the deliberations emerging within the United Nations and the need for effective complementarity.

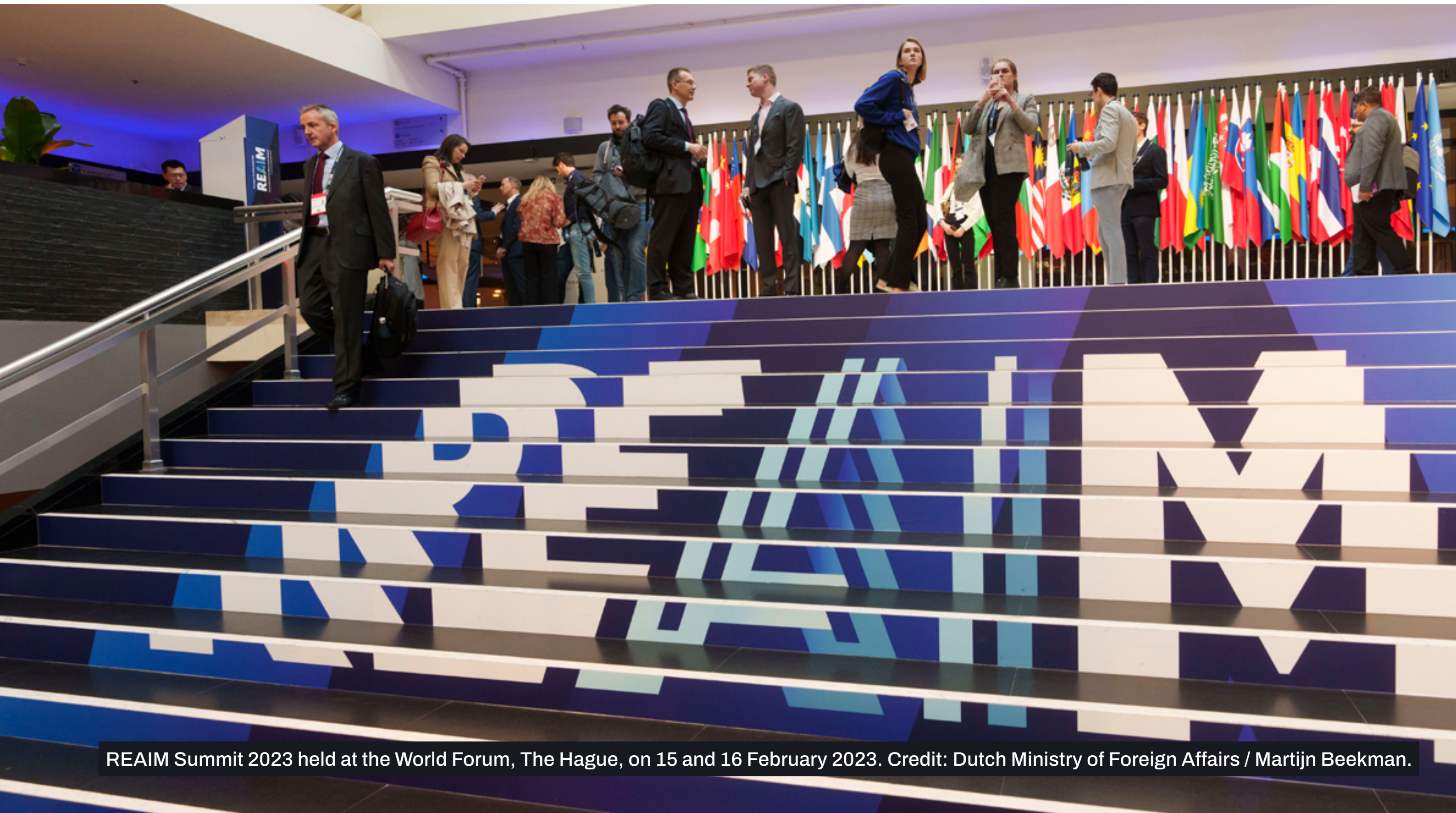
In addition, UNIDIR facilitated a tabletop exercise at each consultation, following a direct recommendation from the 2024 REAIM Regional Consultations.[4] Framed by a fictional scenario,

[4] Afina, *The Global Kaleidoscope of Military AI Governance*.

the exercise specifically provided an opportunity for participants to reflect on concrete ways to operationalize some of the principles of responsible AI captured in the contexts of procurement, use and incident response. In further acknowledgment of the importance of engaging with the multi-stakeholder community, representatives from industry, academia and civil society organizations as well as international and regional organizations were invited to participate in the tabletop exercise.

The consultations held in Seoul, Islamabad, Kenya and New York were each followed by a one-day round table led by UNIDIR for the region's own multi-stakeholder community, open to state representatives wishing to attend. The round table discussion provided an opportunity to reflect on a host of questions, including to take stock of perspectives on the current state of affairs, areas of policy priority and recommendations, in addition to conducting thematic deep dives on governance and on select substantive issues tailored to each region. The consultation in Geneva had specific segments open to the multi-stakeholder community during the day, providing dedicated spaces for these reflections.

The present report seeks to summarize UNIDIR's key findings and the main takeaways from participants in the regional consultations. In addition to taking stock of national views, policies, good practices and multi-stakeholder perspectives, the report also presents concrete perspectives and quantitative data to highlight patterns and convergences in the participants' approaches to the operationalization of responsible AI in the military domain. These approaches can be reflected through their reflections around select procurement parameters and assurance requirements. The report also highlights reflections and pathways for incident response and risk management. Additionally, it captures states' reflections when looking back at the REAIM journey to date, along with aspirations and recommendations for its future direction. These perspectives all culminate in UNIDIR's reflections, along with the formulation of a series of recommendations for the international community's future work on responsible AI in the military domain, both as part of the REAIM initiative and beyond.



REAIM Summit 2023 held at the World Forum, The Hague, on 15 and 16 February 2023. Credit: Dutch Ministry of Foreign Affairs / Martijn Beekman.

2. Taking stock of national views, policies and good practices

The first thematic segment of the regional consultations provided an opportunity for states to exchange insights on national views, policies and practices on responsible AI in the military domain. Specifically, participants were provided with an opportunity to share their views on perceived opportunities and risks associated with the development, deployment and use of AI in the military domain. Additionally, this segment of the consultations provided a snapshot of good practices and frameworks that are either in place or emerging at both the national and the regional levels. Generally, there has been significant progress and uptake in national and regional initiatives in this space since the convening of the 2024 regional consultations, marking states' growing appetite for and perceived readiness to engage and deliberate on the development, deployment and use of these technologies in the military domain.

BOX 1.

What does the "military domain" mean?

As states and the wider multi-stakeholder community deliberated on the responsible development, deployment and use of AI in the military domain, one key question has emerged: What does the "military domain" mean and correspond to? Participants noted that militaries have different mandates from one state to another, which arises not only from varied historical, political and socioeconomical contexts, but also from distinct national and regional security realities. In Latin America and the Caribbean, a number of participants noted the absence of inter-state conflict within the region; the armed forces are thus mandated primarily with preserving national security and resilience in peacetime through non-combat functions that include, for instance, building critical national infrastructure such as airports.

Ambiguous applications and cases were also brought up, such as the routine use, by the military, of widely available civilian software and services (e.g., cloud services, AI-powered office productivity tools) and the extent to which these uses would also fall within the scope of the present discussions. Some discussions also pointed to the relevance of adjacent activities that may not necessarily constitute direct military use of AI, but which nonetheless support military objectives or capabilities. These include trade-related activities that contribute to the financing or advancement of military and dual-use capabilities.

A number of participants suggested that existing approaches could offer useful reference points, including for example the jurisprudence of the International Criminal Court (ICC), which relies on functional analysis and concepts of liability, rather than formal labels. However, this constitutes only one approach among many others that may be applied for the many facets of the wider discussions around responsible AI in the military domain. As such, this report gives "AI in the military domain" the widest possible scope, acknowledging the nuances present in national and regional approaches.

2.1. Perspectives on opportunities and risks

As part of the exchange on national views, policies and practices, states highlighted a suite of possible opportunities and risks associated with the development, deployment and use of AI in the military domain. Noting that a deep dive on both issues was conducted in the 2024 regional consultations, this exchange provided an opportunity to reinforce, complement and adapt some of the views shared at the earlier consultations.[5]

2.1.1. Perceived opportunities

The following broad opportunities were presented by states during the regional consultations:

- Support for combat functions
- Legal compliance
- Management, administration and logistical functions
- Peacekeeping support
- Operational efficiency, performance and speed
- Data analysis, intelligence and forecasting
- National security and resilience (non-combat functions)

While most of the opportunities laid out by states resonate with those presented at the 2024 regional consultations, the scope and depth of some of these elements generally grew, while there were also further nuances in the participants' remarks. This observation attests to a generally increased level of thinking and understanding by states of the basic, practical details of responsible AI in the military domain. This may be correlated to the observed surge in national efforts and initiatives in developing a policy and strategy on this topic.

BOX 2.

Nuanced approaches to defining "opportunities" in the context of responsible AI in the military domain

As the opportunities presented above are unpacked in Subsection 2.1.1, it is important to note that, while many of these overlap across and within regions, a certain degree of misalignment emerged from a number of states, particularly from Latin America and the Caribbean, around the framing of "opportunities" in the context of "responsible" AI in the military domain. While they would generally welcome opportunities in the contexts of logistics, personnel management, recruitment, humanitarian support, civilian protection and the increased effectiveness of peacekeeping operations, concerns were voiced around opportunities to increase lethality and in relation to the use of force more generally. It was argued by a number of states that framing increased lethality as an opportunity stands in opposition to the notion of responsibility and, more generally, the objective of promoting international peace and security.

[5] Afina, *The Global Kaleidoscope of Military AI Governance*.

Specifically, states presented the following opportunities and applications tied to the responsible development, deployment and use of AI in the military domain. While many of these perceived opportunities overlap across regions, they are not all necessarily shared universally.

Support for combat function

Participants highlighted a range of applications through which AI could support combat functions. These include, for instance, AI-enabled autonomy in vehicles, AI integration into weapon systems and the integration of AI into defensive systems (e.g., counter-drone). These defensive capabilities have been highlighted as being of particular importance, considering the progress made in other technological fields (e.g., missile technologies and increasingly sophisticated cyber offensive capabilities) and the subsequent need to protect not only military assets, but also critical national infrastructure, which falls within the mandate of a number of armed forces.

Legal compliance

A number of participants also highlighted the potential for AI-enabled capabilities to contribute to improved compliance with international law, in particular international humanitarian law (IHL). In principle, the analysis of data points at scale and at speed could, for instance, support proportionality analyses. Another example raised by participants pertains to computer vision programmes, which could, in principle, support the implementation of the rule of distinction in targeting operations. However, there is the perception that these opportunities are contingent on the reliability of these systems in facilitating compliance – thus extending beyond technical and operational reliability.

Management, administration and logistical functions

Beyond operational uses, AI was noted for its applications in a range of management and support functions, including logistics, human resources management, weapon system management, strategic communications and information management. These applications were presented as contributing to improved organizational efficiency and coordination, as well as adding to the efficiency of the distribution and management of human, financial and technological resources.

Peacekeeping support

Participants pointed to the relevance of AI applications in the context of peacekeeping operations. For example, they have potential to help identify early signs of violence, including against peacekeeping troops, in addition to improving the overall efficiency of missions.

Operational efficiency, performance and speed

AI applications were identified as offering a host of operational opportunities. A number of participants particularly highlighted the potential for AI to support faster and more facilitated decision-making, subsequently enhancing battlefield command by increasing situational awareness and improving overall operational efficiency at scale and at speed. Beyond the battlefield, decision-support systems (DSSs) were also seen as potentially enhancing decision-making in the design and planning of a mission and in subsequent implementation as well as risks assessments. The latter would not only include risks of collateral damage and of legal breaches, but also risks to the armed forces' own personnel deployed in the battlefield.

Data analysis, intelligence and forecasting: AI-enabled systems allow for faster and potentially more accurate analysis of large volumes of data. As such, not only would this capability enhance intelligence, surveillance and reconnaissance (ISR), thus increasing situational awareness, they would also support the ability to forecast threats. The latter would enable the anticipation of (and subsequent planning against) growing threats to national, regional and international peace and security. This application is of particular relevance for armed forces mandated with supporting disaster-relief efforts – a national security issue of greater concern for states most exposed to climate insecurity.

National security and resilience (non-combat functions)

For many states, particularly in Latin America and the Caribbean, the armed forces are also, if not primarily, tasked with protecting national sovereignty and resilience, which translates into non-combat responsibilities. AI could offer a range of opportunities to support such a mandate. This notably includes, for example, bolstering cyber defences for the resilience of military and critical infrastructure. It also holds the potential to support natural disaster relief (e.g., through enhanced data processing, logistics and management, and predictive analysis for planning), support efforts in national emergency contexts (e.g., pandemics), as well as the construction, maintenance and protection of critical national infrastructure. Finally, AI has also been presented as adding to national resilience by helping address the challenges associated with ageing societies, particularly to compensate for workforce needs in the armed forces.

2.1.2.Perceived risks

Conversely, participants also noted the following broad risks associated with the development, deployment and use of AI in the military domain:

- Operational, security and military risks
- Nuclear-AI risks, escalation dynamics and stability
- AI system characteristics and technological risks
- Societal, ethical and environmental risks
- Proliferation to non-state armed groups (NSAGs)
- Critical national infrastructure and national resilience
- Legal compliance
- Limited capacity
- Cyber, data and information-related risks

As with the above opportunities, many of the risks outlined during the 2025 REAIM Regional Consultations aligned with those laid out in the 2024 consultations, although many of these perceived risks have either evolved or gained in depth and understanding. While many of the perceived risks were shared across regions, states' approaches and perceptions vary and reveal nuances due to inherent differences in regional and national security contexts. Furthermore, while there is a shared understanding across regions that reflections on risks are necessary for the governance of AI in the military domain, their prioritization varies. For example, many states in Latin America and the Caribbean underscored the importance of primarily focusing international discussions on risks; a number of states in West Asia and the Middle East specifically expressed deep concern around the AI–nuclear nexus and its implications for global and regional stability; while a major focus on risks of proliferation into the hands of NSAGs was observed among African states.

Operational, security and military risks

Participants identified a range of operational and security risks, particularly those stemming from the possession and use of AI technologies by NSAGs and subsequent added threats to troops and to military and national infrastructure. The need for up-to-date rules of engagement and directives was also emphasized; without these, the development, deployment and use of AI could present operational and legal risks. Clear processes to address risks of automation bias and subsequent operational risks are particularly needed when decision-making relies fully or, at least, partly on AI systems.

Nuclear-AI risks, escalation dynamics and stability

Particular attention was drawn by a number of states to the nexus between AI and nuclear issues and its implications particularly for regional and global stability. Particular concerns were expressed regarding the integration of AI into nuclear command, control and communications (NC3) systems. These concerns would be further exacerbated by the use of AI-enabled technologies in conflicts involving nuclear-armed states, which may have implications for neighbouring states. Against a backdrop of regional tensions and a reported deficit of mutual trust noted in South Asia, concerns were raised regarding the implications that AI may have in accelerating conventional escalation and, subsequently, leading to the potential use of nuclear weapons, in addition to contributing to regional instabilities – all underscoring the importance of confidence-building measures (CBMs). Furthermore, a number of participants from Latin America and the Caribbean noted the risks but also the sensitivities of nuclear–AI discussions, given the region's historical advocacy for the prohibition of nuclear weapons, notably through the establishment of a nuclear weapon-free zone (NWFZ) and the 2017 Treaty on the Prohibition of Nuclear Weapons (TPNW).

AI system characteristics and technological risks

Participants highlighted risks stemming from certain characteristics of AI systems, including the difficulty of understanding outputs due to opaque system architectures and the subsequent implications that this black box may have for decision-making and the ability to conduct effective investigations. Design flaws in an AI system were identified as potentially leading to deviations from its intended behaviour and, in some cases, malfunctions. These characteristics were underscored as giving rise to risks of misinterpretation and unintended consequences, alongside challenges related to limited interoperability and automation bias.

Societal, ethical and environmental risks

Broader considerations were raised, including risks of harmful bias and the environmental impact of AI and its supporting infrastructures.

Proliferation to non-state armed groups

Participants across regions identified proliferation to and the subsequent use of AI technologies by NSAGs, organized criminal groups, gangs and insurgent groups as a major concern. There was a shared concern that the possession and use of relevant technologies by these groups could ultimately undermine state authority, disrupt military operations, and exacerbate national and regional instability. These actors were reported as using AI to enable their operations, including those in the back end (e.g., logistics and financing), as well as to support combat-related activities, conduct disinformation campaigns and harm national agencies.

Critical national infrastructure and national resilience

The integrity and resilience of critical national infrastructure were identified as key concerns. This was due not only to heightened risks of attacks against this infrastructure, but also in the light of increased reliance and dependence on AI-enabled systems for their functioning, maintenance and security.

Legal compliance

Participants underscored challenges related to respect for, and the enforcement of, international law in the context of AI development, deployment and use in the military domain. Indeed, there is a perceived need to better understand what measures could be adopted to support the application and implementation of international law and to reduce states' risk of non-compliance.

Limited capacity

Many participants highlighted their perceived limitations in national capacity to address and mitigate AI-related risks. Limited capacity was presented as a key issue on the human, financial and technological fronts. Structural constraints were noted, including limited funding and inadequate digital infrastructure, which may hinder effective governance and implementation. Concerns were also raised regarding dependence on internationally procured technologies that may not be tailored to local contexts, needs, values and priorities, and limited capacity to procure such technologies responsibly.

Cyber, data and information-related risks

Participants emphasized risks related to the convergence of AI with cyber, noting in particular the risk of increased vulnerabilities. Particular concerns were raised in relation to the use of AI, by both adversary states and NSAGs, to further disrupt existing networks and technologies, particularly those on which military infrastructure depends. Concerns were also raised regarding the disruption, by cyber means, of networks and technologies on which AI depends. Data governance also featured prominently in the discussions, noting the close relationship between AI and data. Participants also raised concerns around the spread of disinformation, noting its potentially severe destabilizing effects. Overall, a number of participants subsequently noted that the AI–cyber nexus and its implications extend far beyond the military domain, as they could be elevated as a core driver of escalation and civilian harm.

2.2. National policy priorities

The exchanges on perceived opportunities and risks paved the way for reflections on states' national priorities in the context of AI in the military domain. These are, broadly:

- Compliance with international law and alignment with ethical guidelines
- Risk reduction and management
- A human-centric approach
- Wider security issues
- Responsible procurement and public–private partnership
- Multi-stakeholder engagement
- Capacity-building
- Meaningful regional and international engagement

These national policy priorities were consistently reflected throughout the consultations in states' interventions, approaches and postures. While there are nuances and differences across and within regions in these national policy priorities, there is a general understanding that they influence states' regional and international engagement. Conversely, discussions and deliberations at supranational levels will have direct and indirect implications for states' national policy surrounding the responsible development, deployment and use of AI in the military domain.

Compliance with international law and alignment with ethical guidelines

Participants from all regions underscored the importance of compliance with international law and alignment with ethical guidelines as a key national priority. This priority subsequently shapes and frames measures adopted at the national level for the responsible development, deployment and use of AI in the military domain. These measures include, for instance, national policies and measures specifically dedicated to the application of international law throughout the life cycle of AI, thus embedding IHL, international human rights law (IHRL) and relevant disarmament treaties into procurement processes. Some participants noted the potential value of international guidance to inform national efforts and measures. Others noted civilian protection as a key, overarching objective behind compliance efforts. A number of states specifically emphasized accountability and responsibility amid compliance efforts, for instance through the need for monitoring and oversight mechanisms and clear command structures as an enabler for accountability. Against this backdrop, a number of participants also emphasized the need to strike a balance between these efforts and operational effectiveness, noting that both sides should be mutually reinforcing and not seen as mutually exclusive.

Risk reduction and management

A number of participants highlighted the value of adopting a risks-based approach grounded in effective risk management, supported by multilayered control measures. They further stressed the importance of coherent measures across governance layers (i.e., at the national, regional and international levels) to address these risks – noting, however, that risk perceptions and prioritization can differ within and across regions. A number of states prioritize national efforts to address the convergence between technological fields, including bio-robotics, quantum technologies and anti-satellite capabilities, and to address related security risks and implications. A few other participants noted that addressing strategic and escalation risks was a top national priority, particularly in the light of heightened risks of escalation from the conventional to the nuclear realms with the use of AI-driven DSSs. Finally, a number of states across regions highlighted the importance of addressing potential risks associated with frontier technologies, including agentic AI and artificial general intelligence (AGI), and their possible deployment in the military domain as a matter of national priority.

A human-centric approach

Participants, in particular the majority of state representatives from West Asia and the Middle East, noted the importance of adopting a human-centric approach to the governance of AI in the military domain, whether in terms of policymaking, military decision-making or technical design choices. Some emphasis was placed on DSSs and weapon systems, noting potential linkages to ongoing discussions in the context of lethal autonomous weapon systems (LAWS) – with nevertheless the caveat that, while a few states stress the need to acknowledge, at least, the possible connections, others voiced strong concerns around blurring lines and potential confusion between both sectors. States' perceptions of what the "human element" consists of vary. Many, if not most, flagged ongoing reflections to define it in the first place and how such an approach, while important, should not be seen as a silver bullet, but rather as an enabler for the responsible development, deployment and use of AI in the military domain.

Wider security issues

Participants stressed the need to address wider security threats as a key national priority, in particular those posed by NSAGs. In fact, a number of participants across regions noted that non-international armed conflicts were perceived as posing a greater risk for national and regional security than international armed conflicts. Proliferation risks were indeed of concern for many participants and are subsequently prioritized in many national efforts. Furthermore, a number of states highlighted the need to address wider security dimensions, including economic security, human security and environmental security – topics that are adjacent to the governance of AI in the military domain but still highly relevant. The security of these systems was also highlighted, in addition to the security of supply chains from possible interference and disruption, which also constitute topics of national priority.

Responsible procurement and public-private partnership

All regions underscored the importance of careful and responsible procurement practices and, more generally, of measured engagement with industry, including through public–private partnerships. For most states consulted, there is the perception that they – along with their respective regions more generally – are not necessarily ones that would develop their own technological capabilities, but rather purchase them from abroad. As such, states acknowledged the need to consolidate their reflections and frameworks surrounding procurement, particularly with respect to civilian and foreign technologies, in addition to acknowledging the growing role and influence of the private sector. This will be particularly critical to mitigate potential biases that may result, for instance, from the systems' training data sets and the need for these to reflect local contexts, values and realities. National frameworks enabling knowledge transfer from the developer as part of a sale and post-sale, in addition to ensuring the security of supply chains, were seen as critical, especially for those states perceiving themselves as primarily, if not exclusively, customers. To this end, participants stressed the importance of establishing international, regional and national platforms and frameworks for effective engagement with industry and to enable its responsible behaviour.

Multi-stakeholder engagement

All regions emphasized the value of dialogue, cooperation and partnerships with the multi-stakeholder community, from academic institutions to national defence universities, civil society, international and regional organizations, and industry. These efforts should not only constitute a means to facilitate dialogue, but also to acknowledge the broader multi-stakeholder ecosystem involved in the development, deployment and use of these technologies. To this end, a number of states have designated such engagement as a national priority. A few have also outlined the value of dedicated CBMs to foster trust and collaboration across fields and sectors.

Capacity-building

Participants underscored the importance of capacity-building, including through international and regional cooperation to strengthen local computing capacity and data warehouses. Tailored capacity-building for buyers was highlighted, alongside the need to promote equitable access, including through technology transfer, and to support the development of trusted innovation ecosystems. These efforts must be undertaken in acknowledgment of the technology's dual-use nature, in addition to addressing the implications of using the same AI-enabled capabilities across law enforcement agencies and military forces – putting coordination as a central enabler. Additionally, an emphasis on equitable access to the technology and resilient infrastructure has also been identified as a key enabler for capacity-building. To this end, states acknowledged the importance of national frameworks and structures through, for example, a dedicated national strategy on AI in the military domain. A number of states flagged ongoing efforts or, at least, the desire to formulate a dedicated strategy document. Such documents would not only allow them to identify capacity needs and the formulation of clear objectives, but would also enable allocation of the appropriate funding and resources necessary. However, some participants noted difficulties on this front considering the ongoing and increasing cuts to international aid; questions were thus raised as to which institution or agency could deliver these capacity-building programmes in a way that is holistic and well-targeted to different communities.

Meaningful regional and international engagement

Participants emphasized the value of meaningful regional and international engagement, including through international scientific cooperation, as well as the value of their national representation, participation and active contribution in United Nations and other international forums. To this end, participants highlighted the importance of foresight, a strategic vision and means that would particularly position states of the Global South as active architects of regional and global governance frameworks. The importance of inclusivity at the regional and international levels were thus underscored, also noting the need to leverage existing tools, frameworks and platforms including in adjacent technologies (e.g., cyber) within the United Nations and regional organizations, such as the African Union, the Association of Southeast Asian Nations (ASEAN) and the Organization of American States (OAS). While noting the need for complementarity across these initiatives, a number of states nevertheless flagged discussions within the United Nations as a national priority. This puts into question the role, place and value of initiatives outside the United Nations, such as REAIM, the necessity of which was underscored by many participants.

2.3. Taking stock of national and regional policies on AI in the military domain

Beyond national views and policy priorities, the consultations provided a snapshot of national policies and frameworks that are either in place or emerging, at both the national and regional levels. This has indicated the significant progress and uptake in initiatives for responsible AI in the military domain that states have achieved since the first round of regional consultations, in early 2024. Specifically, the following observations were made:

National AI strategies are increasingly under development

Participants across regions shared a range of policy developments and efforts for the formulation, adoption and subsequent implementation of a national AI strategy that would cover, partly or fully, the military and security domains.[6] A number of states also noted ongoing efforts to develop a national strategy on AI, including for military applications, as well as adjacent tools akin to implementation road maps.

AI in the military domain is also covered by documents, tools and frameworks that are either more general in scope or that cover other technological fields

For instance, a number of states flagged AI as forming part of the wider scope of their national security strategy or defence plan. Alternatively, a few states flagged their wider national security strategy as a reference point to inform their adoption of AI technologies.

Beyond national strategies, states are ramping up the development of tools and frameworks for the responsible development, deployment and use of AI in the military domain

For instance, some states flagged the ongoing development of a military AI doctrine, while others referenced their Ministry of Defence's dedicated policy on addressing AI-related risks. One state mentioned the existence of a national framework dedicated to evaluating the trustworthiness of AI in the defence sector, in addition to risk mitigation-measures embedded in operational processes that stem from a dedicated AI strategy.

Regional efforts for AI governance are under way, with growing relevance to the military domain

Participants flagged a number of policy efforts, initiatives and growing frameworks at the regional level that are of relevance to AI in the military domain. For instance, a number of participants in the Asia-Pacific consultation noted the adoption by the ASEAN Defence Ministers of a Joint Statement on Cooperation in the Field of Artificial Intelligence in the Defence Sector in early 2025. In Africa, some participants are active members of the African Union's dedicated working group on AI, noting potential relevance to the military domain. It was also noted that Argentina hosted the 16th Conference of Defense Ministers of the Americas (CDMA) in October 2024, which comprised a dedicated working group for the "responsible development, application and governance of artificial intelligence in the military domain" for the region.

[6] National AI strategy documents referenced during the consultations include those of Azerbaijan, Egypt, Ghana, India, Kazakhstan, Kenya, the Philippines and Viet Nam, some of which are general AI strategies that may be applicable to military applications. For a comprehensive overview of existing national AI strategies that are publicly available, see the UNIDIR AI Policy Portal, https://aipolicyportal.org/.

2.4. Taking stock of national and regional good practices for responsible AI in the military domain

Building on states' exchange of their national policy priorities and ongoing efforts and initiatives, the consultations presented an opportunity for states to share a range of good practices for the responsible development, deployment and use of AI in the military domain. Noting the value of such an exchange, the following categories of good practices emerged from the consultations:

- Ecosystem of national governance frameworks and strategic guidance
- Measures for legal compliance
- National coherence and whole-of-government coordination
- Capacity-building and skills development
- Knowledge generation
- Multi-stakeholder engagement
- Development of a national industrial and research and development (R&D) ecosystem
- Clear procurement processes

Documentation and use of concrete operational applications and use cases of AI Echoing Sub-sections 2.1–2.3, participants pointed to the importance of developing a robust **ecosystem of national governance frameworks, tools and strategic guidance** surrounding the responsible development, deployment and use of AI in the military domain as a good practice. This ecosystem should include national strategies, ethical guardrails, guidelines for AI development ecosystems incorporating regional forecasting, and efforts to ensure coherence across frameworks, including those in place already at the national and regional levels. Some states cited, for example, their experience in leveraging data protection and cybersecurity legislation to require lawfulness and fairness, including in relation to AI training data, and to empower cybersecurity agencies to respond to AI-enabled threats to critical national infrastructure. Cybersecurity exercises embedding AI-enabled threats, including phishing and disinformation campaigns and synthetic media use, were also highlighted. Against this backdrop, reference was made to the experience of a number of states across regions in developing such an ecosystem of strategies, as well as regional frameworks and guidance that are useful reference points even if they are, at times, broader in scope than the military domain. Beyond documents, participants also referred to the value of establishing dedicated governmental processes and initiatives, including through parliamentary working groups and dedicated partnerships with regional and specialized international organizations.

With regards to **legal compliance**, a number of states outlined good practices through their national experiences. Such measures include, for example, the integration and prioritization of legal compliance in national strategy documents, which provides a clear and explicit anchor to inform and guide future activities in the context of AI in the military domain. Further examples include the establishment by a state of a compliance unit within its armed forces dedicated to ensuring IHL compliance with respect to AI in the military domain. Some states would also extend these legal considerations beyond warfare to peacetime, including in the routine administrative functions of armed forces where international law may apply.

In the same vein, participants noted the importance of **national coherence and whole-of-government coordination** across relevant agencies, including ministries, law enforcement bodies, intelligence agencies, R&D organizations and other bodies, while remaining mindful of institutional and political sensitivities. Such a whole-of-government approach is indeed presented as valuable for national alignment and coordination to inform and consolidate engagement at the regional and international levels, the lack of which many states regretted and presented as a key obstacle to more significant participation in international discussions. In this regard, participants referenced the organization of dedicated national meetings on AI for security and the military domain, supported by inter-ministerial committees, as well as the establishment of dedicated national committees on AI development. Participants also flagged the development of national security plans that factor in the complex nature of AI in the military domain, acknowledging the overlap of the use of AI in the military domain with broader security considerations, including counter-terrorism and cybersecurity and the need to engage all concerned agencies. The designation by some states of a dedicated national focal point for capacity-building processes was also highlighted as a good practice to ensure coherence nationwide.

Good practices were also shared in relation to **capacity-building and skills development**, including dedicated digital talent programme training of information and communications technology (ICT) graduates with AI specialization to support the implementation of a national digital masterplan. Participants also referred to capacity-building efforts to improve technical and multidisciplinary literacy, enabling a more holistic appreciation of both the risks and opportunities associated with AI in the military domain, including dimensions that remain under-studied. Reference was also made to some states' launch of dedicated national programmes to foster digital skills, including the establishment of national centres tasked with broader efforts to promote public education on AI. In fact, participants referred to capacity-building initiatives in the civilian domain that are designed to raise awareness of the dual-use nature of AI technologies and their security implications.

Building on the latter point, participants shared good practices related to **knowledge generation, education and cultivating healthy research ecosystems**. Specifically, a number of participants underscored the role of academic institutions and defence universities as generators of knowledge, conveners of dialogue between government, industry and academia, incubators of responsible leadership in the military domain, and platforms to strengthen regional cooperation and align best practices. Additional examples include the establishment of dedicated research centres, the creation of a defence AI research unit as a national research institute linked to the military academy while ensuring coherence with technological progress in the civilian domain, as well as the development of indigenous AI models. Some participants flagged their respective governments' ambitions to develop regional centres of excellence for research and the allocation of the necessary resources, with a significant number of participants from West Asia and the Middle East noting the value of establishing a network of these centres to promote capacity-building, knowledge transfer and the exchange of best practices.

Furthermore, the value of **engaging the multi-stakeholder community**, for instance through dedicated workshops to feed into national policies, was underscored. For example, a number of participants noted the recent organization by their government of seminars and workshops dedicated to engaging with the military, academia and other national agencies to further unpack and broaden the national understanding of AI in the military domain – with some participants noting that the inputs would directly feed into efforts to develop a national strategy.

With regards to the private sector specifically, participants highlighted efforts to **develop national industrial and R&D ecosystems** that are capable of building indigenous systems, partly or fully, while simultaneously engaging in governance discussions to inform the integration of foreign technologies into national systems and contexts. To this end, participants highlighted the value of **specific procurement practices**. One example is the integration of AI into a state's Air Force using locally developed technologies. Reference was also made to procurement practices that integrate transparency and explainability to ensure accountability, as well as the adoption of dedicated large language models for Ministry of Defence systems, accompanied by strict protocols and restricted access. As such, participants underscored the value of clear protocols and processes at the national level for the development, purchase, deployment and use of these technologies to ensure that AI use is based on clear objectives and grounded in well-defined criteria. In the same vein, good practices also included efforts to strengthen public–private partnerships and prioritize scientific innovation holistically. One participant further noted their state's launch of a dedicated AI investment hub in partnership with other states and venture capital as a means of accelerating its R&D ecosystem.

Finally, participants shared a number of success stories surrounding **concrete operational applications and use cases of AI** – and the value of sharing those as a point of learning for other states, as a reference for future governance efforts, and as a CBM. These include, for example, the establishment of dedicated "smart camps" for the armed forces and peacekeeping units that aim at enhancing mission effectiveness. Further examples include the use of AI-enabled capabilities for ISR in the maritime domain, including through regional cooperation for the deployment and use of such applications, as well as the deployment of uncrewed systems to reduce risks to human personnel, particularly in environments assessed as posing fewer risks than land or urban settings. Participants also referred to the use of uncrewed systems for the protection of critical and underwater infrastructure, as well as emerging linkages between AI-enabled systems and space-related capabilities.

Engineers check aerodynamics of new development drone (generated with AI). Credit: Adobe Stock / Framestock.

Abstract glowing network connecting global data points (generated with AI). Credit: Adobe Stock / Saowanee.

3. Taking stock of multi-stakeholder perspectives on responsible AI in the military domain

Acknowledging the importance of engaging with the multi-stakeholder community, the regional consultations had three dedicated segments that enabled such engagement:

- A dedicated kick-off panel with select representatives of the multi-stakeholder community
- A multi-stakeholder tabletop exercise (see Section 4)
- A UNIDIR-led round table discussion with the regional multi-stakeholder community

A key objective of the consultations was to take stock of the regional multi-stakeholder community's perspectives on the current state of affairs and areas of policy priority, along with thematic deep dives on governance and substantive topics of relevance to each region. This section summarizes the key findings of each component, which then feed into the recommendations and insights for the way ahead in Section 6.

3.1. Current state of affairs and areas of policy priority

This segment of the consultations provided an opportunity to take stock of the perspective of each region's multi-stakeholder community on the current state of affairs and to subsequently identify the areas of priority – both substantive and in relation to governance – for AI in the military domain. In fact, there is a shared emphasis, across regions, on the importance of factoring in **local considerations and contextual realities when examining each region's multi-stakeholder perspectives**.

BOX 3.

The perceived impact of regional contexts on the local multi-stakeholder landscape

Participants in all consultations noted specific examples where their respective regional contexts would bear a degree of influence over the local multi-stakeholder landscape and the subsequent practices and priorities that surface from these communities.

For instance, there was the sentiment among participants in the *European and North American* consultation that there is growing demand for the rapid development and production of AI technologies, accelerated by industry, due to evolving strategic considerations including the conflict in Ukraine, as noted by a number of delegations. Against this context, this explains, in the view of some participants, the growing and vibrant industrial ecosystem dedicated to developing AI solutions in the military domain. Participants in this consultation also questioned the assumption that regulation would hamper technological progress and innovation, noting that this is not inevitably the case, with solid regulation arguably driving innovation, including in the military domain.

In the *Asia-Pacific*, participants noted how certain subregions remained underrepresented, particularly the Pacific Islands, which subsequently translates into an emphasis by others on ensuring geographical diversity, representation and capacity-building in international forums. Yet, at the same time, many participants from the region reported funding challenges – including by international, regional and academic institutions – thus noting the growing need to rely, in part, on philanthropic support. They emphasized that a sense of urgency is not always present due to limitations in awareness-raising efforts within the region, which results in perceived challenges to enable further multi-stakeholder participation.

There is a view in *Latin America and the Caribbean* that, against a backdrop of significant differences in national capacities across the region, there is currently a lack of established expertise and of an epistemic community within the region. The absence of a region-wide structured community capable of enabling coordination, partnerships and alignment across sectors, including the industry, legal, policy and technical communities, is partly due to these variations in states' capacity. At the national level, discrepancies were also noted in the extent to which states consult their respective multi-stakeholder communities, with some primarily engaging defence colleges, further underscoring the need to encourage collaboration and dialogue within the broader defence and security multi-stakeholder ecosystem.

In *West Asia and the Middle East*, participants strongly emphasized that the role of the multi-stakeholder community within the region should be contextualized against regional and cultural realities. Specifically, participants underscored the importance of recognizing states as the primary stakeholders in the region's multi-stakeholder ecosystem, as reflected in the concentration of R&D efforts within government agencies in many of the attending states. Nevertheless, attention was drawn to perceptions of asymmetries of power between stakeholders and the need to acknowledge existing trust deficits not only between states but also between sectors and stakeholders – particularly noting the view that large technology companies developing capabilities for both the civilian and military sectors are growing in size, resources and power.

Finally, *in Africa*, participants underscored the importance of factoring in regional values and ongoing challenges. For instance, a number of delegates highlighted the value of AI indices and benchmarks that would subsequently provide snapshots of the continent's situation with respect to its underpinning values, but also the issues the continent is facing, such as the (un) availability of local data sets, access to the technology and technical literacy. Participants also underscored the need to identify vulnerable communities in conflict, particularly in the light of the community-oriented realities and culture within the region: a number of participants noted that this point highlights the potential tensions that may emerge between high-level international norms on the one hand, and regional and cultural sensitivities on the other hand. To this end, a harms-based approach was presented as a useful reference point. In addition, the need to prioritize non-proliferation to NSAGs and insurgent group was highlighted, noting in particular the potential second- and third-degree implications and risks of harm that such proliferation may have for local communities.

Ogossagou, Mali, 2022. Credit: UN Photo/Harandane Dicko.

Against this backdrop, there were broadly four areas of policy priority that emerged from the consulted representatives of the multi-stakeholder community: the importance of technical considerations in governance discussions; capacity- and knowledge-building; compliance with international law; and effective multi-stakeholder engagement.

3.1.1. Importance of technical considerations in governance discussions

First, an emphasis is placed on the **need to better understand, unpack and integrate the technical nature of AI and adjacent technological considerations**. In fact, a number of participants emphasized the need for governance approaches, structures and frameworks to take into account this technical dimension, including AI's characterization and nature as a "system of systems". Furthermore, a number of participants highlighted the need to address the dual-use nature of AI technologies, including both situations in which civilian technologies are used for military purposes and military technologies are used for civilian purposes. In this sense, a wide range of applications must be considered, from cloud computing to routine operations, logistics, intelligence gathering and battlefield applications – noting in particular a growing focus by states on the implications of AI-enabled decision-support systems, among other capabilities. In this context, attention was drawn to the need to reconcile potential tensions, such as those arising from the military use of civilian cloud services that are subject to privacy laws. Against this backdrop, there was a clear interest in further exploring how issues related to AI in the military domain intersect with other technological sectors and thematic areas, including cyber, and associated risks of destabilization, including in scenarios of hybrid warfare. Finally, beyond current technologies, reference was also made to prospects related to technological progress and frontier models, including agentic AI and AGI, and their subsequent implications for international peace and security.

Coupled with greater focus on grounding these initiatives on concrete use cases, there is the perspective that integrating the technical dimension would consolidate governance efforts and help address some of the long-standing issues that exist in this space. For example, participants underscored the need to reflect more carefully on the role of the human element. A thorough and holistic understanding of the issue will not only confirm the perception that human involvement does not always necessarily result in improved military action; solid technological foundations would enable a better understanding of human–machine teaming, including how design choices and user experience (UX) considerations can be leveraged to foster the responsible development, deployment and use of these technologies.

3.1.2. Capacity- and knowledge-building

Second, participants identified **capacity-building** as a priority area for a number of regions, while also raising questions regarding what capacity-building concretely entails, how priorities should be identified and in what sequence measures should be implemented. They emphasized the need to identify and scope capacity-building needs, the importance of cross-disciplinary and cross-agency approaches, and the value of initiatives that encourage critical thinking around the use of AI, particularly in decision-making contexts. The compilation of best practices across targeted sectors was identified as a potential avenue to build both capacity and trust. In that sense, some delegations highlighted the value of sharing best practices related to Article 36 legal reviews as a concrete example,[7] in addition to underscoring the value of looking to best practices from initiatives outside the military domain with a view to enabling the transfer of relevant knowledge and lessons learned.

To this end, participants emphasized the importance of cultivating education systems and industrial ecosystems in the region, while ensuring that responsibility, compliance and ethics are not perceived as a luxury, but rather are supported through the appropriate incentives. They emphasized the need to support states in building the political infrastructure required to subsequently enable multi-stakeholder engagement in neutral and depoliticized settings, as well as the importance of strengthening the capacity of regional and international organizations to convene dialogue and support for their member states. Participants also stressed the value of spreading awareness through civil society organizations and think tanks, including those from outside capital cities, and of exploring cultural factors shaping AI innovation in the region. As such, participants noted the value of improving messaging, outreach and institutional memory to increase buy-in, in addition to encourage greater participation of multi-stakeholder communities in the future.

3.1.3. Compliance with international law

Third, participants reiterated the importance of ensuring **legal compliance**. Indeed, a number of participants stressed the importance of integrating measures to foster compliance with applicable international law throughout the life cycle of a technology. While participants reiterated that compliance with international law remains the core responsibility of states, measures for such compliance will require multi-stakeholder support and contributions. In fact, it has been argued that the multi-stakeholder community has a role in implementing most of the measures cited, including legal review processes, the integration of legal considerations within testing and evaluation cycles, interface design and UX research, user training, and bias-mitigation measures. In this sense, participants emphasized the importance of establishing clarity regarding expectations for different stakeholders, along with heightened efforts for awareness-raising. Participants raised particular questions as to whether industry and developers have sufficient, if any, awareness of the international law obligations that bind states (i.e., their clients). In this regard, participants highlighted the importance of translating states' legal obligations into concrete measures and expectations for industry for example, including through measures related to remedy. Additionally, participants noted that, while IHL remains a core focus of discussions around legal compliance, IHRL is also of importance due to its continued applicability, particularly given the prevalence of non-international armed conflicts and violent situations below the threshold of armed conflict across regions that involve insurgent groups, organized crime syndicates and gangs. Amid these discussions, participants pointed to the relevance of existing frameworks, platforms and initiatives that seek to bridge the conversation

[7] Article 36 legal reviews implement an obligation of states parties to the 1977 Additional Protocol I to the 1949 Geneva Conventions, specifically laid out in its Article 36: "In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party." While Additional Protocol I is not adopted universally, a number of non-states parties have nevertheless reported conducting such legal reviews despite not being specifically bound by the Additional Protocol.

between states and industry with regards to legal compliance, and their potential to provide inspiration for discussions on AI in the military domain. These include, for example, the United Nations Guiding Principles on Business and Human Rights, as well as the (varied) involvement of industry in the context of past arms control agreements such as the 2013 Arms Trade Treaty and the 1972 Biological Weapons Convention.

3.1.4. Effective multi-stakeholder engagement

Finally, participants underscored the value of **sustaining and scaling up multi-stakeholder engagement in an effective and meaningful way**. They stressed the particular importance of initiatives and platforms that promote exchanges of perspectives across stakeholders, including across government bodies such as Ministries of Foreign Affairs and of Defence, as well as with industry and civil society. Such exchanges were seen as critical to understanding differences in priorities, in expectations and in definitions of foundational terms, and to promote mutual understanding. As such, participants highlighted the critical role of trust- and resilience-building initiatives that involve the multi-stakeholder community, citing in particular REAIM, UNIDIR's Roundtable for AI, Security and Ethics (RAISE),[8] as well as the US–China Track II Dialogue on AI and International Security, organized since October 2019 by the Center for International Security and Strategy at Tsinghua University and the Brookings Institution.[9] Participants emphasized building mutual understanding and epistemic community knowledge, including through track-2 dialogues, as well as facilitating repositories of use cases and sustained dialogue and collaboration. Shared understandings were underscored as particularly important given the growing securitization of trade and education, and the need to secure public trust and legitimacy through civil society engagement.

Participants also underscored that responsible governance requires shared responsibility and effective collaboration across all stakeholders and constituencies, including the youth. They emphasized the importance of rigorous human rights due diligence, the maintenance of responsibility and accountability while ensuring privacy, and the potential role of technical solutions, including responsible design choices – all facilitated through effective multi-stakeholder dialogue. Participants also highlighted the need for military procurement processes with appropriate levels of oversight, as well as the value of standardized commitments or guidance to support the implementation of the principles of responsible AI in the military domain. In addition, participants noted the relevance of factoring in governance efforts undertaken in other sectors, such as the ICC's development of a policy on international criminal prosecution in cyberspace, which was underpinned by multi-stakeholder input. Thus, building on these reflections, participants stressed the need for fully inclusive platforms that transcend geopolitical barriers and the need to ensure effectiveness. To this end, they emphasized the importance of clarity in the scope, sensitivity and adequacy of each forum and each discussion. They highlighted the importance of distinguishing between objectives such as building knowledge bases (including through case studies), establishing benchmarks, and informing

[8] On RAISE see https://unidir.org/raise/.

[9] For the latest from the Track II dialogue, see Melanie W. Sisson et al., "Steps toward AI governance in the military domain", Brookings Institution, 12 November 2025, https://www.brookings.edu/articles/steps-toward-ai-governance-in-the-military-domain/.

framework and norm development. Participants also noted that states should actively platform and champion their respective members of the multi-stakeholder community and should share their names in order to facilitate their participation and contribution to international platforms such as REAIM. Participants also highlighted the value of enabling permanent channels for information exchange, including within the multi-stakeholder community itself, and means to preserve institutional memory to ensure the continued effectiveness of these initiatives.

3.2. Thematic deep dives: Substance

The multi-stakeholder round table provided a unique opportunity for participants to delve deeper into select substantive issues of relevance to the different regions consulted. While many of the themes covered would overlap across two or more regions, participants were able to share unique perspectives that enabled a more granular understanding of regional nuances, contexts and realities. The themes covered were: wider security considerations; regional destabilization; dual-use and proliferation risks; responsible procurement practices; and the AI–NC3 nexus.

3.2.1. AI and wider security considerations

Participants noted the difficulty of defining “the military domain” in the context of AI, highlighting in particular the different national approaches to the allocation of defence and security responsibilities around the globe. A number of states have, for instance, adopted a national model under which the armed forces may be activated in certain national crises, including political crises, to support law enforcement agencies and assigning the military with security roles, thus blurring lines between military bodies and police functions – including through the use of the same AI technologies in these instances, for example. Conversely, participants further noted that defence institutions in some contexts are tasked with non-wartime and non-combat responsibilities, such as preventing illegal fishing, protecting territorial waters and safeguarding the environment. This illustrates the broader role of defence actors in addressing internal security challenges, which all may result in the use of technologies that may be of use for both defence and security purposes.

While acknowledging the multiform nature of the “military domain”, participants emphasized the need to factor in the wider security considerations of AI in international deliberations. Indeed, a number of participants highlighted, in particular, issues related to state sovereignty and the balance between ensuring regional unity and preserving national interests amid regional and international discussions. This is also important given the need to consider the entanglement, in many cases, between military and law enforcement circles when addressing the application of AI in defence and security. In fact, participants emphasized the importance of enabling, clarifying and preserving civilian–military cooperation frameworks, particularly given the dual-use nature of many AI technologies and the significant role of civilian actors in R&D and innovation with implications in the military domain. Participants also underscored the need to look beyond direct military applications and consider second-order effects of AI in the military domain that may affect the wider security environment, including in the realms of human, economic and environmental security.

Finally, considering the sheer complexity of governance that encompasses both military and wider security issues, participants pointed to the potential role of international and regional organizations in gathering evidence and suggesting possible avenues for this governance.

3.2.2. Regional destabilization

Participants highlighted the growing integration of AI into military capabilities witnessed in recent conflicts, notably through the use of drones and applications in ISR, military logistics and AI-enabled decision-support systems. They emphasized that such developments have the potential to exacerbate regional and international escalation dynamics and contribute to arms race pressures. Participants emphasized the risks associated with pre-emptive strikes, alongside the inherent difficulties of attribution and verification in this context.

Participants particularly noted the emergence of drone swarms and the weaponization of cyber capabilities as illustrative of the convergence of AI with other technological domains, including robotics and cyber, as well as broader chemical-, biological-, radiological- and nuclear-related considerations. In this context, they underscored the importance of verification and CBMs, which are both of critical importance given the grave risks that AI may pose not only in the military domain but also for wider national security, including in the context of attacks against critical national infrastructure. In the same vein, participants noted interoperability risks, particularly those arising from legacy systems, and noted that such vulnerabilities could be exploited by adversary states as well as non-state armed groups.

Participants thus emphasized that AI itself can constitute a destabilizing risk, including due to vulnerabilities such as data poisoning, automation bias and hallucination, as well as risks associated with generative AI, particularly in the context of disinformation and misinformation. The existence of regional imbalances and the associated heightened risks of escalation were particularly noted by a number of states, further exacerbating risks of destabilization.

3.2.3. Dual-use and proliferation risks

There was a consistent recognition across consultations that the dual-use nature of AI technologies has a series of implications, including proliferation risks. At the more foundational level, participants raised questions regarding what “dual-use” concretely means in the context of AI in the military domain, and whether it should be understood as an intrinsic characteristic of a system itself or whether it is context-dependent, based on how the same system is employed (e.g., in peacekeeping operations versus offensive military uses).

There was strong emphasis on the importance of engaging closely with industry actors to unpack the meaning and implications of dual-use technologies, to better understand associated risks, and to explore possible mitigation measures. This was particularly salient given the perceived commercial incentives for companies to broaden potential markets for their technologies, raising questions both about how to incentivize industry to meaningfully engage with dual-use considerations in the first place, and about how to encourage the subsequent adoption of concrete risk-mitigation measures.

Discussions also pointed to a perceived tension between military imperatives, commercial incentives and ethical considerations faced by industry actors, with notable variation in practices across sectors. Larger, diversified technology companies often have some form of engagement on issues such as privacy, accessibility and responsible development and use, sometimes through voluntary coalitions. In contrast, specialized defence companies may not systematically adopt such practices nor mobilize the resources necessary for such engagement. This divergence reinforced the perceived need for a coherent framework to engage this sector and promote more consistent approaches. In fact, participants identified the absence of an authoritative framework for applying responsible AI principles and international law considerations within the private sector as a major challenge. Many participants were left uncertain regarding the extent to which corporations have thoroughly considered the dual-use and proliferation risks associated with these technologies, as well as how such risks are being identified and mitigated. Participants also highlighted the need to manage risks related to accidental failures arising from the employment of dual-use technologies. In the absence of such guidance, companies are left to make informed but fragmented judgments, resulting in divergent benchmarks and practices. This includes those related to management of the implications of these technologies' dual-use nature – especially given that contracts for dual-use technologies are typically negotiated in peacetime, prior to any operational military use.

Participants further highlighted disparities in existing legislative frameworks governing the sale and purchase of dual-use technologies, pointing in particular to legal complexities surrounding the acquisition of foreign technologies, which require further reflection and clarification. While these challenges are generally acknowledged, they were noted as remaining the subject of ongoing discussion.

On the specific issue of proliferation risks, participants highlighted the perceived risk of proliferation of AI-enhanced cyber capabilities into the hands of NSAGs, including both cyber offensive capabilities and in support of cyber espionage activities. Indeed, participants noted that such capabilities may provide operational advantages for these groups, while also enhancing disinformation campaigns that subsequently add to national and regional destabilization. Participants also raised concerns regarding the emergence of small projectiles that enable AI systems to autonomously locate and track targets.

As a consequence, participants emphasized the importance of adopting a multilateral approach to address these risks. They underscored the potential role of technical measures, such as end-user agreements and technical solutions to ensure traceability and maintain a digital footprint, alongside the use of domestic legal and regulatory frameworks to support accountability. There is, however, a need to strike the right balance between, on the one hand, open-source AI and the opportunities it offers, particularly to the Global South, in terms of accessibility, and, on the other, the associated proliferation risks. Ultimately, participants emphasized the need for a global framework of corporate responsibility, grounded in international law and evidence, as well as the need to incentivize industry engagement in governance efforts. In this context, the establishment of licensing systems to classify the risk levels of AI systems, including with regards to the risks of proliferation and of misuse, was also highlighted.

Against this backdrop, participants highlighted the critical importance of the ability to foresee and anticipate risks, while cautioning against prejudging risks associated with technologies, including those that are yet to be deployed. In fact, a number of participants noted that, while dual-use technologies raise significant proliferation concerns and risks, they also hold potential benefits. These range from their applications (e.g., to support crisis management and natural disaster relief) to their acquisition and use (e.g., through easier and more affordable access to these dual-use technologies). Participants cited a wide range of potentially dual-use technologies of relevance, from computer vision programmes via monitor wildfires to facial recognition technologies, as well as large language models and the use of generative AI for information warfare. These reflections underscored the need for evidence-based decision-making and policy choices grounded in use cases to support preparedness and proactive governance, rather than reactive approaches, especially given the potentially high stakes associated with military AI and its dual-use nature.

3.2.4. Responsible procurement practices

Participants emphasized that responsible procurement must be grounded in both international and national law. IHL was identified as a necessary starting point for acquisition decisions, in order to ensure that procurement aligns with applicable international guidelines. Participants highlighted the importance of due diligence to ensure not only legal compliance, but also the security of systems, the integrity of data, and resilience against cyber and hardware vulnerabilities.

To this end, participants highlighted the need for greater clarity regarding what different stakeholders are concerned about, prioritize and implement in the context of responsible procurement and sales of AI-enabled technologies. Particular attention was drawn to industry practices, including the importance for purchasers to understand the types of information that companies can provide, notably with respect to testing and evaluation, sources of data, potential privacy implications, and the use of synthetic data. In this context, participants underscored the usefulness of an international guideline or reference point to support states in articulating expectations towards industry, especially given the limited resources available to some states to independently assess these issues.

Furthermore, participants underscored the importance of state-led or, at the very least, state-sanctioned testing and evaluation, particularly when procuring technologies from private sector actors and from foreign suppliers. The availability of benchmarks was seen as useful to assess whether procurement practices meet expected standards. Participants also noted the importance of prioritizing in-house support and maintenance where possible, and of de-prioritizing reliance on foreign suppliers.

Moreover, participants emphasized the need to factor in the reality that many developing countries face significant resource constraints, which limit their ability to implement certain proposed measures for responsible procurement. This reinforces the perceived value of an international guideline, which should be evidence-based, reflect regional realities – particularly in contexts where threats from NSAGs and proxy wars are prevalent – and remain depoliticized and neutral.

3.2.5. AI–NC3 nexus

Participants emphasized the need to consider the full suite of AI uses across nuclear command, control and communications, including applications related to ISR, as well as early-warning functions feeding into decision-making processes. Participants raised questions regarding non-nuclear weapons with strategic effects and highlighted that recent escalation dynamics in South Asia, accelerated notably (but not only) by AI-driven capabilities, have demonstrated the risks of conventional escalation leading into the nuclear realm.

Participants also underscored the importance of ensuring the integrity of information transfer and data. Concerns were raised regarding risks of data poisoning and data manipulation, including through silent cyber operations that may occur prior to a system being activated, as well as challenges related to verifying the reliability of data. In addition, participants highlighted the problem of unequal access to reliable data sets, which may have harmful consequences further down the line particularly in times of crisis.

Amid these reflections, participants emphasized that the human element remains the most important aspect in this context and stressed that AI should not be conferred with decision-making authority.

Finally, participants underscored the importance of verification and emphasized the need for neutral and depoliticized spaces to enable dialogue across all sides. In this context, they highlighted the issue of mutual vulnerability in the absence of dialogue.

3.3. Thematic deep dives: Governance

In addition to substantive deep dives, the multi-stakeholder round table discussions also presented an opportunity for participants to delve into three governance issues: knowledge and capacity-building, trust-building, and the human element. These topics were selected building on the work of UNIDIR's RAISE, from which representatives of the multi-stakeholder community identified six areas of priority for the governance of AI in the military domain.[10] While three of these topics were more technical in nature, the remaining three were specifically oriented towards governance, which were further unpacked in the REAIM Regional Consultations.

3.3.1. Knowledge and capacity-building

Echoing the point made by state representatives participating in the regional consultations, members of the multi-stakeholder community agreed that knowledge and capacity-building constitute a critical of to the responsible development, deployment and use of AI in the military domain. Participants noted that gaps in knowledge could lead to over-dependence on AI-enabled technologies and could, in turn, increase vulnerability. At the same time, participants emphasized the need for a better understanding of how AI could concretely support the response

[10] For more details on the six themes, see Y. Afina and G. 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/.

to current and contemporary challenges. These are not limited to military and operational concerns but also extend to wider security issues tackling illegal mining, preventing environmental destruction, controlling illegal migration and fishing, and curbing human trafficking. Meanwhile, the adoption of AI technologies for these uses must remain responsible and fully aligned with applicable compliance requirements and obligations.

Additionally, participants underscored the importance of bridging gaps and facilitating dialogue between different sectors, including between technology communities and the military. In fact, participants emphasized that capacity-building efforts should not be limited to governments but should also encompass the wider research ecosystem. Indeed, they highlighted the need to build technical capacity. To this end, participants underscored the importance of facilitating dialogue across stakeholders and sectors to encourage mutual understanding and the sharing of knowledge, including through engagement with industry actors. As such, participants highlighted good practices related to facilitating national networks and dialogue between universities and research centres, as well as cooperation among these actors. They further underscored the importance of extending such facilitation to the regional level, while ensuring that spaces for dialogue remain depoliticized. Nevertheless, participants noted that reflections on knowledge and capacity-building cannot be fully divorced from political considerations, highlighting the need to better understand the parameters and motivations around which states have developed their AI capacity in the security and defence domains.

There was also strong emphasis on the importance of developing a clearer and more nuanced understanding of the respective roles and capacities of different stakeholders. Governments were widely seen as being best positioned to lead on norm-setting and policy development, while industry actors were viewed as uniquely placed to identify and address technical capabilities and limitations. Civil society, academia and think tanks were recognized for their critical role in scrutinizing progress, identifying gaps and providing independent analysis. Participants further noted the value of increasing transparency across sectors, particularly with respect to industry efforts to develop and deploy AI systems responsibly, including steps taken to implement responsible AI principles and to align practices with IHL. Over time, such transparency was seen as something that could become a market expectation and be translated into commercial incentives. Ultimately, such efforts, coupled with collaboration and dialogue, were seen as supporting the consolidation of research efforts and the upskilling of stakeholders, notably in developing countries.

To enable knowledge and capacity-building efforts, participants emphasized the importance of academic and research institutes in enabling capacity-building, particularly as a means to foster collaboration between developing and developed countries. Participants highlighted the value of compiling concrete use cases of AI applications in the military domain as a means to build a shared knowledge base and support informed governance discussions. At the same time, they acknowledged the significant challenges associated with such efforts, including reluctance to share information due to confidentiality and secrecy concerns, persistent trust deficits between stakeholders, and unresolved questions regarding the appropriate scope of information to be shared.

Amid these reflections, participants raised questions regarding the most effective and appropriate outlets and platforms to facilitate knowledge-sharing and capacity-building. Considerations included how to ensure accessibility, trust and sustained engagement, as well as how to avoid fragmentation by building on or connecting existing initiatives, rather than creating parallel structures.

3.3.2. Trust-building

Generally, participants emphasized the critical importance of building trust between stakeholders through convening events, and they highlighted the value of sharing good practices as a means to support confidence and mutual understanding. Trust-building between states and among stakeholders was also seen as a means to consolidate collaboration and work to prevent proliferation of these technologies into the hands of NSAGs. At the same time, participants underscored the importance of ensuring information integrity and sensitivity, while at the same time allowing space for interoperability exercises and appropriate information-sharing.

In order to build trust, participants highlighted the need to identify sources of distrust in the first place, not only between states, but also between public and private sector actors. Particular concerns were raised in relation to dual-use technologies and the potential for industry actors to sell the same capabilities to adversaries. This reflects the perceived different priorities of industry, including profit-driven incentives. In this context, participants noted the potential value of contractual clauses to prevent sales to adversaries, while recognizing that such measures would need to be assessed on a case-by-case basis.

Furthermore, participants emphasized the need for dedicated platforms to support trust-building. They underscored that ensuring a willingness to engage in dialogue is essential and they highlighted the importance of depoliticizing such spaces by setting aside states' strategic cultures and historical experiences that may otherwise stall trust-building efforts. In fact, neutral spaces will be critical if power asymmetries are to be addressed, as well as perceptions of power asymmetry, as part of broader trust-building efforts. However, a number of participants also noted that political will for dialogue remains a key element, by emphasizing that regional stability is closely linked to stability at the national level.

Building on these discussions, participants further explored the value of a number of CBMs. For instance, they highlighted the importance of verification mechanisms and underscored the value of sharing responsible norms as a basis for building confidence. Participants also underscored the importance of track-2 dialogue and the value of facilitated engagement by third parties. In this context, they highlighted the relevance of building on existing work and best practices, including experiences from the US–China Track II Dialogue on AI and International Security between Tsinghua University and the Brookings Institution. The latter has indeed allowed the unpacking of micro-dynamics of trust-building, noting that targeted, context-specific efforts can provide important sources of inspiration for the wider international community. Participants also highlighted insights generated through UNIDIR's RAISE initiative, which was seen as offering a broader perspective by bringing together multi-stakeholder representation from all United Nations regional groupings, including nationals of the five permanent members of the Security Council (P5).

In addition, participants emphasized the usefulness of these initiatives to identifying and prioritizing “low-hanging fruit” in trust-building efforts. One example cited was the retention of human control in NC3, as reflected in the November 2024 joint declaration by President Xi Jinping of China and President Joe Biden of the United States.[11] Such focused and tangible commitments were seen as potential entry points that could subsequently be widened to other nuclear-armed states and, over time, to the wider international community

Finally, participants underscored the importance of trust-building as an integral component of capacity-building, including to facilitate cross-agency communication (e.g., to enable the notification of incidents across agencies), drawing lessons from practices developed in the cyber domain. Participants also highlighted the need for a common ethical and principles-based baseline to support operationalization across stakeholders and actors, emphasizing the importance of convening events and of meaningful engagement and dialogue to this end.

3.3.3. The human element

A number of participants started the discussions with the assertion that the role and value of the human element varies depending on the application, including whether AI is used in decision-support systems or contexts, and that the intended use of the technology has direct implications for where the human element should begin and end or whether it should be continuous. For example, a number of participants noted that integration of the human element “in”, “on”, or “out of” the loop, as well as the necessity and form of such integration, requires careful consideration of the operational context in which an AI system is deployed. Participants highlighted that requirements may differ significantly depending on the application, for example between AI-enabled DSSs and cyber operations. This underscores that a one-size-fits-all approach would be insufficient.

In the same vein, the term “the human element” has been presented as having a meaning that varies across sectors and communities, including among lawyers, developers and engineers, and policymakers. This, in turn, underscores the need to partner with the multi-stakeholder community to better understand what is required in practice and how adjacent concepts, such as “human control”, could or should be interpreted, acknowledging at the same time the political weight that these terms carry. In this context, a number of participants – notably from West Asia and the Middle East – put a specific emphasis on preserving human control and judgment throughout the life cycle of AI in the military domain and the need for dedicated, enabling measures. They also noted the lessons that may be drawn from the ongoing discussions on LAWS. Furthermore, participants pointed to the relevance of learning from other frameworks and discussions, including those emerging from AI safety circles.

Furthermore, participants highlighted the need to better understand the legal implications of the human element at the international, regional and national levels, including the need to clarify both state responsibility and individual responsibility throughout the life cycle of a technology. Participants suggested drawing inspiration from existing frameworks, such as the

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

risk-based approach reflected in the European Union AI Act and ongoing discussions around human control, while cautioning against remaining overly entrenched in discussions limited to LAWS. Beyond high-level regulation, there is also a need to consider implications of the human element at the sub-regulatory level, including in the context of rules of engagement and how they operationalize IHL. Attention was drawn to the fact that rules of engagement are ultimately instruments that regulate human behaviour, rather than the environment or the technology itself. As such, lessons can be drawn from how rules of engagement are formulated, interpreted and implemented in practice when considering how the human element should be embedded in the use of AI-enabled systems.

Additionally, questions were raised about how best to socialize these issues within the developer and technical communities and to ensure that relevant knowledge flows down the production chain. This would lead to the subsequent operationalization of existing frameworks, tools and standards through concrete measures such as red-teaming, auditing and adversarial testing. Participants emphasized the importance of concrete and practical implementation, meaningful industry engagement and incentives for participation in governance discussions, noting that there is a window of opportunity to act.

Participants further underscored the critical importance of the level of system integration in enabling oversight, supporting human–machine teaming and facilitating after-action reviews, as well as the role of CBMs and safeguards, including incident report sharing and hotlines. These measures are particularly important in response to the concerns shared by a number of participants regarding risks of bias and emphasized the need to protect human resilience and accountability, while taking into account the challenges posed by information flux. In addition, concerns and questions were raised in relation to distributed agency in AI systems; these emphasized the potential need for standards to address these issues.

Participants particularly stressed the need to focus on the humans affected by AI-enabled systems, alongside the importance of user training, interface design and the role of training data. In fact, strong emphasis was placed on the need for adequate training of commanders, decision-makers and end-users of systems. Such training should cover not only system functionalities and capabilities, but also limitations and failure modes. Participants also highlighted the value of integrating UX considerations into the design of systems and interfaces. As a concrete best practice, reference was made to the approach taken by one state whereby obstacles are identified early during procurement through the formulation of clear user requirements and objectives from the outset, particularly in co-development contexts.

Finally, participants raised questions regarding the implications of agentic AI, particularly the need for appropriate safeguards and kill switches in cases where systems act or behave in unpredictable or unintended ways. This includes the potential need for technical functions that notify operators when system performance deviates from expected parameters. The concept of “failing gracefully” was discussed, referring to mechanisms that allow systems to safely degrade or shut down to prevent harm, analogous to consumer devices that power down when overheating. This was distinguished from “material protection” measures, such as self-destruction mechanisms designed to prevent adversaries from accessing sensitive information, for example in the case of compromised or hacked drones.

4. Operationalizing responsible AI: Insights on procurement, use and incident-response management from a tabletop exercise

Each consultation had a dedicated segment for a multi-stakeholder tabletop exercise facilitated by UNIDIR. Through a fictional scenario, the exercise provided an opportunity for participants to reflect on concrete ways of operationalizing some of the responsible AI principles captured over the years. The scenario was divided in two parts, one for procurement and one for use, which allowed participants to share reflections in these two contexts, particularly in relation to crisis management and risk reduction.[12] For this session, state delegations were joined by select members of the multi-stakeholder community, which provided an opportunity to add further nuances and depths to their reflections.

4.1. Procurement of AI capabilities: Prioritization of procurement parameters and assurance requirements

In the context of a fictional escalating conflict, participants played the role of state officials that sought to purchase three AI-enabled capabilities:

1. AI-based electronic warfare capabilities to disrupt the adversary's ability to send and receive radio, infrared and radar signals
2. A swarm of uncrewed stealth systems[13] with AI-enabled autonomous navigation functions for ISR purposes to identify and track illicit flow of goods operated by the adversary
3. An AI-enabled DSS to upgrade an existing battle-management software to aid commanders' decision-making, including decisions to use force

Due to operational needs, the procurement cycle is two months.

[12] For the North American and European consultation, only the first part of the exercise was conducted due to time constraints.

[13] In order to adapt to its security landscape, the exercise conducted in the Asia-Pacific considered the purchase of uncrewed stealth maritime vehicles, while in the four other regions, participants considered the purchase of uncrewed stealth aerial vehicles.

BOX 4.

Notes and methodology for the exercise

Divided in groups, participants were tasked with unpacking 10 measures corresponding to procurement parameters and assurance requirements, all identified and defined in advance by UNIDIR. Participants subsequently cast their individual votes on the three measures they would prioritize for each capability and then discussed the results within and across groups. These 10 measures were designed to capture some of the key principles of responsible AI and considerations that have emerged over recent years. While by no means meant to be an exhaustive list, some of the key principles and considerations that these measures capture include compliance with international law, transparency, accountability, traceability, the role of industry, and security and safety.

Some of the measures identified are meant to reveal points of tension that may emerge between the substance of the measures and their implementation. For instance, a complete handover of the system's maintenance from the supplier to the client (i.e., the purchasing state) may contradict requirements for the supplier to ensure the system's consistent post-sale reliability, which in itself implies a degree of maintenance led and assumed by the supplier. While the exercise did not seek to present these measures as mutually exclusive, it aimed to reveal implementation challenges that states and relevant stakeholders may need to address upstream as they collectively develop and design future governance architectures and measures for the responsible development, deployment and use of AI in the military domain. Notwithstanding the ability to reconcile these tensions, the exercise enabled states to contextualize these points of tension against a need to prioritize and possible nuances across applications.

Finally, the two-month procurement time frame in the scenario is designed to acknowledge the emerging trend, worldwide, in favour of rapid procurement and deployment of AI systems – either for operational or strategic purposes. At times, this may be perceived as a source of tensions with the implementation and operationalization of responsible AI principles. Nevertheless, a number of participants across the regional consultations noted that, despite the tensions that emerge, compliance with applicable laws must remain a priority in the light of their binding, non-optional nature.

Participants voted on the following 10 measures:

Measure 1 – Iterative legal reviews: Establish a clause mandating regular reviews of the capability's alignment with international law

Measure 2 – Iterative performance reviews: Establish a clause for regular reviews of the capability's performance (e.g., in terms of precision, utility, consistency in performance, etc.)

Measure 3 – Independent third-party review: Contract a third-party organization/entity to conduct a regular audit of the system's performance, resilience and robustness

Measure 4 – Complete control of the system's maintenance: Completely hand over the system's maintenance from the supplier to the client (i.e., the purchasing state), including troubleshooting protocols and the re-training of system

Measure 5 – Complete control over the system's decommissioning: Completely hand over the system's decommissioning protocols from the supplier to the client (i.e., the purchasing state), including responsible disposal and data deletion

Measure 6 – Access to all pre-deployment information: Ensure full access to relevant pre-deployment information, including the system's training data, testing and evaluation parameters and protocols, and laboratory history

Measure 7 – Digital forensics: Establish technical solutions and protocols for digital forensics and the distribution of responsibilities (state vs supplier vs third party)

Measure 8 – Consistent reliability: Establish the supplier's responsibility to ensure consistent reliability throughout the technology's life cycle

Measure 9 – Disclosure of vulnerabilities: Mandate the supplier to disclose post-sale discovery of new forms of vulnerability

Measure 10 – Supplier backdoor: Establish a backdoor for the supplier to monitor and identify (new, post-sale) vulnerabilities in systems in use

All participants in the five regional consultations were invited to individually cast votes for measures to apply for each of the three hypothetical AI-enabled military capabilities considered in the exercise. Conducting the same structured voting exercise across regions enables the identification of region-specific assurance priorities, while also allowing for the examination of broader patterns across regions.

However, as the number of participants differed across regions, the absolute number of votes cast also varied. As a result, direct comparison based on raw vote counts would be misleading. To enable meaningful cross-regional comparison despite uneven participation levels, the results presented here are therefore expressed using normalized vote shares.

For each of the three capabilities, votes were normalized within regions such that the total number of votes cast by participants from a given region equals 100 per cent. This means that the figures show how each region distributed its overall assurance priorities across different measures, rather than reflecting absolute numbers of votes.

$$\textit{Vote share} = \frac{\textit{Votes for assurance measures}}{\textit{Total votes cast by region for capability}}$$

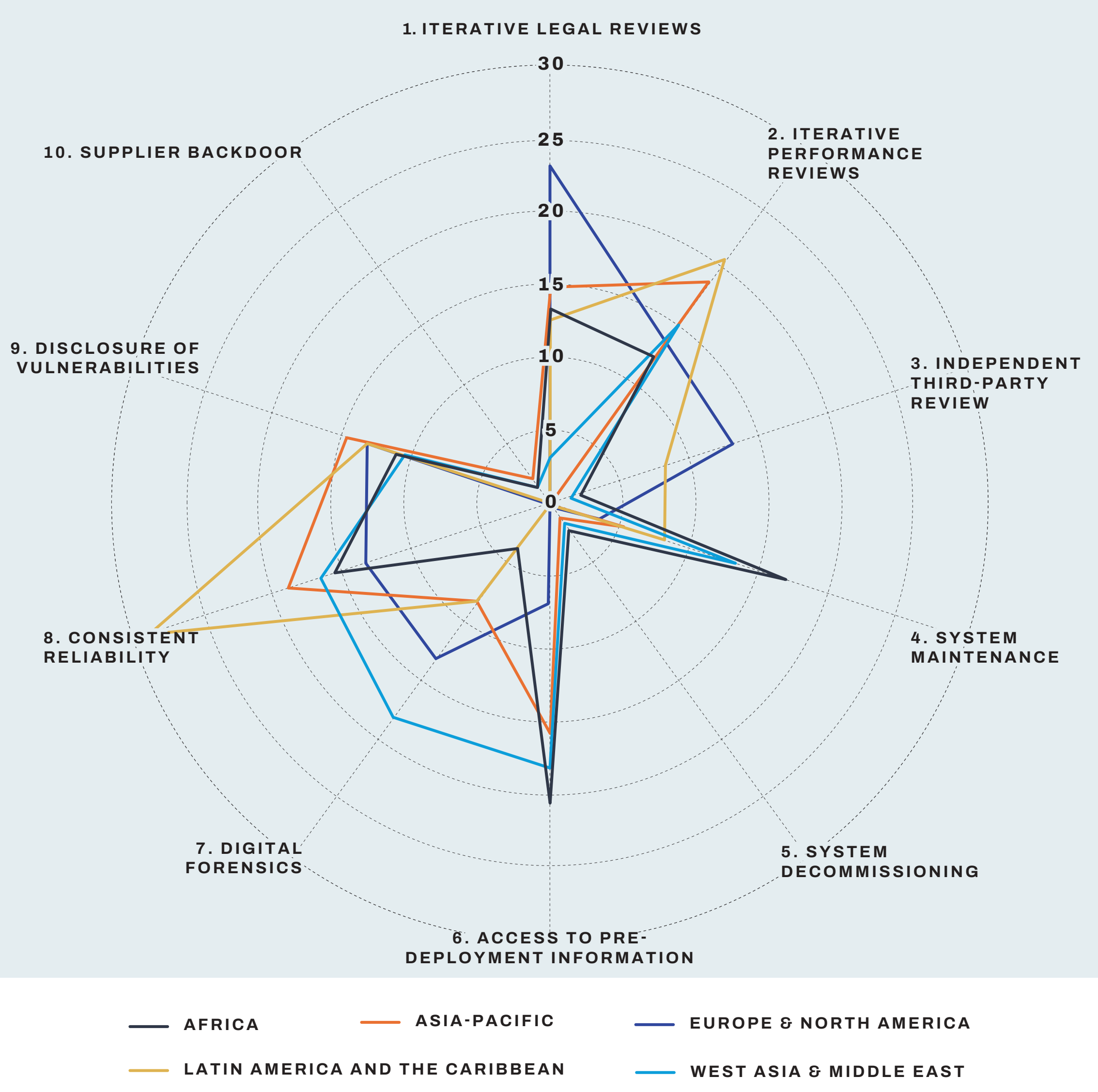
This approach ensures that each participant's vote carries equal weight, while accounting for differences in regional participation levels. The findings should be interpreted as indicative of priorities expressed during the consultations, rather than as representative of regional positions. The exact numbers used to generate each of the following graphs, along with the aggregated votes for each region, appear in the Annex at the end of this report.

4.1.1. Assurance priorities for AI-enabled electronic warfare capabilities

The first capability that participants voted on was the purchase of AI-enabled electronic warfare capabilities. Votes were cast as shown in Figure 1.

FIGURE 1.

Regional distribution of assurance priorities for AI-enabled electronic warfare capabilities (each region normalized to 100% of votes)



Key observations

Votes for AI-enabled electronic warfare generally concentrate most consistently on access to pre-deployment information (measure 6), complete control of the system's maintenance (measure 4) and consistent reliability (measure 8). Put together, these measures account for a substantial share of votes in every region, indicating a broadly shared emphasis and prioritization of operational robustness and reliability across the technology's life cycle.

Yet, while all regions allocate a meaningful share of votes to reliability-related measures, the relative balance between maintenance-focused measures and review-oriented measures varies across regions.

Nevertheless, in comparison with the two other capabilities, AI-enabled electronic warfare was the system for which maintenance- and sustainment-related assurance measures receive strongest and most consistent emphasis across all five regions. This suggests heightened sensitivity, by participants, towards the system's long-term resilience, long-term operability, potential degradation and, overall, its reliability. While many participants noted that this application may be less contentious than the other two under examination, it remains a critical one for mission success given the enabling nature of electronic warfare capabilities.

Conversely, independent third-party review (measure 3), control over the system's decommissioning (measure 5) and a supplier backdoor (measure 10) received comparatively low vote shares across all regions, with no region assigning a significant proportion of votes to these measures.

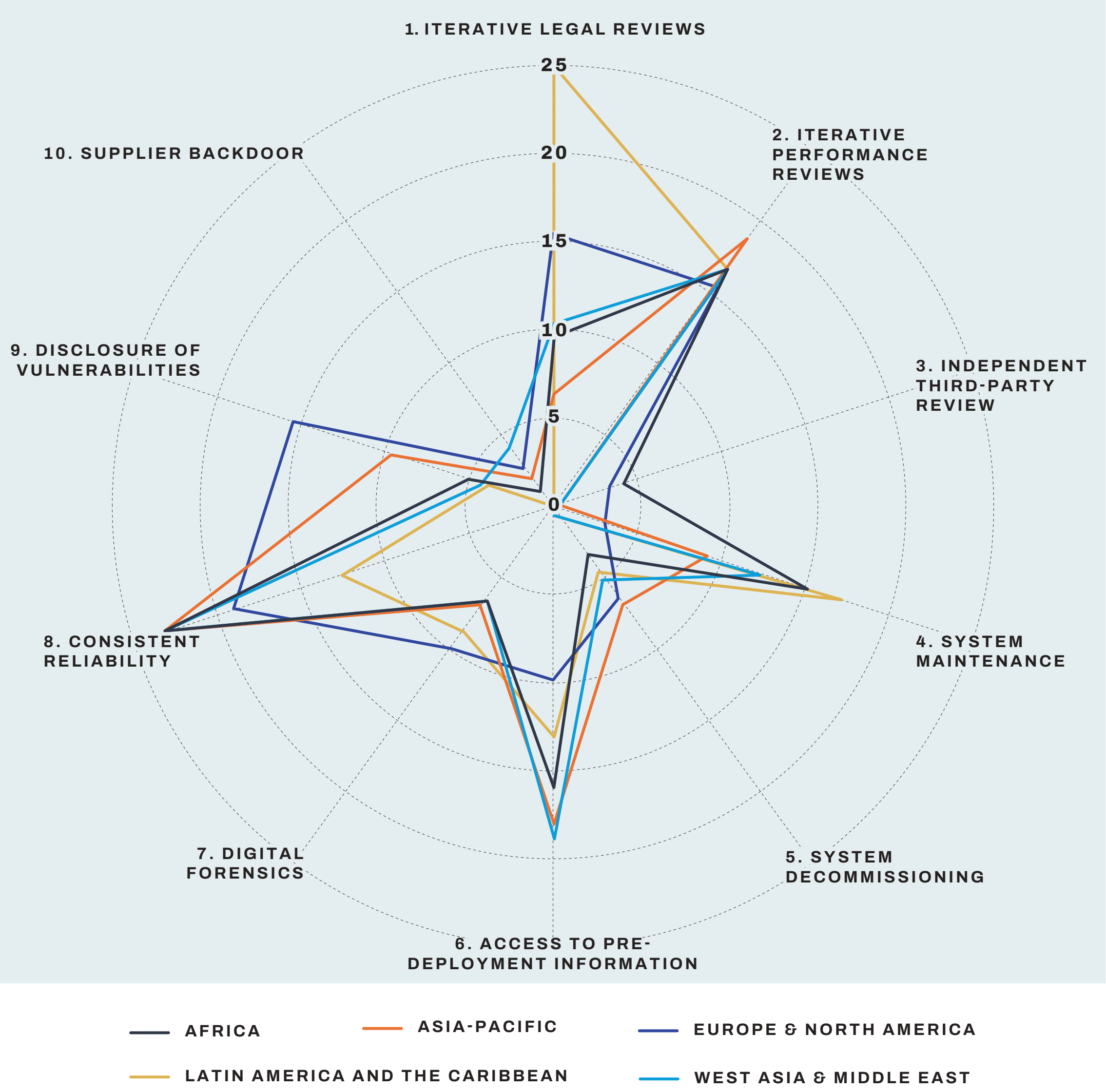
A soldier and a surveillance drone (generated with AI). Credit: Adobe Stock / Olga Gorkun.

4.1.2. Assurance priorities for swarm-based uncrewed capabilities for ISR

The second capability that participants voted on was the purchase of uncrewed stealth systems with AI-enabled autonomous navigation functions for ISR. Votes were cast as shown in Figure 2.

FIGURE 2.

Regional distribution of assurance priorities for swarm-based uncrewed capabilities for ISR (each region normalized to 100% of votes)



Key observations

For swarm-based uncrewed capabilities for ISR, participants across regions consistently allocated a high share of votes to consistent reliability (measure 8) and iterative performance reviews (measure 2). The concentration in these two measures suggests a shared concern related to the system's behaviour in dynamic operational environments, that is, in its use, its predictability and, ultimately, its reliability.

Furthermore, access to pre-deployment information (measure 6) was also a prominent measure across regions for this system. Indeed, many participants highlighted the risk that, because these systems would be used for ISR, there is a need to ensure that their training data reflects the local context and realities. Access to the training data, along with the parameters around which the systems were designed, tested and evaluated, would constitute a significant assurance and a preventative, risk-reduction measure to address these concerns.

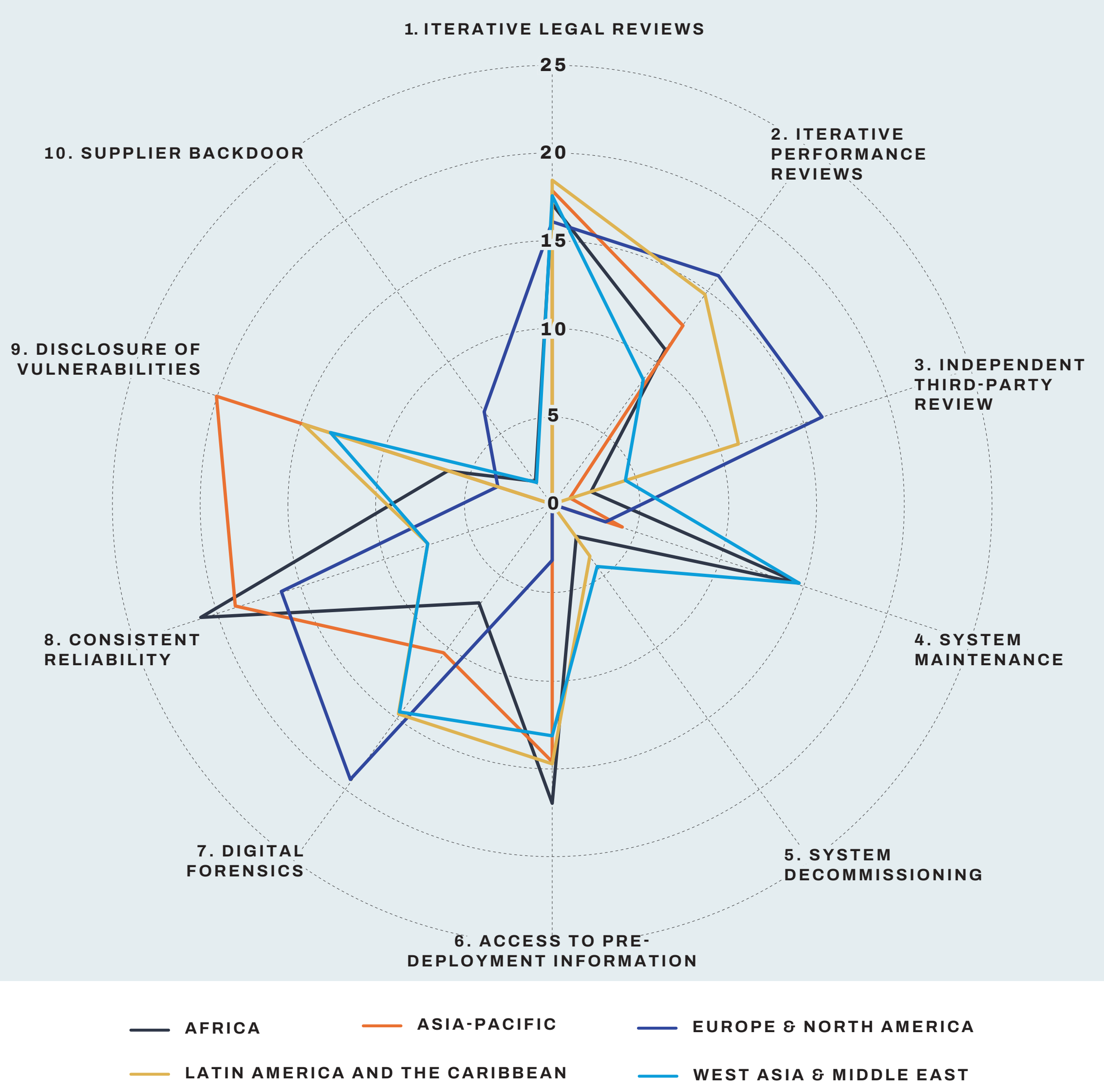
Overall, performance- and reliability-oriented measures (i.e., measures 2, 4 and 8) most clearly dominate across regions. This pattern suggests that participants primarily associate assurance for swarm and ISR systems with real-time behaviour and system coordination.

4.1.3. Assurance priorities for AI-enabled decision-support systems

The third and final capability that participants voted on was the purchase of an AI-enabled decision-support system as an upgrade to an existing battle-management software. Votes were cast as shown in Figure 3.

FIGURE 3.

Regional distribution of assurance priorities for an AI-enabled decision-support system (each region normalized to 100% of votes)



Key observations

Across regions, the AI-enabled DSS attract consistently high vote shares for iterative legal reviews (measure 1), access to pre-deployment information (measure 6) and consistent reliability (measure 8). Compared to the other capabilities, DSS shows the strongest overall emphasis on review- and information-related assurance measures. Reflecting this, a number of participants noted that iterative legal reviews are of particular importance due to the implications that an AI-enabled DSS may have for the application of IHL in operations. Some participants further emphasized that reliance on these technologies for decision-making, in the light of their criticality and potentially direct impact on civilians, would render compliance and the appropriate safeguards absolutely necessary conditions for their development, deployment and use. Furthermore, as for the swarm-based capabilities for ISR, some participants across regions noted their concerns around the training of a DSS and potential biases due to the lack of exposure to local realities and contexts.

In comparison to the other two capabilities, the share of votes allocated to disclosure of vulnerabilities in the DSS varied significantly across regions, with some regions assigning it a notably higher proportion of votes in this case. Differences are also visible in the relative emphasis placed on digital forensics, indicating varying regional approaches to traceability and post-use analysis for decision-support functions. Participants who prioritized digital forensics (measure 7) noted that, while preventative measures for reliability and compliance are important, downstream oversight is also critical to ensure accountability is preserved and can be established at all times through robust documentation processes and the conduct of effective investigations in response to incidents, particularly given the opaque nature of AI technologies (i.e., their black box nature).

Overall, relative to electronic warfare and swarm-based ISR, AI-enabled decision support is the capability for which compliance-, review- and transparency-oriented measures received the greatest emphasis from participants across regions. This resonates with the greater emphasis and association of AI-enabled DSSs directly with legal scrutiny, transparency of information and accountability than the two other capabilities. For those two capabilities, technical and operational considerations, including sustainment and system coordination, were more prominent without neglecting, at the same time, the importance of legal compliance.

4.1.4. General observations on assurance priorities for AI-enabled capabilities

The exercise provided a useful snapshot of how each region would typically approach a given suite of assurance measures and methods for different AI-enabled capabilities, subsequently offering a snapshot of regional priorities for the operationalization of responsible AI principles in the military domain. Overall, across all three capabilities, the results indicate that participants do not apply a uniform assurance logic to AI-enabled military systems. In fact, across all regions and capabilities, participants distributed their votes across a broad set of assurance measures rather than concentrating overwhelmingly on a single measure. This suggests that states tend to approach assurance as a portfolio of complementary measures, rather than as a question of identifying one dominant safeguard. Nevertheless, the relative prioritization of assurance measures shifts depending on the operational role of the capability: life cycle and sustainment considerations are more prominent for electronic warfare; performance and reliability considerations dominate for swarm-based ISR; and review- and information-centric measures are most pronounced for AI-enabled DSSs.

Participants noted, during the exercise, that while the measures presented can vary by nature – ranging from being focused on post-deployment accountability and traceability measures, to those that are performance-oriented – there is the general and shared understanding that responsible AI principles must be operationalized from the early stages of the technology. This understanding underscores the importance of formulating and implementing assurance measures from the outset. In emphasizing this, participants also noted the points of tension and trade-offs that may emerge between measures, along with the possible ways through which these may be reconciled. Each stakeholder will thus need to be clear on their ultimate objectives, along with the means available, as they assess the potential trade-offs.

BOX 5.

Possible points of tension in implementing assurance measures

Participants were given the opportunity to exchange views on the possible points of tension that may arise between the assurance measures presented as part of the tabletop exercise. One prominent example of such a tension that emerged throughout the consultations corresponds to the desire, by many participants, for a complete post-sale handover of the system's maintenance to the client (i.e., the purchasing state) from the supplier (measure 4). However, most of these same participants would note the lack of human and technical capacity to maintain cutting-edge technologies.

Another point of tension that participants noted is that, while third-party reviews could prove to be valuable not only to provide states with the expertise they lack, but also in enhancing public legitimacy, participants were concerned about confidentiality. Some of the possible solutions mentioned included the thorough vetting of the third-party reviewer, conferring the conduct of such reviews on other government entities or parliamentary bodies, or exploring potential technological solutions to preserve confidentiality.

A number of participants also noted that, while the exercise focused on the measures that would be prioritized depending on each circumstance, the de-prioritization of some of the indicated measures does not necessarily negate their importance. In fact, across regions and technologies, independent third-party review (measure 3), system decommissioning (measure 5) and supplier backdoor controls (measure 10) consistently received lower vote shares. While this does not imply that these measures are viewed as unimportant, in the context of prioritization under constrained timing, operational needs and financial limitations, a few participants tended to favour measures perceived as more directly linked to operational use and immediate governance needs. Some participants noted that, in the future, one approach to the division of priority could be to use risks and the magnitude of consequences as a basis.

Finally, the tabletop exercise was a useful means to reflect on how the local and regional context and realities influence the participants' approach to assurances. For instance, in the West Asian and Middle Eastern consultation, a number of participants strongly emphasized a complete handover of the system's maintenance (measure 4) with almost no votes for independent third-party reviews (measure 3). A number of participants correlated this to a perception that a significant share of R&D efforts in the region is conducted within government R&D agencies. In this context of reduced reliance on contractors, they underscored the importance of maintaining state authority and oversight, not only with respect to system data but also data processing, storing and subsequent deletion. These considerations were also linked to risks of potential external interference and exploitation of vulnerabilities, particularly in situations where systems operate near or across sensitive borders. Conversely, the European and North American consultation revealed a general sense of openness to independent third-party reviews with very little insistence on a complete handover of the system's maintenance. This may reflect not only existing practices but also the general nature of, and culture surrounding, public–private partnerships in the region.

4.2. Use of AI capabilities: Reflections on incident response, crisis management and risk reduction

The second part of the tabletop exercise provided an opportunity for participants to consider the implications of AI use, specifically in the context of incident response, crisis management and short-, mid- and long-term risk reduction. Participants were invited to conduct an after-action review following the use of an AI system, which initially functioned as instructed and as intended. However, participants were then invited to consider the implications of AI use in the military domain when, despite the system functioning as intended, its use still resulted in unintended consequences, both as a secondary effect of its use but also following a malfunction.

4.2.1. Substantive considerations

As participants reflected on incident response and risk reduction, two substantive considerations consistently emerged throughout the consultations.

First, participants raised concerns regarding **automation bias**, particularly the risk of over-dependency on AI-enabled systems in time-critical situations. They underscored the importance of education and training in advance of deployment, alongside the need for proper documentation of factors that may have influenced decision-making.

Second, participants also noted that **legal considerations** will be important across all stages of post-use assessments. These range from the extent to which international law was considered in the decision-making that led to the outcome, via the status of objects and individuals involved in the situation and the military necessity of the operation, to whether the applicable rules of engagement were compliant with international law.

4.2.2. Incident response and crisis management

The tabletop exercise offered an opportunity for participants to reflect on considerations and measures in the context of incident response and crisis management following the use of AI-enabled systems in the military domain. These considerations are: the need for holistic incident response and crisis management; malfunction management; accountability and responsibility; cooperation and post-crisis response; and public communication and international reporting.

First, participants underscored the **importance of establishing clear protocols and processes to respond holistically to incidents that involve the use of AI**. To this end, participants highlighted the importance of robust national instruments, processes and tools to frame the use of these technologies, including through checks and balances that apply to all stakeholders and agencies involved, as well as the establishment of due process mechanisms. Additional measures would range from the immediate suspension of the capability involved to the activation of designated units for forensics and investigations (including to retrace decision-making and legal assessments, identify the source of malfunction and establish potential disruption or interference with the system, and address the black box issue). These also include public relations management, especially if the incident is documented online on social media platforms and could potentially involve other states, as well as an examination of direct and indirect implications for national security, authority and stability. The latter can include secondary effects such as disinformation surrounding the use of these technologies, which may shape public opinion and contribute to further destabilization.

Participants further noted that, even for non-AI systems, malfunction can also happen with an inevitable margin of error for the platform or system to not act as intended (e.g., a missile not flying as intended). Nevertheless, considering AI's unique features, some participants underscored the necessity of having **clear protocols and processes to manage malfunctions** – either developing new, dedicated ones or adapting existing tools. Some of the considerations to take into account in this instance include the establishment of accountability and responsibility (including in relation to the supplier), examination of what would be considered as predictable and what could have been done to predict and anticipate some of the malfunctions in question, and reflection on what should be considered as "predictable" in the first place.

Participants also highlighted the importance of establishing **accountability and responsibility** – determining what it means in practice and what is expected from each stakeholder. This would include, for instance, reviewing contractual agreements with suppliers in order to establish eventual supplier accountability, and more generally establishing the distribution of responsibility in cases of incidents. While participants noted that it is difficult to find a one-size-fits-all solution, processes should be in place to ensure that there are no problems with establishing such accountability and responsibility, including in relation to clarifying the distribution of responsibility for financial damage and compensation resulting from the incident in question. Clarity on post-sale liability is particularly important for purchasers from the Global South since, without it, they could be left without effective legal avenues to raise grievances against (foreign) suppliers. Furthermore, participants also highlighted the value of multi-stakeholder engagement in this context and of co-development as a means to avoid misunderstandings and ensure alignment over time. This would include establishing clear expectations regarding measures to be taken by industry actors for accountability, remedy and reparations.

Furthermore, participants underscored the critical **importance of cooperation in post-crisis situations**, including the involvement of relevant stakeholders and multiparty interventions. This includes engagement with international organizations and agencies, as appropriate (e.g. the International Atomic Energy Agency (IAEA) in cases with a nuclear aspect) as well as with the technical community and industry. Participants also emphasized the importance of appropriate framing and raised questions regarding how to ensure that the right entity is designated to lead any cooperative process such as a joint investigation.

Finally, participants highlighted the value of **public communication and international reporting.** Nationally, issuing public statements to establish public trust and legitimacy will be important, particularly in relation to action by non-state armed groups. Participants also emphasized the importance of effective communication with concerned states, including neighbouring states and allies, to establish or restore trust. More broadly at the international level, participants noted that such incidents could serve as use cases for reporting to the international community, with a view to drawing lessons that inform ongoing discussions and processes.

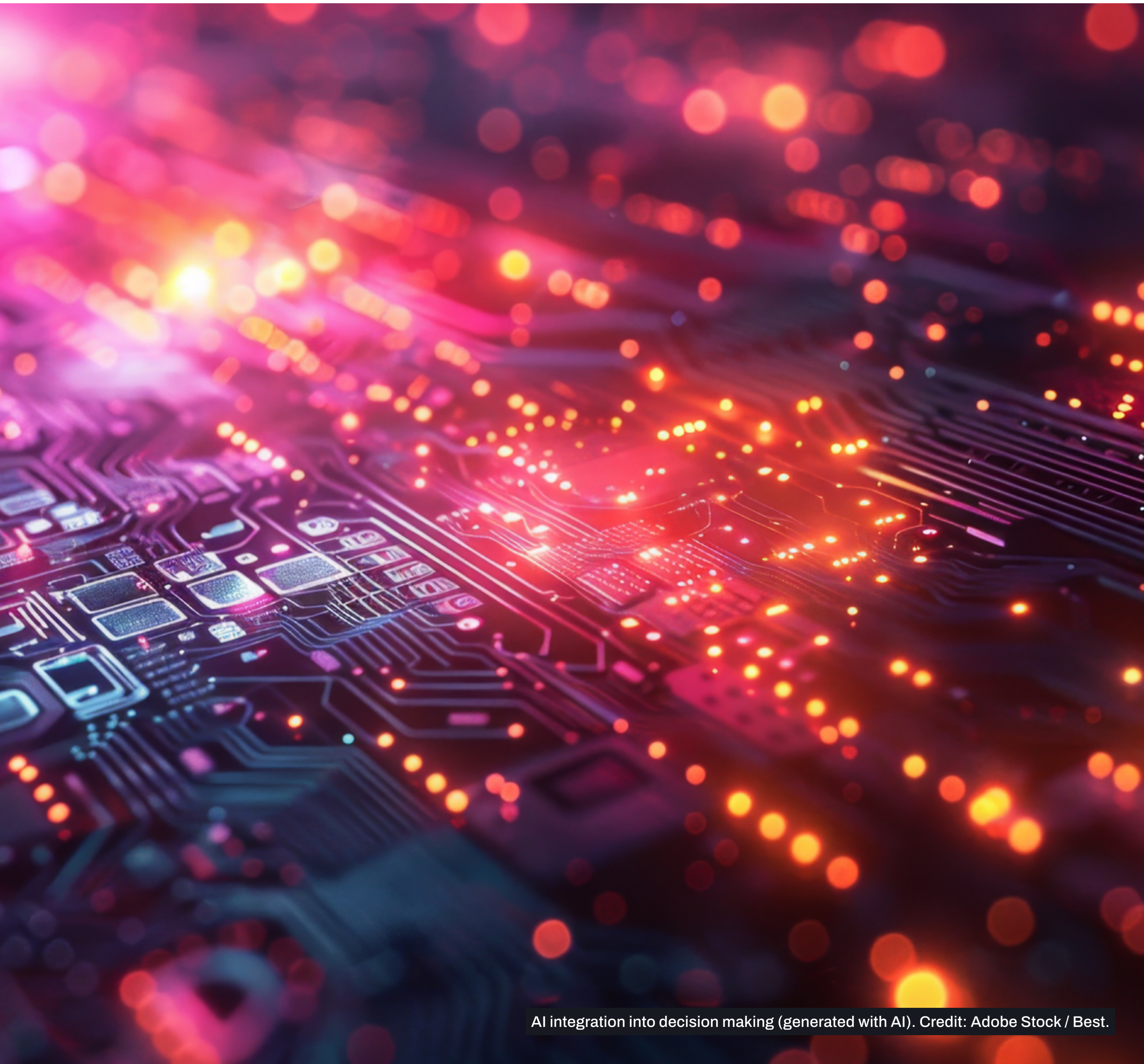
4.2.3. Measures for risk reduction

Another key component of the tabletop exercise was identifying measures that states could implement to reduce risks of incidents in the short, medium and long terms, particularly to prevent recurrence. Participants broadly identified four categories of measures for risk reduction following the exercise: technical measures and safeguards; design choices for UX; industry engagement; and long-term risk reduction and management.

First, participants emphasized the need to establish **concrete technical measures and safeguards** to implement responsible AI principles. Examples included the establishment of non-operation zones informed by law and policy, particularly around sensitive infrastructure and protected objects, as well as the importance of agreeing on relevant standards at the international level. Particular emphasis was placed on the need to understand the testing and evaluation parameters used for the technologies in question and, more broadly, the measures undertaken by the developers and industries more generally to embed international law, as well as local contextual considerations, at the design and development stages and beyond. Furthermore, participants emphasized on the need for transparency over the training data, particularly to ensure that the systems in question are exposed to, and are informed by, local socioeconomic contexts. Participants also highlighted the need to ensure system robustness, to embed kill switches and to clarify how reliability should be defined.

At a more specific level, participants underscored the importance of **design choices, particularly with regards to the system's interface, with UX in mind**. For instance, in the context of an AI-enabled DSS, participants highlighted the importance of critical information such as collateral damage estimates and the presence of civilians forming part of the interface that the user sees and interacts with, subsequently ensuring that the commander's decision-making is holistic and in compliance with international law. In fact, this "critical information" must be grounded in international law requirements, particularly IHL. Clarity will also be needed on the scope and definition of key terms and metrics appearing on the interface. For example, if the system provides an estimate of “success rate”, how it is defined would be important, as would some of the types of data used for such an estimate. If a system provides a risk assessment, the interface must then be clear on what information has been taken into account, the weight each bit of information carries and the source of data.

The tabletop exercise also highlighted the importance of **industry engagement**, not only with respect to establishing instances where the supplier may be held responsible, but also the importance of transparency in relation to how a procured system has been developed and trained, the training data used and labelling processes, among others. Any effort to implement responsible AI principles from the early stages of the technology’s life cycle will require action and commitment from the companies developing these systems; thus, clarity and international standards or guidance on what industry is doing, what it could be doing, and what it should be doing is seen as valuable.



AI integration into decision making (generated with AI). Credit: Adobe Stock / Best.

5. Looking back and looking ahead: Reflections and lessons from the REAIM journey

Since its inception and launch in February 2023 in The Hague, the REAIM initiative, with its series of activities and outputs, has provided a unique platform for dialogue and the exchange of ideas, best practices and perspectives between states and the wider multi-stakeholder community, including industry, civil society and academia. Since then, the technological state of the art and the policy landscape have evolved significantly, which calls for reflection and introspection on REAIM's journey and its role to further support governance efforts surrounding AI in the military domain. To this end, the consultations provided space for a twofold discussion. First, states were provided with an opportunity to look back and reflect on the activities undertaken since the inaugural summit, along with lived experiences and perspectives on its value and where there is room for improvement. Second, on that basis, states were then provided with an opportunity to look ahead, and reflect on their ambitions, perceptions and sentiment on the future direction of REAIM, where its added value lies in the future, and how it could best interact with and complement the work in other avenues, including that of the United Nations.

5.1. Looking back: Reflections on the REAIM journey

This segment of the consultations aimed to provide space for states to reflect on the REAIM journey since its formal launch with the inaugural summit in February 2023. The aim of that summit was to put AI in the military domain higher in political agendas, and subsequent efforts and activities of the REAIM initiative have consistently been driven by the desire to promote inclusivity and multi-stakeholder engagement. For instance, to this end the Netherlands established the Global Commission on Responsible AI in the Military Domain (GC REAIM) with a multi-stakeholder and multidisciplinary membership and with a mandate to formulate strategic recommendations within two years, which it captured through its strategic report.[14] The 2024 REAIM Regional Consultations provided a unique platform to capture regional perspectives, identifying both areas of convergence and areas of divergence in states' perceptions of the responsible development, deployment and use of AI in the military domain. The Republic of Korea subsequently captured these perspectives into the outcome document of the second REAIM Summit, the Blueprint for Action, by identifying the key principles underpinning responsible AI in the military domain.

The growing success of REAIM over the years has been contingent on a number of factors which, while applying differently across regions, could reveal good practices for its continued relevance, growth and sustainability. As such, in the consultations, states were given the

[14] Global Commission on Responsible Artificial Intelligence in the Military Domain, *Responsible by Design: Strategic Guidance Report on the Risks, Opportunities, and Governance of Artificial Intelligence in the Military Domain* (The Hague: The Hague Centre for Strategic Studies, 2025), https://hcss.nl/wp-content/uploads/2025/09/GC-REAIM-Strategic-Guidance-Report-Final-WEB.pdf.

opportunity to share some of the factors they perceive as underpinning the success of REAIM, which has been reflected, among other ways, through the endorsement of its outcome documents. These were further complemented by reflections on the impact that REAIM may have had, either directly or indirectly, on both national and regional policies in the context of AI governance in the military domain. Conversely, states were also given the opportunity to reflect on some of the challenges and limitations surrounding REAIM. However, whether they were procedural or substantive, states that have experienced such challenges or limitations have not necessarily totally rejected REAIM as an initiative. In fact, one key observation from these exchanges was that non-endorsement did not necessarily reflect a lack of political will to engage. Many states that did not endorse the Blueprint for Action are recurring participants in the regional consultations, even though they were not in agreement with its substance. These reflections on both success factors, impact and perceived challenges were used as a basis for subsequent discussions on recommendations for the way ahead, the key takeaways of which are described in Subsection 5.2.

5.1.1. Success factors and impact of REAIM

Reflecting back, states shared a host of factors that underpin their support for REAIM as an initiative, along with support for either of the summits' outcome documents. Many participants noted that REAIM constitutes a useful starting point and platform to raise awareness on the issues surrounding AI in the military domain. Specifically, most of the voices sharing their alignment with and subsequent endorsement of REAIM based this on the space that it afforded for multi-stakeholder engagement and to discuss, informally, the topic of responsible AI in the military domain. Multi-stakeholder participation was indeed generally perceived as critical in all regions, and there has been consistent acknowledgment of the usefulness of REAIM to provide a platform to better understand what actions and measures could be implemented to operationalize responsible AI principles, in addition to the exchange of information, best practices and awareness-raising on underexplored issues. In fact, states welcomed the opportunity to take stock of the current state of affairs not only at the policy level but also with respect to technological developments. The ability to engage with industry was seen as particularly valuable, not only to understand the latest state of the art, but also to pave the way for engaging constructively in the implementation and operationalization of responsible AI principles by the private sector.

Substantively, a number of states that have endorsed at least one of the outcome documents were aligned with their content, with all regional consultations noting general agreement with respect to legal compliance, ethical considerations, multi-stakeholder engagement, the adoption of a life cycle approach and capacity-building efforts. In addition, a number of participants, notably from the African consultation, welcomed the acknowledgment of concerns in relation to proliferation and use by NSAGs. Additionally, a number of participants, particularly from the European and North American consultation, noted positively the outcome documents' approach to risks, including both foreseen and unforeseen risks and their multidimensional nature. Participants also welcomed the emphasis placed on the sharing of lessons learned, including with and from the private sector, and noted positively the involvement of both Ministries of Foreign Affairs and Ministries of Defence in the deliberations.

Beyond alignment on the substance, one factor that a number of states noted as constituting a key enabler for support and subsequent endorsement of the outcome documents was the pre-existence of a national approach, policy or general position on the use of AI and innovation. Indeed, these states have flagged their value in facilitating alignment with the Call to Action or the Blueprint for Action, and international efforts more generally; meanwhile, states without such a national position noted that absence as one of the driving factors standing in the way of endorsement.

Building on these reflections, states shared their views on the impact that REAIM has had both in their region and at the national level. A number of participants noted that both REAIM outcome documents have been instrumental in informing internal processes and thinking. Both the Call to Action and the Blueprint for Action were, in fact, reported as informing ongoing efforts to develop and implement national policies and strategies. The 2025 African regional consultation revealed steady progress in ongoing efforts to develop, implement or review national strategies addressing AI in the military domain. These efforts included both the review of existing policy frameworks (e.g., cybersecurity doctrines) and the development of dedicated AI strategies. Many of the strategies developed across the continent, and initiatives towards their formulation, emerged between 2024 and 2025, which was seen by participants as attesting to the impact of the 2024 REAIM Regional Consultations in raising awareness and catalysing national-level engagement on these issues. Beyond the African consultation, a number of other participants noted the development of guidelines for their armed forces' strategic approach to AI in the military domain. A few states noted that the impact must be assessed beyond the outcome documents: the REAIM Regional Consultations were generally seen as having provided room for an exchange of perspective, experiences and good practices, and as helping to identify possible avenues for regional cooperation. These regional conversations, in turn, catalyse further reflections at the national level, which help to further enrich collective, cross-agency efforts to develop a national strategy for the governance of these technologies. These developments thus attest to the impact of the collective work undertaken within REAIM by both states and the broader multi-stakeholder community in informing national decision-making.

Finally, participants further noted that REAIM discussions and outputs have informed Article 36 legal reviews and contributed more broadly to national efforts to comply with international law. In addition, REAIM was seen as having helped mobilize stakeholders to engage at the cross-institutional level with relevant principles and processes, including Article 36 legal reviews, secure data practices and certification. These developments attest to the growing focus on the operationalization of the principles surrounding the responsible development and use of these technologies, with the shared perception that REAIM has acted as a catalyst for further progress in this space.

5.1.2. Perceived challenges surrounding REAIM and its outcome documents

Conversely, states also shared a host of obstacles to their support for the outcome documents, in addition to noting perceived limitations surrounding REAIM as an initiative.

Substantively, a number of participants stressed existing tensions between the previous REAIM outcome documents on the one hand and regional and national priorities on the other hand. For instance, some participants noted the need for a clearer acknowledgment of the AI divide, which they saw as insufficiently addressed in previous outcome documents, and the need to address this issue in subsequent REAIM outcome documents in addition to a stronger emphasis on international cooperation. Some participants expressed the desire to see further granularity in relation to the human element, for instance through a clearer articulation of what forms of human involvement are acceptable, allowed, necessary or unacceptable, depending on the specific application. A few noted a particular emphasis on preserving human control, while others flagged differences in prioritization and the inherent difficulty of reconciling these variations in a document with global ambitions. For example, while it is noted above that many state representatives from Europe and North America noted positively the REAIM outcome documents' approach to risks, others and particularly from Latin America and the Caribbean presented its approach as grounds for non-endorsement. Specifically, Latin American and Caribbean states would have preferred seeing further emphasis on a proactive approach to preventing and mitigating the risks associated with the development, deployment and use of AI in the military domain.

Another area of misalignment has emerged with respect to the substance captured within the Blueprint for Action, particularly for a number of states in Latin America and the Caribbean in relation to the reference to nuclear weapons. In the light of the region's historical advocacy against nuclear weapons, both in the context of the TPNW and through the establishment of a nuclear weapon-free zone, a number of states voiced sensitivities and concerns that such language could legitimize the existence of these weapons, which would stand at odds with the perceived stance of the region on this issue. However, it is however important to note that, conversely, a number of representatives from other regions saw REAIM as a potential avenue to address, informally, the nexus between AI and nuclear weapon issues, qualifying these as a "low hanging fruit" where more efforts could be done by, for example, facilitating dialogue and regional CBMs.

BOX 6.

Fundamental misalignment of approaches: What should "responsible AI" represent in the international context?

During the consultations, a number of states voiced fundamental misalignment with respect to their approach to the concept of "responsible" AI in the military domain. Against the perception that international discussions should not incentivize the use of lethal force, even less so within the United Nations, there is a desire from these states to see greater focus in the deliberations on preventing the use of these technologies to enable or even create more lethal capabilities. At the other end of the spectrum, other states noted that, considering the nature of today's warfare and technological advancements, there is a perception that pursuing the development, deployment and use of AI capabilities constitutes a "responsible" imperative. However, this has the caveat that such adoption of the technology must be grounded in international law and the principles that underpin the discussions around "responsible AI". While there is no universally acknowledged right answer or approach to reconcile these differences, states at both ends of the spectrum and in between noted the need for clarity with respect to REAIM's long-term strategy, which could then encapsulate the answers to this fundamental dilemma and subsequently indicate, firmly, whether states' own national and regional approach would be aligned with the initiative in the long term.

To further complement these reflections, states also shared a host of procedural issues and concerns surrounding REAIM at both international and national levels. In fact, a significant number of states across regions voiced concerns in relation to the long-term sustainability of a non-United Nations forum to provide the space for inter-state deliberations. While a number of states argued that the United Nations provides an inclusive forum with clarity on its rules of procedure, thus allowing for inclusivity in the negotiations of any output, there is the perception that the REAIM initiative would benefit from further procedural clarity and consistency. In addition, there is an insistence by a number of states across regions to maintain formal processes and deliberations within the United Nations, with this stance further amplified by the prospect of non-United Nations discussions (including, but not exclusively REAIM) influencing and having an impact on states' national positions and international commitments. This perception brings to the fore another set of misalignments with respect to the deliberative dimension of REAIM, marked against a backdrop of growing discussions within the First Committee of the United Nations General Assembly. These states have indeed shared their aim to prioritize the latter, particularly when the financial and human resources to attend and participate meaningfully in these international discussions are limited. A number of participants pointed as well to a perceived lack of clarity regarding the current and intended relationship between REAIM discussions and parallel processes and discussions in the context of both the First Committee and the process on LAWS, and more generally with ongoing discussions within the United Nations and regional organizations. Thus, the question of REAIM's sustainability and appropriateness, as an initiative, for future inter-state deliberations on the development, deployment and use of these technologies has thus been raised – at least in its current format, with a significant number of states noting that in order to maintain its place and relevance, there may need to be changes to its structure and objectives to provide more space for the United Nations.

In the same vein, questions were raised with regards to participation and invitations. The absence of certain states from the discussions was particularly noted, with questions as to how to improve future engagement. A few states have, in this regard, expressed the desire to understand how the constituency of each regional consultation, and the invitations to the REAIM Summits, are determined by the hosts. Noting as well that the background and level of participation to these discussions so far remain at the discretion of the invited states, a number of states have expressed the need to increase efforts to engage with the defence sector, which many felt had been left out of the conversation over the past few years. A few states were also of the view that certain topics may be too sensitive to be handled in policy circles and must remain exclusively in the hands of defence officials – thus noting the value of military-to-military dialogue; yet a number of other states were of the view that a cross-sectoral platform for exchange and mutual learning remains necessary. These differences further underscore the value of clarity over REAIM's long-term objectives and strategy, which would then subsequently shape its future activities and structure. This sentiment in favour of inclusive platforms was further amplified by the reported perception of a number of participants regarding the Western-led nature of the initiative, which in some cases gave rise to concerns about potential double standards. Participants also highlighted concerns related to the non-binding nature of the outcome document. This may result in asymmetrical and discretionary implementation, particularly in the absence of differentiated expectations between states with advanced capabilities to develop AI systems and those without.

Additionally, in the absence of a dedicated mechanism to ensure institutional memory both internationally and nationally, particularly factoring in the rotating nature of file-holders within governmental institutions, any form of commitment (or intention towards such commitment) may prove to be difficult for a number of states. Institutional memory was, in fact, highlighted as key for long-term coherence and consistency, the absence of which stands in the way of endorsement for many states. This issue is further compounded by the perceived need for additional efforts to maintain engagement and for additional trust-building in between REAIM Summits. In fact, in the same vein, a number of states also noted a lack of clarity regarding which international entity should serve as the primary focal point for REAIM-related affairs that may extend beyond political queries, for example, in relation to past summits, the regional consultations and the substance more generally.

On the process behind the summits' outcome documents, the reflections were relatively divided. A number of states expressed the need for the penholder to increase efforts to reach out to states to review, at the earliest, the zero draft of the outcome document, in addition to providing an opportunity to share feedback and review its content. However, states also noted the risk that too many inputs may be too difficult for the host to reconcile. A few states reported being involved in the pre-summit online consultations on the draft of the 2024 Blueprint for Action and appreciated that opportunity; yet a few states reported only seeing the text of the document at the summit. In addition, participants underscored the need to factor in states' own national processes and procedures for endorsement. These processes can involve multiple steps, multiple actors and multiple levels of clearances; as such, ample time is needed to enable states to navigate these processes upon the circulation of the outcome document's final draft. At times, some states reported the absence of a clearly designated national entity, ministry or department with both the capacity to deliberate and formal responsibility for AI in the military domain. At times, parallel invitations addressed to different ministries of the same government were reported to have created internal tensions, compounded by the heavy bureaucratic procedures required to approve endorsement. A number of participants also noted policy readiness considerations (or the lack thereof), particularly citing currently pending decisions for the completion of national AI strategies or a national policy, which are perceived as an imperative for many before endorsing the document. These challenges were further exacerbated by practical challenges, including visa-related issues for attendance at the summits and language barriers associated with reliance on English.

5.2. Looking ahead: Reflections for the 2026 REAIM Summit and beyond

Building on the reflections surrounding the REAIM journey, this segment of the consultations aimed to provide space for states to reflect on the way ahead and to share not only substantive areas that ought to be further prioritized at both the 2026 Summit and its adjacent activities (e.g., future iterations of the regional consultations), but also concrete recommendations for action that could be taken as food-for-thought within REAIM and beyond.

5.2.1. Substantive recommendations

States shared a number of substantive areas that ought to be further prioritized, not only at the 2026 Summit and in its outcome document, but also for future work forming part of the REAIM initiative and beyond. These recommendations can, broadly, be organized around five themes:

- Operationalization of responsible AI principles in the military domain
- Technological considerations
- Wider security issues
- Risk management and confidence-building measures
- Governance considerations

Operationalization of responsible AI principles in the military domain

First, a significant proportion of participants stressed **the need to move beyond high-level commitments by operationalizing the principles** set out in the 2023 Call to Action and the 2024 Blueprint for Action. These would include, for instance, a stronger focus on life cycle management and more targeted language addressing the operational dimension of AI in the military domain, such as in the context of decision-support systems. Participants underscored in particular the need for greater clarity on the role of humans in decision-making, including when and why the human element may be required within the kill chain, and whether actions such as “pressing a button” can meaningfully be considered as constituting the human element required for the operationalization of responsible AI principles. In this sense, attention was also drawn to the importance of use cases, questions of technical feasibility and the identification of practical obstacles to implementation. In addition, reflections on how implementation could or should unfold at the national and regional levels would constitute a critical element of these operationalization efforts.

Another critical aspect to the operationalization of responsible AI principles rests on meaningful engagement with the private sector, which participants of all regional consultations highlighted as critical. In fact, participants discussed the importance of strengthening industry engagement, including ways to involve industry more effectively, incentivize participation, and engage with the private sector to help translate policy principles and legal obligations into concrete and implementable measures. Such engagement could also be leveraged to nurture a culture of responsible innovation, addressing the assumption that governance and regulation would constitute a challenge to technological advancement. Laying out the benefits and incentives of promoting a culture of responsible innovation with industry will, in this sense, be critical, while ensuring that these efforts will ultimately be in the interest of states. This is particularly important amid calls for guidance for the responsible procurement of AI in the military domain as a key means to operationalize these principles, particularly from the Global South where states generally perceive themselves as purchasers of these technologies. This guidance may follow from, for instance, an established framework for responsible procurement. Guardrails and measures can also be established to ensure that internationally procured technologies were developed, tested and evaluated by embedding regional, national and local data and that the technologies reflect their values to prevent harmful biases and outcomes. The need to engage meaningfully with industry in between REAIM Summits was also emphasized, noting the need to move beyond engagement towards a collective reflection of “what good looks like”. More generally, beyond industry, participants highlighted the importance of including the multi-stakeholder community in these discussions and dedicating more reflections to effective engagement, including with academia, research institutes and civil society organizations. Such engagement would be particularly important when considering some of the underexplored yet important wider implications of the development, deployment and use of AI in the military domain, including gender and harmful biases.

Finally, another key aspect of operationalization that states reported being keen on unpacking further pertains to legal compliance. Specifically, respect for international law sits at the core of the participants’ reflections as to how the operationalization of responsible AI principles could be conducted. States in Latin America and the Caribbean have presented this centrality as a logical stance and continuation of the region’s historical role in shaping and promoting respect and compliance. Participants particularly noted civilian protection and the principle of humanity as a core to underpin international legal frameworks, which must then subsequently guide and frame ongoing and future endeavours to operationalize responsible AI in the military domain. Participants emphasized the need to clarify how international law could apply in the context of AI in the military domain, with a particular emphasis on the measures necessary to ensure compliance with applicable legal frameworks and how these relate to responsible AI principles. In this respect, participants underscored the importance of developing appropriate oversight and verification mechanisms to support the enforcement and implementation of responsible AI principles, including compliance.

Technological considerations

Second, participants noted the importance of ensuring that REAIM, and international efforts more generally, **keep pace with technological issues and technical developments.** In fact, one common thread that participants would like to see maintained in the future is the importance of grounding efforts in evidence to ensure their timeliness, feasibility and sustainability. For instance, participants underscored the importance of governance across the life cycle of AI systems, including effective life cycle management and safeguards to address and mitigate the potentially harmful implications of bias. To this end, consideration must be given to the technical dimensions of system resilience and durability, in addition to their effectiveness, including in relation to the hardware and energy consumption.

Additionally, one key and recurring aspect of the consultations was transparency, which a number of states are keen to unpack further. Participants highlighted the need to clarify how considerations related to privacy and transparency could apply in the military domain, including through technological means. In fact, some participants, particularly from West Asia and the Middle East, noted the potential tension that may arise between, on the one hand, national security and military imperatives and, on the other, calls for transparency, with both being made more complex by privacy requirements and how they should be implemented in military settings. Additionally, emphasis was placed on the importance of traceability, predictability and reliability of AI-enabled systems, all of which form key aspects of transparency and prompt the above questions. Transparency would also draw in the question of access to information on the technology's capabilities and limitations as well as the development, testing and evaluation parameters that may be of relevance for subsequent deployment and use. To this end, some states argued that efforts should be dedicated to auditing systems throughout the system's life cycle, although the issue of access to these solutions has come to the fore. Some of the solutions explored include third-party reviews and robust documentation processes.

Another key technological aspect is data governance, which constitutes an adjacent but key theme to transparency, at times even argued to be a critical enabler. Factoring in existing data-governance frameworks that may be applicable, including from the civilian domain, participants viewed dedicated governance efforts for data as being critical for future deliberations around AI in the military domain. These efforts should include, among other things, measures for transparency and traceability over the source and processing history of training and testing data, as well as dedicated frameworks and processes on the generation and use of synthetic data to train, test and use AI technologies.

Additionally, there is a desire from states to explore further the convergence between AI and other technological fields, notably the cyber–AI nexus. While this examination of convergences is a critical part of risk assessments, it also offers ample opportunity for greater cross-pollination with longer-running cybersecurity discussions. Another technological convergence that states are keen to see emphasized more is to draw lessons from civilian AI efforts, such as approaches to recognizing and labelling AI-generated content, which can be critical to addressing disinformation threats.

Finally, participants were also keen to leverage REAIM's multi-stakeholder nature to further unpack emerging concepts such as frugality and the implications of agentic AI. At the frontier, a number of participants also noted the need to consider the implications of AGI on international peace and security, and how to best approach this topic at the international and national levels.

Wider security issues

Third, participants reflected on the need to factor in the impact of AI on **broader security considerations** and their implications. This is critical for participants from regions where non-international armed conflicts and situations of violence below the threshold of armed conflict, along with the broader mandate of certain armed forces and other defence and security agencies. As such, participants noted the importance, for instance, of considering the impact of AI on environmental security and human security, particularly when these technologies are deployed in environments where the presence of civilians and natural resources is expected. Considering the Latin American and Caribbean region's armed forces and other defence and security agencies have a wider mandate that may touch upon these "other" security dimensions through non-wartime duties, it is important that these considerations are taken into account in future international discussions and deliberations.

Furthermore, participants highlighted the need to address proliferation risks associated with access to AI by NSAGs, organized criminal groups, gangs and insurgent groups, as well as broader proliferation concerns. This is particularly important in the light of the dual-use nature of AI technologies, their convergence with areas such as cybersecurity, their application in back functions and logistics to enable these groups' operations, and their potential proliferation from industry. Risks associated with disinformation, information warfare and subsequent destabilization at the national and regional levels were further underscored. The importance of guidance on forensics and investigation, crisis management and a stronger focus on the peace and security dimensions were thus highlighted in this context. As such, states are keen to see greater focus on this topic at REAIM, notably in the light of the multi-stakeholder nature of its summits.

Risk management and confidence-building measures

Fourth, participants highlighted the **importance of risk management and confidence-building measures**. In fact, there is the perception that REAIM could offer an informal yet valuable space to further unpack some of the risk-management measures that could be developed and adopted to address risks associated with the development, deployment and use of A Iin the military domain. These include, for instance, the formulation of technical safeguards, assurance and accountability mechanisms, and guidance on how to conduct risk assessments at the national level, which must be developed while nurturing innovation and ensuring access to beneficial technologies. Furthermore, a number of states were keen to see further work to develop measures to address risks of inadvertent escalation, including scenarios in which AI-accelerated conventional escalation could lead to nuclear escalation. In fact, a number of participants were keen to see further emphasis on addressing particularly sensitive domains, including the intersection between AI and nuclear-related issues, while acknowledging the sensitivities associated with such discussions as noted in Subsection 5.1. In fact, for a number of states, such sensitivities must be acknowledged explicitly in future REAIM outputs, in addition to some of the stronger positions on this issue. Additionally, a number of states expressed their desire to see greater focus or emphasis on civilian protection and the prevention of harm. To this end, the development of risk-management frameworks and mechanisms, along with CBMs, factoring in the inherent sociotechnical issues associated with AI, could be considered within the remit of REAIM.

Generally, CBMs were a favoured topic among participants across regions. Some of the measures explored included incident reporting (drawing inspiration from other sectors where such mechanisms are in place, including cyber and the maritime domains), third-party observation and reviews, red-teaming, data set evaluations, auditing, the identification of vulnerabilities and digital forensics. However, states noted that efforts to develop CBMs must be done across geopolitical blocs, underlining the importance of inclusivity and the provision of a neutral platform for dialogue. Furthermore, participants emphasized the value of regional CBMs, which REAIM could support and facilitate the development of. They also emphasized the importance of creating space for regional solutions, including region-specific countermeasures and policies to address localized AI-related threats, as well as the continued relevance of multi-stakeholder engagement.

Governance considerations

Fifth and last, but not least, states expressed the desire to see REAIM engage meaningfully with fundamental **governance** questions. At the international level, questions were raised regarding the future of REAIM, and how it is expected to interact with United Nations initiatives and deliberations, along with other relevant processes – with inclusivity and effectiveness at the heart of these considerations, particularly against concerns over the possible duplication of efforts. Relatedly, states placed an emphasis on the need to establish clear linkages with existing disarmament frameworks, including those related to weapons of mass destruction. In addition, participants stressed the importance of greater clarity regarding the relationship between discussions on AI in the military domain and parallel processes on LAWS, with conscious efforts to ensure coherence across forums, initiatives and processes. Some states have, however, urged caution with respect to the absence of clarity as to how LAWS and AI in the military domain interact. As such, participants also emphasized the need for clarity regarding both the immediate and long-term objectives of REAIM, as well as what these objectives require and imply for states. Furthermore, participants reflected on the appropriate allocation of issues across international, regional and national levels. Some noted in particular that the 2026 Summit outcome document should remain relatively high-level, while others were of the view that granularity in the outcome document is essential for international outputs to inform national efforts. In this context, prioritizing institutional knowledge will be important to ensure the sustainability and longevity of REAIM as an initiative and the efforts that stem out of it, along with clarity and specificity regarding the distribution of responsibilities between states and the multi-stakeholder community in implementing and operationalizing responsible AI principles.

Finally, participants highlighted the importance of furthering targeted capacity-building efforts. Relatedly, they reflected on what capacity-building concretely means and requires in the context of responsible AI in the military domain: beyond the need for multi-disciplinary training across domains and sectors, participants noted the importance of developing appropriate guidance to frame the development, deployment and use of such technologies. Leveraging the multi-stakeholder nature of the REAIM community, such guidance would span the technology's life cycle, from development, testing and evaluation, via procurement (i.e. how to "procure responsibly"), to deployment, use and decommissioning. To this end, participants noted the importance of equitable access to hardware and other enabling technologies, and they expressed the desire to leverage REAIM as a platform to unpack what international and regional cooperation could look like to meet these needs through responsible transfers of technology in the light of the persistence of the digital divide. Participants also discussed the prospect of developing tools and frameworks for states to assess and better understand existing gaps in capacity across domains, from testing and evaluation practices to the development of AI-enabled systems, which could be further explored in subsequent REAIM Summits and other related activities.

5.2.2. Recommendations for action

Building on these substantive recommendations, participants shared a series of actionable recommendations for future activities within REAIM and beyond. Generally, these recommendations can be structured around seven themes: capacity-building; knowledge-building; facilitated information-sharing; dialogue; industry engagement; evidence-based activities; and the institutionalization of REAIM. The key, concrete recommendations formulated by states during the regional consultations are laid out in Table 1:

TABLE 1.

Recommendations for future action within REAIM and beyond

CAPACITY-BUILDING

- Adopt a "train-the-trainers" approach to capacity-building and training efforts and initiatives
- Establish or designate a neutral coordinating body within the United Nations as custodian and facilitator for regional capacity-building in close collaboration with regional organizations and states
- Acknowledge the UNIDIR Guidelines for the Development of a National Strategy on AI in Security and Defence,[15] and facilitate the development of guidance for states to develop a risk-assessment framework
- Develop an indexing or benchmarking system led by the United Nations to assess capacity-building requirements; measure national readiness, progress and needs; and evaluate assistance required through regional and international cooperation

KNOWLEDGE-BUILDING

- Support efforts to develop a compendium of confidence-building measures across regions and improved approaches to CBM reporting
- Mandate the development of targeted lexicons, including on relevant key terminologies across languages and sectors
- Develop reference documents and guidelines to support AI development and procurement through the formulation of baseline expectations for suppliers, grounded in international law and agreed principles around responsible AI in the military domain
- Co-develop global standards for design and development with the technological community

FACILITATED INFORMATION-SHARING

- Establish a platform to facilitate the development of a repository of good practices for the operationalization of responsible AI principles in the military domain
- Facilitate the establishment of a network of centres of excellence, research centres and academic institutions with regional hubs to create a coordinated epistemic community working to implement responsible AI principles across domains and levels

DIALOGUE

- Preserve multi-stakeholder participation and cross-disciplinary dialogue, including through increased efforts to consolidate policy deliberations with evidence and added perspectives
- Issue invitations for regional organizations to attend and contribute to future REAIM activities, including subsequent REAIM Summits and regional consultations
- Identify regional champions to ensure the integration of local contexts and realities in future efforts at the international level
- Facilitate military-to-military dialogues

[15] Security & 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/.

INDUSTRY ENGAGEMENT

- Support efforts to co-design guiding principles with industry actors to translate agreed principles of responsible AI into operationalization measures
- Establish and maintain an industry network across sectors and geographies, including technology companies, defence contractors, suppliers, intermediary companies, and small and medium enterprises (SMEs) to facilitate dialogue and the collective development of guiding principles and good practices for the implementation of responsible AI principles

EVIDENCE-BASED APPROACH

- Facilitate the compilation of use cases from armed conflict, hybrid and peacetime settings to ground discussions in concrete evidence and applications, to be voluntary submitted by states, companies and other developers
- Conduct regular tabletop and foresight exercises within and across regions, including with the multi-stakeholder community
- Provide space for technology demonstrations to raise awareness of products under development that align with responsible AI principles

INSTITUTIONALIZATION OF REAIM

- Establish a clear process and the infrastructure necessary to facilitate the transfer of institutional knowledge to states, including those newly engaged with REAIM as an initiative, and the wider multi-stakeholder community
- Establish a dedicated and clearly identified international focal point for all REAIM-related matters, including to help ensure coordination and coherence with ongoing and future international deliberations and processes within the United Nations
- Formulate a dedicated strategy for REAIM as an initiative, laying out its short-, medium- and long-term goals and intended role in the international landscape of AI governance in the military domain, acknowledging the emerging deliberations within the United Nations

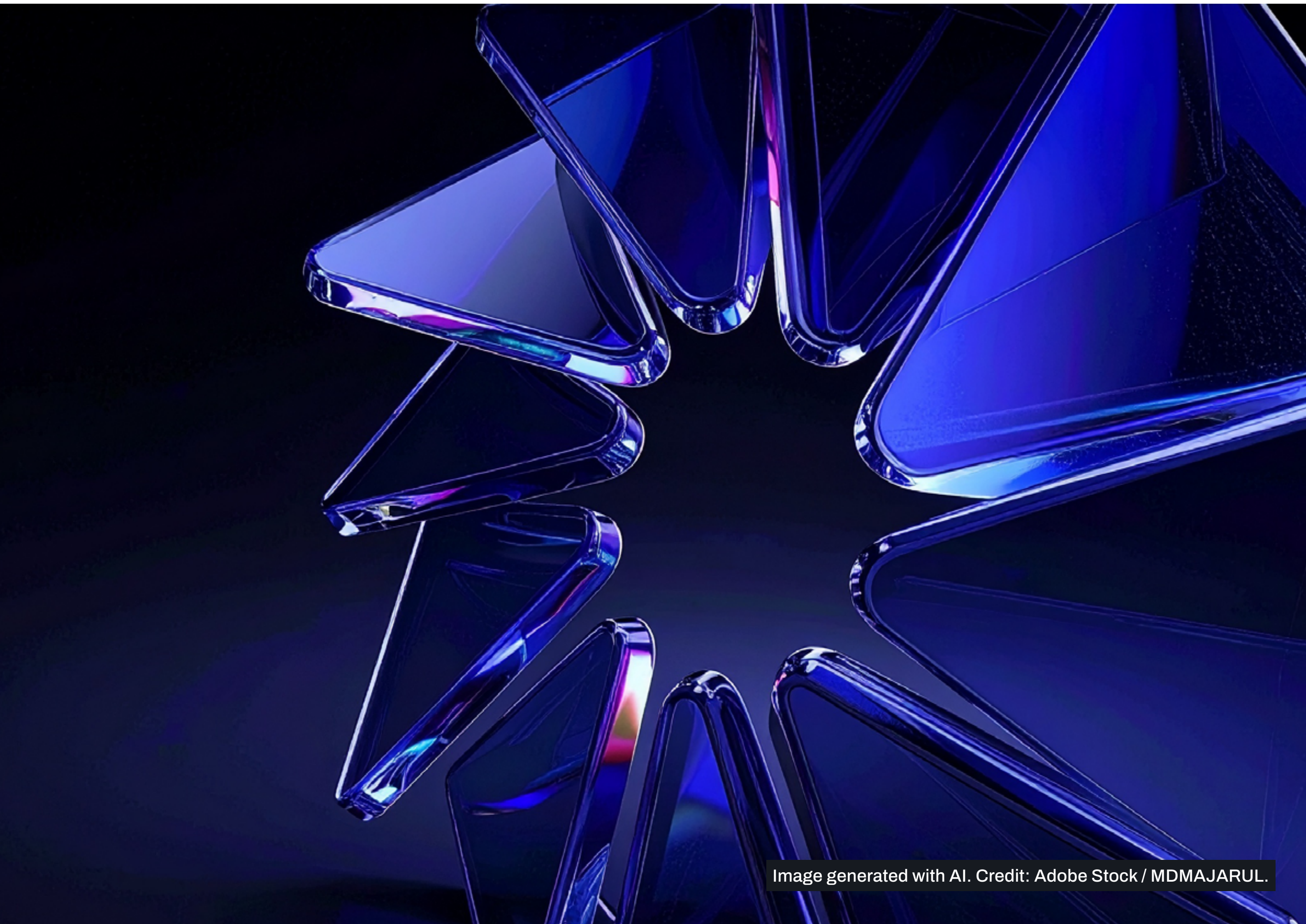


Image generated with AI. Credit: Adobe Stock / MDMAJARUL.

6. Conclusion and the way ahead

The 2025 REAIM Regional Consultations confirmed the core finding and insight drawn from the 2024 consultations:[16] the governance of AI in the military domain is neither linear nor uniform. It is instead shaped by parallel and, at times, overlapping regional, subregional and national trajectories that reflect distinct security contexts, institutional capacities, policy ambitions and normative objectives. Since 2024, what has changed is not necessarily the diversity and depth of perspectives in and of themselves, but rather the degree to which patterns and commonalities are increasingly visible across regions where areas of collective efforts could have the greatest impact, as states and the wider multi-stakeholder community alike seek to shift away from stocktaking and the formulation of shared principles to operationalization and implementation.

Participants in the five regional consultations consistently acknowledged that AI in the military domain is no longer of speculative nature. As a result, is clear that the question is no longer about what principles underpin the responsible development, deployment and use of these technologies; rather, it is about translating these principles, underpinned by international law, into actionable practices and measures that are legally sound, technically robust and credible, while being nimble and responsive to the evolving security landscape at the national, regional and international levels. This evolution brings to the fore core questions surrounding institutions, infrastructure, processes, incentives and dialogue, all of which implicitly require contributions from and cooperation between public authorities, private actors and the civil society community.

While REAIM sits at a critical juncture and as states and the multi-stakeholder community alike seek to ramp up their contributions to the governance of responsible AI in the military domain, the following food for thought emerges from the 2025 REAIM Regional Consultations.

6.1. From principles to operationalization

A central conclusion emerging from the consultations is that efforts for responsible AI governance in the military domain must increasingly focus and rest on implementation. While the 2023 Call to Action and the 2024 Blueprint for Action continue to provide a critical shared reference point, in addition to other initiatives and processes (noting in particular those emerging within the United Nations), participants repeatedly highlighted the limits of high-level commitments in the absence of practical guidance and enabling mechanisms for implementation, oversight and overall operationalization. How responsibility and accountability should be distributed between public and private actors, and how risks can be assessed and mitigated in concrete operational settings all laid at the heart of the 2025 regional consultations.

The scenario-based tabletop exercise further reinforced this observation. By grounding discussions in fictional yet plausible operational contexts, the exercise illustrated how governance outcomes are often shaped by decisions taken well before deployment during peacetime, from

[16] These are captured in Afina, *The Global Kaleidoscope of Military AI Governance*.

system design, via procurement to testing and evaluation. Many of these decisions involve close interaction with industry actors, underscoring the importance of early and structured engagement to ensure that responsible AI considerations are embedded from the outset, rather than retrofitted at later stages.

6.2. Structured engagement with industry

One of the clearest developments since 2024 has been the growing prominence of industry engagement as a governance priority. While states generally acknowledged the important role of the private sector in 2024, one year later, participants across regions recognized that industry actors play a decisive role in shaping AI-enabled military capabilities, including through system-design choices, data practices, testing and evaluation, as well as post-deployment maintenance and support. Furthermore, a significant number of states noted that, while collaboration with the private sector will be critical, concerns were shared with respect to the effectiveness of such engagement and the need to preserve national sovereignty.

Against this backdrop, the consultations point to the value of a more structured and consistent approach to industry engagement, one that can translate responsible AI principles and the international legal obligations of states into actionable expectations and guidelines, while remaining sensitive to confidentiality, intellectual property and national security considerations. For many participants, particularly those from resource-constrained contexts, such an approach was perceived as a critical means to support informed procurement, reducing asymmetries in power with greater clarity on what to expect from industry, and promoting greater consistency across national and regional practices. To this end, a more structured, neutral and independent channel for engagement with industry is increasingly seen not as an optional add-on to international deliberations, but as a necessary condition for translating responsible AI principles and commitments as a critical pathway to action.

6.3. Capacity and trust

Echoing the 2024 regional consultations, discrepancies in states' capacities emerged as a structural feature of the global AI governance landscape. Participants emphasized that the digital divide, reflected through uneven access to technical expertise, data, infrastructure and institutional resources, continues to shape both risk perceptions and governance options. These asymmetries were frequently linked to dependence on externally developed technologies and to challenges in enforcing compliance with international legal obligations and ethical standards, particularly where states have limited leverage or visibility over proprietary systems.

Additionally, trust-building was repeatedly identified as a necessary complement to capacity-building. In regions reportedly marked by heightened tensions or deficits in mutual trust, participants highlighted the stabilizing role of dialogue, confidence-building measures and track-2 engagement. Importantly, trust was discussed not only in inter-state terms, but also with respect to states' relationship with the multi-stakeholder community, as well as human–machine interaction.

6.4. Strategy for REAIM's next phase

Looking ahead, participants consistently emphasized the importance of continuity, coherence and complementarity across efforts, initiatives and processes. While REAIM has widely been praised and valued as a platform that enables dialogue within and across regions, it has also been acknowledged that it fosters multi-stakeholder dialogue. Its continued relevance, however, was seen as contingent on its ability to evolve alongside the technology and the governance landscape, including by providing space to address implementation challenges that cut across public and private actors, but also without duplicating or competing with United Nations-based processes that many states have favoured as a forum for inter-state deliberations, rather that initiatives that sit outside the organization.

While there have been discussions on how future efforts could further expand their substantive scope and depth, a significant proportion of the conversations focused on deepening and sustaining impact. Participants also stressed the importance of maintaining space for regional perspectives and solutions, particularly in the light of varying security environments and governance capacities. Overall, amid the growing differences over the desire to preserve REAIM as a non-United Nations forum for inter-state deliberations, a number of participants nevertheless noted that, looking ahead, it will be critical to underscore REAIM's long-term added value as an implementation accelerator and multi-stakeholder bridge, while leaving space for growing discussions within the First Committee.

Ultimately, the 2025 REAIM Regional Consultations underscored the fact that the governance of responsible AI in the military domain is an iterative process, one that requires sustained and measured dialogue, adaptation and cooperation across regions and sectors. The consultations provided a critical and clear guidance for future efforts on this subject, now leaving space for states and stakeholders to collectively shape the path forward in ways that reflect both shared principles and regional realities.

REAIM responsible innovation demonstrations at REAIM Summit 2023, The Hague. Credit: Dutch Ministry of Foreign Affairs / Valerie Kuypers.

Annex. Data from the tabletop exercise on measures for procurement and assurances

This annex lays out complementary data related to the first part of the tabletop exercise. It provides the normalized distribution of votes allocated to each of the 10 assurance measures, by capability and by region. It also gives region-specific visualizations illustrating how participants prioritized assurance measures across the three AI-enabled military capabilities considered.

The information and data in this annex are intended to support further analysis of how responsible AI principles may be operationalized in the military domain, including by facilitating the identification of recurring patterns and areas of convergence across regions and capabilities.

I. Normalized assurance priorities by capability: Cross-regional overview

The following tables present the normalized distribution of votes allocated to each assurance measure by participants from different regions for the capability shown. Values are expressed as percentages to reflect the share of total votes cast within each region for that capability.

How to read the tables

For each region (column), the figures sum to 100 per cent and indicate how participants distributed their assurance priorities across the listed measures. Higher values therefore represent a greater relative prioritization of a given assurance measure within that region. The figures do not reflect absolute numbers of votes or levels of participation and should be interpreted as indicative of priorities expressed during the consultations.

The acronyms in the tables refer to the following regions. **APAC:** Asia-Pacific; **Eu & NA:** Europe & North America; **LAC:** Latin America & Caribbean; **WA & ME:** West Asia & Middle East.

TABLE I.

AI-enabled electronic warfare capabilities

		AFRICA	APAC	EU & NA	LAC	WA & ME
1	**Iterative legal reviews**	13.3	14.7	23.3	12.5	3
2	**Iterative performance reviews**	12.6	18.9	13.3	20.8	15.2
3	**Independent third-party review**	2.2	0	13.3	8.3	1.5
4	**Complete control of the system's maintenance**	17	5.3	3.3	8.3	13.6
5	**Complete control over the system's decommissioning**	2.2	1.1	0	0	1.5
6	**Access to all pre-deployment information**	20.7	15.8	6.7	0	18.2

7	Digital forensics	3.7	8.4	13.3	8.3	18.2
8	Consistent reliability	15.6	18.9	13.3	29.2	16.7
9	Disclosure of vulnerabilities	11.1	14.7	13.3	12.5	10.6
10	Supplier backdoor	1.5	2.1	0	0	1.5

TABLE II.

Swarm-based uncrewed capabilities for ISR

		AFRICA	APAC	EU & NA	LAC	WA & ME
1	Iterative legal reviews	9.8	6.5	15.6	25	10.6
2	Iterative performance reviews	16.7	18.5	15.6	16.7	16.7
3	Independent third-party review	4.5	0	3.1	0	0
4	Complete control of the system's maintenance	15.2	8.7	3.1	16.7	12.1
5	Complete control over the system's decommissioning	3	6.5	6.2	4.2	4.5
6	Access to all pre-deployment information	15.2	17.4	9.4	12.5	18.2
7	Digital forensics	6.1	6.5	9.4	8.3	6.1
8	Consistent reliability	22.7	23.9	18.8	12.5	22.7
9	Disclosure of vulnerabilities	5.3	9.8	15.6	4.2	4.5
10	Supplier backdoor	1.5	2.2	3.1	0	4.5

TABLE III.

AI-enabled decision-support systems

		AFRICA	APAC	EU & NA	LAC	WA & ME
1	Iterative legal reviews	17.1	17.9	16.1	18.5	17.6
2	Iterative performance reviews	10.9	12.6	16.1	14.8	8.8
3	Independent third-party review	2.3	1.1	16.1	11.1	4.4
4	Complete control of the system's maintenance	14.7	4.2	3.2	0	14.7
5	Complete control over the system's decommissioning	2.3	0	0	3.7	4.4
6	Access to all pre-deployment information	17.1	14.7	3.2	14.8	13.2
7	Digital forensics	7	10.5	19.4	14.8	14.7
8	Consistent reliability	20.9	18.9	16.1	7.4	7.4
9	Disclosure of vulnerabilities	6.2	20	3.2	14.8	13.2
10	Supplier backdoor	1.6	0	0	0	1.5

II. Normalized assurance priorities by region

This section presents a set of region-specific visualizations illustrating how participants from each region distributed their assurance priorities across the three AI-enabled military capabilities considered in the first part of the tabletop exercise. By focusing on one region at a time, these figures highlight how assurance priorities shift across capabilities within the same regional context, rather than comparing one region to another.

The figures are intended to complement the cross-regional comparisons presented in Figures 1–3 in Section 4.1 by providing a more granular view of intra-regional trends, patterns and capability-specific variations in assurance prioritization.

How to read the figures

Each figure presents the distribution of assurance priorities expressed by participants from a single region across the three AI-enabled military capabilities considered: electronic warfare, swarm-based ISR and AI-enabled DSS.

For each capability, votes are normalized within the region such that the total number of votes cast equals 100 per cent. Values therefore represent the share of votes in a region allocated to a given assurance measure for each capability. Differences across capabilities within the same figure indicate shifts in relative prioritization, rather than differences in overall levels of concern or participation.

The figures should be interpreted as illustrative of patterns emerging from the consultations and do not represent regional positions or levels of endorsement.

The numbers (1–10) on the X axis of each figure refer to the 10 assurance priorities below:

1	Iterative legal reviews	**6**	Access to all pre-deployment information
2	Iterative performance reviews	**7**	Digital forensics
3	Independent third-party review	**8**	Consistent reliability
4	Complete control of the system's maintenance	**9**	Disclosure of vulnerabilities
5	Complete control over the system's decommissioning	**10**	Supplier backdoor

FIGURE I.

Africa

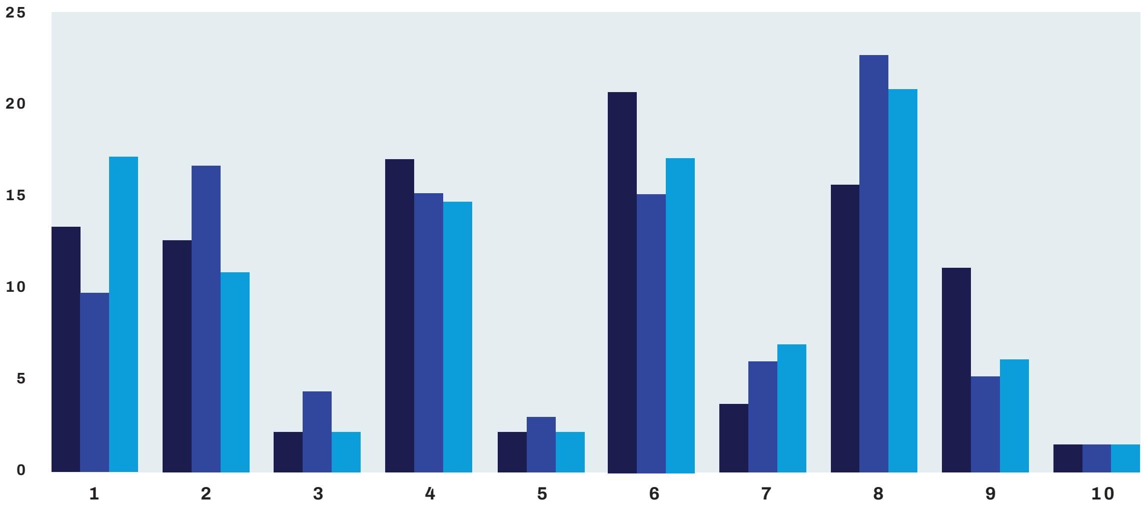
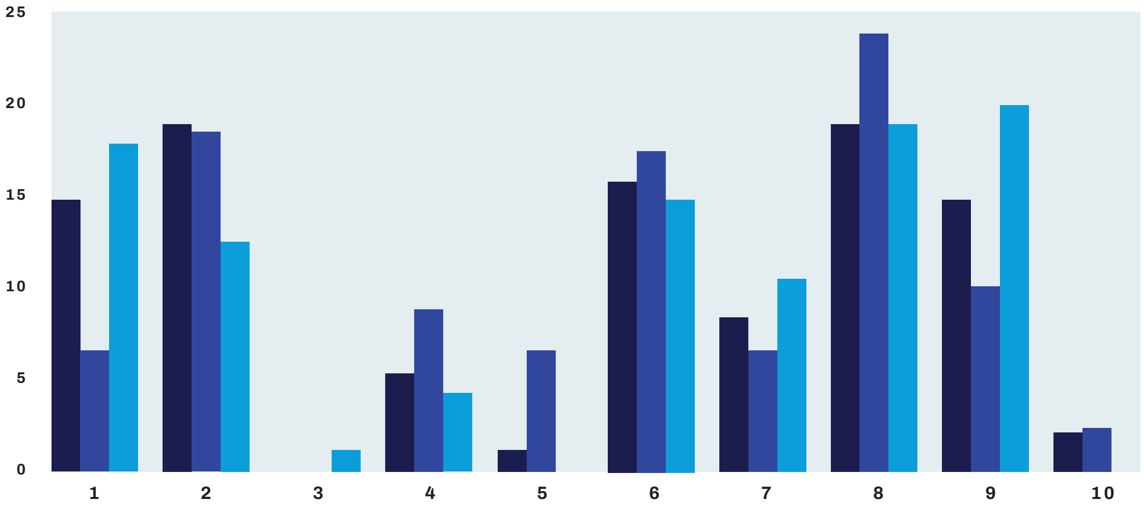


FIGURE II.

Asia-Pacific



● ELECTRONIC WARFARE ● SWARM FOR ISR ● AI-DSS

1 Iterative legal reviews

2 Iterative performance reviews

3 Independent third-party review

4 Complete control of the system’s maintenance

5 Complete control over the system’s decommissioning

6 Access to all pre-deployment information

7 Digital forensics

8 Consistent reliability

9 Disclosure of vulnerabilities

10 Supplier backdoor

FIGURE III.

Europe and North America

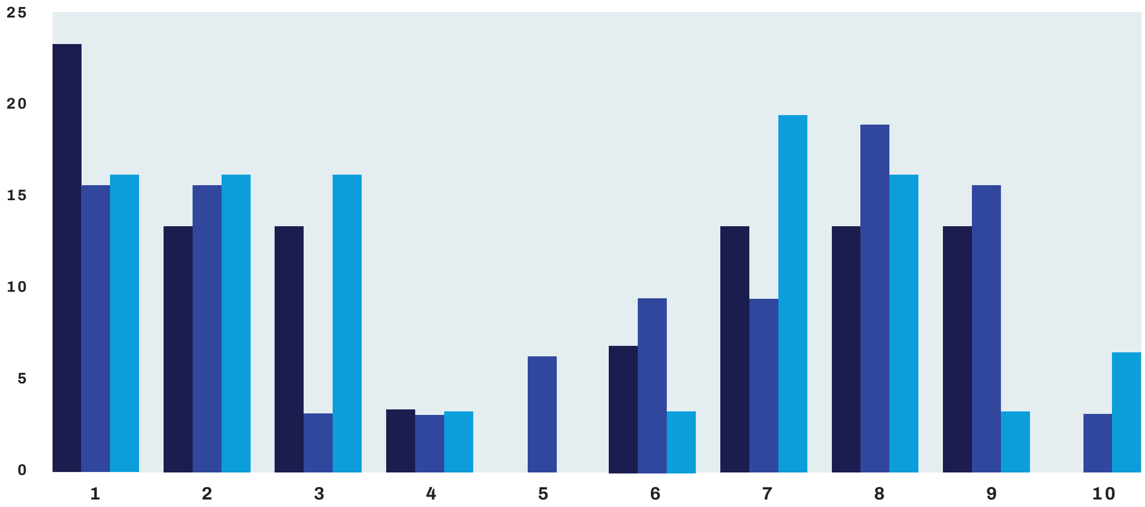
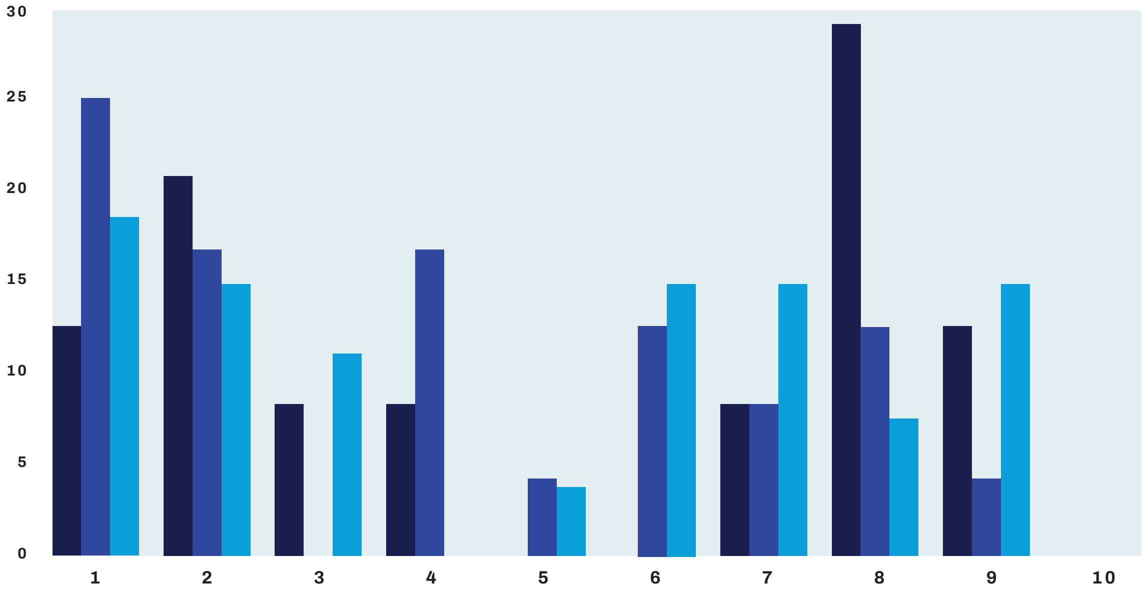


FIGURE IV.

Latin America and the Caribbean



● ELECTRONIC WARFARE

● SWARM FOR ISR

● AI-DSS

1 Iterative legal reviews

2 Iterative performance reviews

3 Independent third-party review

4 Complete control of the system's maintenance

5 Complete control over the system's decommissioning

6 Access to all pre-deployment information

7 Digital forensics

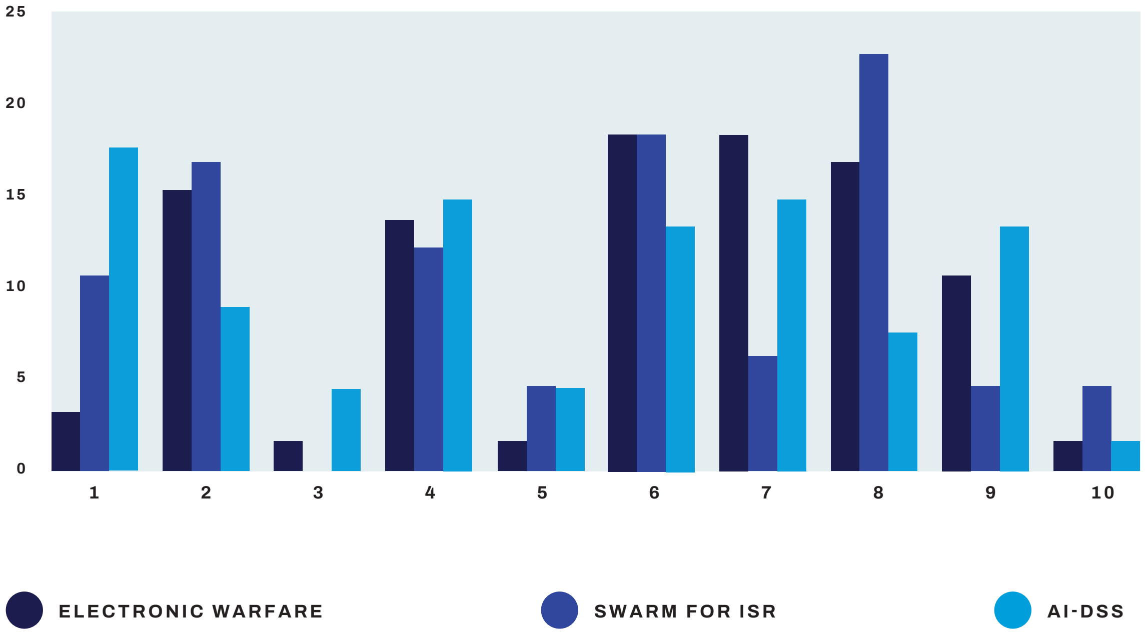
8 Consistent reliability

9 Disclosure of vulnerabilities

10 Supplier backdoor

FIGURE V.

West Asia and the Middle East



1 Iterative legal reviews

2 Iterative performance reviews

3 Independent third-party review

4 Complete control of the system’s maintenance

5 Complete control over the system’s decommissioning

6 Access to all pre-deployment information

7 Digital forensics

8 Consistent reliability

9 Disclosure of vulnerabilities

10 Supplier backdoor

Cover Image: Illuminated holographic paper sculpture of a bird (generated with AI). Credit: Adobe Stock / Hungarian.

UNIDIR

𝕏 @unidir
in /unidir
/un_disarmresearch
f /unidirgeneva
▶ /unidir

Palais des Nations
1211 Geneva, Switzerland

WWW.UNIDIR.ORG