United Nations Office for Disarmament Affairs

Ministry of Foreign Affairs
Republic of Korea

MILITARY AI PEACE & SECURITY DIALOGUES

KEY TAKEAWAYS

The 2025 Military AI, Peace & Security (MAPS) Dialogues, convened by the United Nations Office for Disarmament Affairs (UNODA) and funded by the Republic of Korea, aim to provide a platform for States to share knowledge and raise awareness about the international peace and security implications of AI in the military domain. The initiative serves ongoing multilateral discussions by providing a neutral setting in which States can engage with international organizations, civil society, researchers and industry, exchange views and identify practical areas for cooperation.

This note distils inputs and takeaways from two public webinars, *Military AI: Opportunities, Risks, and International Peace & Security* and *Capacity-Building and International Cooperation on AI in the Military Domain*, respectively held on 12 and 25 March 2025, and a subsequent in-person meeting held on 2 July 2025 under the Chatham House rule. It records themes that may inform subsequent dialogue among Member States, including areas where near-term exchanges could be helpful.

The two public webinars featured nine experts and respectively featured 205 and 187 attendees, from Member States (50%), civil society organizations (30%), and international and regional organizations (20%). The in-person meeting was attended by 33 delegates. All three events had high Global South participation.

Throughout all events, participants recognized that AI technologies may be integrated across a range of functions and tasks, and that associated risks and opportunities depend on the context of use.

Potential contributions to peace and security

Participants identified several areas where the responsible use of AI could contribute to peace and security. They emphasized the importance of maintaining a distinction between strategic and humanitarian opportunities (such as protection of civilians in conflict), while acknowledging that responsible practice can support both types of use cases when pursued deliberately.

First, AI can **enhance information handling and decision support**, enabling operators and commanders to process large volumes of data, identify patterns, and allocate attention more effectively. This was linked to potential gains in situational awareness and informed decision-making across kinetic and virtual environments.

Second, AI may **reduce risks to personnel** by enabling or enhancing tasks in hazardous environments, such as explosive ordnance search, route clearance, or operations where communications are degraded. These functions were seen as offering safety and efficiency benefits when performance limits are understood and respected.

Third, several participants pointed to potential contributions to **protecting civilians**. Examples included tools that help characterize the civilian environment, support planning that minimizes incidental harm, or improve the timeliness and accuracy of warnings. It was underlined that such benefits are not automatic. They depend on data quality, robustness, and control measures that ensure legal and ethical judgments are made by humans.

Finally, areas such as **logistics, maintenance, and training** were frequently highlighted as applicable and comparatively less controversial. Even in these domains, participants emphasized the importance of disciplined practice, including testing and operator training, to guard against over-reliance and unintended effects.

Risk dynamics and operational concerns

Alongside opportunities, participants considered risks that arise or are amplified in military settings. First, **technical reliability and robustness** were central concerns. Probabilistic systems may behave unpredictably when data are sparse, noisy, or unrepresentative of the operating environment. Vulnerabilities to adversarial manipulation and data poisoning were discussed, as were the practical limits of explainability in time-critical contexts. Data confidentiality and broader cybersecurity risks affect both the development and integration of AI in the military domain. Upstream factors were also noted: training on synthetic or non-purpose-built data can entrench bias or degrade performance. The energy demands of training and operating large AI models also introduce environmental impacts, for instance, carbon emissions or power consumption; therefore, it is important for States to undertake sustainability and resilience planning.

Second, participants also examined the risks associated with **human-machine interaction**. Over-trust, automation bias, and degraded vigilance may lead users to defer excessively to system outputs, particularly when interfaces convey unwarranted certainty. Conversely, under-trust can negate potential benefits, such as disregarding accurate AI-generated insights because they contradict the human operator's beliefs or expectations. Several participants, therefore, focused on design choices and training that calibrate trust appropriately, including clear indications of when to disengage or escalate to human review. It was also observed that high efficiency of AI-enabled systems in performing tasks (such as data processing or operational support) can inadvertently encourage overreliance and create a sense of "distance" from the human impact of decisions.

Third, risks of **misuse** were raised across the life cycle of AI systems. These included malicious use at deployment by insider threats or other "bad actors," the prospect of offensive cyber operations enhanced by AI, and use by non-State actors, including private military companies operating outside standard accountability frameworks. Some participants, in particular, cautioned about high-stakes contexts such as nuclear command, control, and communications (NC3) structures, where misinterpretation or miscalculation could have especially severe consequences.

Another recurring theme was **decision-time compression**. While some AI tools aim to accelerate the process of interpreting complex information, faster decision cycles can reduce the time available for human deliberation and impair commanders' ability to comprehend the rationale behind outputs and actions taken by AI-enabled systems. This increases the possibility of error, escalation or miscalculation, especially under stress and in times of crisis. Relatedly, arms racing dynamics were cited as an additional driver of risk. Perceptions of strategic competition and psychological beliefs about gaining advantage over an adversary may create pressures to field capabilities rapidly, potentially outpacing assurance, testing and governance processes.

Legal, ethical and policy anchors

Participants affirmed that relevant existing legal obligations, including the UN Charter, international humanitarian law and international human rights law, apply to the development and use of AI-enabled military capabilities. Emphasis was placed on retaining human judgment and control over decisions on the use of force, and on ensuring that responsibility and accountability remain with human actors throughout the life cycle. States described or referenced legal reviews pursuant to Article 36 of Additional Protocol 1 to the Geneva Conventions, indicating these practices also apply to AI, including follow-up reviews when systems are updated or repurposed, as well as to the role of military legal advisers.

The importance of responsibility and accountability was emphasized. It was cautioned that legal determinations, such as in relation to the principles of necessity, proportionality and distinction, should not be coded into opaque systems, and that decision-making support tools should enable, not replace, legality and ethical reasoning. As systems are updated or composed from other models ("AI training AI"), attribution and traceability can become more complex, underscoring the need for documentation, auditability and clear roles across operators, developers and commanders. Ethical concerns were recorded regarding any delegation of life-and-death decisions to AI. Some contributors also noted the cognitive complexity of real-world decision-making, which often relies on tacit judgment and intuition, and questioned the extent to which current systems can meaningfully replicate such human faculties. It was noted that international humanitarian law required individuals to be held accountable for war crimes.

Assurance and evaluation practices

Assurance measures were often framed as enabling trust in AI systems, including through evidence-based testing, evaluation, verification and validation (TEVV) across the life cycle of AI. Participants emphasized the value of representative test data, red-teaming, and scenario-based trials in understanding performance limits, failure modes, and robustness under distribution shifts. Documentation practices, such as clear articulation of intended use, operational boundaries, and known limitations, were viewed as enablers of oversight and learning.

Participants indicated that assurance is not a label but a continuous process. Updates to models, data, or integration pathways can significantly alter behavior. As such, change management and periodic reassessment were highlighted. Interfaces that convey uncertainty, confidence intervals, and appropriate cautionary cues were considered part of assurance in use.
Operator training was part of this discussion. Training can help mitigate cognitive pitfalls, including automation bias and over-reliance, and should include drills on disengagement criteria and procedures for escalating to human review. Several participants suggested that assurance and training are mutually reinforcing, with evaluation findings feeding into curricula and standard operating procedures.

Capacity-building and cooperation needs

Participants identified priority needs across people, processes and technology.
On **people**, a multidisciplinary workforce was considered essential. This includes operators, commanders and planners, engineers and data specialists, lawyers and policy officials. Training to identify cognitive biases and the limits of AI systems was regarded as important for all these groups, not only technical specialists.

On **processes**, participants discussed shareable elements of legal review practice for AI-enabled systems, baseline TEVV protocols, and procurement approaches that foreground assurance. Several contributions highlighted the value of knowledge and technology sharing, as well as regional peer learning, in narrowing capability gaps and promoting compatible practices. Such approaches could also enable meaningful participation in AI-related multilateral discussions by all States. Cooperation was framed as proportionate and respectful of security contexts and levels of AI development and diffusion, with a focus on practical methods and lessons learned.

On **technology**, several participants highlighted the need for accessible and transparent evaluation tools, as well as secure data access arrangements that respect sovereignty and privacy. The ability to test against representative conditions through datasets, scenarios, or shared methodologies was seen as a practical enabler for many States.

Convergence and practical avenues for further exchange

Object-level avenues

Across discussions, several themes were viewed as practical means to translate high-level principles into workable practices. Areas of convergence include the applicability of international law, the need to retain human responsibility for use-of-force decisions, the importance of assurance and life-cycle management, and the value of inclusive capacity-building. There was broad recognition that responsible practice requires deliberate investment rather than being an automatic by-product of technological modernization.

The intersection of AI and nuclear weapons was identified as a high-consequence area where confidence-building measures and transparency were viewed as essential yet underdeveloped. Suggestions focused on practical, proportionate steps to reduce misperception while being mindful of sensitivities. Some contributions highlighted potential red lines regarding the integration of AI into nuclear decision-making.

Process-level avenues

Several participants suggested that preliminary work to clarify scope and terminology could lead to productive discussions in operationalizing the principles for responsible AI in the military domain. Because AI is dual-use and risks are context-dependent, it was noted that the scope should not be defined solely by technology or by the type of operator, but also by the context of use. Further exchanges could explore practical scoping criteria to ground discussions and help identify where governance gaps may exist.

Participants also observed that the setting and modality of dialogue shape outcomes. Multi-stakeholder formats can foster transparency and trust at various levels, while focused discussions tend to yield more operationally beneficial outcomes. It was suggested that States consider transitioning from broad principles to tailored frameworks – whether application-focused, risk-responsive, or other pragmatic approaches – to examine how specific characteristics of certain technologies, used in specific operational contexts, interact with existing legal obligations and operational practice.

Finally, coordination with parallel processes was encouraged to avoid duplication or the development of divergent norms, as well as to leverage synergies. Exchanges could bridge work on civilian AI governance and discussions on lethal autonomous weapon systems, identifying areas where compatible approaches, practices, and lessons are transferable, and where distinct treatment is warranted. Regional or thematic follow-ups were also suggested to consolidate learning, reflect diverse operating contexts, and deepen collaboration between technical, legal and policy communities.

Conclusion

The meetings underscored both the promise and the risks of AI in the military domain. Opportunities – in information handling, safety and efficiency – are real but contingent on deliberate safeguards. Risks – of error, escalation, brittleness and misuse – are manageable only with sustained investment in assurance, human judgement and control, and cooperation.

UNODA identified several future areas of work under the MAPS Dialogues, including in-depth discussions to build a common understanding of opportunities and risks, potential red lines (such as in relation to disarmament, arms control and nonproliferation regimes), and diplomatic tools to support dialogue among States. To this end, UNODA is currently seeking funding for the 2026 MAPS Dialogue.

For more information on MAPS Dialogues, please contact UNODA.