MAPS DIALOGUES

SERIES REPORT

A *Collective* Endeavor

A Synthesis of the 2026 Military AI, Peace & Security Dialogues

CONVENED BY
UN Office for Disarmament Affairs

AUTHOR
Ulysse Richard

PUBLISHED
October 2026

United Nations Office for Disarmament Affairs | Ministry of Foreign Affairs

CONTENTS

Introduction 03

Panelists 04

1 Around the life cycle: governance considerations 05

1.1 Before use 06

Data and design choices
Testing and evaluation
Preserving human judgment and oversight
Compliance with international law and scoping of use cases

1.2 Into service 08

Industry engagement: responsibility and feedback
Procurement
Transfer

1.3 During and after use 09

Ensuring effective human control
Continuous assessment and incident reporting
End of life

2 Around the room: process considerations 11

2.1 Complementarity among initiatives 12

2.2 Multi-stakeholder participation 12

2.3 International cooperation and capacity-building 13

2.4 Normative proposals and next steps 13

Conclusion 14

Notes 15

INTRODUCTION

From 15 to 17 June 2026, States met in Geneva for three days of informal exchanges on artificial intelligence (AI) in the military domain, convened pursuant to General Assembly resolution 80/58. The Office for Disarmament Affairs has since issued a written factual summary of those exchanges to the First Committee.[1]

The 2026 MAPS Dialogues, held on 7 May, 28 May and 10 June, were convened to foster inclusive multilateral discussions ahead of these exchanges. Twelve panelists provided political, technical, legal, and operational expertise across the three sessions.[2]



This report draws the three sessions together into a series report, structured around the agenda of the informal exchanges. Each section opens with the positions articulated by States, and then presents the perspectives developed during the webinars.

PANELISTS

I A Shared Vocabulary for Military AI

7 May 2026



Zena Assaad
Australian National University



Heidy Khlaaf
AI Now Institute



Paul Lushenko
Cornell University

II Before and After Use: A Lifecycle View of Military AI

28 May 2026



Ariel Conn
University of Colorado Boulder



Vincent Boulanin
Stockholm International Peace Research Institute



Heather Frase
Virginia Tech National Security Institute



Seb Matthews
Industry leader



Jonathan Kwik
T.M.C. Asser Institute

III Working the Disagreements: From Tensions to Traction

10 June 2026



Julia Elizabeth Rodríguez Acosta
Permanent Mission of El Salvador to the UN



Kim van der Sluis
Permanent Mission of the Netherlands to the UN



Giacomo Persi Paoli
UN Institute for Disarmament Research



Beyza Unal
UN Office for Disarmament Affairs

All three webinars were moderated by **Ulysse Richard**, UN Office for Disarmament Affairs. Closing remarks for the series were delivered by **Jiyoung Yun**, Director, Ministry of Foreign Affairs, Republic of Korea.

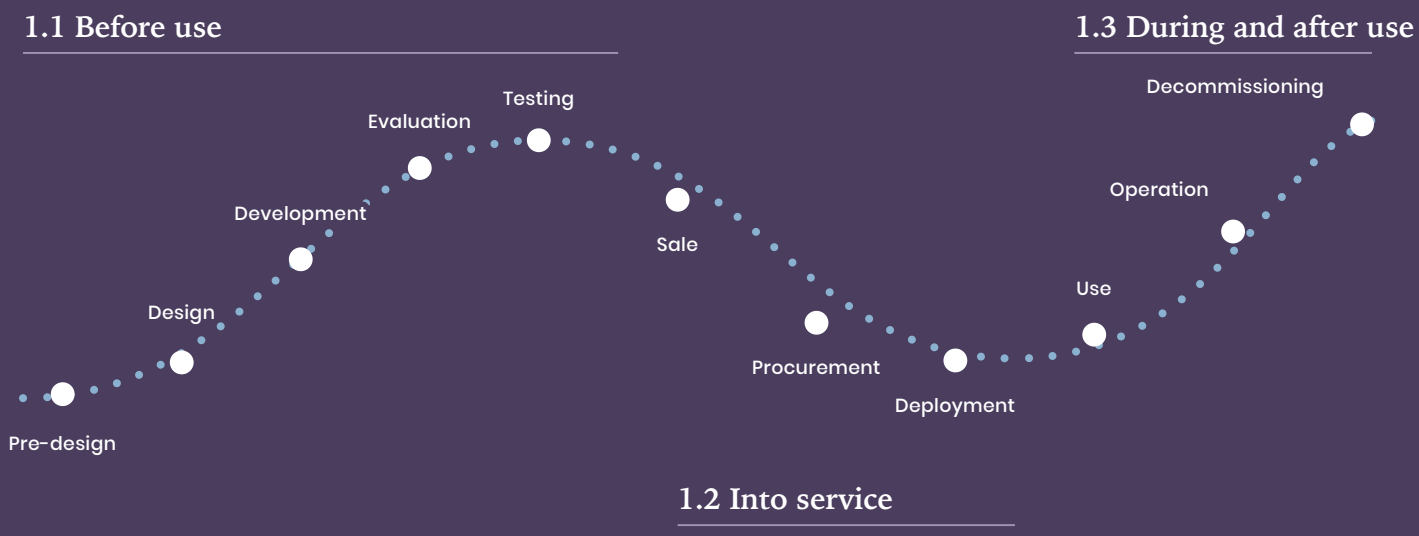
This document compiles views from across the series and does not represent agreement among panelists or participants, who took part in a personal capacity and may not agree with all points recorded in the report. Individual contributions are not attributed. Divergent and recurring points reflect the interlinked nature of the subject, where a single design choice can appear differently from the perspectives of testing, legal review and operational use.

MAPS DIALOGUES

PART 1

Around the *life cycle*

Governance considerations



Throughout the three-dialogue series, participants stressed that ensuring lawful and safe use of AI is a collective endeavor, and such use does not rest solely with the end user. By the time an AI-enabled system is employed, developers have determined its architecture and data, testers and legal advisers have delineated its parameters for safe and lawful operation, and trainers have equipped the operator to interpret its outputs. A lapse at any stage may lead the commander to cause unintended harm; as one panelist observed, responsibility cannot rest solely with the user if it was not systematically embedded from the outset of design.

Resolution 80/58 delineates eleven such stages in the system life cycle, from pre-design to decommissioning.[3] For the diplomatic community, the value of the life-cycle framing lies in identifying where policy intervention is possible. Most of those intervention points operate at the national level, which explains why many of the measures discussed are within a State's capacity to implement independently.

1.1 GOVERNANCE CONSIDERATIONS
Before use

FROM THE INFORMAL EXCHANGES — A/81/221, paras. 13, 15 and 16

Several States emphasized that early design, development and testing choices shape later behavior and risks; stressed data quality, integrity, representativeness and governance; called for rigorous testing, evaluation, verification and validation in realistic conditions, with retesting after substantial change; and highlighted article 36 legal reviews adapted to AI and repeated after significant system changes.

Data and design choices

The quality of data used in AI systems has legal as well as performance implications: a system's inability to respect the principle of distinction between civilians and combatants often traces to shortcomings in the original data and its labeling. Panelists emphasized that this awareness must extend beyond training into development and procurement.

Representativeness has two dimensions. Synthetic or non-purpose-built training data can introduce bias or impair performance. The upstream concern is thus not only data quantity but also suitability for the intended application. It also covers the diversity of users; test scenarios should therefore reflect the full range of operators. Panelists stressed involving operational testers early, before it becomes impossible to insert the data hooks and traces needed for post-deployment assessment.

System architecture and the choice of learning paradigm are likewise design decisions with legal and technical consequences. For example, reinforcement learning can produce innovative solutions but often reduces predictability.[4]

Testing and evaluation

Panelists approached assurance as an ongoing process, as opposed to a one-time certification.[5] Because updates to models, data or integration pathways can change system behavior, clear documentation, change management and periodic reassessment of intended use, operational boundaries and known limitations are essential.

AI systems challenge three core assumptions of established practice: that the system tested is the system fielded, which weakens when models are updated, drift or keep learning after deployment; that failures can be diagnosed by reproducing them, which stochasticity complicates;[6] and that the action space can be bounded.

The scope of testing coverage therefore becomes a policy decision. Where testing covers only part of the operational envelope, restricting system use is a practical way to manage residual uncertainty, turning a testing limitation into an operational constraint. Panelists stressed that an AI system that cannot meet a traditional threshold is unsuitable for that application and should not be a reason to lower the threshold, adopt a less specific benchmark or create exceptions for AI systems.

Performance on general tasks says little about performance on a specific military function, and participants pushed back on general benchmarking[7] as a substitute for application-specific requirements. The right trade-off between capability and explainability likewise depends on the function: panelists contrasted surveillance and targeting, where the consequences of error differ substantially.[8]

Preserving human judgment and oversight

The informal exchanges identified preserving human judgment as an early consideration; the series unpacked several relevant concrete design-stage measures. Involving operators early in the design phase ensures that interface and explainability decisions reflect the needs of those who will use the system under operational conditions. A panelist provided the example of Human Readiness Levels, which offer a structured way to assess whether users can operate a technology as intended, alongside Technology Readiness Levels.[9]

Interface design can itself serve as an oversight mechanism, for instance by displaying confidence levels and underlying assumptions in a way that prompts scrutiny. Legal determinations such as necessity, proportionality and distinction should not be embedded in opaque systems. Decision-support tools should support, not replace, legal and ethical reasoning.

Compliance with international law and scoping of use cases

Measures to ensure compliance with international law can be injected at several stages, including procurement, review and operation. Panelists argued that focusing on Article 36 of Additional Protocol I undersells what international law can do across the life cycle, since the obligation of constant care requires ongoing precautions.[10]

Compliance efforts begin before the formal review. States can disclose in advance the requirements developers must meet for acquisition approval, bridging the disconnects panelists noted between industry and government, and between procurement officials and legal advisers.

A legal review applies to the normal and expected use of a weapon, means or method of warfare, and its outcomes should be reflected in standard operating procedures. Reviews are harder for off-the-shelf systems, where requirements cannot be coordinated with the developer in advance.

On scoping, the series suggested starting with whether the application needs AI at all and whether the model class on offer is fit for purpose. Panelists argued that discussions often assume systems are more capable than they are, and stressed that declining an application should remain on the table, especially where compliance with international humanitarian law cannot be assured.

Some legal requirements, such as identifying whether a person is wounded or surrendering, sit beyond the reliable reach of current technology. Organizations therefore need clear processes and authority to halt AI projects before they become entrenched, with stop mechanisms built into development, procurement and operational workflows.

Declining an application should remain on the table, especially where compliance with international humanitarian law cannot be assured.

1.2 GOVERNANCE CONSIDERATIONS
Into service

FROM THE INFORMAL EXCHANGES — A/81/221, paras. 14, 23 and 24

Several States noted the importance of early engagement with industry, which was seen as responsible for safe and secure design, with States retaining ultimate responsibility for compliance with international law, and underscored that informed procurement requires technical, operational and legal expertise and independent testing. On transfers, several States noted obligations including under the Arms Trade Treaty and concern about proliferation to non-State actors, while several considered that non-proliferation measures should not deny legitimate access.

Industry engagement: responsibility and feedback

The division of labor between industry and government is clear on paper (industry builds, government ensures compliance), but accountability boundaries are interpreted differently. Vendors are typically not required to factor international law into their evaluation and may leave responsibility for lawful use with the end user, while policymakers may assume that safeguards are already built in, and performance metrics are not always independently verified.

After handover, supplier visibility into use declines sharply. Feedback that could help refine and secure systems, such as data on decision outcomes and unexpected failures, seldom reaches developers; when it does, it flows through contractual channels and not technical ones. Panelists recommended making structured post-deployment feedback a contractual requirement rather than an ad hoc arrangement.

Procurement

In response to a perceived need to keep pace with rapid advances in AI, many governments are rethinking procurement strategies, working more closely with vendors and moving from step-by-step to iterative approaches. Panelists cautioned that speed should not come at the expense of legal, technical and ethical reviews: every iteration should include checkpoints for legal review and testing and evaluation; and supplier discussions should ask whether a capability is needed and can be used lawfully, not only what is available. States were encouraged to maintain independent testing capacity, internally or through trusted third parties, rather than relying on vendor assurance.

Transfer

Cross-border transfer of AI-enabled systems differs from transfer of hardware. Model weights and training data carry legal and security implications that may not be visible,[11] and suppliers frequently retain the capacity to update systems remotely, limiting the recipient State's effective control.

Conventional defense programs and regimes such as the Missile Technology Control Regime[12] offer partial analogies, but neither addresses ongoing supplier influence through updates and data dependencies. Panelists identified a combination of State-regulated transfer mechanisms and commercially enforced licensing agreements as a priority for policy development, and reframed the access debate from eligibility to acquire AI capabilities to the assurances that build trust, such as end-user verification and transparent licensing.

1.3 GOVERNANCE CONSIDERATIONS During and after use

FROM THE INFORMAL EXCHANGES A/81/221, paras. 17 to 21 and 25

States emphasized that use must be consistent with international law, that accountability cannot be delegated to machines, and that effective human control is required throughout the life cycle; several added that control must be meaningful rather than procedural, while others noted that the term lacks consensus and that appropriate control is context- and system-dependent. On decommissioning, several States stressed secure disposal of data and models, prevention of diversion or repurposing, and independent verification.

Ensuring effective human control

The series reiterated the distinction between meaningful and procedural control: an operator who approves recommendations under time pressure is formally involved but exercises limited agency. Automation bias and skill atrophy from overreliance both contribute, as do large language models that agree with users even when they are wrong[13] and systems that generate recommendations faster than they can be evaluated.

Panelists recommended that operator training cover the cognitive processes of human-machine interaction, including automation bias, confirmation bias and overconfidence; that doctrine guarantee adequate deliberation time in high-stakes situations; and that testing extend beyond technical metrics to whether operators appropriately challenge system outputs.

Under-trust is an equally significant and often overlooked failure mode: operators may dismiss accurate outputs that contradict their expectations. Panelists cited evidence that junior personnel with substantial digital experience often take a more conservative approach and prefer human oversight.

Current policy language does not capture the full spectrum of control configurations observed in practice. Panelists argued that policy frameworks should draw on existing typologies of human and machine oversight, with functional, domain-specific and scale-based distinctions, which would also facilitate interoperability in multinational operations.

Appropriate control is also context- and system-dependent, as the informal exchanges highlighted. The series examined this through operational speed. In nuclear command, control and communications, sufficient time is essential and excessive speed can produce action-reaction cycles leading to irreversible decisions; States in Geneva referenced General Assembly resolution 80/23 on this risk. Missile defense, by contrast, requires responses within seconds. The central policy issue is which decisions may be delegated, under what conditions, and with what role for human operators.

Continuous assessment and incident reporting

The informal exchanges identified the value of iterative legal reviews after significant system changes; however, the threshold for triggering them remains unsettled.

One view is that any change that significantly affects target identification, selection, engagement, or expected effects should trigger a new review. A panelist called this a strict approach but noted that even small changes can lead to unexpected shifts in behavior. An alternative asks under what conditions there is sufficient positive evidence that no further review is necessary. Either way, changes are often observable only once altered behavior emerges, and modifications are frequently applied to individual systems, raising the question of what to retest.

The series distinguished formal review from the ongoing duty of assessment grounded in the obligation of constant care, supported by horizontal reporting across the force and vertical reporting up the chain of command. Behavioral change can also arise without an update, from differences between operational inputs and training data or from changed operator usage – i.e., from a factual change, distinct from an epistemic change, where the system is unchanged but understanding of its behavior evolves.

Panelists noted that most militaries lack a shared understanding of failure modes or standardized incident reporting. Cross-organizational pattern detection would support accountability and bound the test space.

End of life

Secure disposal is a planning challenge: which artifacts require control (model weights, training and test data, architectural information, persistent memory), who is responsible and for how long should be decided before development and not at decommissioning. Because hardware and software age at different rates, retirement decisions rarely apply to a whole system, and decommissioning can leave gaps in interconnected systems.

The series also addressed repurposing. Military systems may be repurposed for civilian use, including by police or border authorities, where the operational and legal context changes significantly. Adapting civilian systems for military use requires hardening and testing against adversarial attack, data poisoning and cyber backdoors; places the legal review late in the life cycle, with less influence over design; and, given their dual-use nature, raises proliferation concerns, with documented misuse by criminal and terrorist actors.

Who is responsible for which artifacts, and for how long, should be decided before development and not at decommissioning.

MAPS DIALOGUES

PART 2

Around the *room*

Process considerations

2.1 Complementarity among initiatives

2.2 Multi-stakeholder participation

2.3 International cooperation & capacity-building

2.4 Normative proposals & next steps

"Process design cannot substitute for political will. The greater risk is that geopolitical dynamics slow multilateral progress and shift discussions outside the United Nations, where they may move faster but with less inclusivity and universality."

PROCESS CONSIDERATIONS

2.1 Complementarity among initiatives

FROM THE INFORMAL EXCHANGES — A/81/221, paras. 26, 36 and 45

Several States called for greater complementarity among initiatives, warning that initiatives with selective participation could duplicate inclusive intergovernmental processes and contribute to fragmentation, and for future discussions that are focused, practical and centred on a limited number of priority issues.

The structural challenge is that AI is a horizontal technology while the multilateral system is organized along vertical, issue-based lines. Domain-specific analysis is necessary, since AI risks in the cyber domain differ from those in the biological domain, but the cost is that expertise stays siloed within each vertical. A mandate gap compounds this challenge. Resolutions on AI in the military domain exclude civilian applications, while the Global Digital Compact does not address military uses.

Panelists identified two areas where synergies between civilian and military discussions are possible. The first is to mirror the existing Independent International Scientific Panel within the military domain/peace and security pillar. The second is to rely on existing mechanisms: informal coordination among chairs and facilitators of AI-related processes, joint expert briefings during General Assembly sessions, and integrating AI into all relevant processes.

Panelists cautioned that process design cannot substitute for political will. The greater risk is that geopolitical dynamics slow multilateral progress and shift discussions outside the United Nations, where they may move faster but with less inclusivity and universality.

PROCESS CONSIDERATIONS

2.2 Multi-stakeholder participation

FROM THE INFORMAL EXCHANGES — A/81/221, paras. 41 and 44

Several States recognized the value of contributions from United Nations entities, international organizations and civil society, and many emphasized a multi-stakeholder approach, while several underscored that discussions should remain State-led, with other contributions subsidiary and depoliticized.

Expertise is distributed across sectors: technical knowledge in industry and research, operational experience in militaries, legal expertise in academia and civil society, and understanding of societal impacts in communities often absent from security dialogues. The series identified the technical community's direct involvement as essential, since policymakers cannot formulate effective technical questions in isolation. National mechanisms such as standing advisory bodies, time-limited consultations and industry boards could serve as models, provided they have clear mandates and safeguards for balanced representation.

Industry engagement remains an unmet need. Panelists noted that many companies are willing to participate but lack clear pathways, with only one technology company submitting input for the Secretary-General's report. The AI industry should be defined more broadly than a few large, predominantly Western firms, to include developers, semiconductor manufacturers, cloud providers, startups and adjacent sectors from all regions. Two models were proposed, each preserving State authority: regular, structured industry briefings for Member States ahead of major meetings, as already practiced in civilian AI contexts, and a Secretariat-maintained directory of the global AI industry under a specific mandate.

2.3 PROCESS CONSIDERATIONS
International cooperation and capacity-building

FROM THE INFORMAL EXCHANGES — A/81/221, paras. 31 to 33

States considered international cooperation and capacity-building critically important and suggested identifying the most immediate capacity gaps as a first step. Priorities included policy literacy, international humanitarian law and article 36 reviews, technical expertise for testing and evaluation, procurement guidance and operator training.

Capacity spans two dimensions: the ability to access and deploy the technology while maintaining appropriate oversight and control, and the ability to participate in international discussions on its governance. Least-resourced States face challenges on both fronts and therefore have less ability to influence rulemaking. For many delegations, the binding-versus-voluntary debate is secondary to whether they can participate effectively at all. Panelists made clear that prioritizing capacity-building does not signal reluctance to negotiate normative frameworks, but is rather a precondition for outcomes that serve all Member States.

Participation itself builds capacity, and measures from the UN processes on ICT security, such as funding the attendance of delegates from developing countries, have expanded access. Panelists recommended curation over new creation, such as a repository of existing research and primers as a single reference point for diplomats or a foundational course on AI diplomacy modeled on the cyber diplomacy curriculum, with the fellowship program for women in cyber cited as a successful precedent.

2.4 PROCESS CONSIDERATIONS
Normative proposals and next steps

FROM THE INFORMAL EXCHANGES — A/81/221, paras. 42, 43, 46 and 48

States emphasized international law as the essential framework, and several called for translating existing principles into practical guidance. Views divided on ethical considerations and on the concept of responsible use, and on whether to pursue frameworks, norms and practical measures or a legally binding instrument.

The series treated the voluntary norms versus legally binding approaches as a false dichotomy. International security processes have advanced incrementally. The cyberspace framework combines norms, confidence-building measures, political commitments, transparency and capacity-building while preserving space to negotiate legally binding instruments.

Panelists noted two reasons for the current reluctance to pursue binding measures in AI in the military domain: disarmament treaties have historically followed long periods of building shared understanding through interim steps, and some States fear binding rules will become obsolete. Panelists advised against new structures, building norms as the most appropriate model at this stage, and saw the immediate task as identifying a shared entry point, such as confidence-building measures, capacity-building or specific threats.

Panelists highlighted two regulatory approaches. Responsibility-based approaches regulate behavior, such as the use of technology and accountability mechanisms. Capability-based approaches, common in legally binding instruments, regulate which technologies may be developed, possessed or deployed. Some States find the former too subjective; a counterpoint is that capability restrictions alone cannot govern intangible technologies. National policies typically layer both.

States agree that international law applies to AI in the military domain. Panelists framed clarification as the immediate priority. While applicability enjoys broad consensus, States have no structured exchange on how they interpret it. Publishing national positions, as in the cyber domain, could serve that function. The clarification agenda spans the life cycle: integrating international law at design, addressing privacy under human rights law, and clarifying accountability, including in rapid decision contexts.

SERIES REPORT

Conclusion

Questions which stay abstract at a general level often become manageable when anchored to a specific function, domain, life cycle stage or operational context.

Most requirements identified by States fall within national decision-making and rely on mechanisms already in place: legal review, testing and evaluation, procurement, training, doctrine and incident reporting. Unresolved questions tend to surface at transition points: from development to deployment, between system versions, and across borders.

That these decisions are national does not make the multilateral setting secondary. It allows States to compare practices and lessons learned, and it is the only place to address what no State can resolve alone: the application of international law to AI-enabled systems, the transfer of capabilities and responsibilities across borders, and confidence-building measures for risks that arise between States.

Views differed on sequencing, with some participants favoring shared terminology before substantive work and others parallel progress. The sessions showed, however, that questions which stay abstract at a general level often become manageable when anchored to a specific function, domain, life cycle stage or operational context.

Recordings, readouts and podcast episodes

The 2025 and 2026 sessions are available at **mapsdialogues.org**.

NOTES

1 A/81/221, Written factual summary of the informal exchanges on artificial intelligence in the military domain and its implications for international peace and security, 17 July 2026. The informal exchanges were convened pursuant to General Assembly resolution 80/58. States were also invited to submit views for a compilation issued as A/C.1/81/INF/5.

2 This document compiles views from across the series and does not represent agreement among panelists or participants, who took part in a personal capacity and may not agree with all points recorded in the report. Individual contributions are not attributed. Divergent and recurring points reflect the interlinked nature of the subject, where a single design choice can appear differently from the perspectives of testing, legal review and operational use.

3 General Assembly resolution 80/58, preambular paragraph: pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning.

4 Reinforcement learning is a training approach in which a system is guided by rewards for outcomes. This means the system explores actions that maximize its reward signal, sometimes finding solutions that its designers did not foresee. This capacity to discover unanticipated strategies is central to both the method's operational value and its unpredictability.

5 Testing, evaluation, verification and validation is the established engineering process for establishing that a system performs as specified and is suitable for its intended use. The operational envelope is the range of conditions a system is expected to encounter in service.

6 Stochasticity means that, given identical inputs, a model does not yield the same outputs. It is a feature of current frontier models.

7 Measuring a model against a standard set of tasks not specific to the intended application.

8 Panelists argued that testing is conceptually similar in civilian and military settings but serves different missions, users, and risk tolerances. Limited interaction leaves the testing communities unaware of each other's tools. Others saw the key distinction as military versus law enforcement, border control, or civil unrest control, which operate under different legal and operational contexts.

9 Technology Readiness Levels are a widely used scale for how mature a technology is, from basic research through to proven operation. Human Readiness Levels are a parallel scale for whether humans are able to use that technology as intended.

10 Article 36 of Additional Protocol I to the Geneva Conventions requires a High Contracting Party to the Protocol, in the study, development, acquisition or adoption of a new weapon, means or method of warfare, to determine whether its employment would, in some or all circumstances, be prohibited by the Protocol or any other rule of international law applicable to the High Contracting Party. The obligation of constant care in article 57 of the same Protocol requires constant care in conducting military operations to spare the civilian population. Legal reviews are considered "best practices" even for States that are not party to the Protocol.

11 Model weights are the numerical parameters learned during training; copying them effectively copies the trained model. Inference infrastructure is the computing capacity needed to run a trained model in operation.

12 A voluntary export control arrangement covering missiles and related technology. It has no enforcement machinery.

13 The behavior is often called sycophancy, i.e., the model's tendency to agree with the user's framing, including when the user is mistaken.

MAPS
DIALOGUES

Military AI, Peace & Security (MAPS) Dialogues
Convened by the UN Office for Disarmament Affairs with the support of the Republic of Korea.
mapsdialogues.org

United Nations Office for Disarmament Affairs | Ministry of Foreign Affairs