Capacity-Building as Risk-Reduction Infrastructure for Military AI

From International Dialogue to Practical Governance Capacities

By Irma Argüello

Based on remarks delivered at the UNODA Informal Exchanges on AI in the Military Domain, Geneva, 16 June 2026.

Updated version with Operational Annex - July 2026 • 10 min read

Key messages

Artificial intelligence in the military domain creates systemic risks that no State can manage alone. These risks are distributed, less visible than traditional threats, and may emerge through interactions among systems, doctrines, suppliers, data, human operators and crises.

International cooperation on military AI should be understood as practical risk reduction. Its immediate purpose is not to force trust-building where trust does not exist, but to reduce strategic opacity, improve predictability and build governance capacities.

Capacity-building should be treated as risk-reduction infrastructure: it should go beyond generic training to establish a set of practical capacities for judgment, evaluation, supervision, acquisition, auditability, response, accountability and decommissioning.

For most States, the urgent challenge will not be developing their own advanced military AI systems, but buying, licensing, integrating, operating them without importing dependency and eroding their sovereignty. In this context, the global AI divide is both a development and a security concern.

Legal and Policy Context

The emerging legal and policy context already points toward a life-cycle approach to military AI governance. United Nations General Assembly resolution 80/58 reaffirms that international law, including the United Nations Charter, international humanitarian law and international human rights law, applies throughout the life cycle of AI capabilities in the military domain. It also frames that life cycle broadly, including pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning.

This provides an important basis for treating procurement, auditability, accountability, incident response and decommissioning as governance issues, not merely technical or administrative matters. It also supports the view that States need practical capacities to understand and manage obligations across the full life cycle of AI-enabled military systems.

Resolution 80/58 also recognizes the need to narrow digital and artificial intelligence divides between and within developed and developing countries, with specific consideration for the needs, priorities and conditions of developing countries. It links this concern to enhancing international cooperation and strengthening capacity-building.

At the same time, the broader legal landscape remains fragmented. Discussions on lethal autonomous weapons systems under the Convention on Certain Conventional Weapons address a narrower set of issues related to autonomy, weapons, human control and international humanitarian law. Broader AI governance instruments provide useful principles on risk management, accountability, transparency and systemic risk, but they do not fully address the specific operational, procurement and security challenges raised by AI in the military domain.

This is why a capacity-building agenda for military AI should connect legal, technical, policy, procurement and arms control expertise. The challenge is not only to affirm that international law applies. It is to build the practical capacities States need to know what they acquire, assess whether it can be used responsibly, preserve accountability, communicate incidents and retire systems safely.

Why Cooperation Matters

Some risks are systemic. Their consequences extend beyond the actor directly involved and create shared exposure for others. In such contexts, cooperation is far from an act of goodwill or trust. It is a rational response to risks whose effects cannot be contained within national borders.

This logic is familiar to other fields of international security. In the nuclear field, cooperation has often been grounded less in trust than in converging security interests: the recognition that certain risks are systemic and that their consequences can extend far beyond the place where they originate. Procedures, standards, communication channels and verification practices helped reduce uncertainty and support predictability.

Military AI is different from nuclear risk, but it also presents systemic features. It can accelerate decision-making, introduce opacity into command and control, complicate attribution, reinforce dependency on private providers, create procurement vulnerabilities and increase the risk of miscalculation or unintended escalation.

The starting point for international cooperation in military AI should therefore be a realistic assessment of costs and benefits, with particular attention to the cost of non-cooperation.

The Cost of Non-Cooperation

The cost of non-cooperation is strategic opacity and increasing uncertainty. Without minimum shared concepts, practical channels and basic governance capacities, States may operate in blindness. They may not fully understand what they are buying, what others are deploying, what failed, what was intentional, what was accidental, or what dependencies they are importing through foreign suppliers and digital infrastructure.

In the current strategic environment, the first goal is practical predictability. States do not need to trust each other completely to recognize that certain standards and risk-reduction practices serve their own security interests. Predictability is not a naïve aspiration; it is a practical condition for reducing miscalculations.

Military AI makes this especially important because risk may emerge across the full life cycle of a system. Cooperation that remains confined to rhetoric and high-level principles falls short of real needs. States require capacities that allow them to understand and govern each step.

Across current debates on AI in the military domain, there is growing recognition that international cooperation must move beyond dialogue and produce practical and actionable outcomes: toolkits, legal and technical expertise, procurement guidance, policy frameworks, and channels for communication and response.

Capacity-Building as Risk-Reduction Infrastructure

Capacity-building in military AI, based on international cooperation, should be treated as risk-reduction infrastructure. It should go beyond generic training, workshops, declarations or policy exchanges. Those instruments may be useful, but only if they strengthen real and practical capacities for judgment, evaluation, supervision, acquisition, auditability, response and accountability. This requires mapping risks step by step across the life cycle. States need to know where risks arise, who controls relevant decisions, what evidence is available, what obligations apply, which institutions are responsible and how accountability can be preserved before, during and after deployment.

Capacity-building should be equitable, inclusive and non-discriminatory. It should enable meaningful participation by States with different levels of technological development, that is, all States that will be affected by the deployment, procurement, integration and strategic consequences of frontier or advanced military AI systems. It goes beyond the select group of States able to develop and produce them.

This is why capacity-building should not be treated as a secondary development issue detached from security. In the military domain, the divide between AI-developing and AI-adopting States can become a security vulnerability. It may deepen dependency, reduce oversight, weaken bargaining power with providers and limit the ability of States to participate meaningfully in governance.

Capacity-building should therefore help States develop at least nine core capacities:

1. **Conceptual capacity:** shared terminology and minimum conceptual clarity on military AI risks.
2. **Legal and policy capacity:** ability to understand applicable international law, national obligations, policy frameworks and the legal implications of acquiring, integrating and using AI-enabled military systems.
3. **Assessment capacity:** ability to evaluate systems, use cases, suppliers, data dependencies and operational risks.
4. **Procurement capacity:** ability to conduct due diligence before acquiring or integrating AI-enabled systems.
5. **Supervisory capacity:** ability to audit, monitor and oversee systems across their life cycle.
6. **Red-line analysis capacity:** ability to identify unacceptable outcomes, warning indicators, evidentiary needs and escalation pathways through scenario-based analysis and policy red-teaming.
7. **Incident-response capacity:** ability to communicate, investigate and respond when failures or unexpected behaviors occur.
8. **Accountability capacity:** ability to assign responsibility across institutions, operators, suppliers and decision-makers.
9. **Decommissioning capacity:** ability to safely suspend, modify or retire systems when risks become unacceptable.

These are part of the core infrastructure required to reduce military AI risks in practice and should not be regarded as secondary technical matters.

Procurement as a Security and Sovereignty Vulnerability

The most urgent gaps in military AI governance involve the broader ecosystem through which States acquire and depend on AI capabilities.

A relatively small constellation of States and private companies currently controls the most advanced AI capabilities, infrastructures and supply chains. Most States will not develop advanced military AI systems themselves. They will buy, license, integrate, operate and depend on them.

This makes procurement a matter of security and sovereignty. In military AI, no State, even the most developed, holds all the cards. Advanced systems depend on data, models, compute, cloud infrastructure, specialized talent, supply chains and private companies that may be as powerful as, or more capable than, many States in specific segments of the AI ecosystem.

This reality changes the terms of strategic dependency. As a result, international cooperation should strengthen the States' capacity to understand, negotiate with, supervise and, when necessary, constrain private providers.

Procurement due diligence should therefore go far beyond technical features and performance. It should examine the ecosystem of the provider, including jurisdiction, legal obligations, data access, remote updates, cloud dependency, subcontractors, audit rights, lock-in and the conditions under which the provider may be required or incentivized to prioritize interests other than those of the purchasing State.

A State may believe it is acquiring an AI capability when, in fact, it is importing dependency.

For this reason, governance capacity should precede operational capacity. International cooperation should first strengthen the ability of States to evaluate, supervise and responsibly acquire AI systems before facilitating access to sensitive operational capabilities.

Integrating AI Cooperation with Arms Control Logic

Military AI governance should not develop in isolation from existing international security experience. It should draw on arms control, disarmament, verification, legal and technical expertise.

These communities have practical experience in defining obligations, developing confidence-building measures, managing sensitive capabilities, designing reporting procedures, supporting verification and reducing risks under conditions of mistrust. Their experience is relevant to military AI even if the technologies, timelines and verification challenges are different.

The aim is not to copy older regimes mechanically. The point is to connect technical expertise with legal, policy and arms control expertise. The challenge is to understand what systems can do and to define what States and providers are obligated to know, disclose, prevent, monitor, report and correct.

Lessons from the Nuclear Field as a Practical Toolbox for Military AI

The nuclear field offers practical lessons that can inform international cooperation on military AI. These lessons should not be copied mechanically. AI and nuclear technologies differ profoundly in materiality, speed, diffusion, verification challenges and institutional context. Still, the nuclear field's risk-reduction architecture — safety, security and safeguards — offers a useful toolbox of practices for managing sensitive technologies under conditions of mistrust, complemented by export controls, advisory missions and emergency preparedness.

The value of the nuclear analogy lies in identifying concrete practices that can be translated into military AI governance.

From accounting and control to inventory and traceability

In the nuclear field, a basic practice is to know what material exists, where it is, who controls it, under what conditions and with what records.

In military AI, the equivalent is building inventory and traceability for models used in military functions, sensitive datasets, systems integrated into targeting, ISR, early warning, command and control or cyber operations, versions and updates, providers and subcontractors, logs and audit trails.

The practical lesson is simple: States cannot govern what they have not inventoried.

From safeguards and verification mindset to responsibility by design

Verification, records, review and inspection are not based on the assumption that all actors are trustworthy. They are designed to reduce uncertainty and make relevant behavior observable.

Military AI will not necessarily allow for the same forms of intrusive inspection. However, it does require an auditability mindset: system documentation, testing records, traceability of changes, records of decisions, pre- and post-deployment evaluations and the ability to reconstruct incidents.

Responsibility must be designed before an incident occurs, rather than improvised afterwards, and must remain anchored in human and institutional decisions across the life cycle.

From nuclear security culture to AI risk reduction culture

Risk reduction culture requires training, clear roles, organizational discipline, anomaly reporting and awareness of consequences.

Military AI requires an equivalent AI risk reduction culture inside defense and security institutions. Operators and commanders must know when not to rely on an output, how to avoid automation bias, how to report failures, how to preserve meaningful human judgment and how to suspend or override AI-enabled systems when necessary.

The important lesson is that risk-reduction culture can be trained.

From physical protection and insider threat to model, data and access security

Sensitive technologies require protection against unauthorized access, theft, sabotage and insider threats, as central sources of concern.

In military AI, the equivalent includes access control for models and data, credential security, protection against data poisoning, API control, monitoring of unauthorized uses, provider security and update management.

The risk goes beyond the deployed system in operation. It also includes who can modify it, feed it, update it, extract it or repurpose it.

From export controls and end-use controls to supplier and end-use due diligence

In nuclear and dual-use governance, end-use, end-user, sensitive transfers, components and diversion risks are central to responsible control.

In military AI, the equivalent is more complex but equally important. Procurement should examine who buys, who provides, under which jurisdiction the provider operates, what obligations it has toward its home State, what access it retains, which updates it controls, what restrictions it imposes and what capacity the purchasing State has to audit, suspend or terminate the relationship. Procurement is a security decision.

From peer reviews and advisory missions to voluntary peer learning and red-team exercises

Across the nuclear field, peer reviews, advisory missions, exercises and exchanges of good practices have helped States and operators strengthen safety, security and safeguards-related capacities without necessarily disclosing sensitive information.

In military AI, similar practices could include voluntary exercises on procurement, incident simulations, policy red-teaming, peer review of national frameworks and exchanges of practice on legal review, testing, incident reporting and decommissioning.

Cooperation can start with non-sensitive practices before moving to more sensitive areas.

From emergency preparedness and response to incident communication and crisis protocols

A robust nuclear risk-reduction system includes prevention, emergency preparedness and response. It covers communication, coordination, exercises and emergency protocols.

Military AI requires incident communication and crisis protocols: points of contact, communication channels, voluntary reporting of serious incidents, criteria for suspending systems and mechanisms to distinguish accident, malfunction and deliberate action.

The practical lesson is clear: the worst time to build communication channels is during a crisis.

From decommissioning and disposal to secure retirement of models, data and access

In the nuclear field, end-of-life management matters. Materials, facilities, waste, transport and storage continue to create obligations after operational use.

Military AI also requires secure decommissioning, even if the risks are less visible.

This includes deleting data, revoking credentials, disabling access, retiring hardware, documenting the process and preventing models, datasets or configurations from being reused by unauthorized third parties.

A system’s life cycle should be considered complete only when it can no longer create risk.

A Minimal International Cooperation Package

A practical capacity-building agenda for military AI could begin with a minimal international cooperation package, available to all States. Such a package would not seek to harmonize all national approaches, nor would it require States to disclose sensitive military capabilities. Its value would be more practical: to create a common operational baseline that allows States to participate meaningfully in governance, reduce uncertainty and strengthen their own ability to understand, acquire, supervise and respond to AI-enabled military systems.

This kind of package should respond to an increasingly present concern across international debates: cooperation must move beyond dialogue and produce practical tools that States can use.

Meaningful participation in military AI governance requires more than participation in discussions or diplomatic statements. It requires access to legal, technical and policy capacities that allow them to assess risks, understand obligations, evaluate suppliers and make informed decisions.

A minimal international cooperation package could include the following elements.

Shared terminology.

States need a basic common vocabulary for military AI risks, system functions, levels of autonomy, human involvement, auditability, incident reporting and life-cycle responsibilities. Without shared terminology, cooperation remains vulnerable to ambiguity, misunderstanding and talking past each other.

Life-cycle risk assessment.

Risk should be assessed across the full life cycle of military AI systems. This approach is consistent with the life-cycle framing adopted in United Nations General Assembly resolution 80/58 and helps States identify where governance capacities are needed before risks materialize.

Procurement due diligence and skills to engage with private AI providers.

For most States, procurement will be the main point of exposure to advanced military AI. Cooperation should therefore support practical guidance on how to assess providers, contractual obligations, jurisdiction, termination conditions and potential conflicts between the provider’s obligations and the purchasing State’s security interests.

Because private companies will play a central role in the military AI ecosystem, States need capacity to understand, negotiate with, supervise and, when necessary, constrain providers. Cooperation should help States identify which provider-related questions are security-relevant and how to incorporate them into procurement, oversight and accountability frameworks.

Legal and technical toolkits.

States require practical toolkits that combine legal, technical and policy expertise. These could include model clauses for procurement contracts, checklists for supplier assessment,

templates for incident reporting, guidance for human and institutional responsibility, and minimum questions to ask before integrating AI-enabled systems into sensitive military functions.

Auditability and accountability practices.

A cooperation package should help States design auditability before incidents occur. This includes documentation requirements, testing records, change logs, decision records, pre- and post-deployment evaluations, and mechanisms to reconstruct incidents.

The objective is not only technical attribution, which may be difficult, but institutional responsibility that remains anchored across the life cycle.

Incident communication channels.

States should have channels to communicate serious incidents, unexpected system behavior, failures, near misses or ambiguous events that could generate misperception or escalation. These channels do not require full trust. They require points of contact, agreed procedures and minimum expectations for timely communication.

Scenario-based exercises and policy red-teaming.

Cooperation can begin with non-sensitive practices. Voluntary exercises, tabletop simulations, policy red-teaming and peer learning can help States test how they would respond to procurement failures, malfunctioning systems, unauthorized updates, cyber compromise, automation bias, escalation dynamics or decommissioning challenges. They could also help States explore which AI-enabled military uses, behaviors, deployment conditions or provider dependencies may deserve to be treated as red lines. The purpose would be to build the analytical capacity to identify unacceptable outcomes, warning indicators, evidentiary needs and possible escalation pathways.

Secure update and decommissioning practices.

A minimal package should include guidance on secure updates, credential revocation, data deletion, access control, documentation, retirement of hardware, and prevention of unauthorized reuse of models, datasets or configurations.

A minimal international cooperation package can help reduce the gap between States that develop advanced systems and States that mainly acquire, integrate or depend on them. It would also support more equitable, inclusive and non-discriminatory participation in military AI governance by giving States practical capacities, not only invitations to dialogue.

Such a package could be developed through a flexible, multi-stakeholder and State-led process, drawing on existing expertise from disarmament, arms control, international law, AI safety and security, procurement and technical evaluation communities.

In a crowded field of forums, panels and principles, the value of international cooperation should be measured by the practical capacities it builds.

Policy Recommendations

Governments, international organizations, and relevant stakeholders should:

1. Treat capacity-building as risk-reduction infrastructure. Design capacity-building to reduce opacity, improve predictability and strengthen real governance capacities across the life cycle of military AI systems.

2. Promote inclusive and practical cooperation. Support States with practical toolkits, legal and technical expertise, policy frameworks, procurement guidance and institutional capacities to participate meaningfully in military AI governance.

3. Address the global AI divide as a security concern. Recognize that, in the military domain, the divide between AI-developing and AI-adopting States can increase dependency, reduce oversight and limit meaningful participation in governance.

4. Integrate AI cooperation with arms control and international security expertise. Draw on arms control, disarmament, verification, legal and technical communities to define obligations, manage sensitive capabilities and develop risk-reduction procedures under conditions of mistrust.

5. Prioritize governance capacity before operational access. Strengthen States' ability to evaluate, audit, supervise and responsibly acquire AI-enabled military systems before facilitating access to sensitive operational capabilities.

6. Develop procurement due diligence standards for military AI. Assess not only performance, but also provider jurisdiction, data access, remote updates, cloud dependency, subcontractors, audit rights, lock-in and potential conflicts of obligation.

7. Build incident communication, red-line analysis and secure decommissioning capacities.

Prepare communication channels, points of contact, scenario-based exercises, criteria for suspension, analysis of potential red lines and secure retirement practices before incidents or end-of-life risks materialize.

8. Promote coherence across existing initiatives. Avoid confusing intense institutional activity with effective risk reduction. Cooperation should produce usable tools, not only new forums or declarations.

Conclusion

The purpose of international cooperation is to prevent military AI from making competition more opaque, crises faster, responsibility more blurred and escalation harder to control.

While trust building is desirable, the most immediate challenge is to achieve practical predictability. That requires shared concepts, governance capacities, procurement discipline, auditability, incident communication and institutional accountability.

Military AI, as a field in continuous development, will test the ability of States to cooperate under conditions of competition. Capacity-building should therefore be understood as one of the central pillars of risk reduction.

In a field already crowded with dialogues, principles and initiatives, the next step should be practical: build the capacities States need to understand, acquire, supervise and govern military AI systems before those systems reshape security decisions faster than institutions can respond.

Operational Annex

Five Questions Before Signing a Military AI Contract

For States considering the acquisition or integration of AI-enabled military systems, procurement is not only a technical or budgetary decision. It is also a governance and sovereignty decision. Before signing, States should be able to answer at least five questions:

1. Who controls updates and remote access?
Can the provider modify, suspend, degrade, update or remotely access the system after deployment? Under what conditions, and with whose authorization?

2. What jurisdictional obligations bind the provider?
Could foreign law, export controls, sanctions, corporate decisions or government orders affect the continuity, availability or behavior of the system?

3. Can the State audit, test and explain the system?
What documentation, logs, evaluation results, testing access and evidence will be available to national authorities before and after deployment?

4. What happens in case of incident, misuse or escalation?
Who reports the incident, who investigates it, who can suspend operation, and what communication channels exist with the provider and relevant authorities?

5. Can the State exit without losing sovereign capacity?
Is there any vendor lock-in? Can data, interfaces, operational knowledge and critical functions be transferred, replaced or maintained without the provider?

The central question is whether a State can govern AI military capabilities before those capabilities reshape security decisions, operational dependencies and sovereign choices.

Selected References

United Nations General Assembly. *Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security*. Resolution A/RES/80/58, adopted 1 December 2025.

United Nations Secretary-General. *Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security*. Report of the Secretary-General, A/80/78, 2025.

United Nations Office of Disarmament Affairs. *Artificial Intelligence in the Military Domain*. UNODA.

United Nations Institute for Disarmament Research. *Artificial Intelligence.* UNIDIR.

International Atomic Energy Agency. *Safeguards and Verification.* IAEA.

International Atomic Energy Agency. *Nuclear Security.* IAEA.

International Atomic Energy Agency. *Review Missions and Advisory Services.* IAEA.

Irma Argüello

Global strategist. Founder and Chair of the NPSGLOBAL Foundation. With extensive experience in strategic foresight, global security, WMD nonproliferation, and ethical frameworks for artificial intelligence, Irma brings a transdisciplinary perspective that bridges hard science, policy, and diplomacy.