Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security

An Evidence-Based Road Map for Future Policy Action

Giacomo Persi Paoli, Yasmin Afina, Sarah Grand-Clément,
Ioana Puscas, Shimona Mohan and Jessica Abowitz

United Nations Institute for Disarmament Research
Palais des Nations | 1211 Geneva | Switzerland | + 41 22 917 1141 | unidir@un.org | unidir.org

Table of Contents

Abbreviations

AI	Artificial intelligence
C2	Command and control
CBM	Confidence-building measure
IHL	International humanitarian law
ISR	Intelligence, surveillance and reconnaissance
LAWS	Lethal autonomous weapon systems
NC3	Nuclear command, control and communications
REAIM	Responsible Artificial Intelligence in the Military Domain (summits)
SOP	Standard operating procedure
TTP	Tactics, techniques and procedure
WMD	Weapons of mass destruction

Executive Summary

Artificial intelligence (AI) is rapidly transforming the military domain and profoundly influencing international peace and security. Initiatives such as the summits on Responsible AI in the Military Domain (REAIM) and the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, while not being universal processes, have significantly elevated international attention on the military applications of AI. In particular, they have moved the debate beyond lethal autonomous weapon systems (LAWS) and have successfully highlighted the multifaceted impacts of AI, and so have fostered broader international policy engagement. Building on the political momentum generated by these initiatives, resolution 79/239 adopted by the United Nations General Assembly in December 2024 has further expanded the international dialogue around AI in the military context and has offered Member States, international and regional organizations and the multi-stakeholder community the opportunity to share their views on opportunities and risks.

For many years, the United Nations Institute for Disarmament Research (UNIDIR) has played an important role in shaping and informing these discussions and efforts. It has undertaken research, facilitated multilateral dialogues, and offered policy insights that underline AI’s transformative potential for international peace and security.

The international community can now shape how AI is used in the military domain, putting principles of responsible AI at the core. A central challenge is the complexity of defining the “military domain”. States and regions interpret the scope of this domain differently based on their unique security landscapes, realities and operational practices. For some countries, military roles extend to internal security tasks such as policing, border control, combating organized crime, protection of critical infrastructure or humanitarian relief in response to natural disasters. Others maintain a stricter definition, limiting military functions to battlefield engagements. These variations, rather than serving as barriers, offer important context for multilateral discussions. International governance frameworks must remain flexible and inclusive, acknowledging and adapting to diverse national and regional security perspectives.

In the many operational contextswithin the military domain,AI acts as a force multiplier across several military tasks, including command and control (C2), information and intelligence, advanced autonomy, logistics, training and simulation, and organizational and support functions. In C2, AI enhances the speed and quality of decision-making, thereby helping commanders rapidly analyse battlefield scenarios. It has thepotential to improve adherence to international humanitarian law (IHL) by integrating detailed proportionality and other legal assessments. AI-driven intelligence tools analyse large volumes of data at speed, and so improve situational awareness and threat detection. In logistics, AI optimizes supply chains and predictive maintenance, enhancing operational readiness. AI further supports advanced autonomy in drones, cybersecurity, cognitive warfare and information operations. Training and simulation benefit from AI by creating personalized, realistic synthetic environments and scenarios. In short, if developed, deployed and used responsibly, AI could increase operational effectiveness, mitigate risks and reduce harm.

However, integrating AI in military contexts also presents significant risks and challenges – technological, security, legal, policy and ethical.

Technologically, military AI systems face issues related to the quality, availability and inherent biases of data. These may lead to unpredictable and potentially harmful outcomes, including violations of international law. The "black box" nature of AI systems, often coupled with their adaptiveness and highly context-dependent nature, complicates trustworthiness assessments and may, at times, challenge the conduct of effective investigations into alleged violations of IHL. Cybersecurity vulnerabilities also expose AI systems to adversarial attacks, requiring stringent security measures.

Security challenges include risks of miscalculation and unintended escalation, particularly through AI-enabled rapid decision-making processes and AI-enabled autonomy, which may result in escalatory responses. The potential for an AI arms race exacerbates international and regional tensions, possibly leading to destabilizing outcomes similar to historical arms competitions. The proliferation of AI technologies to non-state actors further complicates threat landscapes and necessitates robust life-cycle management of military AI systems. Additionally, AI-generated disinformation threatens societal stability by undermining trust in information and can have a direct impact on military operations.

Legal challenges revolve around ensuring compliance with international law, particularly IHL and international human rights law. Key debates focus on, among other things, accountability and both state and individual responsibility for AI-driven actions, especially regarding lethal decisions. States diverge on whether existing legal frameworks are sufficient or if new, specialized regulations are required. Beyond international law, ethical considerations emphasize maintaining human judgment in critical decision-making and preventing societal biases from infiltrating AI systems. The latter requirement calls for greater diversity and inclusivity in AI development. Additionally, bridging gaps between government, academia and the private sector remains challenging yet crucial for effective governance.

Addressing these challenges requires a comprehensive road map with actions at the multilateral, regional and national levels.

Multilaterally, establishing a United Nations-led comprehensive platform that enables a regular institutional dialogue to address military AI's broader implications on international peace and security is key. This platform could build on the existing internationally developed AI principles and frameworks, such as UNESCO's recommendations or the commitments made in the Global Digital Compact (e.g. safe, secure and trustworthy AI) and further refine them for application in the military domain. In addition, the United Nations could be leveraged as a platform to develop practical confidence-building measures (CBMs), lead inclusive multi-stakeholder engagement, and deliver global capacity-building programmes that enhance global security via transparency, cooperation and predictability.

Regionally, existing organizational frameworks can be used to tailor CBMs and guidelines that reflect local security contexts. Cross-regional dialogues would facilitate mutual learning, prevent information silos, and include diverse perspective which would encourage globally coherent responses.

Nationally, states should develop comprehensive AI strategies that detail vision, priorities and governance frameworks, ensuring compliance with international norms and ethical standards. Robust governance structures (e.g., dedicated AI steering committees and ethics boards), alongside iterative legal reviews, would enhance accountability and safety. Transparent communication and clearly defined accountability protocols would further support responsible AI implementation. High standards of data governance, life-cycle management approaches, rigorous training programmes and updated military operational guidelines complete these proposed national measures, ensuring the responsible integration of AI in the military domain.

In conclusion, AI's integration into military contexts presents both significant opportunities and complex challenges for international peace and security. Through proactive governance, inclusive dialogue and context-sensitive frameworks, states can leverage AI's strategic advantages while mitigating associated risks. Embracing diversity in definitions and operational contexts, alongside concerted multilateral, regional and national actions, will provide a robust foundation for responsible and effective governance of military AI.

1. Introduction

Artificial intelligence (AI) is rapidly transforming the military domain, with profound implications for international peace and security. Until recently, multilateral discussions on military uses of AI were limited to the question of how this technology relates to lethal autonomous weapon systems (LAWS) – an important yet narrow field of application. In late 2024, however, the United Nations General Assembly adopted a landmark resolution that recognized the wide range of military applications of AI and called for the examination of this technology in the military domain beyond weapon systems. This resolution built on the growing awareness of AI in the military domain and the increase in its policy traction over the past 3 years.[1] This has been prompted by initiatives outside the United Nations, such as the Responsible AI in the Military Domain (REAIM) summits and the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.[2] These processes were fundamental in increasing awareness and served as incubators for policy action on the international stage.[3] Against this backdrop, for many years UNIDIR has contributed significantly to initiating and shaping national, regional and international discussions through its research, its capacity-building and its convening power.

The push for responsible AI in the military domain has opened new channels for dialogue among states. The shared recognition of AI's disruptive potential, both positive and negative, has led to international discussions specifically about ensuring its safe, controlled development, deployment and use. The international community now has an opportunity to shape the future of international peace and security in the era of AI, putting principles of responsible AI at the core. Such engagement can build trust and mutual understanding, future-proofing the international peace and security architecture.[4]

To further advance multilateral discussions on this new and fast-evolving issue, it is crucial to clarify what "the military domain" means and entails; to survey key applications of AI in military settings in order to understand the associated opportunities; and to analyse the challenges and consider recommendations for policy development at all levels. This report addresses each of these aspects in turn (in Sections 2–4), drawing on UNIDIR's research and analysis on these topics over the years.

[1] General Assembly resolution 79/239, "Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security", 24 December 2024, https://docs.un.org/en/A/RES/79/239.

[2] The first REAIM summit was organized by the Kingdom of the Netherlands in 2023 and co-hosted by the Republic of Korea, which, alongside the Netherlands, Singapore, Kenya and the United Kingdom, hosted the second summit in Seoul in 2024. The third summit will be hosted by Spain in September 2025. For more information see REAIM 2023, https://www.government.nl/ministries/ministry-of-foreign-affairs/activiteiten/reaim/about-reaim-2023; REAIM 2024 https://reaim2024.kr/reaimeng/index.do; REAIM 2025: https://exteriores.gob.es/campanas/REAIM2025. On the Political Declaration see United States Department of State, "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy", https://www.state.gov/bureau-of-arms-control-deterrence-and-stability/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy.

[3] Government of the Netherlands, "REAIM 2023 Call to Action", 16 February 2023. https://www.government.nl/binaries/government/documenten/publications/2023/02/16/reaim-2023-call-to-action/REAIM+2023+Call+to+Action.pdf; Government of the Republic of Korea, "REAIM Blueprint to Action", 10 September 2024.

[4] Y. Afina, *The Global Kaleidoscope of Military AI Governance* (Geneva: UNIDIR, 2024); G. Persi Paoli et al., *Modernizing Arms Control: Exploring Responses to the Use of AI in Military Decision-Making* (Geneva: UNIDIR, 2020).

It then (in Section 5) proposes a 10-step road map towards effective national and international governance of AI in the military domain.

2. Defining the Military Domain

Different regions and states define the scope of the "military domain" in various ways. For example, in some national and regional contexts, the military's role extends into internal security and public safety functions – such as assisting in law enforcement, border protection or countering organized crime – effectively blurring the line between defence and policing.[5] A number of states entrust their armed forces with facilitating and implementing humanitarian aid operations, including the delivery of medical supplies in rural areas and natural disaster relief efforts. In other contexts, the military domain is understood more narrowly, focusing strictly on defence and the conventional operations of armed forces, in particular combat operations abroad, with internal security handled by separate entities.[6]

These distinctions are rooted in each state's unique security environment: different regions face different threats, perceive risks differently, and deploy their forces under different legal and normative frameworks.[7] This leads to diverse interpretations of what falls in the "military" domain.[8] In short, there is no single universal definition – the military domain can encompass a broad spectrum of activities in one country, while being confined to warfighting duties in another.

Importantly, such variation in defining the military domain should not be seen as an obstacle to international governance of AI or to dialogue on AI. In fact, recognizing and respecting these nuances, and acknowledging the degree of mutual influence that military and non-military uses of AI will have, can lead to a more inclusive policy debate and ultimately strengthen global AI governance. To be effective, governance should therefore remain sensitive to regional security perspectives, ensuring that frameworks for governing military AI are adaptable to the realities of various states and regions.[9]

A possible approach to defining the military domain is to map the different operational contexts in which military forces may be deployed following a structured categorization that can, in turn, inform AI governance discussions. Table 1 summarizes such a categorization.

A first distinction to be made is that between the use of AI capabilities by military forces in armed conflict as defined by international law (both international armed conflicts and non-international armed conflicts) and their use by military forces when deployed on other operations. This distinction is key as the distinct contexts prompt different legal, operational, technical and ethical questions. However, it may not be practical to use beyond the legal assessment given that the same type of operation could fall within or outside the scope of armed conflict based on the operation's intensity or the level of violence.

Within the broader category of military deployments that fall outside armed conflict as defined by international law, a further distinction can be made between (a) military operations that require the use of force (below the threshold of armed conflict) such as counter-terrorism or counter-piracy or counter-insurgency operations, (b) military operations in support of national security and public safety, such as support for national law enforcement, border security or protection of national critical

[5] Afina, *The Global Kaleidoscope*.

[6] Ibid.

[7] Ibid.

[8] G. Persi Paoli and Y. Afina, *AI in the Military Domain: ABriefing Note for States* (Geneva: UNIDIR, 2025).

[9] Afina, *The Global Kaleidoscope*.

infrastructure sites, and (c) military assistance, which includes scenarios such as peace operations, humanitarian and disaster relief, evacuation of civilians, and search and rescue.

It should be noted that there are inevitable areas of overlap between different categories and that where each specific operation fits will be highly dependent on context. Hybrid operations, for example, could have elements spanning across the various categories.

Table 1. Unpacking the Military Domain : An Illustrative Example of Classification of Operations

Armed conflict	Other military operat ions requiring the use of force *	Support to national security	Military assistance
• International armed conflicts • Non-international armed conflicts	• Counter-piracy • Counter-terrorism • Counter-insurgency • Combating organized crime	• Support for national law enforcement • Border security • Protection of national critical infrastructure sites	• Peace operations • Humanitarian and disaster relief • Evacuation of civilians • Search and rescue

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

Embracing th e plurality illustrated by this categorization ensures that no state is left out of the conversation. In sum, diversity in defining the military domain is a reality to be embraced. When international efforts account for these differences from the outset, they can foster trust, enhance buy-in from all regions, and pave the way for effective and sustainable governance of AI in the military sphere.[10]

[10] Ibid.

3. Exploring the Opportunities: Applications of AI in the Military Domain

Across the various operational contexts described above, defense and military organizations[11] are exploring and implementing AI across a broad spectrum of applications, many of which do not involve weapons or lethal force.

AI[12] is seen as a potential *force multiplier* that can enhance efficiency, decision-making and effectiveness in numerous military functions.[13] As technology evolves, new use cases (including both new systems or enhancing existing ones through AI) are developed or refined or – if proven unreliable or not cost-effective compared to non-AI alternatives – abandoned. Based on UNIDIR's research and engagement with states and experts,[14] the opportunities, actual or potential, deriving from the adoption of AI in the military domain can be grouped in the following non-exhaustive categories:

3.1. Command and Control

AI decision-support tools can aid commanders in tasks such as mission planning, target analysis and course-of-action development. For instance, an algorithm might help analyse battlefield data to identify high-value targets or to optimize mission plans by gaming various scenarios with mission parameters and constraints decided by human operators (e.g., limits on geography, time, tolerance for collateral damage, etc.). AI-enabled command and control (C2) systems (including applications that rely on large language models[15]) can collect live data, process information at scale and suggest options faster than human staff alone. This has the potential to improve the tempo and quality of command decisions.

Related to the issue of C2 is the potential that AI carries to enhance legal compliance. If used correctly, AI could consolidate complex assessments of proportionality (e.g., integrating data on blast radius,

[11] Military organizations, like the Army, Navy, and Air Force, are the armed forces responsible for national defence, while defence organizations, such as the Ministry of Defence, are government agencies that oversee and support the military. In essence, military organizations are the operational forces, while defence organizations provide the framework and resources for those forces.

[12] It is important to note that 'AI' does not refer to a single, unified technology; rather, it represents a diverse family of technologies, each tailored and adapted to support specific military applications and operational contexts. Examples include Large Language Models (LLMs) for language processing and analysis, computer vision systems for imagery interpretation and target identification, and machine learning algorithms supporting predictive analytics and autonomous decision-making.

[13] S. Grand-Clément, *Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain* (Geneva: UNIDOR, 2023); Afina, *The Global Kaleidoscope*; Y. Afina, UNIDIR Briefing to the Security Council Arria-Formula Meeting: "Harnessing Safe, Inclusive, Trustworthy AI for the Maintenance of International Peace and Security", 4 April 2025, https://unidir.org/wp-content/uploads/2025/04/UNIDIR_Yasmin-Afina_Briefing_UNSC_Arria_AI_4April-2025-3.pdf.

[14] See, for example, Y. Afina and G. Persi Paoli, *Governance of Artificial Intelligence in the Military Domain: A Multistakeholder Perspective on Priority Areas* (Geneva: UNIDIR, 2024); Afina, *The Global Kaleidoscope*; Grand-Clément, *Artificial Intelligence Beyond Weapons*; Persi Paoli et al., *Modernizing Arms Control*; G. Persi Paoli and S. Dominioni, *Exploring the AI–ICT Security Nexus* (Geneva: UNIDIR, 2024); UNIDIR, "The Roundtable for AI, Security and Ethics: Forging Global Alignment through Multistakeholder Dialogue", https://unidir.org/event/the-roundtable-for-ai-security-and-ethics-forging-global-alignment-through-multistakeholder-dialogue/; UNIDIR, "The Second Roundtable for AI, Security and Ethics (RAISE)", https://unidir.org/event/the-second-roundtable-for-ai-security-and-ethics-raise/.

[15] I. Puscas, *Large Language Models and International Security: A Primer* (Geneva: UNIDIR, 2024).

population density and timing) to advise on whether a strike can be conducted within the limits of international humanitarian law (IHL).[16] Similarly, AI could help enforce precautionary measures, such as suggesting alternate tactics that reduce civilian risk. These uses illustrate how AI might strengthen the adherence of decision-makers and users to IHL by providing commanders with better information and recommendations to minimize harm.[17] In a broader sense, a growing number of states recognize that integrating AI into military planning might allow for more objective, data-driven assessments of proportionality or necessity in the use of force.[18]

3.2. Information and Intelligence

AI can be leveraged for intelligence, surveillance and reconnaissance (ISR) – analysing vast streams of sensor data, satellite imagery and communications traffic to detect patterns or threats. Machine learning can, in principle, automate the processing of anything for which data exists, from reconnaissance footage to cybersecurity logs, and thereby uncover insights that human analysts might miss.

AI systems can also assist in information management, fusing data from multiple sources and disseminating relevant intelligence to units in the field. These applications enhance situational awareness by sifting through “big data” for actionable information, and then processing and digesting the information for the user’s consumption.

3.3. Advanced Autonomy

AI can enable a more advanced level of autonomy in both physical and digital systems. In the physical world, this can mean, for example, uncrewed systems that are more capable of performing various tasks even in communications-denied environments or where direct supervision cannot be guaranteed due to environmental circumstances or adversarial action. Even in a weapon system where the final decision to fire is under direct human control, AI can offer significant operational and tactical advantages by enabling a more sophisticated level of autonomy.

In the digital domain, AI can be used for cybersecurity (e.g., to strengthen national cyber resilience by improving threat intelligence, network monitoring, and incident response and recovery), as well as to strengthen offensive cyber capabilities.[19] Also within the digital domain, AI can reinforce and support cognitive warfare capabilities and information operations more broadly, including intelligence and counter-intelligence.

3.4. Logistics and Support

Behind the front lines, AI can significantly improve the logistical backbone of military structures, a key enabler of the sustainability of military operations. This includes predictive maintenance of equipment (using AI to anticipate failures or servicing needs), managing supply chains and transport, and optimizing the deployment of personnel and materiel. For example, AI algorithms can improve force

[16] Persi Paoli and Afina, *AI in the Military Domain*.

[17] Ibid.

[18] Y. Afina and S. Grand-Clément, *Bytes and Battles: Inclusion of Data Governance in Responsible Military AI* (Geneva: UNIDIR, 2024).

[19] Persi Paoli and Dominioni, *Exploring the AI–ICT Security Nexus*.

readiness by routeing supply convoys more efficiently or allocating spare parts based on predicted demand. Such uses often adapt civilian AI solutions (e.g., in transportation or inventory management) for military purposes, thus constituting a prime example of the blurred lines between civilian and military applications.

3.5. Training and Simulation

AI-driven systems can be used to train military personnel. Intelligent tutoring systems, war-gaming simulators and virtual reality trainers can personalize and optimize scenarios in synthetic environments and provide feedback to trainees. For example, by building on pre-existing intelligence data, lessons and good practices from past operations, AI can generate realistic adversary behaviours in simulators or suggest improvements to training programmes by analysing performance data. These applications help prepare forces for real operations more effectively and efficiently, including from a cost perspective.

3.6. Other Organizational and Support Functions

Militaries can also apply AI in administrative and support roles – sometimes similar to civilian sector applications. This can include AI tools for personnel management (e.g., recruitment or talent-management analytics) or for medical support (e.g., diagnostics and telemedicine for deployed forces). Many armed forces are also experimenting with AI-enabled systems for back-end support tasks such as finance or procurement. While not unique to the military, when such systems are used by armed forces, they fall within the military domain. Notably, even if such applications may seem distant from an operational context, they are strategic components that contribute significantly to the ability of any armed force to mobilize. This exposes them to threats by adversaries in the same way as more front-line targets.

In sum, AI use in the military domain ranges from the tactical to the strategic and from combat to support. The common theme is that AI in the military domain goes well beyond weapons: it encompasses support systems, decision aids and analytical tools intended to improve military effectiveness. These applications are becoming more widespread as the technology matures. Many armed forces consider that adoption of AI is essential to keep pace with evolving forms of warfare – such as cyber and hybrid threats – and to maintain a competitive edge.

AI and Weapons of Mass Destruction

While this policy note is not intended to provide a deep analysis of any specific type of application or system where AI could be implemented, for the sake of completeness it is important to note how the debate on the convergence between AI andweapons of mass destruction (WMD) has taken a different approach based on the type of weapon systems.

In the context ofnuclear weapons, the vast majority of the debate has focused on the integration of AI in nuclear command, control and communications (NC3) systems. While the combination of complexity and secrecy surrounding these systems makes any detailed discussion on the impact of AI speculative to a degree, it is worth noting that there is increased pressure on nuclear-armed states to agreeto exclude or limit the integration of AI in NC3Several of these states have declared that they have no plans to integrate AI into nuclear decision-making. Beyond NC3 and nuclear decision-making, an embryonic yet growing body of research seeks to reflect further on the AI–nuclear nexus. This ranges from the opportunities that AI offers for monitoring and verification to the implications for the nuclear supply chain.

In the context of chemical and biological weapons, discussions remain focused on the use of AI in early phases of research and development (e.g., discovery of new biological or chemical agents). However, there are otherpotential uses of AIto augment production of materials or information for the development of chemical and biological weapons. While these are upstream applications that remain distant and somewhat unrelated to the operational context, they have, at least in principle, military significance. Moreover, the application of AI to accelerate chemical and biological weapon-related mis- and disinformation should not be understated.

4. Unpacking the Challenges

In addition to its promises, the integration of AI into military affairs brings significant challenges. These challenges can be categorized into three broad areas: (a) technological challenges intrinsic to AI systems and their development; (b) security challenges deriving from their use; and (c) legal, policy and ethical challenges regarding governance and responsible use.

4.1. Technological Challenges

The very nature of military AI systems means that they face a host of technical hurdles related to reliability, safety and transparency.

Data Quality and Availability

One fundamental issue is the quality and availability of data. AI algorithms (especially machine learning models) require vast amounts of training data;[20] but, in military contexts, relevant data may be scarce, incomplete or biased.[21]

Failures of AI systems can stem from "known unknowns": hidden vulnerabilities or edge cases in the data and code that designers did not anticipate.[22] If an AI system has not encountered a certain scenario in training data, it may respond unpredictably in the real world. This mutability is dangerous in high-stakes military settings.

Data bias is another concern – if the underlying data reflects biases on the basis of gender, race, age, ability, culture or other demographic qualifiers, the AI system can reproduce or even amplify those biases in its outputs.[23] Evidence from the civilian applications of AI systems provides examples that can be transposed into military contexts and used to foresee potential risks.[24] Biased algorithms might, for example, misidentify targets or civilians based on flawed patterns, undermining both effectiveness, legal compliance and ethical obligations. For instance, an intelligence-collection system may not necessarily be trained on data that factors in specific cultural contexts in which the bearing of arms may be accepted; the system may subsequently misidentify such practices as possible threats.[25] As another example, biased algorithms employed in military systems, such as computer vision for surveillance drones, may misidentify a civilian man as a combatant based on an assumption (drawn from misrepresentative data sets) that most combatants are men. This is especially dangerous because such biases have proven difficult to correct and mitigate.

[20] It should be noted that recent open-source, cost-effective AI models (e.g., DeepSeek-R1) challenge the idea that advanced AI requires massive resources, including computing and data. However, the applicability of such models in the military domain remains to be assessed and validated.

[21] Afina and Grand-Clément, *Bytes and Battles*.

[22] A. H. Michel, *Known Unknowns: Data Issues and Military Autonomous Systems* (Geneva: UNIDIR, 2021).

[23] Gender and Disarmament & Security and Technology Programmes, "Gender and Lethal Autonomous Weapons Systems", Factsheet, UNIDIR, 2024.

[24] K. Chandler, *Does Military AI Have Gender? Understanding Bias and Promoting Ethical Approaches in Military Applications of AI* (Geneva: UNIDIR, 2021).

[25] Afina and Persi Paoli, *Governance of Artificial Intelligence in the Military Domain*

Opacity

Another technical challenge is the opacity or "black box" nature of many AI models. Modern machine learning (e.g., deep neural networks) often operates in ways that are not explainable to human operators. This lack of transparency makes it hard to assess the trustworthiness of AI systems and to diagnose errors. In military use, an inability to understand why an AI system made a recommendation or took an action can erode human confidence and complicate accountability.

Ensuring interpretability, explainability or traceability of AI decisions is an unresolved technical problem. Techniques for "explainable AI" exist but applying them to complex military systems is an ongoing challenge.[26] In addition, traceability in systems or, at the very least, robust documentation and forensic evidence protocols would ensure that states are able to comply with the IHL obligation to conduct effective investigations into alleged violations.[27]

Testing and Evaluation

Testing and evaluation of AI systems pose further difficulties. Traditional military procurement relies on sequential testing (i.e., prototype trials, then operational testing) with deterministic systems. AI systems, in contrast, are adaptive and their performance may change with new data. This requires any such AI system to be continuously evaluated, including iterative legal reviews to ensure its consistent compliance with international law.[28]

This complexity is further exacerbated by the context-dependent nature of AI systems and the related non-transferability of performance: for example, a system that meets the assurance criteria for deployment in a desert environment cannot be assumed to perform equally well in a snowy environment. It is crucial to keep this limitation in mind as many models and much training data will be subject to technology transfer among security allies and partners. This transfer further complicates the interoperability of systems and expectations of performance even in relation to shared regional objectives. Moreover, when AI components are integrated into larger "systems of systems", new failure modes or cyber vulnerabilities can emerge from interactions of the subsystems, making comprehensive testing even more complex.[29]

It is thus essential to assess the trustworthiness of AI systems through rigorous AI assurance processes that are designed to provide authorities with enough evidence-based confidence in the trustworthiness of the system to allow them to authorize its employment in specific contexts.[30]

[26] N. Goussac and M. Pacholska, *The Interpretation and Application of International Humanitarian Laws in Relation to Lethal Autonomous Weapon Systems: Background Paper on the Views of States, Scholars and Other Experts* (Geneva: UNIDIR, 2025).

[27] Y. Afina, *Regional Perspectives on the Application of International Humanitarian Law to Lethal Autonomous Weapon Systems* (Geneva: UNIDIR, 2025).

[28] Afina and Persi Paoli, *Governance of Artificial Intelligence in the Military Domain*.

[29] I. Puscas, *AI and International Security: Understanding the Risks and Paving the Path for ConfidenceBuilding Measures* (Geneva: UNIDIR, 2023).

[30] J. Pinelis and K. Vignard, "Responsible AI vs. AI Assurance: A Semantic Showdown", Presentation, Global Conference on AI Security and Ethics 2025, 27 March 2025, Geneva.

Cybersecurity

Cybersecurity is another technical challenge: AI systems can be attacked or subverted. The purposes of typical cyberattacks on AI systems can be clustered in three main groups: (a) degrading performance (data poisoning, adversarial attacks or sponge attacks); b) accessing data information (model inversion or membership inference); or (c) accessing model information (model extraction).

Cyber-attacks against AI system.

In a data poisoning attack, malicious data is introduced into the training data sets to corrupt the model's learning process and cause inaccurate or biased outputs.

In adversarial evasion attacks, real-time data captured by an AI system is manipulated to trick the model into making incorrect classifications.

A sponge attack specifically targets the resource consumption of a system, causing it to be overwhelmed by legitimate-looking requests or data that the system is unable to process efficiently

Model extraction allows an attacker to duplicate a proprietary AI model by querying it, effectively stealing intellectual property

Model inversion allows attackers to retrieve private data from a model

Membership inference allows attackers to determine whether a certain data sample was part of the training data set

For more information see Puscas, *AI and International Security*

Military AI systems deployed in adversarial environments must contend with deliberate attempts to degrade them. This requires building robustness against spoofing or manipulation, which is an area of active research that has not yet been fully solved. If not hardened, vulnerabilities in AI systems could be exploited by adversaries. This interplay of AI and cyberthreats complicates deployment and necessitates strong safeguards.[31] In addition, AI also enhances traditional cyberthreats to digital systems, requiring governments and military forces to update their cybersecurity postures in the light of this development.[32]

Misuse or Misunderstanding

Finally, the human element intersects with technical issues: an operator may (intentionally or unintentionally) misuse or misunderstand AI outputs. Such problems as automation bias (i.e., over-relying on AI recommendations without question) or algorithm aversion (i.e., distrusting and ignoring AI entirely) can both occur.[33] Poor user interfaces or insufficient training can exacerbate these tendencies.[34]

[31] Ibid.

[32] Persi Paoli and Dominioni, *Exploring the AI–ICT Security Nexus*.

[33] Persi Paoli et al., *Modernizing Arms Control*.

[34] I. Puscas, *Human–Machine Interfaces in Autonomous Weapon Systems* (Geneva: UNIDIR, 2022).

Thus, part of the technical challenge is actually socio-technical – designing AI tools that effectively complement human decision makers and ensuring that users are trained to both know how to use the system and interpret its results correctly. Without this, even a technically sound system may be used inappropriately, leading to errors or accidents.

4.2. Security Challenges

The incorporation of AI into military capabilities raises serious security challenges with the potential to affect peace and stability.

One major concern is the risk of unintended escalation or loss of human agency, including control in conflict. AI-enabled systems, especially those that operate at high speed with no human in the loop, could escalate engagements so fast than humans cannot intervene or de-escalate. If two adversaries deploy AI systems that react to each other in microseconds, a crisis could intensify before commanders have time to negotiate or apply brakes.[35] This perverse yet plausible scenario, sometimes referred to as "flash wars" triggered by algorithmic interactions, is often cited as a new escalation risk. With the prospect of AI Agents, or Agentic AI, on the horizon it becomes more plausible. Yet, even in slower scenarios, AI-enabled decision aids might recommend more aggressive actions or be misinterpreted, leading to inadvertent escalation.

Another security challenge is the prospect of an AI arms race and its impact on global security. Major powers are investing heavily in military AI to avoid falling behind rivals. This competition could lead to a rapid deployment of unproven AI technologies in a bid for superiority, or in the use of the battlefield as a testing ground for novel AI capabilities. History suggests that arms races without guardrails increase mistrust and the likelihood of confrontation. As AI becomes a factor in military balance, some analysts warn of destabilizing effects akin to past arms competitions.[36] In addition, such a narrative may also encourage the rapid and premature adoption of AI at the expense of robust testing, evaluation and acceptance protocols as a result of fears and concerns of falling behind in the supposed AI arms race.

The proliferation of users of AI technology is another security concern. AI tools – many of which are commercially available or open-source – can be repurposed by non-state actors, terrorist groups or other armed groups. The use of AI by non-state armed groups could significantly alter the threat landscape, requiring military forces to develop adequate countermeasures. While the issue of open-source AI and the relative ease of weaponization of commercially available systems are certainly key factors to consider (particularly against concerns that limitations on open-source AI may result in inequities), they are not the only proliferation risks. As military-grade AI systems become more widely available, the risk of their diversion to the illegal market will inevitably increase.[37] This is why adopting full life-cycle management of military AI systems, including strict decommissioning protocols, is of paramount importance.[38]

[35] Puscas, *AI and International Security.*

[36] Ibid.

[37] M. Martinez et al., *Diversion Analysis Framework*, Arms Trade Treaty Issue Brief 3(Geneva: UNIDIR, Conflict Armament Research and Stimson Centre, 2021).

[38] Persi Paoli and Afina, *AI in the Military Domain*; Afina and Persi Paoli, *Governance of Artificial Intelligence in the Military Domain*.

Furthermore, AI has implications for the information environment during both peace and wartime. Generative AI and other tools can produce disinformation at scale.[39] This can erode trust in information and in institutions and has the potential to destabilize societies and to have an impact on the whole conflict life cycle, including peacekeeping operations. For example, during a conflict, AI-generated deepfake videos or fake communications could sow confusion among the civilian population or even among military units. States have raised concerns that AI could be used to disrupt decision-making by flooding the information space with false or misleading data.[40]

4.3. Legal, Policy and Ethical Challenges

The advent of AI in the military also raises profound legal, normative and ethical questions, reinforced by the range and diversity of contexts in which military forces may be operating.

Legal Challenges

A central legal challenge is ensuring that the use of AI complies with existing international law, particularly but not limited to international humanitarian law, as well as international human rights law and international criminal law.

For example, IHL establishes legal provisions and principles such as distinction (i.e., discriminating combatants from civilians) and proportionality (i.e., avoiding excessive harm) in armed conflict. Deploying AI features in combat puts pressure on these principles. The question of how to ascertain state and individual responsibility and accountability if an AI targeting system does not perform or act as intended features prominently in international, regional and national discussions. This challenge is closely related to the perceived risk of an accountability gap. When an incident involving an AI system occurs (e.g., an AI decision-support system misclassifies an object which is then unlawfully engaged), how is responsibility to be attributed? Traditional military command structures assume human intent and control at every level, following the principle of delegated authority. The use of AI may obfuscate the linearity of this process. In addition, establishing corporate liability is an open question with which an increasing number of states and non-state stakeholders are concerned. They are mindful of the nature of public international law, according to which only states and individuals constitute subjects of the law. Avenues such as due diligence, business and human rights frameworks, and contract law are actively investigated as means through which clarifying corporate liability.

Some states argue that existing IHL (i.e., lex lata) is sufficient but needs proper measures to ensure compliance when AI systems are used. Others feel the implications brought by (high levels of) autonomy and the sheer speed of AI pose new legal dilemmas that require new, dedicated rules to establish a certain interpretation of the law as it should be (i.e., lex feranda).[41] Ensuring “meaningful human control” (the legal basis of which remains contested between states and experts[42]) over the use of force is often proposed as a means to satisfy legal requirements and ethical principles. Yet,

[39] Puscas, *Large Language Models and International Security*.
[40] Afina, *The Global Kaleidoscope*.
[41] Goussac and Pacholska, *The Interpretation and Application of International Humanitarian Laws*.
[42] Ibid.

what constitutes "meaningful" and whether "control" is the right concept (compared to judgement, oversight, involvement and other alternatives) remain unsettled.

Finally, there is the issue of conducting legal reviews: Article 36 of Additional Protocol I to the Geneva Conventions mandates state parties, when studying, developing, acquiring or adopting a new weapon or means or method of warfare, to determine whether its employment would be prohibited by the Protocol or any other rule of international law applicable to the state. Applying this to AI systems (at least those that would be classified as a means or method of warfare) may raise a series of challenges; it requires the review of not just hardware but also algorithms and data – a process for which few precedents exist;[43] this must potentially be done over time through iterative legal reviews as the AI system learns from previous deployments and refines its performance.[44] As it stands, there is active debate over how to conduct such a review,[45] to what standard (i.e., simply IHL compliance or system safety approach) and how often.

Policy Challenges

Beyond the legal challenges, the policy and governance domain faces questions on how to regulate military AI at the national and international levels.

At the national level, many countries are only beginning to draft policies and strategies for AI. Such efforts are even more embryonic for applications in the military domain.[46] Not only does this important step provide an opportunity for the national security ecosystem to consult with relevant industry and academic stakeholders, it also serves as a catalyst for additional policy and governance action at the operational and tactical levels.

At the international level, there is no dedicated intergovernmental or multilateral policy process dedicated to AI in the military domain and its implications for international peace and security. This has fragmented discussions on AI and security across different specialist bodies, each looking at a narrow field of application (e.g., cyber or autonomous weapons). While specialist discussions are required given how critical contextual considerations are, the lack of a higher-level, comprehensive policy process on AI in the military domain risks creating governance loopholes that could be exploited by malicious actors. With resolution 79/239, the General Assembly has started down a path that could lead to a broader international process, but differences persist among states on the most appropriate course of action.

Achieving consensus on policy responses might be difficult given the varying perspectives on the military benefits and risks of AI, the different national and regional considerations on its use (see Section 2) and the current geopolitical context that affects all multilateral disarmament discussions. Nonetheless, structured and regular institutional dialogue on this issue is urgent and needed.

[43] Ibid.
[44] Afina and Persi Paoli, *Governance of Artificial Intelligence in the Military Domain.*
[45] Afina, *Regional Perspectives on the Application of International Humanitarian Law.*
[46] Afina, *"Draft Guidelines for the Development of a National Strategy on AI in Security and Defence: Policy Brief*(Geneva: UNIDIR, 2024).

Ethical Challenges

In relation to ethics, AI in the military domain triggers debates about the role of human judgment in decision-making about the use of lethal force and the potential dehumanization of warfare. There is an often-cited concern about delegating life-and-death decisions to algorithms, with the United Nations Secretary-General taking a firm position against this scenario.[47] There are also ethical and societal implications regarding bias and inequities. AI systems can inherit biases from training data sets, unsupervised or uncorrected machine learning algorithms, or human developers with their own biases.[48] Societal biases (on the basis of gender, race, etc.) can be encoded in AI, potentially leading to discriminatory outcomes in targeting or threat assessment.[49] Ethically, it is important to mainstream diversity considerations into military AI development through the life cycle – both to prevent and to correct bias in systems.

Finally, there is the governance challenge of multi-stakeholder involvement. Much AI innovation comes not from government, but from the private sector, research laboratories and academia. Effective governance of military AI will require input and cooperation from industry and research laboratories (which build the technology) and civil society (which articulates ethical norms and public concerns). However, bridging the gap between national security and open technology communities is not straightforward.[50] While some initiatives (e.g., the REAIM Summits or UNIDIR's Roundtable for AI, Security and Ethics, RAISE) include multi-stakeholder participation by design, providing a formal channel through which the multi-stakeholder community can effectively stimulate policy development remains a challenge.

[47] UN Secretary General's message to the inaugural Global Conference on AI, Security and Ethics, https://unidir.org/un-secretary-generals-message-inaugural-global-conference-ai-security-ethics/; United Nations, *A New Agenda for Peace*, Our Common Agenda Policy Brief no. 9 (New York: United Nations, July 2023).

[48] Gender and Disarmament & Security and Technology Programmes, "Gender and Lethal Autonomous Weapons Systems".

[49] Ibid.

[50] W. He and A. Anand, *The 2022 Innovations Dialogue: AI Disruption, Peace and Security* (Geneva: UNIDIR, 2023).

5. Recommendations : A 10-step Road Map

In the light of the above analysis, a range of recommendations emerge that can guide states and stakeholders in maximizing the benefits of military AI while mitigating its risks. These recommendations can be grouped into three tiers: (a) multilateral recommendations for the international community and collective action; (b) regional recommendations tailored to cooperation at the regional level; and (c) national recommendations for individual states to implement in their own policies and institutions. Each recommendation is formulated with a clear indication of *who* should take action, *what* should be done and *why* it is important.

5.1. Multilateral Recommendations

1. **Establish a multilateral process under United Nations auspices to provide a comprehensive platform for discussion on military applications of AI and their impact on international peace and security**

✓ *Who*: United Nations Member States.

✓ *What*: While the integration of AI in the military domain may require context- and application-specific discussions (particularly in relation to the legal debate around existing law versus new law), it is important that such discussions are built on a foundation of common understanding that embraces the impact of AI on international peace and security more broadly. The establishment of a multilateral process under United Nations auspices could be leveraged to add overall coherence in policy discussions, covering all critical aspects and exploring AI convergence across different parts of the international security and disarmament architecture, to achieve one or more of the following objectives:

(a) **Develop a set of overarching, core principles of responsible AI in the military domain** to help align national efforts and reduce risk. This would ensure that core principles of responsible AI are adopted universally as guiding criteria for the development, deployment and use of AI in the military domain. Such principles could draw from the language and concepts adopted by consensus in other multilateral processes such as UNESCO's Recommendation on the Ethics of Artificial Intelligence; the Global Digital Compact, which speaks of safe, secure and trustworthy AI as well as of a responsible, accountable, transparent and human-centric approach to the life cycle of digital and emerging technologies; or the guiding principles adopted as part of the work of the Group of Governmental Experts on LAWS under the Certain Conventional Weapons (CCW) Convention. While acknowledging the specificity of the international peace and security context, the success of the above-mentioned processes demonstrates that it is possible for Member States to achieve consensus on foundational principles for AI. Such principles could provide a strong point of departure, potentially speeding up the process of development of such principles, leaving more time dedicated to their characterization for the international peace and security domain. Moreover, the formulation of such principles would provide a unique opportunity for states to translate and operationalize existing legal requirements specifically in the context of AI in the military domain without prejudice to pre-existing differences in states' legal interpretations and applications (e.g., on the level of transparency, explainability and traceability for compliance with the IHL duty for effective investigations).

(b) In the future, **further develop these core principles into international voluntary norms or guidelines for responsible state behaviour** in the development, deployment and use of AI in the military domain. These guidelines could take the form of a code of conduct or a political declaration supplemented by more technical instruments as required (e.g., on AI assurances, and robust protocols for testing and evaluation), somewhat replicating the approach followed by the United Nations Programme of Action on small arms and its International Tracing Instrument. It should be noted that both the principles and the possible normsor guidelines would not exclude the possibilityof states negotiating legally binding instruments regulating specific use cases in a more restrictive way (e.g., AI in LAWS, or AI in security of information and communications technology). In fact, the presence of an overarching framework of responsible state behaviour could allow more focused, specialist discussions in the appropriate disarmament forums, limiting the risk of potential scope creep.

(c) **Develop confidence-building measures (CBMs) for military AI.** States could agree on and implement practical CBMs to increase transparency and trust regarding AI in the military domain. This menu of options could include voluntary information exchanges; notification regimes for certain AI-enabled exercises or deployments (to avoid misinterpretation); establishment of joint technical expert groups between militaries to share best practices on safety, and creation of a mechanism for states to report incidents or near-misses involving AI systems, to learn collectively from them.

(d) **Promote multi-stakeholder engagement in support of multilateral policy action.** Institutionalizing multi-stakeholder participation in multilateral discussions on military AI at the global level would ensure that governments can benefit from the wealth of knowledge and expertise residing in industry, academia and civil society. Their involvement will bring in fresh ideas (e.g., technology companies can share methods to test AI robustness) and increase buy-in for any solutions proposed. It also widens the perspective beyond purely military considerations to include ethical, legal and societal viewpoints. In practice, this can help ensure that international norms, principles or frameworks for AI are realistic, feasible, comprehensive and sustainable.

(e) **Develop and implement a coherent capacity-building programme.** As AI becomes increasingly widespread in military capabilities, capacity-building will be fundamental to ensure that no one is left behind and that everyone is equipped with the locally owned technical, institutional and operational capacity to adopt this technology responsibly and in compliance with legal requirements. Capacity-building is key to ensuring broader participation in the multilateral process and that rules, norms and principles are developed in an inclusive way. Inclusivity, through effective consultations with the target recipients and close collaboration to ensure national and regional ownership,will provide legitimacy to any international norms or guidelines and help those standards take root universally.

- ✓ *Why*: Collectively, these multilateral actions aim to foster cooperation, set common rules and share knowledge on military AI at the international levelwith a view to increasing predictability. They aim to shape the global landscape so that all states move towards safer and more transparent integration of AI in the military domain, thereby reducing the risks outlined above. While clustered under a single umbrella recommendation, each of the actions above could be implemented on its own , although their mutually reinforcing nature would amplify the impact achieved if they are implemented in combination.

5.2. Regional Recommendations

2. Leverage regional and subregional organizations and dialogues

- ✓ *Who*: Regional bodies (e.g., African Union, European Union, ASEAN Regional Forum, Organization of American States) and groups of states in regions.
- ✓ *What*: **Use existing regional cooperation frameworks to discuss the issue of AI in the military domain .** Regions should incorporate AI into their security agendas and frameworks – for example, convening dedicated sessions or working groups on the regional impact of AI. They could develop region-specific CBMs, norms or guidelines that reflect local contexts and set up networks for information-sharing on AI-related best practices suited to their security landscape. Regional and subregional cooperation could also be leveraged to develop joint AI-development projects, aligning operational, legal and technical requirements.
- ✓ *Why*: Regional and subregional approaches allow tailoring to specific security realities and threat perceptions, which could lead to concrete results that are more aligned with specific needs. In addition, regional and subregional approaches could be leveraged to inform and shape global dialogues and strengthen context-specific capacity-building.

3. Initiate cross -regional dialogues

- ✓ *Who*: Regional organizations in partnership (e.g., European Union–African Union cooperation on technology security).
- ✓ *What*: **Initiate cross -regional dialogues on AI**, where two or more regional groups exchange lessons and possibly align their approaches.
- ✓ *Why*: Cross-regional dialogue can be a useful tool to enable mutual learning and avoid echo chambers. By providing a useful platform for information-sharing and for constructive challenge, this will lead to an overall better preparedness to deal with the risks of AI.

5.3. National Recommendations

4. Formulate and implement a natio nal strategy on AI in security and defen ce

- ✓ *Who*: National governments.
- ✓ *What*: **Develop a comprehensive national strategy or policy on AI in security and defen ce.** This strategy should outline the country's vision on military AI, priority areas (e.g., which applications to pursue or avoid), governance structures, and how it will uphold international law and ethics. It should cover procedural aspects (e.g., how the strategy will be implemented,

reviewed, updated) and substantive aspects (e.g., specific measures on data governance, AI assurance). It should be a whole-of-government effort: it should involve defence, science and technology, justice (legal review), and other relevant departments and agencies, as well as consultations with external stakeholders domestically. Once formulated, the strategy must be implemented via concrete action plans, and it must be monitored and regularly reviewed in the light of technological changes.

- ✓ *Why*: A national strategy ensures that a country does not react ad hoc to AI developments. It enables internal coordination, clarifies roles and responsibilities, and provides a clear direction for the development, acquisition, integration and use of AI in the military domain. It also signals to citizens, regional partners and the international community that the state is committed to developing, deploying and using AI responsibly for international peace and security.

5. Establish robust governance structures and review processes

- ✓ *Who*: National governments.
- ✓ *What*: **Set up permanent governance bodies for AI** within defence institutions (e.g., an AI steering committee at the ministry level to oversee all programmes or ethics review boards to assess high-risk projects) and inter-agency civilian–military working groups to review new developments, anticipate dual-use concerns before deployment and develop potential safeguards against misuse. Additionally, **embed AI considerations into existing processes** such as procurement guidelines to require AI risk assessments, or legal review processes for AI-enabled capabilities (analogous to weapons review). These should possibly be iterative, checking systems at development and again before deployment, to ensure that they are built for consistency and for compliance with IHL and other relevant laws.
- ✓ *Why*: Dedicated structures provide focus and accountability. They create effective checkpoints that AI projects must pass and comply with consistently (e.g., ethical approval, legal clearance, safety certification), reducing chances of unsafe or unlawful deployment. Integrating AI into standard procedures also normalizes the consideration of its implications in all operations planning.

6. Implement transparency and accountability measures

- ✓ *Who*: National governments and defence organizations.
- ✓ *What*: **Adopt measures to foster transparency** at the national level about AI programmes and **accountability** for their outcomes. Internally, this could mean maintaining detailed logs of decisions on AI systems (for auditability) and reporting any incidents or malfunctions up the chain of command. Externally, governments should be transparent (to the extent that security allows) about their approach to military AI – possibly through published strategy documents, press statements on new policies, or engagement with legislative bodies and the public. For accountability, military rules should clarify that commanders are responsible for the actions of AI systems under their command just as with human subordinates, and rules of engagement should be updated accordingly (see recommendation 10). In case of accidents, states should consider having an investigation protocol that includes technical experts to examine the AI-related issues, and they should communicate findings (and corrective actions) publicly when possible.

- ✓ *Why*: Transparency builds public trust and international confidence that a state is using AI responsibly. Accountability ensures that the presence of AI does not create a vacuum of responsibility – maintaining the ethical and legal norm that humans areaccountable for military actions. These measures incentivize careful use and continuous improvement

7. Prioritize data governance and quality

- ✓ *Who*: Defence organizations and national regulators.
- ✓ *What*: **Implement robust data practices and governance frameworks** for all military AI applications. This includes investing in curating high-quality, representative and disaggregated data sets, establishing procedures for data verification and updates, and enforcing data security measures. Before deploying AI systems, guidelines should be in place for data to be responsibly collected, processed, used (covering issues of privacy, minimization of bias and provenance of data), stored and eventual destroyed.
- ✓ *Why*: Good data and good data practices are the backbone of effective AI in the civilian and military domain alike. By prioritizing robust data governance and the provision of the necessary infrastructure to enable it, militaries can improve the performance and trustworthiness of their AI systems and reduce error rates. It will also mitigate such risks as bias and adversarial data manipulation. As an overarching step, treating data as a strategic asset – with appropriate standards and stewardship – is essential to harness AI benefits.

8. Adopt a life-cycle management approach

- ✓ *Who*: Defence organizations and contractors.
- ✓ *What*: **Manage AI capabilities throughout their entire life cycle** – from design and development, through testing and deployment, to updates and decommissioning – with continuous risk assessments and mitigation at each stage. This approach should be documented in policies, strategies or guidelines: for instance, mandating rigorous AI assurance processes, including test, evaluation, verification and validation (TEVV), during development; implementing design choices that promote compliance by design; leveraging procurement processes to reward solutions that prioritize safety and security; establishing monitoring and control measures during operational use (e.g., fallback human-override options); and robustly planning for how systems will be retired or replaced safely.
- ✓ *Why*: A life-cycle view ensures that safety and compliance are not one-time checkboxes but ongoing commitments. This reduces chances of failure in the field and ensures that accountability is maintained throughout the system's use. It also means lessons learned can be fed back into the design of next-generation systems. Robust decommissioning protocols and processes will minimize risks of proliferation and diversion to non-state armed groups and malicious actors, unintended consequences from the system's degradation, and exposure to vulnerabilities.

9. Invest in human capital and training

- ✓ *Who*: Defence organizations and decision makers, in partnership with private sector and educational institutions.

- ✓ *What*: **Develop extensive training programmes for military personnel on AI** and cultivate a new generation of AI-literate officers and specialists. This includes not only technical training but also training on the ethical and legal aspects of AI use in operations. In this regard, strict training requirements should be included in any procurement or government-to-government transfer of military AI-systems. Training should incorporate tailored scenarios into military exercises to enable personnel to gain experience of interacting with AI systems under realistic conditions (including adversarial action) and of test procedures. Military-to-military dialogues could be leveraged to share lessons learned and best practices in order to further build knowledge and capacity. Through training, military personnel should develop an understanding of not only what the technology can and cannot do, but also the parameters used for testing, the test results, the benchmarks used for evaluation and other factors that would contribute to calibrating the trust between human and technology.
- ✓ *Why*: Human expertise and judgment remain critical. Personnel at all levels need to understand AI's strengths and limitations in order to use it properly in any given context. Training reduces misuse (e.g., over-reliance or misinterpretation of AI outputs) and enables more effective human–machine teaming. In the big picture, having knowledgeable staff will help militaries to implement all other recommendations more successfully, from testing to policy and legal compliance.

10. **Review military operational guidelines to strengthen AI governance in military contexts**

- ✓ *Who*: Defence organizations and armed forces leadership.
- ✓ *What*: **Adapt existing or develop new military documentation**, including doctrines, standard operating procedures (SOPs), tactics, techniques and procedures (TTPs), logbooks and after-action reports among others, to account for the impact that AI will have on the conduct of warfare; and **review existing rules of engagement and develop new ones as required** to ensure that the chain of accountability remains clear to operators and decision makers, even after the introduction of AI, and that military operations can be conducted in full compliance with international and national legal frameworks. Militaries already use operational frameworks such as SOPs, TTPs, logbooks and after-action reports to govern behaviour on the battlefield, including the use of systems and technology. These instruments ensure consistency in operations, document best practices, and provide structured learning mechanisms to refine military applications over time.
- ✓ *Why*: Existing military governance tools and instruments can be used to strengthen the governance of AI in the military domain at a more practical, tactical level, thereby offering an impactful complement to the highest levels of governance and the associated obligations emanating from international, regional and national laws and regulations.

6. Conclusion

AI is poised to profoundly influence the military domain, offering new capabilities and efficiencies even as it disrupts established practices and poses novel risks. The analysis above underscores that the implications of military AI are double-edged: on one side, AI can strengthen defence, improve precision and aid human decision makers; on the other, if AI is unchecked, it can introduce instability, uncertainty and ethical dilemmas. As the international community moves forward, the overarching imperative is to maximize AI's benefits for security and peace while minimizing its potential to cause harm or escalation.

The current momentum – from the United Nations General Assembly's resolution acknowledging AI in the military context beyond weapons, to multi-regional consultations and expert dialogues – is a promising sign. It shows a recognition by states and the wider multi-stakeholder community that a proactive and collaborative approach is needed.

Ultimately, the successful integration of AI into the military domain should not be measured just by the capabilities acquired. The degree to which the use of AI upholds or even strengthens the international peace and security architecture should become an important standard of assessment. With deliberate action now, AI's disruptive potential can be managed and directed towards enhancing global security. Conversely, neglecting the governance of military AI could exacerbate arms races or erode the laws of war.

The 10 recommended actions provide a road map for states to establish a robust framework for responsibly governing military AI at the national and international levels. Implementing them will require political will and resources, It will sometimes require cultural change within governments and the military community to embrace a more inclusive approach to governance of military capabilities. However, fully implementing them will greatly enhance a state's readiness to capitalize on the benefits of AI while controlling its dangers.

In conclusion, AI in the military domain is a reality that must be neither overhyped nor underestimated. It is a domain to be carefully shaped. By characterizing its scope, mapping its applications and opportunities, and candidly recognizing its challenges, policymakers and practitioners can chart a path that preserves international security.

Reference List

Selected UNIDIR Resources

Giacomo Persi Paoli and Yasmin Afina, *AI in the Military Domain: A Briefing Note for States* (Geneva: UNIDIR, 2025), https://unidir.org/publication/ai-military-domain-briefing-note-states/.

Yasmin Afina, *Regional Perspectives on the Application of International Humanitarian Law to Lethal Autonomous Weapon Systems* (Geneva: UNIDIR, 2025), https://unidir.org/publication/regional-perspectives-on-the-application-of-international-humanitarian-law-to-lethal-autonomous-weapon-systems/.

Netta Goussac and Magdalena Pacholska, *The Interpretation and Application of International Humanitarian Laws in Relation to Lethal Autonomous Weapon Systems: Background Paper on the views of States, scholars and other experts* (Geneva: UNIDIR, 2025), https://unidir.org/publication/the-interpretation-and-application-of-international-humanitarian-law-in-relation-to-lethal-autonomous-weapon-systems/.

Ioana Puscas, *Large Language Models and International Security: A Primer* (Geneva: UNIDIR, 2024), https://unidir.org/publication/large-language-models-and-international-security-a-primer/.

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

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

Giacomo Persi Paoli and Samuele Dominioni, *Exploring the AI–ICT Security Nexus* (Geneva: UNIDIR, 2024), https://unidir.org/publication/exploring-the-ai-ict-security-nexus/.

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

Gender and Disarmament & Security and Technology Programmes, "Gender and Lethal Autonomous Weapons Systems", Factsheet, UNIDIR, 2024, https://unidir.org/publication/gender-and-lethal-autonomous-weapons-systems/.

Alisha Anand and Henry Deng, *Towards Responsible AI in Defence: A Mapping and Comparative Analysis of AI Principles Adopted by States* (Geneva: UNIDIR, 2023), https://unidir.org/publication/towards-responsible-ai-in-defence-a-mapping-and-comparative-analysis-of-ai-principles-adopted-by-states/.

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

Ioana Puscas, *AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures* (Geneva: UNIDIR, 2023), https://unidir.org/publication/ai-and-international-security-understanding-the-risks-and-paving-the-path-for-confidence-building-measures/.

Wenting He and Alisha Anand, *The 2022 Innovations Dialogue: AI Disruption, Peace and Security* (Geneva: UNIDIR, 2023), https://unidir.org/publication/the-2022-innovations-dialogue-ai-disruption-peace-and-security-conference-report/.

Ioana Puscas, *Human–Machine Interfaces in Autonomous Weapon Systems* (Geneva: UNIDIR, 2022), https://unidir.org/publication/human-machine-interfaces-in-autonomous-weapon-systems/.

Arthur Holland Michel, *Known Unknowns: Data Issues and Military Autonomous Systems* (Geneva: UNIDIR, 2021), https://unidir.org/publication/known-unknowns-data-issues-and-military-autonomous-systems/.

Manuel Martinez, Alfredo Malaret, Erica Mumford and Natalie Briggs,*Diversion Analysis Framework*, Arms Trade Treaty Issue Brief 3(Geneva: UNIDIR, Conflict Armament Research and Stimson Centre, 2021), https://unidir.org/publication/arms-trade-treaty-issue-brief-3-diversion-analysis-framework/.

Giacomo Persi Paoli, Kerstin Vignard, David Danks and Paul Meyer,*Modernizing Arms Control: Exploring Responses to the Use of AI in Military Decision-Making* (Geneva: UNIDIR, 2020), https://unidir.org/publication/modernizing-arms-control/.

Arthur Holland Michel, *The Black Box, Unlocked: Predictability and Understandability in Military AI* (Geneva: UNIDIR, 2020), https://unidir.org/publication/the-black-box-unlocked/.

Selected External Publications

Yasmin Afina and Sarah Grand-Clément, *Bytes and Battles: Inclusion of Data Governance in Responsible Military AI*, Centre for International Governance Innovation (CIGI) Papers no. 38 (Waterloo, ON: CIGI, October 2024), https://www.cigionline.org/static/documents/Afina-Grand_Clement.pdf.

United Nations Documents, Resolutions and Statements

General Assembly resolution 79/239, "Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security", 24 December 2024, https://docs.un.org/en/A/RES/79/239 .

United Nations, *A New Agenda for Peace*, Our Common Agenda Policy Brief no. 9 (New York: United Nations, July 2023), https://peacemaker.un.org/sites/default/files/document/files/2024/08/our-common-agenda-policy-brief-new-agenda-peace-en.pdf.

United Nations Secretary-General, Message to the Inaugural Global Conference on AI, Security and Ethics 2025, Geneva, 27 March 2025, https://unidir.org/un-secretary-generals-message-inaugural-global-conference-ai-security-ethics/.

Presentations

Jane Pinelis and Kerstin Vignard, ‘Responsible AI vs. AI Assurance: A Semantic Showdown”, Presentation, Global Conference on AI Security and Ethics 2025, Geneva, 27 March 2025.

Multistakeholder Dialogues

UNIDIR, “The Roundtable for AI, Security and Ethics: Forging Global Alignment through Multistakeholder Dialogue”, 24 October 2024, https://unidir.org/event/the-roundtable-for-ai-security-and-ethics-forging-global-alignment-through-multistakeholder-dialogue/.

UNIDIR, “The Second Roundtable for AI, Security and Ethics (RAISE)”, 4–6 September 2024, https://unidir.org/event/the-second-roundtable-for-ai-security-and-ethics-raise/.