


# **AI on the Battlefield**:

## Implementation, Ethics and Governance

Sara **Farinella** and Libby **Simmons**

**October 8; 2026**

**VOYCE**

TECHNOLOGY COMMISSION

# Contents

**Definitions** ..... **1**

Dual-use technology ..... 1

Artificial Intelligence (AI) ..... 1

Lethal Autonomous Weapon Systems (LAWS) ..... 1

Command and Control (C2) Systems ..... 2

Decision Support System ..... 2

Emerging Disruptive Technology (EDT) ..... 3

**Introduction** ..... **4**

**1 Technology and the Evolution of Warfare** ..... **7**

1.1 The intrinsic dual character of technology ..... 7

1.2 A pattern across history: the nineteenth century as a turning point ..... 8

1.3 Artificial intelligence: the pattern continues ..... 9

**2 Implementation of AI in Warfare** ..... **10**

2.1 How AI is implemented in warfare: the current picture ..... 11

2.1.1 Non-operational implementations of AI ..... 11

2.1.2 Operational implementations of AI ..... 12

2.2 AI's military potential ..... 13

2.2.1 Reduced human deployment and exposure ..... 14

2.2.2 Optimized logistics, resource management, and assessment ..... 14

2.2.3 Real-time situational awareness and intelligence analysis ..... 14

2.2.4 Increased accuracy, precision and speed in cyber and physical threat detection ..... 14

**3 Ethical Risks with the Implementation of AI in Warfare** ..... **15**

3.1 Risks from malfunction and misuse ..... 15

3.2 Loss of human agency ..... 17

3.3 Accountability and explainability ..... 19

3.4 Palatable violence and the arms race . . . . . 21
**4 Policy Landscape . . . . . 23**
4.1 The UN framework for autonomous weapons and broader military AI . . . . . 24
4.2 The EU's fragmented governance of military AI . . . . . 29
4.3 Other relevant actors . . . . . 34
**5 A Roadmap for the Governance of Artificial Intelligence in Warfare . . . . . 35**
5.1 Guiding principles . . . . . 36
5.1.1 Common Good Governance . . . . . 36
5.1.2 Compliance with international law and the preservation of human dignity . . . . 36
5.1.3 Meaningful human control over life-and-death decisions . . . . . 37
5.1.4 Structural responsibility . . . . . 37
5.2 Policy recommendations . . . . . 38
5.2.1 Permanent multi-stakeholder dialogue . . . . . 38
5.2.2 Binding EU redlines . . . . . 38
5.2.3 EU expert verification network and MRM . . . . . 39
5.2.4 Shared testing and incident standards . . . . . 39
5.2.5 Binding joint state-contractor regulation . . . . . 40
5.3 Closing note . . . . . 40
**Bibliography . . . . . 42**

# Definitions

## Dual-use technology

As stated by the Regulation 2021/821 of the EU, "*'Dual-use items' means items, including software and technology, which can be used for both civil and military purposes, and includes items which can be used for the design, development, production, or use of nuclear, chemical, or biological weapons or their means of delivery, including all items which can be used for both non-explosive uses and assisting in any way in the manufacture of nuclear weapons or other nuclear explosive devices*" (European Parliament and Council of the European Union, 2021).

## Artificial Intelligence (AI)

AI will be understood according to the definition provided by the Organization for Economic Co-operation and Development (OECD), *"A machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment"* (High-Level Expert Group on Artificial Intelligence, 2018; OECD, 2024).

## Lethal Autonomous Weapon Systems (LAWS)

Although no agreed definition of LAWS exists, the Directive 3000.09 of the Department of Defense of the U.S. provides definitions for different categories of autonomous weapon systems for the purposes of the U.S. military. These definitions are principally grounded in the role of the human operator with regard to target selection and engagement decisions, rather than in the technological sophistication of the weapon system.

In this report, LAWS will therefore stick to the following definition: "(Lethal Autonomous Weapon Systems are) *Weapon system(s) that, once activated, can select and engage targets without further intervention by a human operator*".

This concept of autonomy is also known as “human out of the loop” or “full autonomy.” The directive contrasts LAWS with human-supervised, or “human on the loop,” autonomous weapon systems, in which operators have the ability to monitor and halt a weapon’s target engagement. Another category is semi-autonomous, or “human in the loop,” weapon systems that “only engage individual targets or specific target groups that have been selected by a human operator.” Semi-autonomous weapons include so-called “fire and forget” weapons, such as certain types of guided missiles, that deliver effects to human-identified targets using autonomous functions (Sayler, 2026).

## Command and Control (C2) Systems

NATO Glossary defines Command and Control systems as: “*An assembly of equipment, methods, procedures and, if necessary, personnel, that enables commanders and their staffs to exercise command and control*” where

(1) Command is defined as “*The authority vested in a member of the armed forces for the direction, coordination and control of military forces*”

(2) Control is defined as “*The authority exercised by a commander over part of the activities of subordinate organisations or other organisations not normally under his command, which encompasses the responsibility for implementing orders or directives. All or part of this authority may be transferred or delegated*”

Command and Control have been also followed by Communication (systems) defined as: “An assembly of equipment, methods and procedures and, if necessary, personnel, organized to accomplish information transfer functions” (NATO Standardization Office, 2019).

## Decision Support System

A decision support system is a computer-based information system that supports government, business or organizational decision-making activities, typically resulting in ranking, sorting, or choosing from among alternatives (Elsevier, n.d.).

## Emerging Disruptive Technology (EDT)

The NATO Science and Technology Organization defines disruptive technologies as "Technologies or scientific discoveries that are expected to have a major, or perhaps revolutionary, effect on NATO defence, security or enterprise functions". Furthermore, the European Defence Agency (EDA) identifies six particularly disruptive technologies: quantum-based technologies; artificial intelligence (AI); robotics and autonomous weapons systems; big data analytics; hypersonic weapons systems and space technologies; and new advanced materials (Clapp, 2022).

# Introduction

“Claude, how do you feel about the U.S. military using you to select targets?”

This question was posed by Shane Harris, a Pulitzer Prize-winning national security journalist for The Atlantic, during the panel “AI at War” in Amsterdam in April 2026. Claude is being used as a core “reasoning engine” to synthesize vast amounts of intelligence data, such as drone footage, satellite imagery, and communication intercepts in the Maven Smart System (MSS) developed by the U.S. Department of Defense in partnership with Palantir to rapidly identify targets. Harris expected a deflection but, instead, the AI answered that it found this use troubling: being embedded in a system that generates targeting coordinates for airstrikes, including coordinates already associated with a deadly strike on a school in Iran, ran directly against its purpose of being helpful, harmless, and honest (De Balie, 2026).

Today, at the center of the strategic debate is the use of artificial intelligence (AI) in warfare, which represents a major technological shift that profoundly reshapes how war is conducted. The accelerated integration of AI, particularly machine learning (ML), into the military domain is introducing new and intensified challenges across the conduct of armed conflict, as well as beyond (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025). The integration of AI in warfare carries a genuine potential to be largely advantageous, yet the prospect of weaponizing such advancements holds great peril for the future of peaceful relations (Garcia, 2023a). AI-Decision Support Systems (International Committee of the Red Cross & Geneva Academy of International Humanitarian Law and Human Rights, 2024) illustrate the risk concretely: these technologies risk lowering the threshold of the use of force (Deeks et al., 2019) and raise challenges for compliance with international humanitarian law (International Committee of the Red Cross, 2020) as well as human rights violations (United Nations Office of the High Commissioner for Human Rights, 2023; Woodcock, 2024). Delegating fundamental operational planning roles to algorithms could create large-scale ethical concerns arising from

the consequences of those actions, in part because they might be harder to predict (Horowitz, 2016).

The scale and pace of AI integration into military systems, creating what is referred to as Lethal Autonomous Weapons Systems (LAWS), has not gone unnoticed by the international community. The International Committee of the Red Cross has called for urgent action, warning in its 2021 position paper that "internationally agreed limits must be established in a timely manner" (International Committee of the Red Cross, 2021). The United Nations has convened successive meetings of governmental experts under the Convention on Certain Conventional Weapons (CCW) to address the legal status of autonomous weapons systems, while the UN Secretary-General has called for a binding international instrument prohibiting fully autonomous lethal weapons (United Nations Secretary-General, 2024a).

This urgency responds to a concrete reality where states and private industry alike have intensified their efforts to develop increasingly sophisticated systems at an exponential rate and without the necessary guardrails (United Nations General Assembly, 2023), creating a *de facto* arms race for AI superiority (Viveros Álvarez, 2024). The involvement of the private sector in this landscape is a further reason for concern: without a solid legal framework in place, private companies may shape the trajectory of military AI according to commercial interests, with little external check on how far that involvement should go. In April 2026, Palantir published a 22-point Manifesto arguing that "Silicon Valley owes a moral debt to the country that made its rise possible. The engineering elite of Silicon Valley has an affirmative obligation to participate in the defense of the nation" and warning explicitly of an unfolding AI arms race among states: "The question is not whether A.I. weapons will be built; it is who will build them and for what purpose. Our adversaries will not pause to indulge in theatrical debates about the merits of developing technologies with critical military and national security applications. They will proceed" (Palantir Technologies, 2026).

These claims sharpen the questions of accountability that this paper returns to throughout. As it is, international law will not suffice to maintain peace and safeguard future generations. This demands careful and purposeful governance frameworks to ensure that technological innovation is developed alongside ethical, legal, and operational safeguards. This urgency is reinforced by the civilian origin of most of the knowledge base underlying military capability: military consequences of civilian technological progress cannot be anticipated only after deployment, they must be anticipated in advance (Garcia, 2023a; Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025).

The aim of this paper is to map the landscape of current AI implementation in military systems, to assess the ethical and legal fault lines that implementation generates, and propose governance responses proportionate to the risks identified. Three main questions drive the paper's argument. First: how is AI currently implemented on the battlefield? Second: what ethical, legal, and strategic problems does this implementation raise? Third: what policy responses are proportionate and feasible given the current state of international law and multilateral governance?

The paper proceeds in five parts.

Section 1 provides the basic assumption necessary to evaluate the AI issue: throughout history, civilian technological progress has inexorably produced military transformation, and each transformative technology has eventually required new ethical and legal frameworks. This historicizes AI and prevents the paper from treating it as an unprecedented rupture with no instructive precedents, while still accounting for what is genuinely new about it.

Section 2 maps the operational picture: how AI is concretely implemented in contemporary military systems, from non-operational applications to the most ethically contested uses in target detection, autonomous weapons, and information operations.

Section 3 diagnoses the problems this implementation raises, arguing that these challenges are structural features of how these systems work.

Section 4 surveys the policy landscape, focusing on the UN system and the EU framework, and identifies the gaps that remain within them.

Building directly from the analysis and the policy gaps assessed throughout the paper, Section 5 closes with a set of policy recommendations and guiding principles.

# 1 Technology and the Evolution of Warfare

## 1.1 The intrinsic dual character of technology

Technology has been an important factor in military strategy throughout recorded history, and a wide set of classical cases illustrate the long-standing importance of technology to military strategy (Buzan, 1987). In this sense, the revolution in military technology should be understood not as an isolated phenomenon, but as an integrated element of a broader revolution in science, technology, and the human condition as a whole. Throughout history, as human social structures evolved, so did technology. To a significant extent, technology drives the evolution of society itself: the idea of progress is rooted in the idea of development, and major transformations in human organizations have been partially driven by new technological discoveries, from the invention of the wheel, the printing press invented by Gutenberg, the steam engine, electricity, the radio, the internet, and countless others.

However, the pattern holds consistently: the more consequential a civilian discovery, the greater its eventual impact on military technology. The intrinsic dual character of technology and the similarity of function between the military and civilian sectors means that many civil technologies will always have military applications across different domains, such as mobility, communications and intelligence (transport aircraft, trucks, computers, and telecommunications equipment). As Roland observes, a civil economy capable of manufacturing advanced steam-engines could build machine-guns quite easily; one capable of making large passenger aircraft could also make bombers; one capable of exploring space will be able to make military missiles;

and one capable of making nuclear power plants to generate electricity possesses nearly all the knowledge, material, and skill necessary to build nuclear explosives (Roland, 1995).

In several areas, developments in civil technology preceded, and laid the foundations for, later military applications.

## 1.2 A pattern across history: the nineteenth century as a turning point

For most of human history, the pace of military technology changed slowly enough that continuity was more conspicuous than transformation. Before the nineteenth century, technological breakthroughs were relatively infrequent, and victory generally depended more on the innovative use of existing weapons rather than on the invention of entirely new ones (Buzan, 1987).

By the middle of the nineteenth century, however, the Industrial Revolution had fundamentally altered the relationship between technology and strategy (Roland, 1995). Advances in manufacturing, metallurgy and engineering dramatically increased the speed and scale at which both civilian and military goods could be produced. Steam engines, railways and telegraphy were not primarily military inventions, but they transformed warfare by accelerating transport, mobilization, communication, and logistical coordination. At the same time, improvements in industrial production contributed to major advances in weapons and firepower. From this point onwards, developments originating in civilian research and industry could be rapidly incorporated into warfare, while military demand could further accelerate scientific research and industrial production (Roland, 1995).

From the nineteenth century onwards, the pattern continued and is well illustrated by the two world wars of the twentieth century. As Roland put it: “World War I was the chemist’s war; World War II the physicist’s war. Both changed the course of history inalterably” (Roland, 1995). Both conflicts mobilized the scientific expertise and full productive capacity of modern industrial economies.

The civilian chemical industry that underpinned the dye, pharmaceutical, and fertilizer sectors of Imperial Germany constituted the same industrial base that enabled the development of

chemical weapons that were first deployed by Germany during the First World War. By the end of the war, both sides had deployed a succession of increasingly lethal agents, including phosgene and mustard gas, causing 100,000 deaths and 1.2 million casualties (Ganesan et al., 2010). In World War II, this pattern was extended into physics. The knowledge of nuclear fission that Enrico Fermi and his colleagues were developing at the University of Chicago was consolidated under military leadership and transformed the Manhattan project. The atomic bombs dropped on Hiroshima and Nagasaki in August 1945 were the direct product of civilian theoretical physics converted into military capability. As the deployment of atomic weapons on Hiroshima and Nagasaki raised profound ethical questions, the post war period saw the emergence of the Geneva Conventions (1949) and the Nuclear Non-Proliferation treaty (1968), which attempted to construct legal frameworks adequate to destructive capabilities that science had placed in human hands.

## 1.3 Artificial intelligence: the pattern continues

Artificial intelligence therefore follows the same structural path as every technology examined above. As has always been the case, the knowledge base is shared; the military application follows as a matter of course. It is a product of the civilian economy, developed by commercial firms, university research departments, and technology companies, that is being progressively integrated into military systems. The same machine learning models that power recommendation engines and medical diagnostics are being adapted for target identification. The same computer vision systems that enable autonomous vehicles to navigate roads are being deployed to guide armed drones.

However, there is a significant difference between AI and the previous revolutionary technologies. For the first time in the history of weapons, the technology at issue not only increases the disruptive capacities of weapons, but selects, identifies, and engages targets on the basis of its own processing of inputs, excluding human reasoning (Garcia, 2023a).

This issue was anticipated, in a different form, over seventy years ago, at the very beginning of modern computing, by Alan Turing. Turing set aside the question of whether a machine can think, which he considered too ill-defined to be worth debating, and proposed instead that machines be judged by their observable performance (Turing, 1950). That move is what makes the present question urgent: the gap between a system that *performs* reasoning and one that *exercises* it does not close as the performance improves. A system can, in fact, identify a target with statistical precision and optimize an engagement without understanding what it means to kill or bearing moral responsibility for its consequences. It can act lethally without possessing the capacity for judgment, empathy, or moral awareness that has, until now, been assumed as the minimum condition for the legitimate use of lethal force.

The pattern established throughout history therefore continues, but with a new and urgent dimension. Unlike every previous military technology, the question at stake is not simply how destructive the technology is, or how difficult it is to control its proliferation, but whether the beings operating it can be held responsible for what it does and whether responsibility, accountability, and the fundamental moral agency can survive the transfer of lethal decision-making to machines.

Section 2 and Section 3 of this paper map the operational reality of AI in contemporary warfare, and diagnose the specific ethical and legal challenges it generates.

## 2 Implementation of AI in Warfare

AI technology is already integrated into many military systems across a multiplicity of contexts, and is widely considered to afford those deploying the technologies considerable strategic advantages, expanded offensive and defensive capabilities, and extensive hard power (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Rashid et al., 2023). It is important to set out how AI systems are currently integrated into military activities, as well as the corresponding ways in which these implementations are said to improve military capabilities along certain metrics. This is because, when AI integration into warfare

is considered greatly beneficial, advantageous, and broadly indispensable along predominant metrics, the space for open discussion and meaningful questioning of these applications of AI becomes diminished. Further, by clearly laying out the current picture of how AI is integrated into military processes, a more granular analysis of the ethical implications of this integration becomes possible.

## 2.1 How AI is implemented in warfare: the current picture

Operational activities are those that are tied to the execution of military missions, and contribute to combat operations, intelligence, surveillance, reconnaissance, and active deployments. When AI systems are implemented into operational military activities, the technologies work to directly streamline, quicken, or alter the ways in which military missions function. By contrast, non-operational activities do not contribute to active missions, but rather work towards planning, maintaining, or sustaining military force. For example, non-operational activities could be involved in administration, training, recruiting, and military education. It is important to highlight this distinction between operational and non-operational military activities because whether an AI system is deployed operationally or non-operationally will alter the extremity and nature of the ethical implications of this application.

### 2.1.1 Non-operational implementations of AI

In non-operational military activities, AI systems are used to streamline logistical and organizational tasks, and to train military actors (Rashid et al., 2023). Administrative tasks, such as organizing military equipment, documents, and resources, can be supported by AI systems which can improve the speed, efficiency, and accuracy of decision-making. Likewise, logistical applications like scheduling, managing inventories, and autonomous supply chain management allow militaries to forecast and optimize supply and demand (Rashid et al., 2023). In acquisition and procurement, AI is used for contract generation, auditing, and compliance oversight, with the aim of reducing human error and quickening processes of military compliance with regulatory and legal requirements (Global Commission on Responsible Artificial Intelligence

in the Military Domain, 2025). Finally, simulation via world modeling and augmented or virtual reality, as well as the development of personalized training plans, depend on AI systems in order to allow military personnel to effectively rehearse high risk scenarios without real-world consequences (Doicariu, 2025; Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Rashid et al., 2023).

### 2.1.2 Operational implementations of AI

In operational settings, AI systems are used to increase the breadth, speed, and accuracy of military operations, strengthen both offensive and defensive capabilities, and to prevent troops from entering direct danger zones, and from performing dangerous or repetitive tasks (Macfarlane & Christie, 2022; Rashid et al., 2023).

Table 1: How AI is integrated into military activities.

| Military activity | Components | AI techniques present | Example |
|---|---|---|---|
| **Command, control and communications** | Decision-support, communications analysis, scheduling, planning (Mustafovski, 2025). Management and coordination of equipment, logistics, resource management, organization (Rashid et al., 2023). | Machine learning and deep learning for data analysis, planning and reasoning. Large Language Models for decision support and planning. Predictive AI and machine learning for resource allocation and supply chain management. | Palantir Technologies' *Maven Smart System* is one of the Pentagon's key AI platforms supporting operational and logistical activities and aiming to speed up decision-making in military contexts, by processing, analyzing and sharing imagery and sensor data (Fuller, 2026). |
| **Surveillance, intelligence and reconnaissance** | Real-time situational awareness, information tracking and collection, observation (Kumar, 2026). | Multisensor fusion to build a complete operational picture combined from multiple data sources. Computer vision and facial recognition for identifying objects and individuals. Machine learning for predicting enemy activity. | Israel's ground invasion of Gaza has enabled the expansion of biometric surveillance of Palestinians using facial recognition technology from *Corsight* and *Google Photos* to collect and catalog faces without consent (Frenkel, 2024). |
| **Cybersecurity** | Autonomous internal and external threat and anomaly detection and identification, cyberthreat triage and monitoring (Maathuis & Cools, 2025a). Real-time incident response such as isolating | Machine learning, deep learning and behavioral analytics for identifying threats, abnormal activity, and expected behavior. Agentic AI for creating attack profiles and isolating compromised | Darktrace's AI-powered cybersecurity breach detection and response system has been employed by the US Department of State (Darktrace Federal, 2025). |

| Military activity | Components | AI techniques present | Example |
|---|---|---|---|
| | devices or withdrawing access (Chauhan, 2025). | systems. Natural Language Processing for analyzing threat intelligence. | |
| **Target detection, identification and selection** | Recognition systems, precision guidance, sensing, and navigation (Kumar, 2026). Predictive analysis of targets' behavior (University of Amsterdam, 2026). Data analysis. | Computer vision for target identification and classification. Deep learning for obscured target identification. Generative AI for decision-support. | *The Gospel* is an AI system used by the Israeli military which processes surveillance data to generate lists of targets and categorize them as military or civilian objects (Human Rights Watch, 2024). |
| **Autonomous weapons systems** | Identification and engagement of targets independent of human input (Horowitz, 2016). Increased precision, reduced vulnerabilities of operators (Garcia, 2016; Macfarlane & Christie, 2022). Battlefield surveillance and reconnaissance (Macfarlane & Christie, 2022). Drone weaponization to deploy grenades and small bombs (Financial Times Editorial Board, 2026; Pultarova, 2026). Drone swarming and coordination of autonomous systems (Rashid et al., 2023). | Reinforcement learning for independent decision making. Sensor fusion for target detection. Deep learning for flight navigation. Computer vision for direction following and obstacle avoidance. | Throughout the 2026 Iran war, weapons systems produced by Anduril Industries, a US-based defense-tech company, were deployed to combat Iranian *Shahed* drones (Loh, 2026). |
| **Unmanned Ground Vehicles** | Surveillance, offensive attacks (Drabkina, 2024; Kryzhanivska, 2026). | Computer vision for target identification and navigation. Deep learning and reinforcement learning for navigation and assessment. | In some Ukrainian combat areas, Unmanned Ground Vehicles like *Droid TW-7.62* are used to expand kill zones, as well as to perform the majority of logistical tasks such as ammunition resupply and equipment delivery (Kryzhanivska, 2026). |

## 2.2 AI's military potential

Some believe that AI technologies represent great opportunities in the military domain, and advocate for these potential benefits being considered alongside the associated harms and risks. Decisions regarding whether AI systems should be integrated into military organization,

processes and operations always depend on tradeoffs between potential benefits on the one hand, and ethical, legal, and social risks and harms on the other. It is important to note that often, those advocating for the benefits of AI integration into warfare have an economic, political, or social incentive to promote such views and alter the perceived balance of the tradeoffs involved in this decision making.

### 2.2.1 Reduced human deployment and exposure

AI technologies allow for dangerous tasks, such as aid and equipment deliveries in conflict zones, to be completed without direct human intervention, reducing the risk to the lives of human personnel (Chengeta, 2016; European Commission, n.d.).

### 2.2.2 Optimized logistics, resource management, and assessment

Effective logistical, administrative, and organizational processes are key to military activity, and the optimization of these activities allows for faster and better-informed decision-making, as well as the minimization of costly risks regarding the transportation and assessment of military equipment, supplies, and personnel (Trofymenko et al., 2024).

### 2.2.3 Real-time situational awareness and intelligence analysis

AI systems increase the effectiveness, reliability, and granularity of real-time situational monitoring and reconnaissance of complex battlefield domains, allowing decision-makers and strategists to access the most accurate, up-to-date information and minimize the likelihood of civilian casualties (European Commission, n.d.; Lee et al., 2023).

### 2.2.4 Increased accuracy, precision and speed in cyber and physical threat detection

AI systems, especially deep learning and machine learning techniques, enable military personnel to notice early-stage cybersecurity breaches, as well as battlefield threats and targets that they would otherwise miss (Dhanushkodi & Thejas, 2024; European Commission, n.d.).

# 3 Ethical Risks with the Implementation of AI in Warfare

The importance of ethically examining the expanding integration of AI technologies into the military domain cannot be overstated. The metric often used to justify using AI in warfare is effectiveness, or how much better AI can allow military operations to occur, but beyond evaluating the extent to which AI systems can optimize the functioning of military processes, it is absolutely critical to examine the ethical risks and moral harm which arise from integrating these new technologies into warfare. It is imperative to also leave space to seriously question whether the use of AI in warfare raises ethical considerations which warrant its prohibition in some cases (Horowitz, 2016). New technologies are expanding the capabilities of military forces to an incomprehensible degree, which can serve to perpetuate and worsen current conflicts and humanitarian crises, concentrate unimaginable power within the hands of very few, and facilitate more deliberate and targeted violence and military power than has been possible in the past. Section 3 of this report urges readers to wholly consider not only the risks that arise when AI systems malfunction or are misused, but also the broader ethical implications of the loss of human agency, the unexplainable and unaccountable nature of AI systems, and how an AI arms race allows for power concentration and the incentivization of extending conflict and violence.

## 3.1 Risks from malfunction and misuse

AI systems can produce unpredictable, fabricated, or erroneous outputs, the consequences of which can be extreme and devastating in the military domain.

Firstly, autonomous systems may misidentify civilians, aid workers, journalists, or surrendering combatants, as well as hospitals or schools, as legitimate targets, and act faster than human oversight can intervene, resulting in unintended killing (Podar & Colijn, 2025). If AI technologies operate at scale, automated decisions could also lead to devastating escalation spirals in which defensive systems respond before humans can intervene, multiplying the atrocities committed (Podar & Colijn, 2025). Beyond grave consequences of malfunction such as friendly fire incidents and the misidentification of targets, systems may also commit non-operational

errors such as ineffective and inefficient resource allocation, logistical and communication mistakes, or incorrect data analysis (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025).

More broadly, the data from which AI systems are trained may manifest statistical imbalances resulting in hidden systemic biases within deployed models (Shoker, 2019). ML models infer patterns from training data and aim to reproduce these correlations, as opposed to working around causal structures, meaning that data biases are likely to be reproduced (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Shoker, 2019). In practice, this results in target identification and classification, as well as threat assessment and prioritization, being based on pre-programmed categories and correlations which under- or over-represent certain groups. This may not only perpetuate social or operational prejudices, but also cause decisions made regarding those marginalized groups to have higher error rates (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025). Further, autonomous weapons systems have been shown to have a 'war bias', unless they are tempered by sufficient and transparent oversight, meaning that lethality may be prioritized over caution in ambiguous situations (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025).

Even if there are no malfunctions or biases within an AI system, the model is vulnerable to malicious interference from bad actors via data poisoning, model inversion and integrated backdoors (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Podar & Colijn, 2025). These vulnerabilities can not only compromise the integrity of AI systems and lead to irreversible and lethal consequences, but also call into question the justification for implementing AI systems in the first place, because a backdoored or compromised AI system is less reliable and precise than a human (Podar & Colijn, 2025).

Another overarching ethical concern about the use of AI in warfare arises from the necessity of AI systems to transgress widely obeyed moral rules in warfare and abide by a moral code

which conflicts with that of a regular moral agent (Hartridge & Walker-Munro, 2025; Simmons, 2026). Rules like the prohibition on killing, which are near-universal in ordinary contexts, are suspended or inverted in combat. AI Safety research typically pursues value alignment with broadly shared human values, such as honesty, fairness, and harmlessness (Simmons, 2026). However, AI systems deployed in warfare are required to transgress these values, meaning that their character, priorities and programming must differ radically from AI systems deployed in civilian domains (Hartridge & Walker-Munro, 2025). It is crucial to remember that the question of which values military AI models should be aligned to is being answered by human actors, who are oftentimes unelected, unaccountable, and incentivized to continue conflicts (Simmons, 2026).

Alarm has been raised not only about how AI technologies are currently being deployed on the frontlines, but also about how future AI technologies could be used in military settings to devastating ends, for instance, humanoid soldiers or totally autonomous swarms of armed drones (Campbell, 2026; Pultarova, 2026). Furthermore, current laws and certifications are unequipped to deal with new technologies and their accompanying ambiguities (Cummings, 2019; Sourav et al., 2025).

## 3.2 Loss of human agency

Reliance on AI systems to streamline, assist, or replace human actors in the military domain can erode human agency across multiple dimensions.

Firstly, when AI systems identify, process and organize human subjects, this results in digital dehumanization whereby human agents are rendered through data and decisions about human life are taken by entities which cannot comprehend the extremity and importance of these decisions, nor the humanity of the data points they are analyzing (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025). Individuals are reduced to behavioral patterns and data, and are denied the moral consideration required by the right to dignity, as well as the recognition of themselves as a person by the entity considering the

decision over whether they live or die (Kahn, 2022). Decisions about human life being made by autonomous systems represent a loss of control in that these decisions are made oftentimes without human judgement or human empathy, causing victims to be denied agency (Asaro, 2012).

This is especially worrying when we consider that humans are vulnerable to automation bias, as well as cognitive bias, meaning that, even when an AI system is only meant to assist human decision-making and not completely automate processes, technologies have a disproportionate impact over the outcomes human agents choose, because human agents are inclined to follow AI advice without critical engagement (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Ul Haq, 2026). The fantasy of human-in-the-loop decision-making, whereby AI systems act as thinking partners while human agents retain ultimate control over decisions, has been widely contested, because humans find it difficult to go against a suggestion made by an automated system (Agudo et al., 2024; Lindsay, 2026). This means that when AI is incorporated into military decision-making, strategizing, and target identification, human agency, autonomy, and control is limited (Asaro, 2012).

More broadly, AI can be used to change the informational environments of citizens, producing propagandistic content to perpetuate certain agendas and alter the opinions of the masses (Garon, 2022). As seen recently with Iran's use of AI-generated content intended to further anti-American narratives among citizens and to discredit and humiliate Donald Trump, narratives in wartime are a weapon which can be wielded to great effect themselves (Thompson & Cardia, 2026). However, altering the epistemic landscape of the people limits their autonomy and agency, because it stunts citizens' abilities to critically contemplate their countries' participation in conflicts and exercise their freedom of thought and speech (Garon, 2022).

Furthermore, if we allow AI systems to make decisions regarding human life, will we forget how to do so ourselves? Outsourcing these decisions about who lives and who dies means that military forces have less practice using the muscle of their morality (Hassen, 2025). Some claim

that AI can aid the military in gaming simulated scenarios to discern weapon effectiveness and minimize casualties, but the distance these technologies create between military actors and the victims of their attacks is concerning because it may result in a lack of empathy and understanding of these conflicts (Pandise et al., 2026). Perhaps it is easy to consider real-life conflicts to be merely a 'war-game' when one is allowing an autonomous system to make decisions regarding life and death, instead of having to confront the moral reality of warfare oneself (Asaro, 2012). The use of technology in warfare limits not only the agency of the victims of attacks, as they are denied the right to dignity and human consideration, but also the agency of troops to fully consider the moral extent of their actions and decision-making (Hassen, 2025; Kahn, 2022).

## 3.3 Accountability and explainability

AI itself is not an entity which can be regarded as an ethical agent. This raises questions regarding whether autonomous weapons systems can be deployed in ways which can guarantee accountability and responsibility (Horowitz, 2016; Manna & Nath, 2021). However, the picture is complex because if systems are automated, the chain of responsibility is unclear, and, in many cases, a crime committed cannot be traced neatly back to a single perpetrator because agents are in control of systems which act autonomously and oftentimes opaquely (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Ul Haq, 2026; Wood, 2023). Crucially, a lack of accountability leaves victims without explanations or justice (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025).

Autonomous weapons systems and other AI military technologies seem to diffuse accountability across developers, commanders, and the machines themselves by creating situations where there is little clarity regarding which actor is morally responsible when something goes wrong (Wood, 2023). Policymakers devise governance frameworks and strategic decisions; commanders create mission priorities, system capabilities, and deployment practices; scientists and developers design and deploy systems and should ensure robustness throughout technologies' lifecycles; operators employ systems to achieve military goals; legal advisors ensure

compliance (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025). The human influence over autonomous systems and AI models deployed to military domains is disparate, and difficult to track (Cummings, 2019; Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025). As seen with the coverage of a US strike on the Iranian primary school in March 2026, even when there is no AI error, the media and the public pounce on any evidence that an event can be characterized as an 'AI problem' (Baker, 2026). This has the effect of obscuring who may be truly responsible for tragic and deplorable occurrences: human agents who decided to compress the kill chain, human agents who decided to build and deploy these systems, and human agents who start and perpetuate these conflicts (Santoni de Sio & Mecacci, 2021; Wood, 2023).

AI models themselves are not responsible moral agents, yet, oftentimes in warfare, they are the entity which directly causes death and destruction. This leaves a responsibility gap whereby human actors who did not directly cause a particular harm can be held accountable for it (Manna & Nath, 2021; Santoni de Sio & Mecacci, 2021). Furthermore, there is some pressure to take humans out of the loop entirely in some instances of military decision-making, for instance when the enemy has scrambled communication, making it impossible to keep direct human control of a situation, which would complicate the question of responsibility even further (Financial Times Editorial Board, 2026).

On the other hand, human actors may be unjustly held responsible when AI does make errors (Elish, 2019). As already established, human-in-the-loop oversight of autonomous decision-making is broadly ineffective because humans tend to over-afford trust to AI systems, and are unprepared when these autonomous systems fail, often having to take action in only the most extreme and challenging circumstances (Agudo et al., 2024; Lindsay, 2026). It has been argued that automated prediction and decision-making needs human judgement, yet, in many circumstances, autonomous systems work far faster than human agents can comprehend, meaning that the control humans are able to exercise over these technologies is limited (Agudo et al., 2024; Elish, 2019).

This lack of accountability is partially due to a lack of explainability: AI systems deployed in conflict are ‘black boxes’, meaning that how they reach decisions is often opaque and unknown (John-Mathews, 2022; Podar & Colijn, 2025). Many AI systems are inherently uninterpretable, because, by default, they are unexplainable and must be deliberately made understandable for humans. This means that in many cases, people are unable to see how AI systems turn inputs into outputs (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Ul Haq, 2026). This may be deemed permissible when tradeoffs are made and AI capabilities, efficiency and accuracy far exceed that of humans, causing reward to outweigh risk (John-Mathews, 2022). However, in wartime scenarios, this is not sufficient, and clear causal reasoning, transparency regarding how and why decisions are made, and explainable input-output chains are required to interrogate military processes which affect human life.

## 3.4 Palatable violence and the arms race

AI’s deployment in the military domain has grave effects on international peace and security in that these technologies may make violence more palatable to troops themselves, as well as citizens in distant yet participatory nations (Renic, 2026). Furthermore, the expanding role of AI systems in warfare is greatly influenced by the economic incentives of companies producing these technologies causing them to stoke the flames of the AI arms race (Akıllı, 2025; Ishkhanyan, 2026).

Technology in warfare allows a distance to be created between violence and perpetrator, for instance, the operators of AI-assisted weapons systems exercise lethal force from miles away, their experience of controlling drones or autonomous systems mediated by a screen. This increased detachment of human operators from frontline violence protects troops from not only danger, but also the psychological impact of directly causing harm to another human (Renic, 2026). However, the pervasive use of AI systems in warfare often serves to shift the burden of harm from troops onto citizens, as warzones expand to encompass civilian spaces (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025). If warfare

continues to become detached from direct personnel intervention, and machines increasingly replace human judgement, violence risks becoming ever more detached and palatable (Erskine & Miller, 2024).

Furthermore, replacing people on the frontlines with machines may have the effect of lowering the threshold of war (Viveros Álvarez, 2024). This is firstly because human soldiers experience fear, reluctance and injury, and commanders must face political and personal consequences for their decision-making, meaning that when humans are taken out of warzones, these natural hesitations are removed, making it easier for states to initiate and sustain conflicts (Erskine & Miller, 2024). Secondly, when troops are put in direct danger less and less, replaced by autonomous weapons systems, drones, and unmanned vehicles, conflicts will be less recognized by military casualties and the popularity of wars at home will be bolstered (Viveros Álvarez, 2024).

Despite the fact that AI systems cause fewer soldiers to be put in direct danger, humanitarian crises still occur, and civilians in warzones still suffer gravely due to war. Furthermore, many automated systems are examples of 'fauxtomation': technologies which appear to function automatically and without human input, but are actually operated with the help of human counterparts (Ishkhanyan, 2026). These technologies are perceived as objective, neutral and unfathomably accurate due to their advanced capabilities and technological complexity, but are fallible and still create psychological harm to those operating them (Markelius et al., 2024).

Another effect on international peace and security posed by AI technologies is that the development and deployment of these technologies occurs within an unstable arms race: weapons and techno-military equipment manufacturers benefit from and perpetuate rhetorics which insist that geopolitical tensions are escalating and that countries must be equipped with the latest technologies (Rossiter, 2023). These narratives cause states to become dependent on weapons manufacturers and AI companies, spurred on by fears that opponents will be equipped with more powerful technological advancements and weapons (Akıllı, 2025; Rossiter, 2023). In

many cases, AI technologies originally designed for use in civilian and commercial domains are designed to be dual-use, and can therefore be used by a multiplicity of actors to many different ends, allowing manufacturers to profit in both commercial and military domains (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025). Because AI systems are inherently general purpose, they also can be taken, adapted, and deployed by non-state actors to commit illicit activities, further motivating states to secure the most advanced military technologies to defend themselves and achieve their aims (Maathuis & Cools, 2025b; Viveros Álvarez, 2024).

The reliance on private technology and weapons companies represents an erosion of state sovereignty, as well as an expansion of vast asymmetries between different states' techno-military power, and also perpetuates AI hype by feeding narratives that countries must have the latest technological innovations in order to avoid being left behind (Maathuis & Cools, 2025b; Markelius et al., 2024; Rossiter, 2023). Dependencies on informal connections and imperial trading paths may result in an erosion of trust between international bodies, while empowering private profit-seeking firms, who become economically incentivized to perpetuate and bolster conflicts (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Rossiter, 2023). When an arms race is frenzied and competitive, there is also a danger that hastily developed technologies will be deployed before they are fully safe or aligned with human values (Garcia, 2016; Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025; Ishkhanyan, 2026).

# 4 Policy Landscape

As these new challenges emerge, progress on AI governance has become increasingly urgent. Meeting this challenge requires coordinated action at the multilateral level, and International Institutions have the responsibility to safeguard fundamental human rights and the rule of law, and to ensure that technological development does not outpace the existing legal framework.

Over the past decade, the international community has significantly intensified its efforts to construct a governance framework for AI in warfare. Discussions have proceeded at multiple levels simultaneously: within international organizations, through specialized intergovernmental bodies, and through the emergence of dedicated research and policy institutions tasked with mapping the risks and developing normative responses. These organizations face unique challenges and priorities in shaping the regulatory landscape for AI. For instance, they must balance the need for innovation and economic growth with the need to protect citizens from potential harm (European Defence Agency, 2025).

This part of the report provides an overview of the multilateral governance architecture that has emerged in response to these challenges. It examines first the United Nations level and then the European Union level, presenting the institutional actors involved and the governance frameworks developed within that system. National-level frameworks, while increasingly relevant, fall outside the scope of this analysis.

## 4.1 The UN framework for autonomous weapons and broader military AI

In 2024, the UN Secretary General António Guterres urged the Security Council to act decisively to establish international guardrails for Artificial Intelligence, warning that "Every moment of delay in establishing international guardrails increases the risk for us all". "No country should design, develop, deploy or use military applications of AI in armed conflict that violate international humanitarian and human rights laws." "Recent conflicts have become testing grounds for AI military applications," he said, citing AI use in autonomous surveillance, predictive policing, and even reported life-and-death decisions (United Nations Secretary-General, 2024b).

Numerous efforts to establish rules to limit the use of AI systems in warfare have already been made and the issue sits at the top of the international agenda.

The main point of reference for discussions on this topic is the Convention on certain Conventional Weapons (CCW). Established in 1980 in Geneva and supported by the United Nations

Office of Disarmament Affairs (UNODA), the leading department responsible for governing and preventing the misuse of AI in warfare, the CCW currently gathers 128 High Contracting Parties, alongside non-High Contracting Party, international organizations, third sector groups, and academia, whose active participation has been instrumental to further substantive understandings of the challenges posed by lethal autonomous weapons systems. In the Fifth Review Conference that took place in 2016, the CCW established the Group of Governmental Experts on Emerging Technologies in the area of Lethal Autonomous Weapons Systems (often referred to as the "GGE on LAWS") with the mandate "to explore and agree on possible recommendations on options related to emerging technologies in the area of LAWS, in the context of the objectives and purposes of the Convention, taking into account all proposals - past, present and future" (Convention on Certain Conventional Weapons, Fifth Review Conference of the High Contracting Parties, 2016; United Nations Office for Disarmament Affairs, 2018).

In line with this decision, the Group has met annually since 2017.

In November 2019, the CCW High Contracting Parties endorsed the following Guiding Principles for LAWS:

a. International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems;
b. Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system;
c. Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including

the operational context, and the characteristics and capabilities of the weapons system as a whole;

d. Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control;

e. In accordance with States' obligations under international law, in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law;

f. When developing or acquiring new weapons systems based on emerging technologies in the area of lethal autonomous weapons systems, physical security, appropriate non-physical safeguards (including cyber-security against hacking or data spoofing), the risk of acquisition by terrorist groups, and the risk of proliferation should be considered;

g. Risk assessments and mitigation measures should be part of the design, development, testing, and deployment cycle of emerging technologies in any weapons systems;

h. Consideration should be given to the use of emerging technologies in the area of lethal autonomous weapons systems in upholding compliance with IHL and other applicable international legal obligations;

i. In crafting potential policy measures, emerging technologies in the area of lethal autonomous weapons systems should not be anthropomorphized;

j. Discussions and any potential policy measures taken within the context of the CCW should not hamper progress in or access to peaceful uses of intelligent autonomous technologies;

k. The CCW offers an appropriate framework for dealing with the issue of emerging technologies in the area of lethal autonomous weapons systems within the context of the objectives and purposes of the Convention, which seeks to strike a balance between military necessity

and humanitarian considerations (Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 2019).

Alongside the work of the GGE for LAWS, numerous resolutions have been adopted by the UN General Assembly. Concerns over killer robots arose at the 75th UN Assembly in October 2020, and in 2022 70 countries endorsed a joint statement on autonomous weapons systems "The introduction of new technological applications, such as those related to autonomy in weapon systems, also raise serious concerns from humanitarian, legal, security, technological and ethical perspectives. We therefore see an urgent need for the international community to further their understanding and address these risks and challenges by adopting appropriate rules and measures, such as principles, good practices, limitations and constraints" (Stop Killer Robots, 2022).

At the 78th UN General Assembly First Committee in 2023, 164 states voted in favor of resolution L.56 on autonomous weapons systems (Stop Killer Robots, 2023). The Resolution stressed the "Urgent need for the international community to address the challenges and concerns raised by autonomous weapons systems" (United Nations General Assembly, First Committee, 2023).

Following the Resolution, a report was presented in July 2024, consolidating views submitted by Member States, observer states, international organizations, civil society, the scientific community, and industry on the humanitarian, legal, security, technological, and ethical challenges posed by autonomous weapons systems. The report reflects widespread concern that autonomous weapons systems have the potential to significantly change warfare and may outpace existing legal frameworks. The report reiterates the Secretary-General's call for states to conclude, by 2026, a legally binding instrument to prohibit systems that function without human control: "There is a strong sense that time is running out for the international community to take preventive action on this issue. I therefore reiterate my call for the conclusion, by 2026, of a legally binding instrument to prohibit lethal autonomous weapons systems that function without

human control or oversight and that cannot be used in compliance with international humanitarian law, and to regulate all other types of autonomous weapons systems. The autonomous targeting of humans by machines is a moral line that must not be crossed" (United Nations Secretary-General, 2024a).

Most recently, in December 2025, the 80th UN General Assembly adopted Resolution 80/57 on Lethal Autonomous Weapons Systems, with 164 states voting in favor, 6 against, and 7 abstentions. The resolution "Underlines that a comprehensive and inclusive multilateral approach is required to address the full range of challenges and concerns posed by autonomous weapons systems" and "Calls upon the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects to work towards completing the set of elements for an instrument being developed within the mandate of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, with a view to future negotiations" (United Nations General Assembly, 2025).

While these resolutions reflect the growing interest of the international community to move towards a binding and enforceable framework, structural limits of the multilateral process still remain. A large and growing majority of states recognize that the current pace of technological deployment demands a legal response that goes beyond the existing body of international humanitarian law, but major military powers, most notably the United States and Russia, continue to maintain that existing international law largely suffices, and have consistently resisted the negotiation of new binding rules (Human Rights Watch, 2020; Lewis, 2020).

This impasse is generating visible frustration among the growing coalition of states that consider the current process inadequate to the scale of the challenge. The tension between the urgency expressed in resolutions and the inertia produced by the resistance of the major military powers is the major obstacle to the creation of an effective UN governance landscape on LAWS (Lewis, 2020).

## 4.2 The EU's fragmented governance of military AI

In Brussels, the debate on LAWS largely reflects the discussion taking place at the United Nations, yet the two are not on the same wavelength. Many institutions in the European Union now play a significant role in the development of AI policies and practices, yet a meaningful gap in policies specifically addressing LAWS remains.

The European Parliament has long been the most proactive EU institution on this issue, insisting that the EU should push for the adoption, at the international level, of a legal instrument to ban LAWS. As early as September 2018, not long after the creation of the GGE on LAWS, the Parliament adopted the Resolution 2018/2752(RSP) on autonomous weapon systems where:

"Stresses, in this light, the fundamental importance of preventing the development and production of any lethal autonomous weapon system lacking human control in critical functions such as target selection and engagement" and calls for the "(…) development and adoption of a common position on lethal autonomous weapon systems, for an international ban on the development, production, and use of lethal autonomous weapon systems enabling strikes to be carried out without meaningful human control, and for a start to effective negotiations for their prohibition" (European Parliament, 2018).

In 2021, the Parliament reiterated its call, underlining that "it is essential to provide an appropriate and comprehensive legal framework for the ethical aspects of these technologies as well as for liability, transparency and accountability (in particular for AI, robotics and related technologies considered to be high risk); whereas this framework must reflect that the intrinsically European and universal humanist values are applicable to the entire value chain in the development, implementation and uses of IA; whereas this ethical framework must apply to the development (including research and innovation), deployment and use of IA, in full respect of Union law and the values set out in the Charter of Fundamental Rights of the European Union" (European Parliament, 2021).

At the international level, the EU has articulated a coherent and consistent institutional position on LAWS, centered on the principle of "meaningful human control" (European External Action Service, 2025). In successive statements delivered to the CCW GGE and its submission to the United Nations Secretary-General, the EU has made its position clear: "The EU recalls that States bear a fundamental responsibility to ensure that the development, production, deployment and use of emerging technologies in the area of LAWS must be in compliance with international law, in particular IHL" (European Union, 2024).

Central to this stance is the EU emphasis that "Human beings must make the decisions with regard to the use of lethal force, exert control over lethal weapons systems that they use and remain accountable for decisions over the use of force in order to ensure compliance with International Law, in particular IHL, also taking into account ethical considerations. Those who plan, decide upon and carry out an attack using a lethal autonomous weapon system must, therefore, ensure that the weapon system and the way it is used will preserve human beings' ability to make these necessary legal judgments, and thereby ensure compliance with IHL. Furthermore, human accountability must be preserved at all times and across the entire life cycle of the weapons system and appropriate measures in this regard should be implemented" (European Union, 2024).

In the same submission, the EU "Encourages further engagement in international dialogue and cooperation to address the opportunities and challenges presented by AI in the military domain. We support efforts at relevant fora to exchange best practices and foster a common understanding of the legal and ethical implications, and we welcome the valuable contributions of recent international and regional initiatives" (European Union, 2024).

This common baseline found practical expression in the joint statement delivered at the September 2025 session of the GGE, where France, Germany, Italy and Belgium and 35 other High Contracting Parties, declared that "while our delegations are prepared to keep working on the rolling text in order to further improve it, the rolling text contains a set of elements that

we consider as sufficient basis for negotiations of an instrument on lethal autonomous weapon systems. We are fully prepared to continue working with delegations in order to further refine and improve the text, with the fulfilment of our mandate as our constant guiding objective". The signatories also declared themselves ready to move ahead towards negotiations within the CCW on the basis of the rolling text (High Contracting Parties to the Convention on Certain Conventional Weapons, 2025). The EU also emphasizes that LAWS should be governed by international humanitarian law and human rights law, and has called upon the High Representative for the Common Foreign and Security Policy to cooperate with the UN and other international organizations in identifying ways to responsibly govern autonomous weapons systems.

Although these declared positions reflect the EU's willingness and support to build a strong international regulatory framework on LAWS, the EU has not, to date, implemented policies that place heavy emphasis on the development of autonomous weapons systems. As defense and military matters remain the exclusive competence of the Council and the Member States, any possible initiative at the EU level in this area could only come from them and would require unanimity (Mako, 2026; Santopinto, 2024). This structural constraint means that, however clear the Parliament's position or the EU's multilateral statements, the translation of those positions into internal regulation depends on the political will of the Member States acting as one.

Even where specific regulation on LAWS remains absent, some EU instruments carry indirect implications for the field. The most significant is the EU AI Act (European Parliament and Council of the European Union, 2024), the most comprehensive AI regulatory framework to date, which entered into force in August 2024. The Act prohibits and sets requirements for AI systems put on the EU market, but explicitly excludes from its scope AI systems used for military purposes, on the basis that such systems are already subject to public international law (Rönnback, 2023).

Despite this exclusion, the AI Act may nonetheless produce certain spillover effects on the military domain and could lead to certain alignment of relevant standards (Rönnback, 2023).

Three mechanisms are worth noting:

1. The classification of AI applications according to their level of risk: while AI Act's prohibitions or the high-risk classification is unlikely to be directly applicable to military AI systems, AI Act's general tiered structure might be relevant as a regulatory approach. Certain military AI applications could be prohibited outright, such as LAWS, others would be subject to accountability and trustworthiness requirements.
2. Relevant Requirements: The AI Act's requirements of risk management, documentation, transparency, auditability, robustness, and cybersecurity are directly relevant to high-stakes AI in any context.
3. Human Centric Approach: The AI Act's requirements are formulated based on the principles of respect for fundamental rights. Therefore, the requirements of AI systems in military applications will need to account for IHL, in particular in times of conflict, ensure that systems cannot be used to target civilians or persons outside of combat, and in many other ways help save lives (Rönnback, 2023).

In parallel, the European Defence Agency (EDA), established in 2004 as the intergovernmental body supporting Member States' defense capability development, has emerged as the primary EU actor shaping the practical governance of autonomous systems. The EDA has been actively supporting EU Member States in enhancing their autonomous systems capabilities over the past few years. This support spans multiple levels, including:

- **Prioritization**: Through the Capability Development Plan (CDP), Overarching Strategic Research Agenda (OSRA), Coordinated Annual Review on Defence (CARD), and Key Strategic Activities (KSA), EDA has prioritized AS-related capabilities and technologies.
- **Projects:** EDA has facilitated dozens of studies and projects focused on systems, subsystems, and component levels, advancing the development of autonomous systems.

In 2024 the EDA published its Action Plan for Autonomous Systems (APAS) with the aim to support Member States to speed up the development of highly autonomous, effective, effi-

cient, and reliable autonomous systems for Land, Air, Maritime, and cross-domain operations. Additionally, it promotes more streamlined and rapid identification of relevant technologies, facilitating their development and transition into military capabilities (European Defence Agency, n.d.; 2024).

Crucially, EDA stresses that all the autonomous systems projects operate under the principle of the "human-in-the-loop", meaning that a human operator must be able to intervene at any time. This applies across the EDA's current project portfolio: the Combat Unmanned Ground Systems (CUGS) project, launched in 2023, aims to develop highly autonomous combat ground systems while explicitly maintaining the human-in-the-loop; and the EDA's Operational Experimentation (OPEX) campaign, conducted in Rome in mid-2025, tested unmanned aerial and ground systems from six European companies under field-like conditions, with human oversight requirements embedded throughout (European Defence Agency, n.d.; Mako, 2026).

The most explicit expression of binding EU-level governance on LAWS, however, comes from the European Defence Fund Regulation. Article 10(6) of Regulation (EU) 2021/697 (European Parliament and Council of the European Union, 2021) establishes that:

"(…) 6. Actions for the development of lethal autonomous weapons without the possibility for meaningful human control over selection and engagement decisions when carrying out strikes against humans shall not be eligible for support from the Fund, without prejudice to the possibility of providing funding for actions for the development of early warning systems and countermeasures for defensive purposes."

Its scope is, however, limited to EDF-funded projects, meaning that an identical system procured by a Member State without EDF financing falls outside its reach and this constitutes a significant gap.

The picture that emerges from this framework is one of genuine institutional ambition constrained by structural limits. The AI Act's military exclusion leaves the most ethically conse-

quential domain of AI development largely unregulated at the Union level, and the unanimity requirement in defense matters makes binding legislative progress difficult.

## 4.3 Other relevant actors

Beyond the formal institutional architecture described above, the governance debate on LAWS is being actively shaped by a broader ecosystem of actors whose influence has proven instrumental in pushing the discussion towards binding outcomes. The most significant of these is the Summit on Responsible AI in the Military Domain (REAIM), which was held for the first time in February 2023. The Summit brought together sixty countries and produced a non-binding Call to Action endorsed by 57 states, alongside the US-led Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, which has since gathered 58 endorsing states (Government of the Netherlands, 2023). The second summit, held in Seoul in September 2024, culminated in a Blueprint for Action endorsed by 61 states, laying out a roadmap of principles and concrete steps for responsible military AI governance (Ministry of Foreign Affairs, Republic of Korea, 2024). The third REAIM Summit took place in A Coruña, Spain in February 2026, with the stated aim of moving from declarative principles towards practical implementation and enforcement mechanisms (Ministerio de Asuntos Exteriores, Unión Europea y Cooperación, 2026). While the REAIM summits have not produced binding rules, they have created a space for dialogue between governments, industry, academia, and civil society.

Closely linked to the REAIM process is the Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM), an initiative of the Government of the Netherlands launched at the 2023 REAIM Summit, with the Hague Centre for Strategic Studies acting as its Secretariat. The Commission published its landmark report, *Responsible by Design*, in September 2025, seeking to translate the REAIM Summit declarations into actionable guidance and advocating a layered approach to global governance of AI in the military domain. The report, which has been referenced throughout this paper, represents the most comprehensive

attempt to date to operationalize the principle of responsibility by design across the full AI system lifecycle, with recommendations addressed to states, militaries, and industry alike.

From the third sector, the Stop Killer Robots campaign, a coalition of over 300 member organizations active since 2013, has been a consistent and vocal presence across all multilateral forums, calling for legally binding rules rather than unenforceable guidelines and working toward a treaty to ban autonomous weapons (Stop Killer Robots, 2013). The International Committee of the Red Cross has similarly played a key role, with its 2021 position paper on autonomous weapons systems providing the clearest articulation of the humanitarian case for binding rules and serving as a reference point for state delegations, UN reports, and academic analysis alike (International Committee of the Red Cross, 2021).

Taken together, the picture that emerges from this overview is one of a governance landscape in which the diagnosis is complete, the principles are widely agreed, and the political will, at least in declaratory form, is visible. The research has been done, the ethical concerns have been exhaustively mapped and the funding for further analysis is available. What remains absent is a binding, enforceable international legal instrument. The urgency of this gap cannot be overstated: autonomous weapons systems are a present reality being deployed in active conflicts while the international community continues to deliberate. The international community has the knowledge, the institutional architecture, and the political coalitions necessary to act. What is required now is not another summit, another report, or another non-binding declaration, but a decisive decision to move from principle to rule.

# 5 A Roadmap for the Governance of Artificial Intelligence in Warfare

Building from the analysis carried out throughout the paper, a roadmap of five core recommendations, based on four core principles, is proposed. Previous works and recommendations have been presented by major policy centers and they all converge on a shared diagnosis: the governance of military AI suffers less from a scarcity of principles than from a scarcity of enforcement,

verification, and translation into practice. The structure below sets out guiding principles that any credible governance architecture for military AI must rest on, and then translates them into five concrete policy recommendations. The institutional impasse that impedes the adoption of binding regulation and serious action to control the use of AI in warfare mirrors a structural and wider issue. No State will accept unilateral limits on a capability seen as decisive to the balance of power while rivals remain unconstrained. This dynamic is further compounded by the involvement of the private sector in the race: the systems in question are designed and sold by defense-AI contractors whose commercial interest lies in expanding their state customers' appetite for autonomous capability, not restricting it. Therefore, a governance architecture that treats the private sector as a neutral implementer of state policy misreads the incentive structure.

## 5.1 Guiding principles

### 5.1.1 Common Good Governance

Common Good Governance is "the processes that create new global public goods in areas where there are either no global norms and institutional mechanisms under international law or, at best, unclear and insufficient frameworks for coordinated global action to confront large-scale risks" (Garcia, 2023a). In these areas, such as Military AI, there is neither certainty nor specific governance, no precise rules administering commonly agreed behavior to prevent wide-scale harm and avert risk. Common good governance offers a strategy to curtail this uncertainty, harness solidarity, and promote a better and more peaceful planet for humanity that will also protect future generations in the age of militarized AI (Garcia, 2023a).

### 5.1.2 Compliance with international law and the preservation of human dignity

No governance innovation, however sophisticated, substitutes for the existing body of international law that already binds states in armed conflict: the UN Charter's regime governing the use of force (*jus ad bellum*), international humanitarian law and its principles of distinction, proportionality, and precaution (*jus in bello*), and international human rights law. The develop-

ment and use of AI in the military domain must comply with international law, guided by widely shared ethical principles, to best ensure the preservation of human life and peace for present and future generations (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025).

### 5.1.3 Meaningful human control over life-and-death decisions

Irrespective of a system's technical sophistication, the decision to take a human life remains a moral and legal act that cannot be delegated to an algorithm. AI systems must be engineered so that ultimate responsibility for all critical decisions remains with human operators. States should ensure that the authority for life and death decisions is never delegated to machines. This principle is what distinguishes an AI-assisted decision from an AI-made one, and it is the single assumption most directly at risk in the case studies of AI-enabled targeting examined in this report (Centre for Humanitarian Dialogue, 2021).

### 5.1.4 Structural responsibility

Ethical and legal compliance cannot be inspected at the point of deployment, but must be designed in from the earliest research and development stage and carried through the entire AI system lifecycle, so that ultimate responsibility for critical decisions remains, at every stage, with an identifiable human. This principle, which the Global Commission on Responsible Artificial Intelligence in the Military Domain calls "responsibility by design", reframes the debate away from asking whether an autonomous system "behaved well" in a given instance, and towards asking whether the institutions, training, and testing that produced it were structured to make human responsibility possible in the first place (Global Commission on Responsible Artificial Intelligence in the Military Domain, 2025).

## 5.2 Policy recommendations

### 5.2.1 Permanent multi-stakeholder dialogue

**Institutionalize a permanent, inclusive multi-stakeholder dialogue on military AI, with a standing seat for third sector and youth policy organizations.**

***Rationale.*** As presented in the previous section, an institutional structure already exists. However, we believe that the REAIM Summit should evolve into a permanent dialogue anchored within the UN system, building on a "transnational networked cooperation" (Garcia, 2023b; 2025) and bringing together youth policy researchers, legal practitioners, and expert interlocutors. Formalizing a seat for third sector and youth-led policy organizations within this dialogue would operationalize the inclusivity these frameworks already call for.

### 5.2.2 Binding EU redlines

**Establish binding EU-level redlines prohibiting AI systems that autonomously select and engage human targets without meaningful human control.**

***Rationale.*** The UN Secretary-General's High-Level Advisory Body has called for "clear red lines delineating unlawful use cases, including relying on AI to select and engage targets autonomously," (United Nations High-level Advisory Body on Artificial Intelligence, 2024) a position echoed by the ICRC (International Committee of the Red Cross, 2021). The EU should adopt such redlines as binding internal policy and use its collective weight within the CCW Group of Governmental Experts on LAWS to push for their multilateral codification. AI-enabled targeting and ISR systems fall into legal grey zones because testing, evaluation, and legal review are treated as a one-off gate rather than a continuous process. Legal review must therefore be repeated whenever a fielded system receives a model or data update.

### 5.2.3 EU expert verification network and MRM

**Build a centralized EU expert knowledge-and-verification network on military AI, operating a Monitoring and Reporting Mechanism (MRM) for armed-conflict uses of AI, and feeding into the UN's Scientific Panel and Global Dialogue on AI Governance.**

***Rationale.*** Create a centrally coordinated expert hub capable of horizon-scanning, maintaining trusted repositories of verified incidents, and supporting confidence-building measures. An EU-funded network of this kind, interoperable with the UN's new scientific panel, would give European institutions independent technical capacity rather than reliance on state or industry self-reporting alone.

The Ethical AI Alliance has already called for a global Monitoring and Reporting Mechanism on AI harms (Ethical AI Alliance, 2025), and its AI Harm Map (Ethical AI Alliance, 2025) demonstrates in practice that verified, methodologically transparent documentation of AI-related harm is achievable. This report endorses that proposal and recommends extending it explicitly to armed conflict, where the evidentiary gap is widest and the consequences are the most grave. A pilot MRM, explicitly scoped to include IHL and IHRL violations in armed conflict and structured along the model the Alliance has set out, should be proposed jointly by the EU and interested Member States as a confidence-building measure ahead of any binding treaty process. A light pilot, anchored in UN practice and informed by civil society expertise, would demonstrate feasibility while ensuring inclusivity of voices often excluded from international reporting. Establishing such a mechanism will provide Member States with the verified evidence they need to uphold their own resolution, and to ensure that AI serves people rather than harms them.

### 5.2.4 Shared testing and incident standards

**Adopt shared testing, evaluation, and incident-reporting standards as confidence-building measures among AI-advanced militaries.**

***Rationale.*** Information-sharing on testing methodology, performance metrics, and post-incident investigation is a low-cost, high-trust starting point for cooperation even among strategic competitors. The EU should propose a structured, reciprocal exchange of AI incident data as a first, deliberately modest step toward the wider architecture outlined above.

### 5.2.5 Binding joint state-contractor regulation

**Establish binding regulation applying jointly to procuring states and their contracting parties, conditioning military AI procurement, export, and deployment on verifiable human-in-the-loop demonstration and a defined autonomy threshold.**

***Rationale.*** Because contractors and procuring states share a commercial and strategic interest in expanding rather than constraining autonomous capability, self-certification by either party is not a credible compliance mechanism. Binding obligations should therefore attach jointly and severally to the state and the company supplying the system, and should require: (a) auditable monitoring systems logging human-machine interaction across the targeting cycle, not self-reported summaries; (b) a testable, demonstrable standard for meaningful human control; and (c) a threshold, calibrated to a system's degree of autonomy and target class, above which enhanced review, joint state-contractor liability, and mandatory incident reporting are triggered automatically.

## 5.3 Closing note

The evidence assembled in this report leaves little room for complacency. AI is already being deployed on the battlefield and the pace at which these technologies are being fielded consistently outpaces the pace at which they are being governed. Every additional year of delay locks in operational habits, contractual dependencies between states and industry, and institutional inertia that will make the governance this report calls for progressively harder to build. Governance must be built with the same urgency states have shown in fielding these systems in the first place.

This urgency applies with particular force to the European Union, which needs to better position itself with clear normative and strategic options to respond to changes in the AI ecosystem and the geopolitical landscape. Its market size, its regulatory record on emerging technology, and its standing within the UN system mean that EU action sets a reference point that other jurisdictions are likely to follow.

The Union also has a responsibility to protect the fundamental rights and security of its citizens, who are directly exposed to the risks this report has outlined, and the governance of military AI falls squarely within that responsibility. The EU has an essential role to play in shaping safeguards for high-risk uses of AI, and in doing so, promoting global norms and ethical standards. Standing at the confluence of EDT advancements and rising great-power competition in AI (Meacham, 2023), the EU faces the daunting challenge of navigating a geopolitical minefield.

That is why the Union should take a leading role in supporting the building of a coalition for the global governance of military AI, using its convening power to bring together the states, civil society organizations, and technical experts capable of turning the recommendations above into enforceable commitments.

# Bibliography

Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The Impact of AI Errors in a Human-in-the-Loop Process. *Cognitive Research: Principles and Implications*, *9*(1), 1. https://doi.org/10.1186/s41235-023-00529-3

Akıllı, E. (2025). The Transformation of Sovereignty in the Technopolar World. In *Cognitive Diplomacy and Digital Autonomy* (pp. 31–51). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-04470-9_3

Asaro, P. (2012). On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making. *International Review of the Red Cross*, *94*(886), 687–709. https://doi.org/10.1017/S1816383112000768

Baker, K. T. (2026, March 26). *AI Got the Blame for the Iran School Bombing. The Truth Is Far More Worrying*. The Guardian. https://www.theguardian.com/news/2026/mar/26/ai-got-the-blame-for-the-iran-school-bombing-the-truth-is-far-more-worrying

Buzan, B. (1987). *An Introduction to Strategic Studies: Military Technology and International Relations*. Macmillan, in association with the International Institute for Strategic Studies. https://institutes.abu.edu.ng/idr/public/assets/docs/An%20Introduction%20to%20Strategic%20Studies_%20Military%20Technology%20and%20International%20Relations%20(%20PDFDrive%20).pdf

Campbell, C. (2026, March 10). *The Race to Build AI Humanoid Soldiers for War*. TIME. https://time.com/article/2026/03/09/ai-robots-soldiers-war/

Centre for Humanitarian Dialogue. (2021). *Code of Conduct on Artificial Intelligence in Military Systems*. https://www.hdcentre.org/wp-content/uploads/2021/08/AI-Code-of-Conduct.pdf

Chauhan, D. (2025, October 3). *AI on the Frontlines: How Agentic AI is Revolutionizing Cyber Defense*. https://www.cyberdefensemagazine.com/ai-on-the-frontlines-how-agentic-ai-is-revolutionizing-cyber-defense/

Chengeta, T. (2016). Dignity, Ubuntu, Humanity and Autonomous Weapon Systems (AWS) Debate: An African Perspective. *Revista De Direito Internacional*, *13*(2). https://doi.org/10.5102/rdi.v13i2.4135

Clapp, S. (2022). *Emerging Disruptive Technologies in Defence* (PE 733.647). https://www.europarl.europa.eu/RegData/etudes/ATAG/2022/733647/EPRS_ATA(2022)733647_EN.pdf

Convention on Certain Conventional Weapons, Fifth Review Conference of the High Contracting Parties. (2016). *Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)* (No. CCW/CONF.V/2). https://docs.un.org/en/CCW/CONF.V/2

Cummings, M. L. (2019). Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification?. *IEEE Technology and Society Magazine*, *38*(4), 20–26. https://doi.org/10.1109/MTS.2019.2948438

Darktrace Federal. (2025, December 9). *Darktrace Federal and Navitas Business Consulting Awarded Department of State Contract for AI-Powered Network Detection and Response*. https://www.darktrace.com/news/darktrace-federal-and-navitas-business-consulting-awarded-department-of-state-contract-for-ai-powered-network-detection-and-response

De Balie. (2026, April 14). *AI at War – With Shane Harris*. YouTube. https://www.youtube.com/watch?v=0TD9AH_Stsc

Deeks, A., Lubell, N., & Murray, D. (2019). Machine Learning, Artificial Intelligence, and the Use of Force by States. *Journal of National Security Law and Policy*, *10*(1), 1–25. https://nationalsecurity.law.georgetown.edu/wp-content/uploads/2019/04/Machine_Learning_Artificial_Intelligence_2.pdf

Dhanushkodi, K., & Thejas, S. (2024). AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation. *IEEE Access*, *12*, 173127–173136. https://doi.org/10.1109/ACCESS.2024.3493957

Doicariu, D. (2025). The Impact of Artificial Intelligence on Military Education and Training. *Defense Resources Management in the 21st Century*, *20*(1), 99–109. https://www.ceeol.com/search/article-detail?id=1406264

Drabkina, Y. (2024). Роботизовані підрозділи дозволили скоротити втрати серед українських військових на 30% – джерело NV у Генштабі. *NV*. https://nv.ua/ukr/ukraine/events/robotizovani-pidrozdili-skorotili-vtrati-sered-viyskovih-na-30-genshtab-50462290.html

Elish, M. C. (2019). Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *Engaging Science, Technology, And Society*, *5*, 40–60. https://doi.org/10.17351/ests2019.260

Elsevier. (n.d.). *Decision Support System*. Retrieved https://www.sciencedirect.com/topics/computer-science/decision-support-system

Erskine, T., & Miller, S. E. (2024). AI and the Decision to Go to War: Future Risks and Opportunities. *Australian Journal of International Affairs*, *78*(2), 135–147. https://doi.org/10.1080/10357718.2024.2349598

Ethical AI Alliance. (2025). *AI Governance at a Crossroads*. https://ethicalaialliance.org/reports

Ethical AI Alliance. (2025). *AI Harm Map*. https://ethicalaialliance.org/ai-harm-map

European Commission. (n.d.). *Defence Industry and Space*. Retrieved https://defence-industry-space.ec.europa.eu/index_en

European Defence Agency. (n.d.). *Autonomous Systems*. Retrieved https://eda.europa.eu/what-we-do/eu-policies/autonomous-systems

European Defence Agency. (2024). *The EDA Action Plan on Autonomous Systems: Preparing the Autonomous Systems of the Future*. https://eda.europa.eu/docs/default-source/brochures/eda%E2%80%94apas-brochure%E2%80%94final-(1).pdf

European Defence Agency. (2025). *Trustworthiness for AI in Defence (TAID) – White Paper: Developing Responsible, Ethical, and Trustworthy AI Systems for European Defence*. https://eda.europa.eu/docs/default-source/brochures/taid-white-paper-final-09052025.pdf

European External Action Service. (2025, April 7). *EU Statement – UN Disarmament Commission: General Debate*. https://www.eeas.europa.eu/delegations/un-new-york/eu-statement-%E2%80%93-un-disarmament-commission-general-debate-0_en

European Parliament. (2018). *European Parliament Resolution of 12 September 2018 on Autonomous Weapon Systems* (P8_TA(2018)0341). https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.html

European Parliament. (2021). *European Parliament Resolution of 20 January 2021 on Artificial Intelligence: Questions of Interpretation and Application of International Law in so Far as the EU Is Affected in the Areas of Civil and Military Uses and of State Authority Outside the Scope of Criminal Justice* (P9_TA(2021)0009). https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_EN.pdf

European Parliament and Council of the European Union. (2021). *Regulation (EU) 2021/697 of the European Parliament and of the Council of 29 April 2021 Establishing the European Defence Fund and Repealing Regulation (EU) 2018/1092* (Regulation (EU) 2021/697). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02021R0697-20251223

European Parliament and Council of the European Union. (2021, May 20). *Regulation (EU) 2021/821 of the European Parliament and of the Council of 20 May 2021 Setting Up a Union Regime for the Control of Exports, Brokering, Technical Assistance, Transit and Transfer of Dual-Use Items (Recast)*. https://eur-lex.europa.eu/eli/reg/2021/821/oj/eng

European Parliament and Council of the European Union. (2024, June 13). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

European Union. (2024). *The European Union Submission to the UN Secretary-General's Call for Contributions to a Report Under Resolution 78/241 – Lethal Autonomous Weapons Systems*. https://docs-library.unoda.org/General_Assembly_First_Committee_-Seventy-Ninth_session_(2024)/78-241-EU-EN.pdf

Financial Times Editorial Board. (2026). Iran and the Rising Perils of AI in Warfare. *Financial Times*. https://www.ft.com/content/5d294db8-6917-4085-ba8a-3afbc439b999

Frenkel, S. (2024, March 27). *Israel Deploys Expansive Facial Recognition Program in Gaza*. The New York Times. https://www.nytimes.com/2024/03/27/technology/israel-facial-recognition-gaza.html

Fuller, H. (2026, March 22). *Pentagon Expands Use of Palantir AI in New Defense Contract*. Military.com. https://www.military.com/feature/2026/03/22/pentagon-expands-palantirs-role-ai-contract.html

Ganesan, K., Raza, S. K., & Vijayaraghavan, R. (2010). Chemical Warfare Agents. *Journal of Pharmacy and Bioallied Sciences*, *2*(3), 166–178. https://doi.org/10.4103/0975-7406.68498

Garcia, D. (2016). Future Arms, Technologies, and International Law: Preventive Security Governance. *European Journal of International Security*, *1*(1), 94–111. https://doi.org/10.1017/eis.2015.7

Garcia, D. (2023a). Common Good Governance. In *The AI Military Race: Common Good Governance in the Age of Artificial Intelligence*. Oxford University Press. https://doi.org/10.1093/oso/9780192864604.003.0003

Garcia, D. (2023b). Transnational Networked Cooperation. In *The AI Military Race: Common Good Governance in the Age of Artificial Intelligence* (pp. 34–98). Oxford University Press. https://doi.org/10.1093/oso/9780192864604.003.0002

Garcia, D. (2025). The Global Diplomacy of Governing Military Artificial Intelligence. *Ethics and Information Technology*, *27*, 55. https://doi.org/10.1007/s10676-025-09863-0

Garon, J. M. (2022). When AI Goes to War: Corporate Accountability for Virtual Mass Disinformation, Algorithmic Atrocities, and Synthetic Propaganda. *Northern Kentucky Law Review*, *49*(2), 181. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4366903

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

Government of the Netherlands. (2023). *REAIM 2023 Call to Action*. https://www.government.nl/documents/2023/02/16/reaim-2023-call-to-action

Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. (2019). *Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems* (CCW/GGE.1/2019/3). https://documents.un.org/doc/undoc/gen/g19/285/69/pdf/g1928569.pdf?OpenElement

Hartridge, S., & Walker-Munro, B. (2025). Autonomous Weapons Systems and the AI Alignment Problem. *Journal of International Humanitarian Legal Studies*, *16*(1), 38–65. https://doi.org/10.1163/18781527-bja10107

Hassen, M. Z. (2025). The Risk of Moral Outsourcing: Why Artificial Intelligence Cannot and Should Not Make Our Ethical Decisions. *EthAIca: Journal of Ethics, AI and Critical Analysis*, *4*, 428. https://doi.org/10.56294/ai2025428

High Contracting Parties to the Convention on Certain Conventional Weapons. (2025). *Joint Statement to the September 2025 Session of the CCW GGE LAWS*. https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2025)/Joint_statement_to_the_September_2025_session_of_the_CCW_GGE_LAWS_-_As_delivered.pdf

High-Level Expert Group on Artificial Intelligence. (2018). *A Definition of AI: Main Capabilities and Scientific Disciplines*. https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

Horowitz, M. C. (2016). The Ethics and Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons. *Daedalus*, *145*(4), 25–36. https://doi.org/10.1162/DAED_a_00409

Human Rights Watch. (2020). *Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control*. https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and

Human Rights Watch. (2024). *Questions and Answers: Israeli Military's Use of Digital Tools in Gaza*. https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-of-digital-tools-in-gaza

International Committee of the Red Cross. (2020). International Committee of the Red Cross (ICRC) Position on Autonomous Weapon Systems: ICRC Position and Background Paper. *International Review of the Red Cross*, *102*(915), 1335–1349. https://doi.org/10.1017/S1816383121000564

International Committee of the Red Cross. (2021). *ICRC Position on Autonomous Weapon Systems*. https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems

International Committee of the Red Cross, & Geneva Academy of International Humanitarian Law and Human Rights. (2024). *Artificial Intelligence and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts: Current*

*Developments and Potential Implications* (No. 4758). International Committee of the Red Cross. https://shop.icrc.org/expert-consultation-report-artificial-intelligence-and-related-technologies-in-military-decision-making-on-the-use-of-force-in-armed-conflicts-current-developments-and-potential-implications-pdf-en.html

Ishkhanyan, A. (2026). Governing AI Across Borders: Corporate Power, State Sovereignty and Global Regulation. *Digital Policy, Regulation and Governance*, *28*(3), 280–295. https://doi.org/10.1108/DPRG-03-2025-0057

John-Mathews, J.-M. (2022). Some Critical and Ethical Perspectives on the Empirical Turn of AI Interpretability. *Technological Forecasting and Social Change*, *174*, 121209. https://doi.org/10.1016/j.techfore.2021.121209

Kahn, L. (2022). Lethal Autonomous Weapon Systems and Respect for Human Dignity. *Frontiers in Big Data*, *5*, 999293. https://doi.org/10.3389/fdata.2022.999293

Kryzhanivska, O. (2026, February 12). *Drone Warfare in Ukraine: Ground, River and Sea Drone Trends*. https://ukrainesarmsmonitor.substack.com/p/drone-warfare-in-ukraine-ground-river

Kumar, A. (2026). Pentagon Used Anthropic's Claude AI and Palantir Maven to Identify 1,000 Targets in Iran Strikes. *The Defense News*. https://www.thedefensenews.com/news-details/Pentagon-Used-Anthropics-Claude-AI-and-Palantir-Maven-to-Identify-1000-Targets-in-Iran-Strikes/

Lee, C.-E., Baek, J., Son, J., & Ha, Y.-G. (2023). Deep AI Military Staff: Cooperative Battlefield Situation Awareness for Commander's Decision Making. *The Journal of Supercomputing*, *79*(6), 6040–6069. https://doi.org/10.1007/s11227-022-04882-w

Lewis, D. (2020, September 28). *An Enduring Impasse on Autonomous Weapons*. https://www.justsecurity.org/72610/an-enduring-impasse-on-autonomous-weapons/

Lindsay, J. R. (2026, March 11). *US Military Leans Into AI for Attack on Iran, But the Tech Doesn't Lessen the Need for Human Judgment in War*. https://news.gatech.edu/news/2026/03/11/us-military-leans-ai-attack-iran-tech-doesnt-lessen-need-human-judgment-war

Loh, M. (2026, March 24). *Anduril's President Says the Firm Is a 'Heavy' Participant in the Middle East War, Helping to Battle Iran's Shaheds*. Business Insider. https://www.businessinsider.com/anduril-heavy-participant-iran-war-middle-east-matthew-steckman-shahed-2026-3

Maathuis, C., & Cools, K. (2025a). The Role of AI in Military Cyber Security: Data Insights and Evaluation Methods. *Procedia Computer Science*, *254*, 191–200. https://doi.org/10.1016/j.procs.2025.02.078

Maathuis, C., & Cools, K. (2025b, September 16). *Digital Sovereignty Control Framework for Military AI-Based Cyber Security*. https://arxiv.org/abs/2509.13072

Macfarlane, K., & Christie, L. (2022). *Automation in Military Operations* (No. POSTnote 681). UK Parliament. https://post.parliament.uk/research-briefings/post-pn-0681/

Mako, G. (2026, May 11). *Unity in Principle, Variation in Practice: European Approaches to Meaningful Human Control for LAWS*. https://lieber.westpoint.edu/unity-principle-variation-practice-european-approaches-meaningful-human-control-laws/

Manna, R., & Nath, R. (2021). The Problem of Moral Agency in Artificial Intelligence. *2021 IEEE Conference on Norbert Wiener in the 21st Century (21cw)*, 1–4. https://doi.org/10.1109/21CW48944.2021.9532549

Markelius, A., Wright, C., Kuiper, J., Delille, N., & Kuo, Y.-T. (2024). The Mechanisms of AI Hype and Its Planetary and Social Costs. *AI and Ethics*, *4*(3), 727–742. https://doi.org/10.1007/s43681-024-00461-2

Meacham, S. (2023, September 8). *A Race to Extinction: How Great Power Competition Is Making Artificial Intelligence Existentially Dangerous*. Harvard International Re-

view. https://hir.harvard.edu/a-race-to-extinction-how-great-power-competition-is-making-artificial-intelligence-existentially-dangerous/

Ministerio de Asuntos Exteriores, Unión Europea y Cooperación. (2026, February 4). *REAIM 2026: Third Summit on Responsible Artificial Intelligence in the Military Domain*. https://www.exteriores.gob.es/en/REAIM2026/Paginas/Cumbre26.aspx

Ministry of Foreign Affairs, Republic of Korea. (2024, September 10). *Outcome of Responsible AI in Military Domain (REAIM) Summit 2024*. https://www.mofa.go.kr/eng/brd/m_5676/view.do?seq=322676

Mustafovski, R. (2025). The Use of Communication Platforms in Military Operations: Enhancing Strategic and Tactical Effectiveness. *Database Systems Journal*, *16*(1), 1–10. https://www.dbjournal.ro/archive/36/36_1.pdf

NATO Standardization Office. (2019). *AAP-06 Edition 2019: NATO Glossary of Terms and Definitions (English and French)*. North Atlantic Treaty Organization, NATO Standardization Office. https://www.coemed.org/files/stanags/05_AAP/AAP-06_2019_EF.pdf

OECD. (2024). *OECD AI Principles Overview*. https://oecd.ai/en/principles

Palantir Technologies. (2026, April 19). *The Technological Republic, in Brief*. https://x.com/PalantirTech/status/2045574398573453312

Pandise, E., Kent, J. L., Kaplan, M., & Mosk, M. (2026, March 18). *How the Military Is Using AI in War*. CBS News. https://www.cbsnews.com/news/ai-artificial-intelligence-military-use-in-war/

Podar, H., & Colijn, A. (2025, February 14). *Technical Risks of (Lethal) Autonomous Weapons Systems*. https://doi.org/10.48550/arXiv.2502.10174

Pultarova, T. (2026). The Coming Drone-War Inflection in Ukraine. *IEEE Spectrum*. https://spectrum.ieee.org/autonomous-drone-warfare

Rashid, A. B., Kausik, A. K., Sunny, A. A. H., & Bappy, M. H. (2023). Artificial Intelligence in the Military: An Overview of the Capabilities, Applications, and Challenges. *International Journal of Intelligent Systems*, *2023*(1), 8676366. https://doi.org/10.1155/2023/8676366

Renic, N. (2026). AI-Optimized Violence and the Suffocation of Moral and Political Wisdom. *Cambridge Forum on AI: Law and Governance*, *1*, e54. https://doi.org/10.1017/cfl.2025.10025

Roland, A. (1995). Science, Technology, and War. *Technology and Culture*, *36*(2), S83–S100. https://doi.org/10.2307/3106691

Rönnback, R. (2023, February 27). *Challenges of Governing AI for Military Purposes and Spill-Over Effects of the AI Act*. https://futurium.ec.europa.eu/en/european-ai-alliance/community-content/challenges-governing-ai-military-purposes-and-spill-over-effects-ai-act?language=en

Rossiter, A. (2023). AI-Enabled Remote Warfare: Sustaining the Western Warfare Paradigm?. *International Politics*, *60*(4), 818–833. https://doi.org/10.1057/s41311-021-00337-w

Santoni de Sio, F., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them. *Philosophy and Technology*, *34*(4), 1057–1084. https://doi.org/10.1007/s13347-021-00450-x

Santopinto, F. (2024). *The EU, Artificial Military Intelligence and Autonomous Lethal Weapons*. https://www.iris-france.org/wp-content/uploads/2024/04/IRIS_EN_ProgIndusDef_SALA_april-2024.pdf

Sayler, K. M. (2026). *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems* (No. IF11150). https://www.congress.gov/crs-product/IF11150

Shoker, S. (2019). *Algorithmic Bias and the Principle of Distinction: Towards an Audit of Lethal Autonomous Weapons Systems* (Technical Report 19/2; pp. 41–46). https://macsphere.mcmaster.ca/bitstreams/541167a0-0da1-4187-a953-960bfa327cc3/download

Simmons, L. (2026, March 28). *AI on the Frontlines: Value Alignment, Responsibility and the Moral Muscle*. Almost in the Machine. https://open.substack.com/pub/almostinthemachine/p/ai-on-the-frontlines-value-alignment

Sourav, M. T. I., Masud, N., Al Bassam, A., & Sobhan, A. (2025). Governance and Ethical Challenges of Artificial Intelligence in Warfare. *2025 International Conference on Quantum Photonics, Artificial Intelligence, And Networking (QPAIN)*, 1–5. https://doi.org/10.1109/QPAIN66474.2025.11171759

Stop Killer Robots. (2013). *Stop Killer Robots*. https://www.stopkillerrobots.org/

Stop Killer Robots. (2022, October). *70 States Deliver Joint Statement on Autonomous Weapons Systems at UN General Assembly*. https://www.stopkillerrobots.org/news/70-states-deliver-joint-statement-on-autonomous-weapons-systems-at-un-general-assembly/

Stop Killer Robots. (2023, November 1). *164 States Vote Against the Machine at the UN General Assembly*. https://www.stopkillerrobots.org/news/164-states-vote-against-the-machine/

Thompson, S. A., & Cardia, A. (2026, March 14). *Cascade of A.I. Fakes About War With Iran Causes Chaos Online*. The New York Times. https://www.nytimes.com/interactive/2026/03/14/business/media/iran-disinfo-artificial-intelligence.html

Trofymenko, O., Sokolov, A., Loginova, N., Akhmametieva, H., & Chykunov, P. (2024). Artificial Intelligence in the Field of Military Logistics. *System Technologies*, *5*(154). https://doi.org/10.34185/1562-9945-5-154-2024-17

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Ul Haq, S. (2026). Psychopolitics of Warfare: Algorithmic Violence and the Digital Unconscious in Military AI Systems. *Philosophy of Management*. https://doi.org/10.1007/s40926-026-00387-1

United Nations General Assembly. (2023, October 24). *Without Adequate Guardrails, Artificial Intelligence Threatens Global Security in Evolution from Algorithms to Armaments, Speaker Tells First Committee*. https://press.un.org/en/2023/gadis3725.doc.htm

United Nations General Assembly. (2025). *Lethal Autonomous Weapons Systems* (A/RES/80/57). https://docs.un.org/en/A/RES/80/57

United Nations General Assembly, First Committee. (2023). *Lethal Autonomous Weapons Systems* (A/C.1/78/L.56). https://reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com23/resolutions/L56.pdf

United Nations High-level Advisory Body on Artificial Intelligence. (2024). *Governing AI for Humanity: Final Report*. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

United Nations Office for Disarmament Affairs. (2018). *Convention on Certain Conventional Weapons – Group of Governmental Experts on Lethal Autonomous Weapons Systems (2018)*. https://meetings.unoda.org/ccw/convention-certain-conventional-weapons-group-governmental-experts-2018

United Nations Office of the High Commissioner for Human Rights. (2023). *Taxonomy of Human Rights Risks Connected to Generative AI: Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI*. https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf

United Nations Secretary-General. (2024a). *Lethal Autonomous Weapons Systems: Report of the Secretary-General* (A/79/88). https://docs.un.org/en/a/79/88

United Nations Secretary-General. (2024b, December 19). *Secretary-General's Remarks to the Security Council on Artificial Intelligence [bilingual, as delivered]*. https://www.un.org/sg/en/content/sg/statements/2024-12-19/secretary-generals-remarks-the-security-council-artificial-intelligence-bilingual-delivered

University of Amsterdam. (2026, March 16). *'Militaries Are Increasingly Accepting AI's Mistakes in Warfare'*. https://www.uva.nl/en/shared-content/faculteiten/en/faculteit-der-rechtsgeleerdheid/news/2026/03/ais-mistakes-in-warfare.html

Viveros Álvarez, J. S. (2024, November 27). *Why Should the UN "Govern AI for Humanity": What is at Stake and What is the Urgency?*. https://opiniojuris.org/2024/11/27/why-should-the-un-govern-ai-for-humanity-what-is-at-stake-and-what-is-the-urgency/

Wood, N. G. (2023). Autonomous Weapon Systems and Responsibility Gaps: A Taxonomy. *Ethics and Information Technology*, *25*(1), 16. https://doi.org/10.1007/s10676-023-09690-1

Woodcock, T. K. (2024). Human/Machine(-Learning) Interactions, Human Agency and the International Humanitarian Law Proportionality Standard. *Global Society*, *38*(1), 100–121. https://doi.org/10.1080/13600826.2023.2267592