Accountability Gaps in Autonomous Warfare: Rethinking the Law of Armed Conflict in the Age of AI

Abdullah Hemmet
International Islamic University Malaysia IIUM (Malaysia), hemmet.abdullah@hotmail.com
https://orcid.org/0009-0001-6650-9788

Received: .20/04/2026 Accepted: .17/05 /2026 Published: 01/06/2026

Abstract

Background. Autonomous and AI-enabled weapon systems have moved from speculative debate into the center of contemporary security governance. International humanitarian law (IHL) unquestionably applies to their development and use, yet the accelerating integration of machine learning, sensor fusion, autonomous navigation, and automated target-recognition functions has exposed a persistent uncertainty: not whether law applies, but whether responsibility for unlawful harm can still be attributed in a sufficiently precise, fair, and operationally meaningful way (ICRC, 2021, 2026; UNIDIR, 2025).

Purpose. This paper examines the accountability gaps that emerge when lethal force is mediated by autonomous or AI-enabled systems and argues that Europe requires a more explicit legal and institutional architecture than the current patchwork of general IHL rules, national weapons reviews, and non-binding AI principles. It asks how accountability diffuses across the lifecycle of such systems, why existing doctrines struggle to absorb that diffusion, and what a European response should look like.

Methodology. The paper employs a doctrinal and policy-analytical methodology. It interprets treaty law, especially Additional Protocol I and the Rome Statute, in light of contemporary institutional practice, and it comparatively analyzes authoritative materials from the ICRC, the United Nations, UNIDIR, NATO, the European Parliament, the Council of Europe, the European Union, and SIPRI, alongside peer-reviewed scholarship on autonomy, targeting, command responsibility, and human control (Additional Protocol I, 1977, art. 36; Rome Statute, 1998, art. 28; NATO, 2024; SIPRI, 2023).

Findings. The paper identifies five interlocking accountability gaps: an epistemic gap, a lifecycle-diffusion gap, a review-to-use gap, a command-responsibility gap, and a coalition-and-contractor gap. It argues that

existing law remains applicable but is insufficiently specific to preserve meaningful accountability under conditions of algorithmic opacity, probabilistic targeting, software updates, and multinational interoperability. In response, the paper proposes a three-layer European accountability architecture combining substantive prohibitions and restrictions, operational assurance duties, and institutional oversight mechanisms.

Research limitations. This study is doctrinal and policy-oriented rather than empirical. It does not rely on classified operational data, proprietary system-testing results, or privileged procurement documentation. Its conclusions therefore concern legal structure and governance design rather than the performance of any single platform.

Originality/value. The paper contributes by moving beyond the binary debate between "law already suffices" and "all autonomy must be banned." It shows that the central problem is the under-specification of accountability across the full sociotechnical lifecycle of military AI, and it develops a distinctly European pathway for preserving both operational effectiveness and legal legitimacy.

Keywords: autonomous weapon systems; law of armed conflict; international humanitarian law; artificial intelligence; accountability; command responsibility; Article 36 review; European security governance.

1. Introduction

Europe's contemporary security environment is increasingly shaped by a conjunction of strategic volatility and technological acceleration. Military planners now operate in a setting marked by attritional interstate war on the continent's periphery, renewed great-power competition, contested supply chains, degraded arms-control architectures, and rapid diffusion of artificial intelligence across surveillance, logistics, decision support, and weapons functions. In this setting, autonomous and AI-enabled weapon systems are no longer a remote legal curiosity. They have become a central site at which strategic necessity, technological ambition, humanitarian protection, and legal responsibility intersect. The debate is therefore not reducible to a philosophical dispute about "killer robots." It concerns whether the law of armed conflict can still perform its core civilizational function—disciplining violence through attributed responsibility—when lethal decisions are increasingly mediated by complex machine processes (Bode, 2023; ICRC, 2024, 2026; NATO, 2024).

The present debate often proceeds from two unsatisfactory extremes. One position assumes that because IHL is technologically neutral, existing law already resolves the issue. A different position treats autonomy as inherently incompatible with the legal and moral architecture of warfare and therefore concludes that prohibition is the only coherent response. Both positions capture something important, but neither is sufficient. The first underestimates the extent to which legal responsibility depends not only on formal rules but on institutional pathways of knowledge, foreseeability, and attribution. The second risks ignoring meaningful distinctions between weapon functions, operational contexts, and degrees of human involvement. The task is not merely to ask whether autonomous weapons are lawful in the abstract. It is to determine how legal accountability can remain intelligible, demonstrable, and enforceable when decision-making is distributed across designers, data curators, procurement authorities, commanders, operators, and machine-learning-enabled systems themselves (Asaro, 2012; Ekelhof, 2018, 2019

This paper argues that the most serious challenge posed by autonomous warfare is not the disappearance of law but the erosion of accountable legal agency. Existing IHL and international criminal law unquestionably remain applicable. Yet their application becomes progressively thinner when those who authorize force cannot fully understand the causal logic of the system they employ, when the legality of a platform depends on context-sensitive constraints that may change after deployment, when software and data are supplied by multiple private actors across jurisdictions, and when coalition operations combine divergent national review standards. In those circumstances, responsibility does not vanish; it diffuses. The result is not a metaphysical "gap" in the sense of complete normlessness, but a doctrinal and institutional under-specification that threatens to turn accountability into a formal assertion rather than an operational reality (ICRC, 2021, 2026; SIPRI, 2022, 2023; UNIDIR, 2025).

The paper's central research question is therefore as follows: where, precisely, do accountability gaps arise in the development and use of autonomous or AI-enabled weapon systems, and what legal and governance reforms should European institutions and states adopt in order to close them? The argument advanced here is threefold. First, the legal difficulties associated with autonomous warfare arise less from machine "agency" as such than from the redistribution of epistemic and decisional authority across a sociotechnical lifecycle. Second, the conventional reliance on Article 36 weapons review and ex post command responsibility is necessary but insufficient, because neither doctrine alone was designed to manage opacity, model adaptation, or multinational software-dependent targeting chains. Third, Europe is unusually well positioned to articulate a layered response because it combines a strong legalist tradition, dense institutional networks, and a growing defense-AI agenda; yet Europe has not yet translated those assets into a coherent accountability architecture (European Parliament Research Service [EPRS], 2025; European Union, 2024; NATO, 2024).

The paper makes four specific contributions. At the doctrinal level, it distinguishes more carefully than much of the existing literature between legality and accountability, showing that the latter cannot be assumed from the former. At the conceptual level, it reframes accountability as a lifecycle problem rather than a battlefield-only problem. At the policy level, it identifies a European regulatory paradox: the Union and associated European institutions speak in the language of human-centric and rights-based AI governance, yet military AI remains only partially addressed or explicitly excluded from key instruments. At the prescriptive level, it proposes a three-layer European accountability architecture that integrates substantive legal limits, operational assurance obligations, and institutional oversight mechanisms. The purpose is not to invent a wholly new law of war for AI. It is to clarify where existing doctrine is strained, where interpretive supplementation is needed, and where new regulation is justified (Council of Europe, 2024; European Parliament, 2018, 2021; Perrin, 2025).

1.1. Methodology, sources, and analytical posture

This study adopts a doctrinal legal method combined with comparative policy analysis. It begins from the positive law of armed conflict and international criminal law, particularly Additional Protocol I, the Rome Statute, and the customary structure of IHL concerning weapons, targeting, and precautions. It then interprets these rules in light of contemporary institutional materials: ICRC position papers, UN and UNIDIR reports, NATO strategy documents, European parliamentary and regulatory materials, and SIPRI analyses. These sources are treated not as equivalent authorities, but as distinct forms of legal and policy evidence. Treaty law and customary norms provide the legal baseline; institutional documents illuminate contemporary interpretation; peer-reviewed scholarship helps expose conceptual tensions and unresolved doctrinal assumptions (Additional Protocol I, 1977; Rome Statute, 1998; ICRC, 2021, 2026; UNIDIR, 2025).

The paper is intentionally neither technophobic nor technologically romantic. It does not assume that any degree of autonomy automatically violates IHL. As Schmitt and Thurnher (2013) persuasively argued, autonomous weapon systems are not unlawful per se. Nor does the paper assume that human decision-making is always superior. Human soldiers and commanders are prone to fatigue, bias, panic, tunnel vision, and poor judgment; in some circumstances, machine assistance may improve discrimination or reduce reaction time. What matters, however, is that legal responsibility in armed conflict is not a purely functional question of outcome optimization. It is also a normative question of who is answerable, on what basis, and through what procedures, when force is used unlawfully. That is the axis on which this paper turns.

1.2. Research limitations

The analysis is subject to three limitations. First, many of the most operationally significant military AI systems are classified or only partially disclosed; no public study can reconstruct the full performance envelope or command architecture of such systems. Second, doctrinal analysis cannot substitute for detailed empirical testing. A legal conclusion about foreseeability or control may differ depending on the technical properties of a given model or platform. Third, although the paper focuses on Europe, the relevant legal field is international and comparative. European approaches are therefore analyzed in conversation with non-European practice, especially U.S. and NATO materials. These limitations do not weaken the central claim. They reinforce it: if accountability is difficult to establish even from the standpoint of public legal analysis, then the burden of creating traceable and reviewable governance mechanisms becomes more—not less—important.

2. Conceptual and legal foundations of autonomy in warfare

2.1. Definitional contestation and the spectrum of autonomy

One reason the legal debate has often stalled is definitional instability. "Autonomous weapon systems," "lethal autonomous weapon systems," "AI-enabled weapons," "algorithmic targeting," and "human-machine teaming" are frequently used as though they were interchangeable. They are not. Some systems operate through highly deterministic automated rules; others incorporate machine learning; some merely support target identification; others can select and engage targets after activation. The ICRC's 2021 position paper offered one of the most influential formulations by describing autonomous weapon systems as those that select and apply force to targets without human intervention after activation, so that the user does not choose, or may not even know, the specific target, timing, or place of the strike (ICRC, 2021). The ICRC's 2026 paper likewise stresses that the essential problem lies in the system's ability, once activated, to select and engage one or more targets without further human intervention (ICRC, 2026).

That basic formulation is useful, but it remains insufficient unless disaggregated. Autonomy may be present in navigation, sensing, prioritization, target recognition, route planning, or terminal engagement. A system may be "autonomous" for one function and not for another. It is therefore more precise to think in terms of an autonomy spectrum rather than a single category. The familiar language of human-in-the-loop, human-on-the-loop, and human-out-of-the-loop is a helpful shorthand, but it can also mislead if it suggests that the legal

question is exhausted by the mere physical presence of a human somewhere in the chain. A nominally "in-the-loop" approval can be legally thin if the operator lacks the information, time, or technical understanding needed to make a meaningful judgment. Conversely, a system with high automation in narrow defensive environments may remain more governable than a seemingly modest AI tool used in open-ended civilian-rich spaces (Boulanin & Verbruggen, 2017; Ekelhof, 2018, 2019).

This distinction matters because accountability problems often arise not only from the final effector platform but from the larger targeting stack. Modern military AI may be used to sort intelligence, generate strike recommendations, identify patterns of life, fuse sensor feeds, rank targets, or calculate routes and timing. If a commander relies on an AI-enabled decision-support system that shapes the target selection process in opaque ways, the legal problem may be substantially similar to that posed by a weapon with autonomous targeting. For this reason, recent SIPRI and UNIDIR work has increasingly emphasized that human-machine interaction across the entire lifecycle and decision chain is the proper analytic frame, rather than a narrow fixation on whether the final trigger event was technically autonomous (SIPRI, 2023, 2025; UNIDIR, 2025). This paper follows that broader but disciplined approach. It focuses on weapons and targeting architectures in which AI or autonomy materially mediates lethal force.

2.2. IHL as baseline: distinction, proportionality, precautions, and humanity

The legal baseline is clear. Autonomous or AI-enabled weapon systems do not exist in a vacuum outside the law of armed conflict. They are governed by the same general corpus of IHL that regulates means and methods of warfare more broadly. At a minimum, this includes the rule that attacks may be directed only against military objectives, the prohibition of indiscriminate attacks, the requirement to take feasible precautions in attack, and the proportionality rule barring attacks expected to cause excessive incidental civilian harm relative to the concrete and direct military advantage anticipated (Additional Protocol I, 1977, arts. 48, 51, 52, 57). The legal challenge is not the absence of rules. It is their translation into operational practice when the relevant assessments are increasingly mediated by software whose outputs may be probabilistic, non-transparent, or sensitive to changing data environments.

The rule of distinction requires reliable discrimination between lawful and unlawful objects of attack and, when persons are concerned, the identification of those targetable under IHL. In structured defensive environments—such as point defense against incoming munitions—this may be relatively tractable. In more open and dynamic contexts, however, distinction is not a matter of simple visual classification. It may require interpretive judgment about surrender, civilian status, direct participation in hostilities, or hors de combat situations. Those judgments are relational and contextual. They often depend on behavior,

environment, and rapidly evolving circumstances that do not map neatly onto generalized target profiles. This is one reason the ICRC has consistently argued that autonomous weapon systems used against persons raise acute legal and ethical concerns and should be subject to prohibition (ICRC, 2021, 2026; Asaro, 2012; Sharkey, 2012).

The proportionality rule presents an equally significant challenge. Proportionality is not a mathematical formula but a legal judgment requiring the attacker to weigh expected incidental civilian harm against anticipated military advantage. That assessment presupposes human valuation: the meaning of "excessive" is contextual, normative, and bound up with military judgment and legal responsibility. A system may calculate blast radius, model collateral effects, or estimate probabilities more rapidly than a human analyst. Yet the ultimate proportionality determination involves qualitative and contestable judgments that cannot be reduced to mere optimization without risking a deformation of the legal standard. To say this is not to deny the usefulness of computational assistance. It is to insist that computational assistance must remain nested within a structure of accountable human evaluation (Davison, 2018; Ekelhof, 2018; SIPRI, 2023).

The duty of precautions in attack is especially important for autonomous warfare because it operates as a bridge between general legality and practical conduct. Article 57 of Additional Protocol I requires attackers to do everything feasible to verify that targets are military objectives, choose means and methods that minimize incidental harm, and cancel or suspend attacks if it becomes apparent that the target is not lawful or the strike would be disproportionate. These duties imply a continuing relation between knowledge, control, and intervention. The more a system operates over extended time, across wider geographic space, or in conditions of changing civilian presence, the more difficult it becomes for a commander or operator to satisfy the requirement that all reasonably foreseeable strikes remain lawful throughout the weapon's period of activation. The ICRC's 2026 paper makes this point directly: commanders and other users of autonomous weapon systems must ensure that any and all possible strikes during the system's active period comply with IHL, accounting for reasonably foreseeable changes in circumstances (ICRC, 2026).

Finally, the debate cannot be reduced to black-letter rules alone. Concerns about humanity and public conscience, long associated with the Martens Clause, continue to shape the normative field. European and international discussions repeatedly return to the intuition that delegating life-and-death choices to machine processes is not merely a technical matter of performance. It raises deeper questions about dignity, moral agency, and the public legitimacy of organized violence. Not all such concerns generate immediate positive law, but they help explain why the debate has moved toward proposals for new legal limits even while states continue to affirm the applicability of existing IHL (ICRC, 2021; European Parliament, 2018; Council of Europe Parliamentary Assembly, 2023).

2.3. Article 36 weapons review and its promise

Article 36 of Additional Protocol I is often invoked as the doctrinal anchor for new technologies in warfare, and rightly so. It provides that "[i]n the study, development, acquisition or adoption of a new weapon, means or method of warfare," a state party must determine whether its employment would be prohibited in some or all circumstances by the Protocol or any other applicable rule of international law (Additional Protocol I, 1977, art. 36). This provision embodies a preventive logic. Rather than waiting for unlawful harm to occur, states must assess new weapons ex ante. In principle, this is a powerful mechanism for handling innovation without constant treaty revision.

For autonomous or AI-enabled systems, Article 36 remains indispensable. It is the point at which legal review should engage not only kinetic effects but also software behavior, operational envelopes, sensor reliability, data provenance, model adaptiveness, explainability, and the conditions under which human supervision can remain effective. It is also the doctrinal site at which reviewers should ask whether a system's effects can be sufficiently understood, predicted, and explained. In that sense, Article 36 review offers the best existing legal foothold for integrating technical assurance into the law of armed conflict (Davison, 2018; SIPRI, 2023; UNIDIR, 2025).

Yet Article 36 is not a complete solution. It is a framework obligation, not a detailed review protocol. States vary considerably in how they implement it, and many do not disclose their review methodologies. More fundamentally, Article 36 is designed to assess whether a weapon can lawfully be fielded, not to substitute for the attack-specific judgments required during operations. A system may be lawful for one narrowly bounded context and unlawful when used in another. A legal review may also become stale if the system is updated, retrained, networked with other software, or repurposed. For deterministic conventional weapons, that problem is manageable. For software-intensive systems whose performance may depend on data, machine-learning behavior, or distributed integration, one-off review is plainly inadequate. A central claim of this paper is therefore that Article 36 review must be reconceived as iterative, lifecycle-based, and evidentially robust if it is to remain meaningful for autonomous warfare.

2.4. Command responsibility and the architecture of attribution

If Article 36 addresses ex ante lawfulness, command responsibility and related doctrines address ex post attribution. The Rome Statute provides that a military commander is criminally responsible for crimes within the Court's jurisdiction committed by forces under his or her effective command and control where the commander knew or, owing to the circumstances, should have known of the crimes and failed to take all

necessary and reasonable measures to prevent or repress them or submit the matter for investigation (Rome Statute, 1998, art. 28). This doctrine reflects a foundational principle: military responsibility is inseparable from command.

In the context of autonomous or AI-enabled weapon systems, this principle remains crucial. A weapon is not a legal subject, and it cannot bear criminal responsibility. Machines do not replace human legal obligation. This is the central insight of both ICRC and SIPRI analyses, and it animates Kraska's important argument that military accountability remains pervasive even when AI is used in warfare (ICRC, 2021; Kraska, 2021; SIPRI, 2022). The law's addressees remain human beings and states.

Nevertheless, the applicability of command responsibility does not settle the accountability problem. Responsibility in law is not only about identifying a norm holder; it is also about demonstrating control, knowledge, foreseeability, and failure in relation to a specific harmful outcome. When the causal chain of harm passes through opaque models, adaptive software, contractor-supplied code, or coalition interoperability layers, the evidentiary path from commander to unlawful result becomes more difficult to establish. That difficulty does not abolish responsibility, but it can render the doctrine less effective, less fair, or less predictable in practice. One of the paper's core arguments is that autonomous warfare puts pressure on the surrounding infrastructure that makes command responsibility workable: logging, auditability, technical literacy, testing, supervision, and clear allocation of authority. Without that infrastructure, formal doctrine risks becoming either symbolic or excessively blunt.

Table 1. Core IHL principles and autonomy-specific stress points.

IHL principle or doctrine	Classical function in attack law	Autonomy-specific stress point	Implication for accountability
Distinction	Requires attacks to be directed only at lawful military objectives and combatants.	Automated target-recognition functions may rely on probabilistic classifications, incomplete sensor data, or environmental assumptions that are difficult for operators to verify in real time.	Unlawful harm becomes harder to attribute when neither the operator nor the commander can fully reconstruct why the system treated an object or person as a lawful target.
Proportionalit	Prohibits attacks	Machine-assisted	The evidentiary basis for the

IHL principle or doctrine	Classical function in attack law	Autonomy-specific stress point	Implication for accountability
y	expected to cause excessive incidental civilian harm relative to the concrete and direct military advantage anticipated.	estimations may compress qualitative judgments into model outputs whose assumptions are not transparent to commanders.	proportionality judgment may be under-documented or technically opaque, weakening both ex ante review and ex post investigation.
Precautions in attack	Requires feasible verification of targets, choice of means and methods that reduce civilian harm, and cancellation or suspension when circumstances change.	Autonomous functions may continue to operate after the battlefield context has changed, while supervision burdens exceed human monitoring capacity.	Responsibility turns on whether intervention was realistically possible and whether the system was designed with meaningful abort, override, and logging features.
Article 36 review	Obliges states to assess the legality of new weapons, means, or methods of warfare during study, development, acquisition, or adoption.	Software updates, retraining, sensor changes, or new data environments may alter system behavior after initial review.	A one-time legal review is insufficient unless review is refreshed across the lifecycle and tied to technical change management.
Command responsibility	Attributes criminal responsibility to commanders who knew or should	Knowledge may depend on dispersed technical evidence, vendor disclosures, or audit logs that	Formal authority remains, but practical attribution weakens unless legal command is matched by

IHL principle or doctrine	Classical function in attack law	Autonomy-specific stress point	Implication for accountability
	have known of crimes and failed to prevent or punish them.	are unavailable or uninterpretable at the operational level.	technical traceability and institutional support.

Source: Author's synthesis based on Additional Protocol I, Article 36 review practice, and the accountability literature on autonomous weapon systems.

3. Why accountability becomes difficult in AI-enabled hostilities

3.1. From direct causation to distributed sociotechnical agency

Traditional legal analysis of weapon use often assumes a relatively linear chain: a state fields a weapon, a commander orders its use, an operator employs it, and a harmful result follows. The law can then assess whether the weapon was lawful, whether the attack complied with IHL, and whether commanders or operators failed in their duties. AI-enabled warfare complicates this chain by distributing agency across a wider network of actors, artifacts, and temporal stages. System behavior may depend on data-labeling decisions made months earlier, on model architecture choices made by software engineers, on integration decisions by defense contractors, on contextual parameter-setting by operators, on mission design by commanders, and on real-time environmental inputs beyond direct human perception. The effect is not that the machine becomes a subject of law. It is that the human contribution to the final lethal event becomes mediated, layered, and sometimes epistemically compressed.

Bode's work is especially useful here because it shows that human control is not a binary switch but a normatively loaded practice shaped over time by design, training, and operation (Bode, 2023). If military institutions normalize reduced human involvement in specific targeting decisions as "appropriate" or "normal," then the legal analysis cannot stop at the moment of trigger pull—or its algorithmic equivalent. It must ask how that normality was produced and whether it remains compatible with the structure of IHL accountability. This broader frame helps explain why debates about autonomous weapons so often seem to talk past one another. Technologists emphasize capability boundaries, lawyers emphasize formal rules, and policymakers emphasize strategic competition. The real accountability problem lies in the relation among these domains

3.2. Opacity, explainability, and the epistemic burden of lawful attack

The difficulty is magnified when systems incorporate forms of machine learning that are powerful but not readily interpretable. Even where outputs are statistically impressive, they may not be explainable in ways that support legal judgment. In civilian sectors, this problem is typically framed as one of transparency, fairness, or due process. In armed conflict, the stakes are more acute because lives may be lost before the basis of a decision can be reconstructed. NATO's revised AI strategy identifies explainability and traceability as principles of responsible use precisely because military institutions recognize that legal and operational confidence depend on more than raw performance (NATO, 2024).

The problem is not that every lawful attack requires full scientific comprehension of a system's internal mathematics. Commanders have long used complex weapons they could not personally engineer. The problem is rather whether the relevant military and legal decision-maker can understand the operational conditions, confidence limits, likely failure modes, and contextual assumptions of the system sufficiently to make an accountable judgment. If a target recommendation is generated by a model trained on datasets that encode bias, degraded battlefield imagery, or misleading signatures, then operator approval may amount to little more than ritual ratification. SIPRI's 2025 study on bias in military AI is instructive here: it shows that concerns about bias are not peripheral ethical complaints but integral to the legality of distinction, proportionality, and precautions (SIPRI, 2025). In legal terms, opacity is not merely a technical inconvenience. It is an epistemic burden that bears directly on whether attack decisions are genuinely reasoned and therefore attributable.

3.3. Temporal distancing and the problem of changing circumstances

Autonomous and AI-enabled systems can also generate a temporal distancing effect. A commander may authorize activation in one set of circumstances, while the system acts later in another. The longer the period between authorization and strike, the wider the operational area, and the more fluid the environment, the more attenuated the commander's situational awareness becomes. This matters because IHL obligations are not static. A strike lawful at one moment may become unlawful if civilians enter the area, if the target status changes, or if collateral expectations shift. Human supervision becomes less meaningful when the window for intervention is narrow or when the system's operational logic is not visible in real time.

Here again, the issue is not novelty for novelty's sake. Similar concerns existed with loitering munitions, automated air-defense systems, and other earlier technologies. But AI-enabled autonomy can intensify them by enabling greater persistence, faster reaction, broader search areas, and more complex pattern-recognition

functions. The ICRC's 2026 paper correctly emphasizes that commanders and users must ensure the lawfulness of all possible strikes throughout the time and space in which the system is active (ICRC, 2026). That requirement is doctrinally sound but operationally demanding. It implies a threshold of predictability and contextual constraint that many envisioned uses of autonomous systems may not satisfy.

3.4. Supply chains, contractors, and the externalization of military judgment

The distributed nature of military AI development adds a further difficulty. Unlike many traditional weapons, AI-enabled systems often depend on private-sector software, cloud infrastructure, model components, commercially sourced data, and rapid update cycles. The political economy of military AI is therefore structurally dual-use and transnational. A European state may deploy a system whose relevant subsystems were coded in one jurisdiction, trained on data assembled in another, integrated by a prime contractor in a third, and updated under coalition software pipelines thereafter. This reality creates both legal and practical friction.

From the standpoint of IHL, the state remains responsible for the weapons it fields and the attacks it conducts. Yet the evidentiary and governance pathways needed to assess failure become harder when the knowledge required to explain system behavior is dispersed across contractors and supply-chain actors not easily integrated into battlefield decision-making or post-strike investigation. Procurement law, export controls, cybersecurity obligations, and contract architecture thus become part of the accountability problem, even though they are not traditionally treated as core topics of the law of armed conflict. A serious European response must therefore connect IHL governance to procurement transparency, vendor disclosure duties, and traceable configuration management.

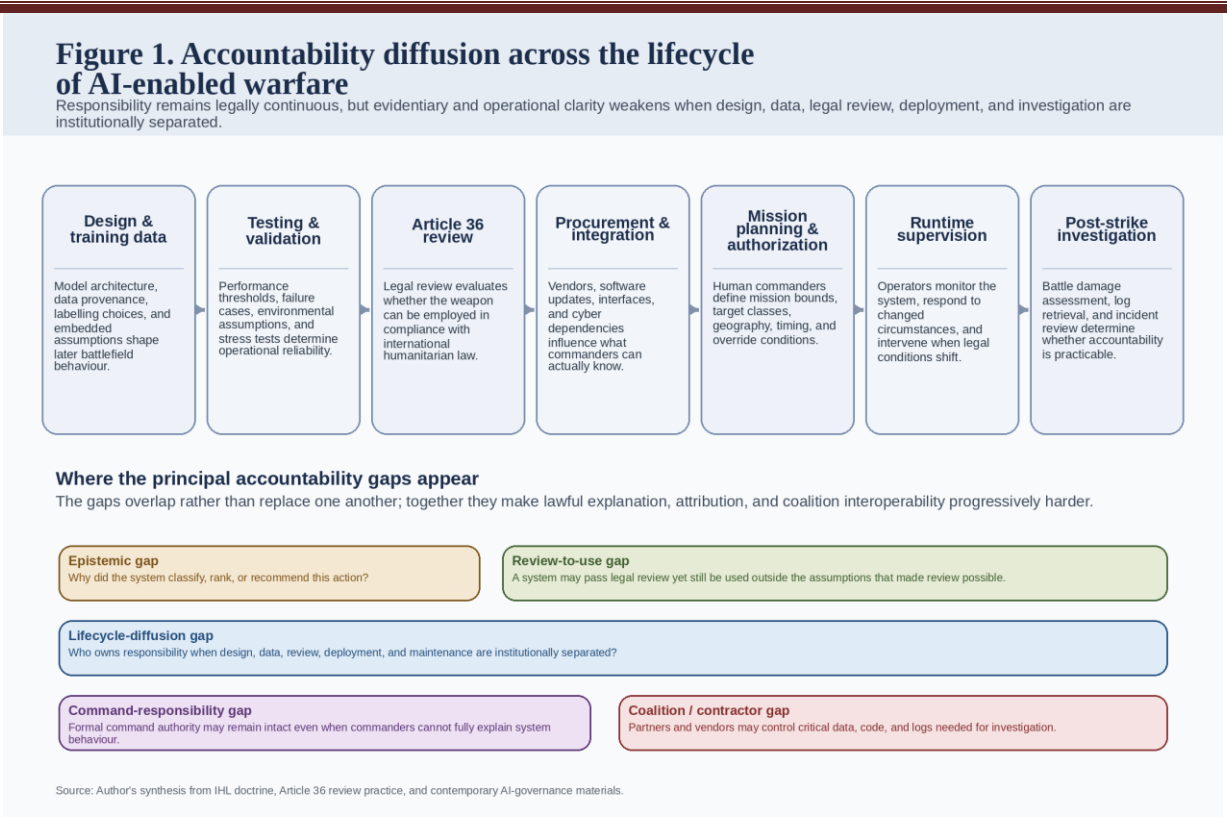

Figure 1. Accountability diffusion across the lifecycle of AI-enabled warfare.

Source: Author-generated synthesis.

4. The accountability gaps

Before turning to the specific gaps, one clarification is necessary. Some scholars have warned that "accountability gap" rhetoric can be overused. McDougall (2019), for example, cautions against assuming gaps before carefully identifying which legal regime is allegedly failing and in what way. This caution is salutary. The concept should not be used as a dramatic slogan to imply that autonomous warfare takes place in a lawless void. The claim advanced in this paper is more precise. The gaps identified below are not zones of total norm absence. They are sites where existing law allocates responsibility at a high level of abstraction but where the operational, evidentiary, or institutional means needed to implement that responsibility remain underdeveloped

4.1. The epistemic gap

The first accountability gap is epistemic. IHL presupposes that those who employ force possess a level of understanding adequate to make and defend legal judgments. They need not know every internal mechanism

of a weapon, but they must understand enough about the system's expected behavior, performance boundaries, and failure conditions to assess lawfulness. With AI-enabled systems, that understanding can become fragile. A commander may be told that a model has passed technical validation, but not know whether its performance degrades in urban clutter, adverse weather, multi-language civilian signage, or camouflage-rich environments. An operator may receive a confidence score without knowing what training assumptions underlie it. A lawyer may review a system after laboratory testing while remaining unable to assess model drift after subsequent field updates.

The consequences are profound. If the decision-maker cannot explain why the system's output was trustworthy in the operational context, accountability risks collapsing into deference to technical authority. The chain of responsibility remains formally human, but substantively it becomes opaque. This is especially problematic for proportionality and precautions, where legal compliance turns on reasoned judgment under uncertainty. In such settings, explainability and traceability are not optional ethical add-ons. They are conditions for meaningful attribution. Europe's repeated invocation of human-centric AI governance will remain rhetorically attractive but legally shallow unless it is translated into concrete duties of technical comprehensibility for military AI systems (European Parliament, 2021; NATO, 2024; SIPRI, 2025).

4.2. The lifecycle-diffusion gap

The second gap is lifecycle diffusion. Responsibility relevant to autonomous warfare is distributed across the entire lifecycle of the system: research, design, data selection, testing, legal review, procurement, doctrine, training, mission planning, activation, supervision, battle damage assessment, and incident investigation. Yet legal discourse still tends to concentrate either on ex ante weapons review or on ex post battlefield conduct. What falls between them is a broad field of decisions that materially shape whether unlawful harm becomes likely, foreseeable, or difficult to investigate.

This lifecycle perspective matters because some of the most consequential legal risks are created long before deployment. Choices about target ontology, labeling standards, excluded scenarios, retraining authority, system overrides, and log retention may determine whether later investigations can reconstruct what happened. Likewise, post-strike processes matter because accountability depends on preserving machine-readable evidence, event logs, operator inputs, sensor streams, and model-version histories. If these are absent, the state may still be internationally responsible in principle, but meaningful fact-finding and individual liability become significantly harder. What is needed, therefore, is a legal conception of responsibility that does not wait until the moment of attack to become relevant.

4.3. The review-to-use gap

The third gap lies between ex ante weapons review and ex post use. Article 36 review is essential, but it cannot answer every question required for lawful employment. A system might be lawful in principle when constrained to object targets in isolated maritime or anti-materiel environments, yet unlawful when used against persons or in mixed civilian settings. It may also be lawful only if accompanied by certain safeguards—short activation windows, narrow geofencing, constant communication links, robust override capability, and pre-mission scenario testing. The problem is that such contextual conditions are not always translated into binding operational doctrine or real-time command practice.

Accordingly, a system may pass legal review but later be used outside the conditions that made the review plausible in the first place. This is not a speculative concern. The ICRC has repeatedly stressed that the lawfulness of autonomous weapons cannot be assessed in the abstract alone; it depends on predictable effects and tightly bounded contexts of use (ICRC, 2021, 2026). The review-to-use gap emerges when weapons review is treated as a compliance certificate rather than as the opening step in a continuing chain of constraint. Closing the gap requires Europe to reconceive Article 36 review as a dynamic process linked to doctrine, training, authorization thresholds, and post-deployment modification controls.

4.4. The command-responsibility gap

The fourth gap concerns command responsibility. Kraska's important intervention is correct to insist that military commanders remain directly and individually accountable for the methods and means of warfare they employ (Kraska, 2021). This principle must be preserved. But the normative assertion that commanders remain accountable does not by itself solve the evidentiary and fairness difficulties that autonomy introduces. Article 28 of the Rome Statute turns on effective control, knowledge or constructive knowledge, and the failure to take necessary and reasonable measures. In conventional settings, these elements may be assessed through orders, unit relationships, operational awareness, and command negligence. In AI-mediated settings, however, knowledge may be filtered through technical abstractions, contractor interfaces, or misleading confidence claims.

The consequence is not that command responsibility becomes obsolete. Rather, its practical application becomes vulnerable to both overextension and underenforcement. It may be overextended if commanders are held liable for system behavior they could not reasonably foresee or control. It may be underenforced if prosecutors or investigators cannot show what the commander actually knew about the system's risks or why

the warning signs should have been legible. A workable accountability regime must therefore specify the evidentiary and governance preconditions of command responsibility: documentation duties, technical briefings, red-team reports, mission logs, override standards, and clear chains of software authority. Without such scaffolding, the law risks oscillating between symbolic blame and practical impunity

4.5. The coalition-and-contractor gap

The fifth gap is increasingly central for Europe: coalition and contractor opacity. European security is deeply networked. Operations may involve NATO allies, EU defense-industrial cooperation, intelligence-sharing, joint procurement, and federated command structures. At the same time, military AI development is often outsourced or co-developed with private industry. This produces a specific accountability problem: the actor authorizing force is not always the actor who best understands the system, and the actor who best understands the system may not fall neatly within operational command structures. Divergent national review standards further complicate matters. A system reviewed in one state may be deployed in a multinational operation in which another state provides data, communications, or target libraries.

This creates opportunities for responsibility to be deflected laterally. After an incident, states may emphasize contractor fault, software integration problems, intelligence failure, partner-state data quality, or unexpected machine behavior. None of this removes state responsibility under international law. But it does create practical barriers to establishing what happened and who failed in what way. Europe's distinctive challenge is that its operational effectiveness depends on interoperability. It therefore cannot close accountability gaps solely through isolated national review procedures. It requires common baseline standards for traceability, interoperability certification, incident investigation, and contractual transparency.

Table 2. The five accountability gaps in AI-enabled hostilities.

Accountability gap	Core description	Typical manifestation in operations	Primary legal / institutional pressure point	Indicative European remedy
Epistemic gap	The decision path leading to a recommendation, classification, or strike outcome cannot be	Commanders rely on confidence scores or outputs they cannot meaningfully interrogate under time	Distinction, proportionality, and evidentiary sufficiency for individual	Require audit logs, model cards, data provenance, and explainability

Accountability gap	Core description	Typical manifestation in operations	Primary legal / institutional pressure point	Indicative European remedy
	adequately reconstructed or explained by responsible humans.	pressure.	responsibility.	thresholds proportionate to operational risk.
Lifecycle-diffusion gap	Responsibility-relevant choices are distributed across design, testing, review, procurement, deployment, and maintenance.	Different institutions control data, code, legal review, and mission settings; no single actor sees the whole chain.	Article 36 review, procurement governance, and post-incident investigation.	Adopt lifecycle assurance duties, mandatory re-review triggers, and chain-of-custody rules for technical evidence.
Review-to-use gap	A system that passed legal review is fielded in a context materially different from the assumptions that supported that review.	Systems validated for bounded environments are used under degraded communications, different target sets, or altered software configurations.	Feasible precautions and the continuing validity of weapons review conclusions.	Link operational authorization to configuration control, context restrictions, and re-certification after substantial change.
Command-responsibility gap	Commanders remain legally responsible in principle but lack the practical information needed to exercise, document, or defend lawful judgment.	Legal authority is intact, but the commander cannot access or understand the technical basis for a system's behaviour.	Rome Statute Article 28 and operational doctrines of control and supervision.	Invest in legal-technical literacy, clear override doctrines, and staff structures that translate technical

Accountability gap	Core description	Typical manifestation in operations	Primary legal / institutional pressure point	Indicative European remedy
				evidence into command-relevant knowledge.
Coalition / contractor gap	Critical information needed for prevention or investigation is controlled by partner states, vendors, or external service providers.	A coalition strike relies on allied sensor data or proprietary vendor code unavailable to investigators after civilian harm.	Interoperability, evidence retention, and access to records needed for attribution.	Create coalition standards for evidence access and procurement clauses that defeat black-box opacity in military AI.

Source: Author's synthesis.

5. The European security context and the regulatory paradox

5.1. Why Europe cannot rely on minimal compliance

The accountability problem is especially acute in Europe because the continent's security institutions are simultaneously accelerating military AI adoption and affirming a human-centric legal identity. The European Parliament Research Service has noted that AI-powered defense innovation is advancing through European Defence Fund and PESCO projects and that AI in warfare raises concerns about accountability, IHL compliance, and conflict escalation due to reduced human oversight (EPRS, 2025). NATO's revised 2024 AI strategy likewise seeks to accelerate adoption while reaffirming six principles of responsible use: lawfulness, responsibility and accountability, explainability and traceability, reliability, governability, and bias mitigation (NATO, 2024). These developments mean that Europe is not merely a forum for abstract norm entrepreneurship. It is becoming a domain of actual military AI integration.

This matters because Europe's strategic predicament encourages speed. Democracies facing attritional threats and rapid technological competition are tempted to treat legal governance as something that can be backfilled once deployment is already underway. That temptation is dangerous. In the context of autonomous warfare, governance cannot be treated as an after-market add-on, because responsibility depends on ex ante design choices and institutional record-keeping. If Europe waits until incidents occur before clarifying accountability, it will inherit both the operational risks of poorly governed systems and the reputational costs of normatively inconsistent practice. Minimal compliance is therefore not only legally risky. It is strategically self-defeating.

5.2. Europe's human-centric rhetoric and military exclusion

Europe's regulatory discourse on AI is often framed in human-centric terms, but military AI exposes the limits of that paradigm. The Council of Europe's 2024 Framework Convention on Artificial Intelligence establishes principles of human dignity, transparency and oversight, and accountability and responsibility for adverse impacts; yet matters relating to national defense fall outside its scope (Council of Europe, 2024). The EU AI Act similarly stands as the world's most ambitious horizontal AI regulation, but military, defense, and national security uses are excluded from its direct application (European Union, 2024; EPRS, 2025). In parallel, the European Parliament has insisted that AI must remain subject to human control and be correctable or disableable in the event of unforeseen behavior (European Parliament, 2021).

Taken together, these instruments reveal a regulatory paradox. Europe's general AI governance philosophy strongly supports the principles most relevant to autonomous warfare—oversight, accountability, transparency, human autonomy—yet the military domain remains underregulated precisely where the stakes are highest. One can understand the institutional reasons for this. Defense remains closely tied to national sovereignty, and general-purpose civilian AI regulation is not easily transplanted into military contexts. Yet the result is a normative asymmetry: Europe articulates strong cross-sector principles while leaving military AI to a looser combination of national policy, alliance norms, and generic IHL. That asymmetry is no longer sustainable. The strategic integration of AI into defense has outpaced the institutional courage required to govern it.

5.3. Diplomatic momentum: from the UNGA to the 2025 CCW process

The broader international picture reinforces this conclusion. In December 2024, the United Nations General Assembly adopted Resolution 79/62 on lethal autonomous weapons systems with overwhelming support—

166 votes in favor, 3 opposed, and 15 abstentions—signaling that concern about LAWS has moved beyond specialist disarmament circles into the center of multilateral diplomacy (Perrin, 2025). The resolution followed the earlier Resolution 78/241 and reflects growing momentum for a more structured international response (United Nations General Assembly, 2023, 2024).

At the CCW level, the 2025 Group of Governmental Experts continued work on a set of elements of an instrument, without prejudging its nature, to address emerging technologies in the area of lethal autonomous weapon systems (United Nations Office for Disarmament Affairs [UNODA], 2025a). Particularly significant was the joint statement delivered in September 2025 on behalf of thirty-nine High Contracting Parties, which declared that the revised rolling text constituted a sufficient basis to fulfil the GGE's mandate in its current form (UNODA, 2025b). This does not mean that agreement on a treaty is imminent. But it does show that the debate has shifted from whether a normative framework is needed to what kind of instrument and what content it should contain.

The ICRC and the UN Secretary-General have pressed in the same direction. Their joint call for new prohibitions and restrictions on autonomous weapon systems by 2026 framed the issue as an urgent humanitarian priority and underscored the need for a legally binding instrument (ICRC, 2026; Perrin, 2025). For Europe, this diplomatic context matters because it creates an opportunity for leadership. The continent's states are well represented among those pressing for progress, yet European influence will remain limited if it is not backed by a coherent regional accountability model that can function both as policy and as example.

5.4. Why Europe should prefer a layered accountability model

Europe's comparative advantage does not lie in outpacing authoritarian competitors at any cost. Its long-term advantage lies in trusted, interoperable, and legally legitimate military innovation. In coalition warfare, trust is a strategic capability. Partners must be able to rely on one another's systems, review procedures, and incident-investigation standards. Public legitimacy also matters. Democratic states cannot sustainably wield increasingly autonomous force while appearing unable to explain, investigate, or attribute harmful outcomes. A layered accountability model therefore serves both normative and strategic ends. It protects civilians, strengthens coalition cohesion, disciplines procurement, and helps prevent the erosion of legal standards under the pressure of technological novelty.

Table 3. Selected governance instruments and their relevance to European accountability design.

Instrument or process	What it contributes	Why it matters for European security	Main limitation for autonomous warfare
ICRC position on autonomous weapon systems (2021) and selected issues paper (2026)	Articulates the need for new legally binding rules, including prohibitions on unpredictable systems and strong restrictions on other forms of autonomy.	Provides a precise humanitarian benchmark against which European policy can be measured.	ICRC guidance is authoritative but not binding on states absent treaty adoption or domestic implementation.
Council of Europe Framework Convention on AI (2024)	Reinforces transparency, oversight, accountability, and remedy as core AI-governance principles.	Confirms Europe's broader normative commitment to human-rights-centred AI governance.	National defence activities fall outside its scope, leaving military AI largely untouched.
EU AI Act (2024) and related parliamentary positions	Establishes a comprehensive civil AI governance model and affirms the principle of meaningful human control in public debates.	Shapes European regulatory expectations and procurement cultures for trustworthy AI.	Military uses are excluded, generating a gap between Europe's civilian AI philosophy and defence practice.
NATO AI strategies and principles of responsible use (2021; revised 2024)	Specify lawfulness, responsibility and accountability, explainability and traceability, reliability, governability, and bias mitigation.	Directly relevant to the interoperability environment in which many European forces operate.	These are policy principles rather than binding legal rules and do not by themselves resolve attribution in unlawful strikes.

Instrument or process	What it contributes	Why it matters for European security	Main limitation for autonomous warfare
UNGA resolutions on LAWS and CCW GGE process (2023–2025)	Create diplomatic momentum, consolidate state positions, and advance work toward an international instrument.	Offer a multilateral pathway for Europe to shape global standards rather than relying on fragmented national solutions.	Consensus remains incomplete, and negotiations may lag behind technological deployment.
National Article 36 reviews and procurement systems	Remain the primary formal site for ex ante legality assessment and risk management.	They are the most immediate tools available to European states while treaty processes continue.	Practices are uneven, often opaque, and insufficiently adapted to software-driven lifecycle change.

Source: Author's synthesis from ICRC, Council of Europe, European Union, NATO, UNGA, and CCW materials.

6. Toward a European accountability architecture

The preceding analysis suggests that Europe does not need a single magical concept—whether "meaningful human control," "human oversight," or "trustworthy AI"—to resolve the accountability problem. What it needs is an architecture. That architecture should operate across three layers: a substantive legal layer that identifies prohibited and restricted forms of autonomy; an operational assurance layer that translates law into lifecycle controls; and an institutional oversight layer that makes responsibility traceable across states, alliances, and contractors. These layers are complementary. Without substantive limits, operational assurance can become a technocratic rationalization for unacceptable uses. Without operational assurance, substantive legal norms remain too abstract. Without institutional oversight, both law and assurance lack durable enforcement pathways.

6.1. Layer one: substantive legal prohibitions and restrictions

The first layer should begin from a two-tier approach. Certain autonomous weapon systems should be prohibited because their effects cannot be sufficiently understood, predicted, or explained, or because they are designed or used to apply force against persons without context-appropriate human judgment. The

ICRC's 2021 and 2026 positions provide a persuasive normative baseline here: unpredictable systems and systems used to target human beings autonomously should be ruled out, while other systems should be strictly regulated through limits on targets, duration, geography, situations of use, and human-machine interaction (ICRC, 2021, 2026). Europe should support this two-tier model at the CCW and related forums rather than retreating into the ambiguous comfort of purely generic compliance language.

Substantive restrictions should also be more specific than current slogans allow. "Human control" is not meaningful if it is detached from operational content. At minimum, Europe should advocate a rule that the decision to employ lethal force through an AI-enabled or autonomous system must remain bounded by human determination of the mission objective, target class, geographic scope, temporal duration, and conditions for deactivation. Systems should not be permitted to operate indefinitely, across unbounded civilian-rich spaces, or with authority to select and engage persons based solely on generalized profiles. The legal category of prohibited use should include open-ended person-targeting, self-learning targeting behavior after deployment, and deployment in contexts where civilian presence or status cannot be reliably accounted for.

6.2. Layer two: operational assurance across the lifecycle

The second layer concerns operational assurance. Europe should treat Article 36 review as a lifecycle process rather than a single event. Reviews should be updated after substantial software modifications, retraining, new sensor integration, changes in intended operational context, or networking with new decision-support systems. Review protocols should require scenario-based testing, red-teaming, bias analysis, adversarial robustness evaluation, and explicit legal examination of foreseeable misuse or degradation. States should also maintain "configuration discipline": the exact model version, dataset lineage, safety parameters, and authorized mission envelope used in operations must be recorded and auditable.

Operational assurance should extend to use conditions. AI-enabled systems should be fielded only with clear override procedures, communication-failure contingencies, kill-switch functionality where technically feasible, mission time limits, and mandatory logging of critical decision events. Operators and commanders should receive not merely generic training but system-specific legal training focused on confidence limits, known failure modes, and the circumstances in which activation must be suspended or terminated. The point is not to demand impossible certainty. IHL already operates under conditions of uncertainty. The point is to ensure that uncertainty is managed by accountable human judgment rather than displaced onto opaque machine output.

A further element concerns evidentiary design. If accountability is to remain meaningful, systems must be designed for post-strike reconstruction. Europe should require secure retention of sensor inputs, operator interactions, model outputs, mission logs, and relevant system-health data for a legally appropriate period. In effect, military AI systems should have the equivalent of mission data recorders. The absence of such records should not be treated as a mere technical shortcoming; it should weigh against authorization of deployment in contexts where lethal consequences are foreseeable. This proposal mirrors a broader lesson from safety-critical domains: accountability depends not only on decision rules but on recordable, reviewable traces of how those decisions were produced.

6.3. Layer three: institutional oversight, interoperability, and investigation

The third layer is institutional. Europe should not rely on fragmented national practice alone. A credible regional framework would include at least four institutional mechanisms. First, a networked platform for national Article 36 review authorities should be established to share methodologies, good practices, and lessons learned in relation to military AI and autonomous systems. The aim is not to abolish national sovereignty over review, but to prevent opaque divergence and the emergence of weakest-link standards. Second, Europe should develop common interoperability and certification standards—ideally compatible with NATO practice—covering traceability, logging, override capability, software assurance, and incident reporting.

Third, Europe should create independent investigatory mechanisms for incidents involving autonomous or AI-enabled weapons. These mechanisms need not be supranational courts. They could take the form of multinational fact-finding panels, technical review boards, or interoperable national investigation standards. What matters is that incidents are not left to ad hoc internal military review without sufficient technical expertise or independence. Fourth, procurement and contracting rules must be incorporated into the accountability architecture. Vendors supplying AI-enabled military capabilities should be subject to disclosure duties concerning model updates, testing limitations, data provenance, explainability claims, and known failure conditions. Contract clauses should guarantee state access to the technical information required for legal review and post-incident investigation.

6.4. A European roadmap

How, concretely, might this architecture be advanced? A realistic roadmap would include five steps. First, European states should consolidate a common negotiating position in the CCW around a two-tier legally binding instrument. Second, national defense ministries should update weapons-review procedures to

include AI-specific review modules and mandatory review triggers for software and data changes. Third, EU and NATO institutions should develop shared technical baselines for auditability, traceability, and lifecycle assurance in coalition environments. Fourth, procurement policy should be revised so that "black box" vendor claims cannot be used to shield critical military AI systems from legal scrutiny. Fifth, Europe should invest in legal-technical literacy at the command level. The accountability problem cannot be solved by lawyers or engineers working in isolation; it requires institutions capable of integrating both.

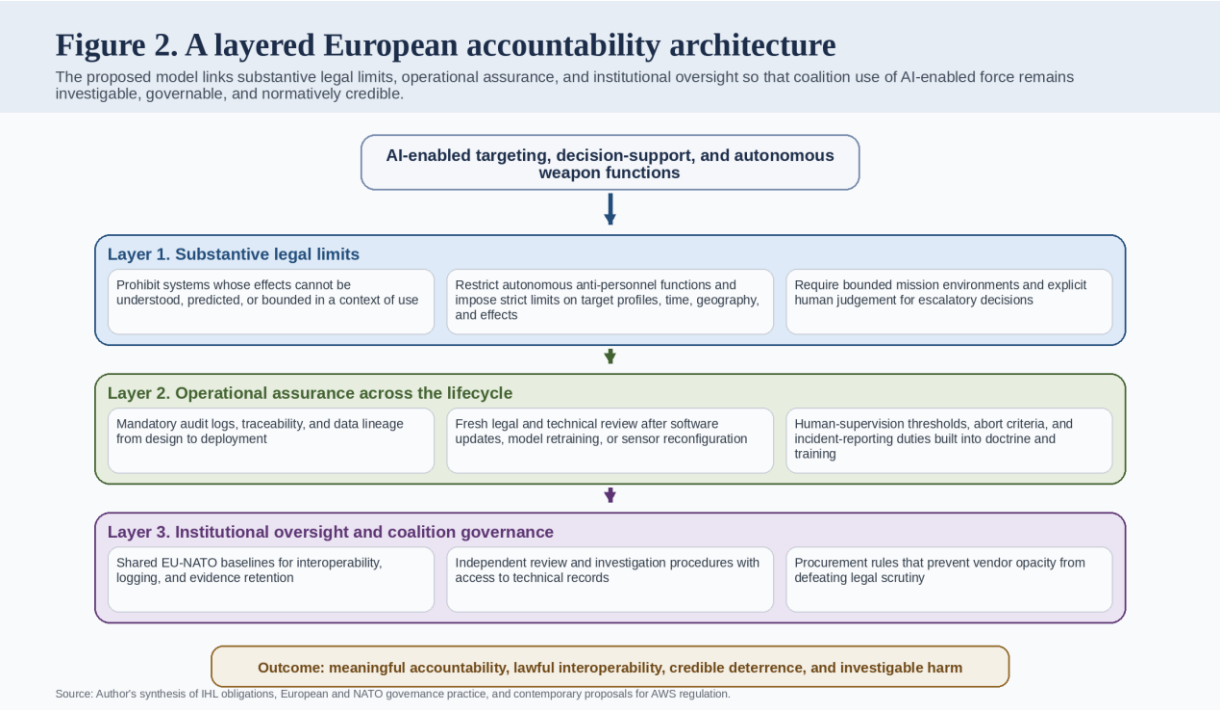

Figure 2. A layered European accountability architecture.

Source: Author-generated synthesis.

7. Objections and responses

7.1. "Existing law already suffices"

The most common objection is that no new regulation is required because IHL is technologically neutral and already applies to any weapon, old or new. This objection is partly correct and wholly insufficient. It is correct in the sense that autonomous or AI-enabled weapons do not fall outside the law. The principles of distinction, proportionality, precautions, and weapons review remain fully applicable. But technological neutrality does not mean doctrinal sufficiency in every practical respect. The fact that a rule applies does not mean that its

content is specific enough to govern new forms of distributed and opaque decision-making without supplementation. The contemporary debate, as UNIDIR has shown, is marked by persistent divergence regarding how exactly IHL applies to LAWS and what measures are required to ensure compliance (UNIDIR, 2025). Europe should therefore reject the false choice between "new law" and "no law." The real question is where clarification, codification, or new restriction is necessary to preserve the existing protective structure.

7.2. "Commanders are always accountable, so there is no gap"

A second objection, associated in a rigorous way with Kraska's work, holds that there is no accountability gap because commanders remain answerable for all means and methods of warfare used under their authority (Kraska, 2021). As a statement of legal principle, this is indispensable. Yet as an institutional design principle, it is incomplete. The problem is not that the law lacks a person to hold accountable in theory. The problem is that the evidentiary and organizational conditions needed to make that accountability real may be absent. A commander cannot meaningfully discharge or be judged against responsibilities that the system architecture itself obscures. Nor is it normatively attractive to respond by imposing strict or quasi-strict criminal responsibility on commanders for harms that could not reasonably be known or prevented. The better response is to strengthen the upstream architecture—traceability, technical literacy, mission constraints, and investigation pathways—so that command responsibility can function as intended.

7.3. "Strong regulation will slow democratic innovation"

A third objection is strategic. In an era of intense competition, the argument goes, democracies cannot afford heavy regulatory burdens that adversaries will ignore. This concern cannot simply be dismissed. Legal governance that is indifferent to operational tempo will fail politically. Yet the premise is flawed. The choice is not between fast innovation and accountable innovation. Poorly governed military AI creates its own strategic liabilities: unreliable targeting, coalition distrust, procurement lock-in, public backlash, escalation risk, and legal exposure. Democracies derive part of their strategic strength from the legitimacy and interoperability of their military practice. Europe in particular cannot base durable security on systems that allies do not trust, courts cannot review, or publics cannot morally defend. Regulation that is intelligently designed to embed assurance and accountability is therefore not a handicap. It is a force multiplier for lawful coalition warfare.

7.4. "Prohibition and regulation are mutually exclusive"

A final objection presents prohibition and regulation as mutually exclusive options. On this view, one must either support a complete ban or accept case-by-case governance under existing law. The two-tier approach advanced here rejects this dichotomy. Some forms of autonomy are so incompatible with predictability, human judgment, or the protection of persons that prohibition is justified. Other systems may be used lawfully under strict constraints. This differentiated approach is more analytically rigorous and more politically realistic than either absolutist camp. It allows Europe to defend red lines while preserving space for genuinely governed innovation in narrowly bounded contexts such as anti-materiel or point-defense applications.

8. Conclusion

The debate over autonomous warfare is often framed as a contest between technological inevitability and humanitarian alarm. That framing is too shallow. The deeper issue is whether the law of armed conflict can continue to individualize and institutionalize responsibility when lethal force is mediated by increasingly complex AI-enabled systems. This paper has argued that the principal danger is not the disappearance of law but the dilution of accountable agency across a sociotechnical lifecycle. Existing IHL and international criminal law remain the indispensable baseline. Yet they do not, by themselves, specify the full set of operational and institutional conditions needed to preserve meaningful accountability under conditions of opacity, adaptation, distributed design, and coalition interoperability.

Five accountability gaps structure the problem: an epistemic gap, a lifecycle-diffusion gap, a review-to-use gap, a command-responsibility gap, and a coalition-and-contractor gap. These are not zones of normlessness. They are sites at which law's addressees remain human, but law's practical pathways of knowledge, proof, and control become increasingly fragile. That fragility matters because the credibility of IHL depends not only on substantive norms but on the real possibility of attributing unlawful harm to responsible actors through intelligible procedures.

Europe has both a strategic need and a normative opportunity to respond. Its defense institutions are integrating AI more deeply, its alliance structures depend on interoperability and trust, and its legal-political identity remains bound up with human dignity, accountability, and the rule of law. Yet the present European framework is paradoxical: robust human-centric rhetoric coexists with significant military exclusions and institutional fragmentation. The answer is not to abandon innovation, nor to retreat into vague invocations of

human control. It is to construct an accountability architecture that matches the reality of contemporary military AI.

That architecture should rest on three layers. First, Europe should support substantive legal prohibitions and restrictions, especially against unpredictable systems and systems that autonomously target persons without context-appropriate human judgment. Second, it should operationalize accountability through lifecycle-based review, scenario testing, traceability, logging, update controls, and legally informed training. Third, it should build institutional pathways for common review standards, procurement transparency, coalition certification, and independent investigation. Such an architecture would not solve every problem posed by autonomous warfare. But it would do something more fundamental: it would help ensure that as war becomes more computational, responsibility does not become more fictional.

Disclosure Statements:

- **Ethical approval and consent to participate:** Participation in the research was approved in accordance with the journal's guidelines.
- **Availability of data and materials:** All data and materials are available upon request.
- **Authors' contributions:** The authors are responsible for all aspects of the research, including content, analysis, methodology, and the final review.
- **Conflicts of interest:** The authors declare that there are no conflicts of interest related to the design, submission, or evaluation of this research.
- **Funding:** This research received no specific funding.
- **Acknowledgements:** The authors would like to express their sincere appreciation to the *Journal of Scientific Development "for Studies and Research" (JSD)* for its support and guidance (https://jsd.sdasmart.org).

References

Additional Protocol I to the Geneva Conventions. (1977). Protocol additional to the Geneva Conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (Protocol I), 8 June 1977.

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.

Bode, I. (2023). Practice-based and public-deliberative normativity: Retaining human control over the use of force. *European Journal of International Relations, 29*(4), 990–1016. https://doi.org/10.1177/13540661231163392

Boulanin, V., & Verbruggen, M. (2017). *Mapping the development of autonomy in weapon systems*. Stockholm International Peace Research Institute.

Council of Europe. (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225).

Crootof, R. (2015). The killer robots are here: Legal and policy implications. *Cardozo Law Review, 36*(5), 1837–1915.

Davison, N. (2018). A legal perspective: Autonomous weapon systems under international humanitarian law. In United Nations Office for Disarmament Affairs (Ed.), *Perspectives on lethal autonomous weapon systems* (UNODA Occasional Papers No. 30, pp. 5–18). United Nations.

Department of Defense. (2023). *Directive 3000.09: Autonomy in weapon systems*. U.S. Department of Defense.

Department of Defense. (2024). *Responsible artificial intelligence strategy and implementation pathway*. U.S. Department of Defense.

Ekelhof, M. A. C. (2018). Lifting the fog of targeting: "Autonomous weapons" and human control through the lens of military targeting. *Naval War College Review, 71*(3), 61–94.

Ekelhof, M. A. C. (2019). Moving beyond semantics on autonomous weapons: Meaningful human control in operation. *Global Policy, 10*(3), 343–348. https://doi.org/10.1111/1758-5899.12665

European Parliament. (2018). *European Parliament resolution of 12 September 2018 on autonomous weapon systems (2018/2752(RSP))*.

European Parliament. (2021). *European Parliament resolution of 20 January 2021 on artificial intelligence: Questions of interpretation and application of international law insofar as the EU is affected in the areas of civil and military uses and state authority outside the scope of criminal justice (2020/2013(INI))*.

European Parliament Research Service. (2025). *Defence and artificial intelligence*. European Parliament.

European Union. (2024). *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)*.

Heyns, C. (2013). *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions* (A/HRC/23/47). United Nations Human Rights Council.

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. (2024). *International humanitarian law and the challenges of contemporary armed conflicts: Building a culture of compliance for IHL to protect humanity in today's and future conflicts*. ICRC.

International Committee of the Red Cross. (2026). *Autonomous weapon systems and international humanitarian law: Selected issues*. ICRC.

Kraska, J. (2021). Command accountability for AI weapon systems in the law of armed conflict. *International Law Studies, 97*(1), 408–445.

McDougall, C. (2019). Autonomous weapon systems and accountability: Putting the cart before the horse. *Melbourne Journal of International Law, 20*(1), 58–87.

NATO. (2021). *Summary of the NATO Artificial Intelligence Strategy*. North Atlantic Treaty Organization.

NATO. (2024). *Summary of NATO's revised artificial intelligence strategy*. North Atlantic Treaty Organization.

Perrin, B. (2025). Lethal autonomous weapons systems & international law: Growing momentum towards a new international treaty. *ASIL Insights, 29*(1).

Rome Statute of the International Criminal Court. (1998). 17 July 1998.

Schmitt, M. N., & Thurnher, J. S. (2013). "Out of the loop": Autonomous weapon systems and the law of armed conflict. *Harvard National Security Journal, 4*, 231–281.

Sharkey, N. E. (2012). The evitability of autonomous robot warfare. *International Review of the Red Cross, 94*(886), 787–799.

Stockholm International Peace Research Institute. (2022). *Retaining human responsibility in the development and use of autonomous weapon systems: On accountability for violations of international humanitarian law involving AWS*. SIPRI.

Stockholm International Peace Research Institute. (2023). *Compliance with international humanitarian law in the development and use of autonomous weapon systems: What does IHL permit, prohibit and require?* SIPRI. https://doi.org/10.55163/DFXR3984

Stockholm International Peace Research Institute. (2025). *Bias in military artificial intelligence and compliance with international humanitarian law*. SIPRI. https://doi.org/10.55163/NLWV5347

United Nations General Assembly. (2023). *Resolution 78/241: Lethal autonomous weapons systems*.

United Nations General Assembly. (2024). *Resolution 79/62: Lethal autonomous weapons systems*.

United Nations Institute for Disarmament Research. (2025). *The interpretation and application of international humanitarian law in relation to lethal autonomous weapon systems* (background paper). UNIDIR.

United Nations Office for Disarmament Affairs. (2025a). *Convention on Certain Conventional Weapons—Group of Governmental Experts on Lethal Autonomous Weapons Systems (2025)*. United Nations.

United Nations Office for Disarmament Affairs. (2025b). *Joint statement by 39 high contracting parties to the September 2025 session of the CCW Group of Governmental Experts on Lethal Autonomous Weapons Systems*. United Nations.

Appendix A. Submission evaluation matrix (author self-assessment)

Note: *This appendix is included for author use in light of the requested evaluation rubric. It may be removed before submission if the venue does not permit appendices.*

Appendix Table A1. Self-assessment against the conference paper submission criteria.

Criterion	Self-rating	Reasoned assessment tied to the manuscript
Background	Excellent	The structured abstract succinctly states what is already known, identifies the unresolved accountability problem, and clearly positions the study's purpose.
Purpose	Excellent	The manuscript frames the research purpose systematically by identifying a concrete contemporary phenomenon: the diffusion of responsibility in AI-enabled hostilities and its implications for European security governance.
Research methodology	Excellent	Section 1.1 explains the doctrinal and policy-analytical design, the source base, and the relationship between legal interpretation and institutional practice.
Findings	Excellent	The paper develops five interlocking accountability gaps and presents them systematically across Sections 3–6, supported by tables and figures that synthesize the analysis.
Research limitations	Excellent	Section 1.2 and the abstract explicitly acknowledge the doctrinal nature of the study and the absence of classified operational or proprietary technical data.
Originality/value	Excellent	The manuscript moves beyond a simple ban-versus-sufficiency dichotomy and offers a layered European accountability architecture that contributes both conceptual novelty and policy utility.
Readability and writing style	Excellent	The paper is structured, numbered, and progressively argued. Tables and visual syntheses assist navigation without diluting scholarly depth.

Criterion	Self-rating	Reasoned assessment tied to the manuscript
Quality of English language	Excellent	The manuscript is written in formal academic English with cohesive transitions, terminological consistency, and explicit analytical signposting.
Topic suitability	Relevant	The subject aligns directly with the Law of Armed Conflict track and with the wider European security agenda around AI, autonomy, military governance, and accountability.
Decision	Minor revision	Substantively, the manuscript is strong. For strict compliance with the European Governance Lab call, however, it should be condensed into the venue's shorter policy-brief format and anonymization requirements should be checked before submission.

The self-assessment is provided to align the manuscript with the rubric supplied by the author. It should be removed if the submission venue requests a blind or fully anonymized manuscript.