International Journal of Law
www.lawjournals.org
ISSN: 2455-2194
Received: 29-12-2025, Accepted: 28-01-2026, Published: 12-02-2026
Volume 12, Issue 1, 2026, Page No. 266-269

Reassessing international legal norms on autonomy and accountability in the laws of war in the age of Artificial Intelligence

Muthulakshmi A
Department of International Law and Organization, Tamilnadu Dr. Ambedkar Law University, Tamil Nadu, India

Abstract
The advent of Artificial Intelligence (AI) in modern warfare signals a profound transformation in both the conduct of hostilities and the interpretation of legal norms governing armed conflict. Autonomous weapon systems, which can operate with varying degrees of human oversight, pose complex questions for international humanitarian law (IHL) particularly regarding the principles of distinction, proportionality, and military necessity. These technologies disrupt traditional accountability paradigms by diffusing responsibility across military commanders, states, and private developers, thereby challenging established doctrines of state responsibility and individual criminal liability. This paper interrogates whether the existing corpus of IHL, codified primarily in the Geneva Conventions and Additional Protocol I, remains adequate to address the ethical and legal dilemmas raised by AI-enabled hostilities. Through a doctrinal and comparative approach, it evaluates treaty provisions, customary norms, and interpretive mechanisms such as the Martens Clause and Article 36 weapons reviews. It further examines contested state practices and soft-law initiatives emerging from multilateral forums, notably the UN Group of Governmental Experts on Lethal Autonomous Weapons Systems. By situating this inquiry within contemporary conflicts including the Russia, Ukraine war and hostilities in Gaza. The paper highlights concrete manifestations of these challenges, such as the opacity of algorithmic targeting and the evidentiary hurdles confronting accountability mechanisms. Rather than advocating for wholesale normative overhaul, it advances a recalibration strategy: reaffirming enduring IHL principles while supplementing them with AI-specific safeguards, transparency obligations, and verification mechanisms. In doing so, the study contributes to an evolving discourse on aligning humanitarian imperatives with the realities of technologically mediated warfare.

Keywords: Artificial Intelligence, international humanitarian law, autonomous weapon systems, accountability, Geneva Conventions, state responsibility

Introduction

Artificial Intelligence (AI) has evolved from a theoretical concept into an operational cornerstone of modern warfare. In recent years, conflicts such as the Russia–Ukraine war (2022–present) and the Gaza hostilities (2023–2024) have demonstrated the deployment of AI-enabled drones, predictive targeting algorithms, and autonomous surveillance platforms capable of prioritising and, in some cases, engaging targets with minimal human oversight [1]. These technological advancements promise improved precision and reduced risk to combatants; however, they simultaneously challenge the normative bedrock of international humanitarian law (IHL). IHL, codified principally in the Geneva Conventions of 1949 and Additional Protocol I of 1977 (AP I), rests on the principles of distinction, proportionality, and military necessity [2]. These principles were drafted for human decision-makers and presuppose human judgment in assessing targets and incidental harm. The advent of AI-driven autonomous weapon systems raises pressing questions: can algorithms reliably distinguish between combatants and civilians? Do proportionality assessments conducted by machines account for humanitarian considerations? Most crucially, where does accountability lie when autonomous systems err in ways not foreseen by their human operators or developers [3].

Research Problem

This study interrogates whether existing international legal norms are sufficient to regulate autonomy and accountability in warfare as militaries increasingly adopt AI-enabled systems. It asks two interrelated questions:

1. Do the principles of IHL, conceived for human-centred warfare, remain adequate for regulating algorithmic decision-making?
2. How should responsibility be allocated among states, commanders, and private developers when harm arises from AI-driven targeting?

Objectives

The objectives of this research are threefold. First, it aims to reassess the adequacy of existing IHL norms for AI-enabled hostilities. Second, it seeks examine accountability mechanisms applicable to real-world conflicts where AI has been employed, particularly Ukraine and Gaza. Third, it endeavours to propose forward-looking legal and policy frameworks capable of balancing military innovation with humanitarian imperatives.

Principles and frameworks governing autonomy in warfare

Autonomy in weapon systems, particularly those enhanced by Artificial Intelligence (AI), refers to the degree of human involvement in their critical functions, most notably in processes of target selection and engagement. The International Committee of the Red Cross (ICRC) distinguishes three levels of autonomy: *human-in-the-loop*, where human approval is required for every action; *human-on-the-loop*, where human oversight enables intervention in real time; and *human-out-of-the-loop*, where once activated, the system functions entirely independently. This typology has become central to deliberations within the United

Nations Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS). Yet, consensus on what constitutes "meaningful human control" remains elusive, leaving states to interpret obligations inconsistently and complicating both regulation and accountability [4].

The core principles of International Humanitarian Law (IHL) codified in the four Geneva Conventions of 1949 and the 1977 Additional Protocol I (AP I) remain the legal foundation for assessing AI in warfare. The principle of distinction, enshrined in Article 48 of AP I, obliges belligerents to differentiate between civilians and combatants, and between civilian objects and military objectives. Closely linked is the principle of proportionality under Article 51(5)(b) of AP I, which prohibits attacks expected to cause incidental civilian harm excessive in relation to anticipated military advantage. These principles are complemented by military necessity, permitting only those measures indispensable for achieving legitimate military objectives, and by the obligation of precaution in attack under Article 57 of AP I, which requires constant care to spare civilians and civilian objects.

Equally important is Article 36 of AP I, which obligates states to conduct legal reviews of new weapons, means, or methods of warfare to ensure their compliance with IHL. In the context of AI, these reviews are indispensable but rarely transparent; few states have disclosed whether they conduct Article 36 reviews for autonomous weapons, raising concerns about accountability and verification [5]. Where treaty law remains silent, the Martens Clause in Article 1(2) of AP I stipulate that civilians and combatants remain under the protection of principles of humanity and the dictates of public conscience. This clause has gained renewed relevance as an interpretive safeguard for regulating emerging technologies such as AI, ensuring humanitarian considerations persist despite legal lacunae.

Accountability for violations involving AI-enabled systems is traditionally framed through two complementary legal doctrines. First, state responsibility, articulated in the Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA 2001), attributes liability to states for internationally wrongful acts committed by their organs or agents, including those using autonomous weapons (Articles 4–11). Second, individual criminal liability under the Rome Statute of the International Criminal Court (1998) encompasses war crimes such as indiscriminate attacks (Article 8) and extends command responsibility to military leaders who fail to prevent or punish violations by subordinates (Article 28). The opacity of AI decision-making challenges both frameworks: attributing wrongful acts becomes problematic when harm results from self-learning algorithms rather than direct human intent, raising difficult questions about foreseeability, control, and due diligence obligations.

Autonomous systems in contemporary warfare: ihl challenges from Ukraine and Gaza

1. Ukraine Conflict (2022–present)

The Russia–Ukraine conflict represents one of the most significant real-world testing grounds for AI-enabled military technologies. Both parties have deployed AI-powered drones and algorithmic targeting systems to conduct surveillance, reconnaissance, and strike operations. Russian forces have utilized *Lancet* loitering munitions and automated artillery correction systems, while Ukrainian forces have employed Western-supplied *Switchblade* drones and predictive software to enhance targeting efficiency. These developments have accelerated decision cycles on the battlefield, reducing the time between target detection and engagement to mere seconds.

This operational speed, while tactically advantageous, raises pressing legal and ethical concerns. Civilian harm reported in Mariupol and Bakhmut areas heavily affected by automated drone strikes has sparked questions about whether AI systems can reliably apply proportionality assessments as required by Article 51(5)(b) of Additional Protocol. The International Criminal Court (ICC), currently investigating alleged violations in Ukraine, faces novel evidentiary challenges: proving intent or knowledge under the Rome Statute becomes increasingly complex when decisions are informed by machine-learning algorithms rather than human orders.

Equally significant is the lack of transparency in Article 36 weapons reviews, which are legally mandated to ensure new weapons comply with IHL before deployment. Neither Russia nor Ukraine has publicly disclosed whether such reviews have been conducted for their AI-enabled systems, undermining both domestic accountability and international oversight. This opacity risks eroding confidence in humanitarian compliance and complicates future post-conflict investigations.

2. Gaza Hostilities (2023–2024)

The 2023–2024 hostilities in Gaza marked another watershed moment in the use of AI during armed conflict, particularly through Israel's deployment of the "Habsora" (Gospel) system. This platform reportedly integrates vast datasets including satellite imagery, electronic intercepts, and pattern-of-life analyses to generate prioritized target lists for air strikes [6]. Israeli authorities have asserted that such systems enhance precision and mitigate civilian harm; however, United Nations fact-finding bodies and human rights organizations have expressed concern over their opacity and the potential for algorithmic bias in densely populated environments.

Legally, the system's reliance on algorithmic outputs raises questions about meaningful human control. While human commanders reportedly authorize AI-generated targets, the speed and scale of operations suggest that oversight may be nominal rather than substantive, blurring the line between human judgment and machine autonomy [7]. This dynamic complicates the application of the principle of distinction under Article 48 of Additional Protocol I, as well as the precautionary obligations codified in Article 57. The absence of disclosed Article 36 reviews for Habsora similarly frustrates international scrutiny and hinders the assessment of compliance with IHL.

3. Synthesis of Case Findings

The case studies of Ukraine and Gaza reveals several converging trends. First, AI enhances military efficiency and operational tempo but simultaneously amplifies accountability gaps, especially where decision-making processes are opaque or adaptive. Second, transparency deficits particularly the failure to disclose weapons reviews or operational data impede meaningful evaluation of proportionality and distinction in practice. Finally, the fragmented state practice among major military powers underscores the urgent need for clarified international

norms, either through treaty development or interpretive guidance, to address the humanitarian challenges posed by AI-enabled warfare.

Accountability challenges in ai warfare

The incorporation of Artificial Intelligence into weapon systems disrupts long-established accountability frameworks in international humanitarian law (IHL). Traditional mechanisms for assigning responsibility state responsibility under the Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA 2001) and individual criminal liability under the Rome Statute of the International Criminal Court presume that humans remain the primary decision-makers in armed conflict. [8] AI, particularly in systems that learn and adapt beyond their initial programming, fragments decision-making across developers, commanders, and algorithms, thereby complicating attribution of wrongful acts.

1. The Attribution Problem

State responsibility under ARSIWA hinges on the attribution of conduct to state organs or agents (Articles 4–11). However, when harm is caused by autonomous systems that self-modify or operate semi-independently, establishing attribution becomes fraught. For instance, if an AI-enabled drone misidentifies a civilian convoy due to flawed training data, does liability attach to the programmer, the commanding officer, or the state deploying the system? International tribunals have yet to articulate clear standards for attributing actions involving algorithmic decision-making, leaving a normative gap [9]. This ambiguity risks enabling a form of "responsibility vacuum," where no actor is held accountable for harm inflicted by autonomous systems.

2. Evidentiary and Procedural Barriers

Prosecuting war crimes involving AI systems faces unprecedented evidentiary challenges. Machine-learning algorithms often function as "black boxes", providing outputs without transparent reasoning. In judicial proceedings, establishing mens rea intent, knowledge, or recklessness is critical under Articles 30 and 8 of the Rome Statute. Yet proving that a commander foresaw or should have foreseen unlawful outcomes becomes exceedingly difficult when algorithms evolve in unpredictable ways. Additionally, technical complexities impede forensic analysis; investigating authorities may lack the expertise to audit algorithms or access proprietary source code held by private defense contractors, further undermining accountability.

3. Command Responsibility and Human Oversight

The doctrine of command responsibility, articulated in Article 28 of the Rome Statute, imposes liability on military leaders who knew or should have known of subordinates' unlawful acts and failed to prevent or repress them. However, in AI-mediated warfare, the notion of a "subordinate" is blurred autonomous systems are neither human soldiers nor fully predictable tools. Commanders may struggle to understand how algorithmic recommendations are generated, leading to questions about whether "effective control," a precondition for liability, can be meaningfully exercised over AI systems. This doctrinal tension raises profound challenges for both deterrence and accountability.

4. Transparency and Article 36 Reviews

Transparency deficits further compound accountability concerns. Article 36 of Additional Protocol I obligate states to review new weapons to ensure compliance with IHL prior to deployment. However, few states have disclosed whether such reviews encompass AI-enabled systems, and none have made the methodologies or findings public. This lack of disclosure undermines international oversight and prevents meaningful peer review of compliance claims. Without standardized review mechanisms, states risk adopting inconsistent or even perfunctory approaches to evaluating AI's legality, further eroding confidence in humanitarian protection.

5. Corporate and Developer Responsibility

The growing role of private actors in developing military AI introduces a further layer of complexity. Although not traditionally subject to IHL obligations, private defense contractors and technology companies design algorithms that directly shape targeting decisions. The Arms Trade Treaty (2014) and principles of corporate social responsibility offer potential analogies for regulating such actors, but no binding framework currently exists. Scholars have proposed hybrid accountability models extending due diligence obligations to developers, yet these remain largely theoretical [10]. Bridging this regulatory gap is essential to ensure that accountability does not end at the military chain of command.

Normative recalibration and policy proposals

1. Clarifying Human Control Through Treaty Mechanisms

A critical priority is defining and operationalising "meaningful human control" a concept widely endorsed in diplomatic and scholarly debates yet lacking precise legal articulation. One approach is to adopt a new protocol to Additional Protocol I under the Geneva Conventions. This protocol could:

1. Mandate that all AI-enabled weapons maintain verifiable human oversight at key decision points.
2. Prohibit "human-out-of-the-loop" systems where lethal decisions are executed without real-time human intervention.
3. Require states to demonstrate compliance through Article 36 reviews and publish summaries of those assessments for transparency and peer review [11].

Such a protocol would mirror historical precedents where technological advancements (e.g., blinding lasers, anti-personnel mines) prompted supplementary treaties to reinforce humanitarian safeguards.

2. Strengthening Accountability Frameworks

Existing accountability regimes must be expanded and harmonised to address the distributed nature of AI decision-making:

1. **State Responsibility:** The Articles on Responsibility of States for Internationally Wrongful Acts (2001) could be supplemented with commentary clarifying attribution in cases where autonomous systems self-learn or operate unpredictably. Strict liability models, already familiar in environmental law, could inform accountability for malfunctions in AI weapons.

2. **Individual Criminal Liability:** The Rome Statute (1998) could be interpreted or amended to explicitly encompass commanders' obligations to oversee algorithmic systems, extending command responsibility (Art. 28) to include failures in supervising AI operations.

3. **Corporate and Developer Responsibility:** Normative frameworks could impose due diligence duties on private actors designing military AI, drawing on precedents from the Arms Trade Treaty (2014) and evolving business and human rights principles.

3. Enhancing Transparency and Verification

Transparency deficits surrounding AI-enabled weapons undermine compliance and erode public trust. Addressing this requires multilateral verification mechanisms akin to the International Atomic Energy Agency (IAEA) model:

1. States would submit AI weapon systems for independent technical review prior to deployment.
2. A registry of AI-enabled weapons could be maintained under the auspices of the UN Office for Disarmament Affairs, fostering confidence-building among states. [12]
3. Confidence-building measures (CBMs) could include voluntary data-sharing on system testing, failure rates, and civilian harm assessments.

4. Integrating Ethical AI Design Principles

Incorporating ethics-by-design into military AI is essential for aligning technology with humanitarian imperatives. This involves embedding bias mitigation, explainability, and safeguard protocols at the design stage, rather than relying solely on legal compliance at deployment. Ethical frameworks such as the OECD AI Principles (2019) and the EU AI Act (2023) could inform military adaptations, ensuring systems are auditable and accountable throughout their lifecycle.

Conclusion

The challenge posed by Artificial Intelligence (AI) in warfare lies not in its novelty as a weapon, but in how it disrupts long-standing assumptions about human judgment and accountability in the conduct of hostilities. The rise of autonomous systems demands a recalibration of interpretive approaches and institutional practices rather than wholesale legal reinvention.

The imperative is twofold. First, human control must be preserved and operationalised: ensuring that algorithms do not become de facto decision-makers in lethal targeting. This requires embedding oversight mechanisms within weapon design, operational doctrines, and review procedures, not as symbolic assurances but as enforceable obligations. Second, accountability frameworks must evolve to address the distributed nature of AI decision-making.

International law has historically adapted to disruptive technologies from chemical weapons to cyber operations through a combination of treaty development, interpretive guidance, and soft-law initiatives. The same incremental yet principled evolution is now required for AI in armed conflict. States, international organisations, and civil society must seize this moment to clarify and strengthen the normative architecture, ensuring that humanitarian imperatives remain central even as warfare becomes increasingly algorithmic.

References

1. Geneva Conventions (1949). Conventions I–IV relative to the protection of victims of war. Geneva, 12 August 1949.
2. Additional Protocol I (1977). Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I). Geneva, 8 June 1977.
3. Rome Statute of the International Criminal Court (1998). UN Doc A/CONF.183/9. Rome, 17 July 1998.
4. Articles on Responsibility of States for Internationally Wrongful Acts (2001). UN General Assembly Resolution 56/83, 12 December 2001.
5. Arms Trade Treaty (2014). UN General Assembly Resolution 67/234 B, 2 April 2013 (entered into force 24 December 2014).
6. International Committee of the Red Cross (ICRC). Commentary on the First Geneva Convention. Geneva: ICRC, 2016.
7. International Committee of the Red Cross (ICRC). Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons. Geneva: ICRC, 2021.
8. United Nations Group of Governmental Experts. Report on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. New York: United Nations, 2023.
9. United Nations Office for Disarmament Affairs (UNODA). Transparency and Confidence-Building Measures in Emerging Technologies. New York: UNODA, 2022.