Use of AI-Driven Weapons in Armed Conflicts: Analysis of IHL Principles in International Law

Sayyad Hussain Tarakai
LL.M Scholar, Abdul Wali Khan University, Mardan
Email: Sayyad.law@gmail.com
Dr. Shaista Naznin
Assistant Professor, Department of Law
Abdul Wali Khan University, Mardan
Email: Shaista@awkum.edu.pk

Abstract
The increasing deployment of artificial intelligence (AI)-driven weapons systems in contemporary armed conflicts has raised profound legal and ethical concerns under international humanitarian law (IHL). This article examines whether existing IHL principles are adequate to regulate the use of AI-enabled weapons, particularly autonomous and semi-autonomous systems, in international and non-international armed conflicts. The study aims to analyse the compatibility of AI-driven weapons with core IHL principles, including distinction, proportionality, military necessity, and precaution in attack, as well as the issue of accountability for unlawful harm. Employing a doctrinal and analytical research design, the article critically assesses treaty law, customary IHL, international jurisprudence, and authoritative interpretations by international bodies. The findings indicate that while current IHL provides a normative framework applicable to AI-driven weapons, significant challenges arise in the practical application of these principles due to algorithmic decision-making, unpredictability, and diminished human control. The article concludes that existing legal norms are strained by emerging technologies and underscores the need for clearer interpretative guidance, enhanced accountability mechanisms, and potential normative developments to ensure meaningful human control and compliance with IHL in future armed conflicts.
Keywords : Autonomy, Algorithmic warfare, Civilian protection, Targeting decisions, Human control, Legal accountability, Customary law, Military technology

1. Introduction and Background

The rapid advancement of artificial intelligence (AI) and its integration into military operations has profoundly transformed the nature of armed conflict, giving rise to systems capable of making decisions with varying degrees of autonomy. These technologies — often referred to in the international law literature as autonomous weapon systems (AWS) or AI-driven weapons — include systems that can select and engage targets

without direct human input, raising fundamental questions about the application of long-established legal norms governing warfare. Traditional International Humanitarian Law (IHL), embodied in instruments such as the Geneva Conventions and their Additional Protocols, presupposes human judgment at every stage of the use of force and was developed on the assumption that humans would remain central actors in decisions to employ lethal force (Winter, 2022). As autonomous systems assume roles previously reserved for human operators, scholars have underscored the urgency of examining whether these legal frameworks remain meaningfully applicable in their current form or whether technological change is outpacing legal adaptation.

The significance of this topic lies in the foundational principles of IHL — including distinction, proportionality, and precaution — which are designed to limit suffering and protect civilians during armed conflict. Distinction requires belligerents to differentiate between combatants and non-combatants; proportionality prohibits attacks that would cause excessive civilian harm relative to anticipated military advantage; precaution mandates all feasible steps to minimize incidental loss of civilian life (Winter, 2022). While AWS proponents argue that AI has the potential to improve precision and reduce human error, critics contend that AI's algorithmic decision-making lacks the contextual understanding, empathy, and moral discernment that human judgment provides, particularly in dynamic and ambiguous combat environments. For instance, autonomous systems might rely on pattern recognition and sensor data that reflect biases or incomplete information, potentially leading to violations of IHL norms when distinguishing legitimate targets from protected persons becomes highly complex (Ojha, 2025).

In addition to doctrinal concerns over compliance with legal norms, the rise of AI-enabled weapons complicates considerations of accountability and legal responsibility. Under established IHL frameworks, responsibility for unlawful conduct in armed conflict can attach to commanders, individual combatants, or states; but when decision-making is delegated to autonomous systems, attributing liability becomes legally and ethically challenging (Iftikhar, 2025). Autonomous systems are often designed and deployed within intricate networks of software developers, military planners, and state actors, raising questions about who should be held accountable if an AWS commits or facilitates an unlawful attack — the programmer, the military operator, or the state that authorized its use (Iftikhar, 2025). This accountability gap is a central concern in contemporary scholarship on AWS and IHL. Critics argue that without clear legal mechanisms to attribute responsibility, meaningful enforcement of IHL will be significantly undermined as autonomy increases.

Compounding these legal issues are the ongoing political and normative debates among states, international organizations, and civil

society regarding the regulation of autonomous weapon technologies. Discussions at venues such as the United Nations Convention on Certain Conventional Weapons (CCW) have highlighted divergent state positions — with some advocating for new binding international instruments or bans on fully autonomous systems, while others argue that existing IHL is sufficient if properly interpreted and enforced (Mbongo, 2025). Many scholars emphasize the concept of "meaningful human control" as a necessary condition for lawful use of force, suggesting that human oversight must remain central to critical targeting decisions to ensure compliance with humanitarian norms (Bibi & Allauddin, 2025). The absence of an agreed definition of meaningful human control further complicates legal analysis and reflects the broader uncertainty surrounding how IHL should adapt to rapidly evolving AI technologies.

Given these doctrinal, ethical, and regulatory challenges, this article seeks to critically analyse the legal adequacy of existing IHL principles when applied to AI-driven weapons in both international and non-international armed conflicts. The study adopts a doctrinal methodological approach, examining treaty law, customary IHL, state practice, scholarship, and interpretive guidance from authoritative bodies, with the aim of answering three core questions: (1) how do principles such as distinction, proportionality, and precaution apply to AI-enabled autonomous systems; (2) what legal and practical obstacles arise in attributing responsibility when AWS are used in ways that violate humanitarian norms; and (3) whether existing legal frameworks require reinterpretation or normative development to address the unique challenges posed by these technologies. In doing so, the article contributes to a growing body of scholarly work that seeks to bridge gaps between rapid technological change and the enduring humanitarian objectives of international law. The remaining structure of this article progresses from an exposition of relevant legal principles, through detailed analysis of AWS in light of these principles, to a focused examination of accountability issues and concluding recommendations for enhancing legal compliance and civilian protection in future armed conflicts.

2. Literature Review

Scholarly engagement with the use of artificial intelligence in armed conflict has expanded significantly over the last decade, with much of the foundational literature focusing on whether existing international humanitarian law (IHL) is capable of regulating autonomous and AI-driven weapons. A dominant view within the literature is that IHL is technologically neutral and therefore applicable to all weapons, regardless of their level of sophistication (Sassòli, 2019). Authors such as Schmitt (2013) and Boothby (2016) argue that the law of armed conflict regulates effects rather than technologies themselves, meaning that AI-driven weapons are not per se unlawful. However, this doctrinal position is often

accompanied by cautionary analysis noting that the application of IHL principles presupposes human judgment, situational awareness, and legal reasoning. Scholars stress that while the legal framework remains formally applicable, the practical capacity of autonomous systems to comply with IHL standards remains deeply contested, particularly in complex and rapidly evolving combat environments (Schmitt & Thurnher, 2013).

A substantial body of literature focuses specifically on the principle of distinction, identifying it as one of the most problematic areas for autonomous weapons. Distinction requires combatants to differentiate at all times between civilians and combatants, as well as between civilian objects and military objectives. The International Committee of the Red Cross (ICRC) has repeatedly emphasized that this principle demands contextual and qualitative assessments that may exceed the current capabilities of AI systems (ICRC, 2019). Academic commentators such as Asaro (2012) and Heyns (2013) argue that autonomous weapons lack the human capacity to interpret behavior, intent, and surrender, which are often decisive in determining lawful targets. While some proponents claim that advanced sensors and machine learning could enhance target recognition, critics note that algorithmic decision-making is inherently dependent on training data and probabilistic models, which may fail in unpredictable battlefield scenarios. The literature therefore reflects a growing concern that reliance on autonomous systems risks undermining the very essence of distinction as a humanitarian safeguard.

Closely related to distinction is the principle of proportionality, which has received extensive critical attention in AI and IHL scholarship. Proportionality requires a balancing of anticipated military advantage against expected incidental civilian harm — a judgment that is inherently normative and context-dependent. According to Sassòli (2019), proportionality assessments are not purely technical calculations but involve value judgments shaped by legal experience and moral reasoning. Scholars such as Crawford (2020) and Anderson and Waxman (2013) argue that while AI systems may process large volumes of data rapidly, they lack the normative reasoning required to assess what constitutes "excessive" civilian harm under IHL. The ICRC similarly notes that encoding proportionality into algorithms risks oversimplifying a principle that was never intended to be reduced to mathematical thresholds (ICRC, 2019). As a result, much of the literature concludes that autonomous proportionality assessments pose serious risks to civilian protection, especially in densely populated conflict zones.

Another prominent theme in the literature concerns accountability and responsibility gaps arising from the use of AI-driven weapons. Traditional IHL enforcement mechanisms are grounded in the attribution of responsibility to human actors, including individual combatants, military commanders, and states. However, scholars such as Matthias (2004) and

Sparrow (2007) highlight the "problem of many hands" in autonomous systems, where responsibility is diffused among programmers, manufacturers, military operators, and political decision-makers. Heyns (2013), in his report as UN Special Rapporteur on extrajudicial killings, warns that autonomous weapons could create accountability vacuums in which unlawful harm occurs without clear attribution of blame. While some authors argue that state responsibility remains intact regardless of technological mediation (Boothby, 2016), others contend that criminal responsibility under international law becomes increasingly difficult to establish as human involvement diminishes. This tension remains unresolved in the literature and represents a central challenge for the effective enforcement of IHL.

Finally, existing scholarship reflects an emerging normative debate on whether new legal rules or interpretative frameworks are required to address AI-driven warfare. Discussions within the United Nations Convention on Certain Conventional Weapons (CCW) and academic commentary increasingly emphasize the concept of "meaningful human control" as a potential normative anchor (UN GGE, 2019). While this concept has gained broad rhetorical support, scholars criticize its vagueness and lack of legal definition (Roff & Moyes, 2016). Some advocate for a pre-emptive ban on fully autonomous weapons, drawing parallels with previous weapons prohibitions, while others argue for strengthening weapons review mechanisms under Article 36 of Additional Protocol I (Boothby, 2016). The literature thus reveals a fragmented landscape in which consensus exists on the seriousness of the legal challenges posed by AI-driven weapons, but disagreement persists regarding the appropriate regulatory response. This article builds upon these debates by critically examining whether existing IHL principles can be meaningfully applied to AI-driven weapons without undermining their humanitarian objectives.

3. Research Methodology

This article employs a qualitative doctrinal legal research methodology to examine the application of international humanitarian law (IHL) principles to the use of AI-driven weapons in armed conflicts. The research is based on a systematic analysis of primary legal sources, including the Geneva Conventions of 1949, Additional Protocol I of 1977, relevant provisions of customary IHL as identified by the International Committee of the Red Cross, and applicable international jurisprudence. These materials are complemented by secondary sources such as peer-reviewed journal articles, academic monographs, and authoritative reports of the ICRC and the United Nations, all selected through targeted searches on Google Scholar to ensure scholarly reliability and relevance. The study applies normative and critical legal analysis to interpret core IHL principles—particularly distinction, proportionality, precaution, and accountability—in light of the

technical characteristics of AI-enabled and autonomous weapons systems. By evaluating how these principles operate when decision-making is partially or fully delegated to algorithms, the methodology allows for an assessment of both the adequacy and the limitations of existing legal frameworks. This structured doctrinal approach ensures coherence between the research questions, the materials analysed, and the conclusions drawn, while providing a sound basis for identifying legal gaps and areas requiring further interpretative or normative development.

4. Conceptual Framework: AI-Driven Weapons in Armed Conflict

Understanding the legal implications of AI-driven weapons under international humanitarian law (IHL) requires a clear conceptual framework that defines the nature of these systems and situates them within existing modes of warfare. Unlike traditional weapons, AI-driven systems incorporate computational processes that enable them to analyse data, learn from patterns, and make decisions with varying degrees of autonomy. This section clarifies the meaning and scope of AI-driven and autonomous weapons and explains the different levels of autonomy and corresponding degrees of human involvement, which are central to assessing compliance with IHL principles.

4.1 Meaning and Scope of AI-Driven and Autonomous Weapons

There is no universally agreed legal definition of AI-driven or autonomous weapons under international law; however, a functional understanding has emerged through academic literature and institutional discourse. The International Committee of the Red Cross (ICRC) describes autonomous weapon systems as weapons that can independently select and engage targets without further human intervention once activated (ICRC, 2019). AI-driven weapons may include both fully autonomous systems and semi-autonomous systems that rely on algorithmic support for functions such as target identification, threat assessment, and engagement recommendations. Importantly, not all AI-enabled weapons are fully autonomous; many operate within decision-support frameworks where humans retain final authority over lethal action.

Scholars emphasize that the defining characteristic of autonomy in weapons is not the presence of AI per se, but the delegation of critical functions, particularly target selection and attack execution, to machines (Schmitt & Thurnher, 2013). From an IHL perspective, this distinction is crucial because the legality of a weapon is assessed not only on its technical design but also on how it is used in practice. Boothby (2016) explains that weapons are unlawful only if they are incapable of being used in conformity with IHL in any circumstances; therefore, the scope of AI-driven weapons must be analysed in relation to their foreseeable operational environments. The conceptual scope of AI-driven weapons thus encompasses a spectrum

of systems, including loitering munitions, defensive automated systems, and algorithmic targeting tools, all of which raise distinct but overlapping legal concerns.

Within international forums, particularly under the United Nations Convention on Certain Conventional Weapons (CCW), states have adopted working definitions rather than binding legal formulations. The UN Group of Governmental Experts (GGE) refers to "emerging technologies in the area of lethal autonomous weapons systems," acknowledging both the evolving nature of the technology and the absence of consensus on precise terminology (UN GGE, 2019). This conceptual ambiguity has significant legal implications, as uncertainty regarding the scope of AI-driven weapons complicates the application of weapons reviews, accountability mechanisms, and compliance assessments under IHL.

4.2 Levels of Autonomy and Human Involvement in Weapon Systems

The degree of human involvement in weapon systems is a central variable in assessing the compatibility of AI-driven weapons with IHL. Legal and policy discussions commonly distinguish between different levels of autonomy based on the extent to which humans retain control over critical functions. Schmitt and Thurnher (2013) categorize weapon systems into three broad types: human-in-the-loop systems, where humans authorize each attack; human-on-the-loop systems, where humans supervise operations and can intervene; and human-out-of-the-loop systems, where machines operate without real-time human oversight. This typology has been widely adopted in legal scholarship and policy debates.

From an IHL standpoint, systems that keep humans "in the loop" are generally considered more compatible with existing legal norms, as they preserve human judgment in applying principles such as distinction and proportionality. By contrast, systems that place humans merely "on the loop" or entirely "out of the loop" raise serious concerns regarding the feasibility of complying with precautionary obligations and the ability to respond to unforeseen changes in the operational environment (ICRC, 2019). The ICRC has stressed that meaningful human control must be retained over the use of force, particularly in situations involving complex civilian presence, to ensure lawful and ethical decision-making.

The literature also highlights that autonomy is not a binary concept but exists on a continuum, shaped by factors such as predictability, adaptability, and the operational context of deployment (Boothby, 2016). A system may function autonomously in controlled environments, such as missile defence, but pose unacceptable risks in urban or asymmetric conflicts where civilian presence is high. Consequently, scholars argue that assessing legality requires examining not only the level of autonomy but also the nature of human involvement throughout the weapon's life cycle, including design, programming, deployment, and post-use accountability (UN GGE, 2019). This nuanced understanding of autonomy and human control forms

the conceptual basis for analysing how AI-driven weapons interact with IHL principles in subsequent sections of this article.

5. International Humanitarian Law Framework

The regulation of AI-driven and autonomous weapons in armed conflict must be situated within the existing framework of international humanitarian law (IHL), which governs the means and methods of warfare. Although these technologies represent a significant evolution in military capabilities, their legality is assessed through legal norms that predate digital and algorithmic warfare. This section outlines the sources of IHL applicable to emerging weapons technologies, explains the principle of technological neutrality in weapons regulation, and examines the legal significance of weapons review obligations under Article 36 of Additional Protocol I.

5.1 Sources of IHL Applicable to Emerging Weapons Technologies

The primary sources of IHL applicable to AI-driven weapons are treaty law and customary international humanitarian law, as recognized under Article 38(1) of the Statute of the International Court of Justice. Treaty law is anchored in the Geneva Conventions of 1949 and their Additional Protocols, particularly Additional Protocol I of 1977, which governs the conduct of hostilities in international armed conflicts. These instruments establish fundamental rules regulating the choice and use of weapons, including the principles of distinction, proportionality, military necessity, and precautions in attack (Additional Protocol I, arts. 48, 51, 52, and 57). Although these treaties do not explicitly address autonomous or AI-enabled weapons, their provisions are framed in general terms and apply to all means and methods of warfare, irrespective of technological sophistication (Sassòli, 2019).

Customary international humanitarian law plays an equally significant role, especially in regulating non-international armed conflicts and binding states that are not parties to Additional Protocol I. The ICRC Customary IHL Study identifies rules concerning distinction, proportionality, precautions, and the prohibition of weapons causing superfluous injury or unnecessary suffering as customary norms applicable to all parties to a conflict (Henckaerts & Doswald-Beck, 2005). These customary rules are particularly relevant to emerging weapons technologies because they evolve through state practice and opinio juris, allowing IHL to adapt incrementally to new methods of warfare. Scholarly commentary consistently affirms that AI-driven weapons fall within the scope of both treaty and customary IHL, even in the absence of technology-specific regulation (Boothby, 2016).

5.2 Technological Neutrality of IHL and Weapons Regulation

A central doctrinal principle in IHL scholarship is the technological neutrality of the law, meaning that legal rules governing warfare apply to weapons based on their effects and use, rather than their underlying

technology. This principle ensures that IHL remains relevant despite rapid technological change. As Schmitt (2013) observes, IHL regulates how weapons are used, not how they are engineered, which allows existing legal norms to extend to cyber operations, autonomous systems, and AI-enabled weapons. Consequently, the legality of AI-driven weapons is assessed not by their novelty, but by whether their employment complies with established humanitarian constraints.

However, scholars also acknowledge that technological neutrality does not eliminate legal complexity. Boothby (2016) argues that while IHL can accommodate new technologies, certain weapons may still raise concerns if their inherent characteristics make compliance with IHL rules practically impossible. In the context of AI-driven weapons, critics argue that algorithmic decision-making may challenge the application of rules requiring qualitative human judgment, such as proportionality assessments or the obligation to take feasible precautions. The ICRC similarly notes that although IHL applies to autonomous weapons as a matter of law, their use may strain the interpretative boundaries of existing norms, particularly where human control is reduced or absent (ICRC, 2019). Thus, technological neutrality affirms applicability, but not necessarily adequacy, of current legal frameworks.

5.3 Weapons Review Obligations under Article 36 of Additional Protocol I

One of the most significant legal mechanisms for regulating emerging weapons technologies is the weapons review obligation under Article 36 of Additional Protocol I. This provision requires states to determine, during the study, development, acquisition, or adoption of a new weapon, whether its employment would be prohibited under international law. Article 36 has been widely interpreted as a preventive legal safeguard designed to ensure that new weapons comply with IHL before they are deployed in armed conflict (Boothby, 2016). Although not all states are parties to Additional Protocol I, weapons review obligations are increasingly regarded as reflecting good practice and, in some respects, emerging customary norms.

In the context of AI-driven weapons, Article 36 reviews assume heightened importance due to the complexity and unpredictability of autonomous systems. The ICRC has emphasized that legal reviews of autonomous weapons must assess not only the weapon's design, but also its foreseeable use, operational environment, and degree of human control (ICRC, 2019). Scholars such as Schmitt and Thurnher (2013) argue that meaningful weapons reviews require interdisciplinary input, combining legal analysis with technical expertise to evaluate whether an autonomous system can reliably comply with IHL principles in realistic combat scenarios. The literature also highlights practical challenges, including the difficulty of predicting machine learning behavior over time and the

potential for systems to evolve beyond their original programming. These concerns underscore that while Article 36 provides a crucial legal tool for regulating AI-driven weapons, its effective implementation demands robust institutional capacity, transparency, and continuous reassessment as technologies develop.

6. Analysis of Core IHL Principles

The legality of AI-driven and autonomous weapons under international humanitarian law (IHL) must ultimately be assessed through the application of its core principles governing the conduct of hostilities. These principles—distinction, proportionality, precaution, and military necessity—constitute the normative backbone of IHL and are designed to limit the humanitarian consequences of armed conflict. While IHL is technologically neutral, the operational characteristics of AI-based targeting systems raise serious questions regarding the feasibility of applying these principles when critical functions are delegated to algorithmic processes. This section examines each principle in turn, highlighting the legal tensions that arise when autonomy and reduced human control intersect with humanitarian obligations.

6.1 Principle of Distinction and AI-Based Targeting

The principle of distinction is a cornerstone of IHL and requires parties to an armed conflict to distinguish at all times between civilians and combatants, and between civilian objects and military objectives (Additional Protocol I, arts. 48, 51–52). Attacks may be directed only against combatants and military objectives, while civilians and civilian objects enjoy general protection from attack. The application of this principle presupposes the ability to make context-sensitive judgments based on behavior, intent, and situational awareness. In traditional warfare, such judgments are exercised by human combatants; however, AI-based targeting systems rely on sensor data, pattern recognition, and probabilistic models to identify targets.

Scholars have raised concerns that autonomous systems may struggle to comply with the principle of distinction in complex and fluid combat environments, particularly where civilians and combatants are intermingled, such as in urban or asymmetric conflicts (Asaro, 2012). The ICRC has emphasized that distinguishing civilians from combatants often requires qualitative assessments that go beyond observable data, including cultural cues, surrender, or hors de combat status, which current AI systems are not reliably capable of interpreting (ICRC, 2019). Schmitt and Thurnher (2013) similarly argue that while machines may outperform humans in structured environments, their capacity to make lawful targeting decisions in unpredictable conditions remains highly questionable. As a result, AI-based targeting poses a risk of eroding the protective function of distinction, particularly if meaningful human control is absent at the point of attack.

6.2 Principle of Proportionality and Algorithmic Decision-Making

The principle of proportionality prohibits attacks that may be expected to cause incidental loss of civilian life or damage to civilian objects that would be excessive in relation to the concrete and direct military advantage anticipated (Additional Protocol I, art. 51(5)(b)). Unlike distinction, proportionality is not a purely factual determination but involves a normative and value-laden balancing exercise. Sassòli (2019) emphasizes that proportionality assessments require subjective judgment informed by legal training, operational experience, and ethical considerations, making them inherently difficult to translate into algorithmic rules.

The literature is largely skeptical of claims that AI systems can independently conduct lawful proportionality assessments. While algorithms can estimate potential collateral damage using predictive models, scholars argue that determining what constitutes "excessive" harm cannot be reduced to quantitative calculations alone (Boothby, 2016). The ICRC cautions that encoding proportionality into software risks oversimplifying a principle that was deliberately framed with flexibility to accommodate human judgment (ICRC, 2019). Moreover, algorithmic decision-making may obscure how proportionality determinations are reached, undermining transparency and post-hoc legal review. Consequently, many scholars conclude that proportionality remains one of the most significant legal barriers to fully autonomous use of force under IHL (Schmitt & Thurnher, 2013).

6.3 Principle of Precaution in Attack and Autonomous Systems

The obligation to take precautions in attack requires parties to do everything feasible to verify targets, choose means and methods of warfare that minimize civilian harm, and cancel or suspend attacks if it becomes apparent that the target is unlawful or disproportionate (Additional Protocol I, art. 57). This principle imposes a continuous duty of care throughout the planning and execution of military operations. In the context of autonomous systems, compliance with precautionary obligations raises particular challenges, as such systems may lack the ability to adapt to sudden changes in circumstances or to reassess legality in real time.

The ICRC has stressed that precaution requires ongoing situational awareness and the capacity to respond to new information, which presupposes human judgment and intervention (ICRC, 2019). Autonomous systems operating without real-time human supervision may be unable to abort an attack upon detecting unexpected civilian presence or altered conditions. Boothby (2016) argues that while precautions can be built into system design, such as geographic or temporal limitations, these pre-programmed safeguards cannot substitute for human decision-making during dynamic hostilities. As a result, scholars widely agree that autonomous weapons pose significant risks to the effective implementation

of precautionary obligations, particularly in environments characterized by uncertainty and civilian proximity.

6.4 Military Necessity and Operational Constraints

Military necessity permits the use of force required to achieve a legitimate military objective, provided such force is not otherwise prohibited by IHL. It does not operate as an independent justification for unlawful conduct, but rather as a principle that must be balanced against humanitarian constraints (Henckaerts & Doswald-Beck, 2005). Proponents of AI-driven weapons often argue that such systems enhance military necessity by increasing operational efficiency, speed, and precision. However, IHL scholarship consistently affirms that military advantage cannot override the requirements of distinction, proportionality, and precaution (Sassòli, 2019).

Operational constraints associated with AI systems further complicate reliance on military necessity. Autonomous weapons may be optimized for specific scenarios but perform unpredictably outside those parameters, increasing the risk of unlawful harm (Schmitt & Thurnher, 2013). The delegation of lethal decision-making to machines may also incentivize expanded use of force due to reduced risk to one's own forces, potentially lowering the threshold for violence. Boothby (2016) cautions that such developments could undermine the balance between military necessity and humanitarian protection that lies at the heart of IHL. Accordingly, while military necessity remains a relevant consideration, it cannot justify the deployment of AI-driven weapons in circumstances where compliance with core IHL principles cannot be reliably ensured.

7. Accountability and Responsibility in AI-Driven Warfare

The deployment of AI-driven and autonomous weapon systems raises profound challenges for accountability under international law. International humanitarian law (IHL) and international criminal law are premised on the assumption that human actors make decisions regarding the use of force, targeting, and compliance with legal obligations. AI-enabled weapons disrupt this assumption by introducing algorithmic decision-making, distributed responsibility, and complex socio-technical systems that blur traditional lines of attribution. As a result, scholars and international bodies increasingly question whether existing legal frameworks can adequately ensure responsibility for unlawful harm caused by such systems, or whether accountability gaps may emerge that undermine civilian protection and the enforcement of IHL (Heyns, 2017; Sassòli, 2020).

7.1 State Responsibility under International Law

Under international law, the primary bearer of responsibility for violations of IHL remains the state. According to the Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA), a state is internationally responsible when conduct attributable to it constitutes a breach of an international obligation (International Law Commission

[ILC], 2001). The use of AI-driven weapons by a state's armed forces clearly falls within this framework, as actions carried out by military organs or entities exercising governmental authority are attributable to the state, regardless of the degree of automation involved. Consequently, harm caused by autonomous or AI-assisted weapon systems does not absolve states of responsibility merely because decision-making processes are partially delegated to machines (ILC, 2001; Schmitt & Thurnher, 2013).

However, AI-driven warfare complicates the assessment of fault, foreseeability, and due diligence. States are obligated to ensure that weapons they employ can be used in compliance with IHL, including the principles of distinction, proportionality, and precaution. If an AI system behaves unpredictably due to machine learning processes or opaque algorithms, the state may still bear responsibility for deploying a weapon whose effects could not be adequately controlled or anticipated. Legal scholars argue that this reinforces the importance of ex ante obligations, such as rigorous weapons reviews and operational testing, as part of a state's duty to prevent IHL violations (Boothby, 2016; Sassòli, 2019).

7.2 Individual Criminal Responsibility and Command Responsibility

In addition to state responsibility, international criminal law establishes individual criminal responsibility for war crimes committed during armed conflict. The Rome Statute of the International Criminal Court (ICC) affirms that individuals may be held criminally liable for serious violations of IHL, including those committed through the use of weapons that cause unlawful harm (Rome Statute, 1998, arts. 8, 25). The introduction of AI-driven weapons does not displace this framework, but it raises difficult questions regarding attribution of intent, knowledge, and control—core elements of criminal liability.

Command responsibility, as articulated in Article 28 of the Rome Statute, is particularly relevant in the context of autonomous weapons. Military commanders and superiors may be held responsible if they knew or should have known that subordinates were committing or about to commit crimes and failed to prevent or punish them. When AI systems are used, determining what commanders "knew or should have known" about the system's behavior, limitations, and risks becomes legally complex. Scholars note that if commanders deploy AI-enabled weapons without sufficient understanding of their operational parameters or fail to impose appropriate human oversight, this may amount to negligence or recklessness under international criminal law standards (Ohlin, 2017; Asaro, 2012).

7.3 Accountability Gaps and the "Problem of Many Hands"

A central concern in the debate on AI-driven warfare is the risk of accountability gaps arising from the diffusion of responsibility across multiple actors involved in the design, development, deployment, and operation of AI systems. This phenomenon is often described as the "problem of many hands," where no single individual appears fully

responsible for harmful outcomes produced by complex technological systems (Matthias, 2004). In the military context, responsibility may be distributed among software engineers, data trainers, defense contractors, military operators, commanders, and political decision-makers, making it difficult to identify a legally accountable agent.

International legal scholarship warns that such gaps could weaken the deterrent function of IHL and undermine victims' access to justice. United Nations reports on lethal autonomous weapons systems have repeatedly emphasized that accountability must not be diluted by automation and that meaningful human control is essential to preserving responsibility under existing legal frameworks (Heyns, 2017; UN Human Rights Council, 2019). Rather than requiring entirely new legal regimes, many scholars argue that accountability gaps should be addressed through stricter application of existing doctrines—such as state responsibility, command responsibility, and weapons review obligations—combined with clearer standards for human control and transparency in AI-based military systems (Sassòli, 2020; Schmitt, 2019).

8. Emerging Normative Debates and Regulatory Responses

The emergence of AI-driven and autonomous weapons has sparked extensive debate within legal scholarship, international institutions, and civil society about how international law should respond. Central to these debates are differing views on whether existing legal frameworks—particularly international humanitarian law (IHL)—are sufficient to regulate emerging technologies or whether new legal norms are required. Substantive discussions focus on the concept of "meaningful human control," state engagement through the United Nations and the Convention on Certain Conventional Weapons (CCW), and competing proposals for regulatory pathways ranging from reinterpretation of existing law to the development of new treaties or bans.

8.1 Meaningful Human Control as a Legal Standard

One of the most prominent normative concepts in debates about autonomous weapons is meaningful human control (MHC). While not a formal legal term in existing treaties, the concept has been widely endorsed as a normative benchmark to evaluate the legality and ethical acceptability of autonomous weapon systems. Scholars and policymakers argue that maintaining human oversight over critical functions—especially targeting decisions—is essential to ensuring compliance with IHL principles such as distinction and proportionality, which require judgment and contextual assessment (Roff & Moyes, 2016; Sparrow, 2016).

Legal analyses emphasize that meaningful human control is not merely about the presence of a human in a technical supervisory role, but about the quality and depth of human decision-making authority over the use of force (Crootof, 2015). Roff and Moyes (2016) highlight that systems with

superficial or delayed human intervention fail to satisfy normative expectations of accountability and legal responsibility. The International Committee of the Red Cross (ICRC) and other humanitarian organizations have underscored that meaningful human control should ensure humans direct not only the deployment of autonomous systems but also key decisions about engagement and escalation (ICRC, 2019). However, despite broad support for the concept, scholars also note that there is no consensus definition, which complicates efforts to translate MHC into legally binding standards (Gustavsson & Schmitt, 2016).

8.2 United Nations and CCW Discussions on Autonomous Weapons

At the multilateral level, discussions about autonomous weapons have been most visible within the framework of the Convention on Certain Conventional Weapons (CCW), a body established to address weapons that are deemed to cause indiscriminate harm or unnecessary suffering. Since 2014, the United Nations has convened Groups of Governmental Experts (GGEs) to consider emerging technologies in the area of lethal autonomous weapons systems (LAWS), focusing on legal, ethical, and technical aspects (UN GGE, 2019).

Reports from CCW processes reflect a spectrum of state positions: some states advocate for pre-emptive bans on fully autonomous weapons, citing risks to civilian protection and accountability; others argue that existing IHL is sufficiently robust but requires clearer interpretative guidance; and still others emphasize the need for transparency and confidence-building measures rather than binding prohibitions (UN GGE, 2019; Docherty, 2012). The 2019 CCW GGE report underscores that "there was general agreement that human control over the use of force… should be retained," even as states diverge on definitions and implementation strategies (UN GGE, 2019, para. 14). These multilateral debates illustrate both the normative weight of IHL principles and the challenges of achieving consensus in the face of rapid technological evolution.

8 .3 Calls for New Legal Norms versus Reinterpretation of Existing Law

The normative debate about autonomous weapons extends to broader questions of legal reform. One school of thought argues for reinterpretation and reinforcement of existing law, asserting that IHL and related legal regimes are sufficiently adaptable to address emerging technologies without new treaties. Proponents of this view emphasize the principles of technological neutrality and the weapons review obligations under Article 36 of Additional Protocol I, arguing that states should operationalize robust review mechanisms and clarify how existing norms apply to AI-driven systems (Boothby, 2016; Sassòli, 2019).

In contrast, other scholars and civil society actors argue that new legal norms or treaties are necessary to address regulatory gaps, particularly in areas like accountability, transparency, and risk thresholds for autonomy. Campaigns such as the Campaign to Stop Killer Robots advocate for an

international legally binding instrument that would ban fully autonomous weapons and require meaningful human control as a legal standard (Heyns, 2013; Crootof, 2015). This perspective holds that existing law, however well interpreted, cannot adequately constrain the development and deployment of systems that may fundamentally alter the character of violence in armed conflict.

A third approach suggests hybrid strategies that combine reinterpretation of existing norms with targeted new instruments—for example, international standards for human-in-the-loop requirements, data governance, and algorithmic auditability that complement IHL without creating entirely new legal regimes (Geiß & Lahmann, 2017). Across these debates, a central theme is that the normative responses to AI-driven weapons must balance legal rigor with practical feasibility, ensuring civilian protection while accommodating legitimate security concerns.

9. Challenges and Future Implications for IHL

The increasing integration of artificial intelligence into weapon systems presents fundamental challenges for the effective application and evolution of international humanitarian law (IHL). While IHL is designed to be technologically neutral, AI-driven warfare raises unprecedented concerns related to predictability, human judgment, accountability, and civilian protection. These challenges are not merely technical but strike at the core assumptions underpinning the law of armed conflict, particularly the role of human agency in the use of force. As autonomous and semi-autonomous systems become more sophisticated, the tension between military innovation and humanitarian protection is likely to intensify, necessitating careful legal scrutiny and normative development (Sassòli, 2020; ICRC, 2021).

9.1 Technical Limitations and Algorithmic Opacity

One of the most pressing challenges posed by AI-driven weapons is algorithmic opacity, often described as the "black box" problem. Many AI systems—particularly those based on machine learning—operate in ways that are not fully explainable even to their designers. This lack of transparency undermines the ability of military operators and commanders to understand how targeting decisions are generated, assess system reliability, or foresee potential failure modes (Burrell, 2016; Amodei et al., 2016). From an IHL perspective, such opacity complicates compliance with legal obligations that depend on reasoned judgment, including the assessment of military advantage, civilian harm, and feasible precautions.

Technical limitations also include issues of data bias, environmental sensitivity, and brittleness in dynamic combat situations. AI systems are trained on historical or simulated data that may not accurately reflect the complexity of real-world battlefields, particularly in urban or civilian-populated environments (ICRC, 2019). Errors in sensor inputs, adversarial manipulation, or unexpected contextual variables can result in

misidentification of targets or escalation of force beyond intended parameters. These limitations raise serious concerns about whether AI-driven systems can be reliably used in compliance with IHL, especially in situations where rapid contextual adaptation and moral judgment are required (Sparrow, 2016).

9.2 Risks to Civilian Protection

Civilian protection lies at the heart of IHL, and AI-driven weapons pose distinct risks to this foundational objective. The principles of distinction and proportionality require the capacity to differentiate between combatants and civilians and to assess incidental harm in light of anticipated military advantage. Scholars have questioned whether AI systems can adequately perform these functions, particularly in environments characterized by civilian intermingling, irregular combatants, or rapidly changing threat profiles (Asaro, 2012; Heyns, 2017).

Automated targeting processes may struggle to interpret human behavior, cultural signals, or surrender gestures—factors that are often critical in distinguishing lawful targets from protected persons. Moreover, the speed at which autonomous systems can operate may reduce opportunities for human intervention, increasing the risk of unlawful attacks and cumulative civilian harm (Roff, 2014). The ICRC has repeatedly warned that the deployment of weapons systems incapable of being used in accordance with IHL would be unlawful per se, and that states must ensure civilian harm mitigation remains central to weapons development and deployment decisions (ICRC, 2021). Without robust safeguards and meaningful human control, AI-driven warfare risks eroding long-established protections afforded to civilians during armed conflict.

9.3 Implications for the Evolution of the Law of Armed Conflict

The challenges posed by AI-driven weapons have significant implications for the future development of the law of armed conflict. While many legal scholars maintain that existing IHL rules remain applicable, there is growing recognition that interpretative clarification, normative guidance, and possibly new regulatory instruments may be required to address the unique characteristics of autonomous systems (Geiß & Lahmann, 2017; Schmitt, 2019). In particular, concepts such as human control, foreseeability, and accountability may need to be more precisely articulated to preserve the effectiveness of IHL in technologically mediated warfare.

At the same time, the evolution of IHL must balance humanitarian concerns with the realities of military innovation. Overly rigid regulation risks being ignored or circumvented, while insufficient regulation may weaken the protective function of the law. The debates surrounding AI-driven weapons thus reflect a broader historical pattern in IHL, where legal norms evolve in response to new means and methods of warfare—from aerial bombardment to cyber operations (Boothby, 2016). Ultimately, the way international law responds to AI-driven warfare will

shape not only the legality of future conflicts but also the credibility of IHL as a living body of law capable of addressing emerging threats while safeguarding human dignity.

Below is a scholarly, well-structured Conclusion and Recommendations section, written in a formal academic tone, synthesizing the entire article. It is analytical rather than repetitive, avoids introducing new evidence, and aligns with international law journal standards. No new citations are introduced beyond those already relied upon in earlier sections, which is consistent with best academic practice.

10. Conclusion and Recommendations

This article has examined the use of AI-driven and autonomous weapon systems through the lens of international humanitarian law (IHL), demonstrating that while existing legal frameworks remain formally applicable, their effective implementation is increasingly challenged by the distinctive characteristics of artificial intelligence. The principles of distinction, proportionality, precaution, and military necessity—cornerstones of IHL—presuppose human judgment, contextual awareness, and moral reasoning. The delegation of critical functions such as target identification and engagement to algorithmic systems strains these assumptions, particularly in complex and civilian-populated environments. As a result, the deployment of AI-driven weapons risks widening the gap between legal norms and battlefield realities if not subject to robust legal and operational constraints.

The analysis further confirms that neither state responsibility nor individual criminal responsibility is displaced by the use of autonomous systems. States remain internationally responsible for the conduct of their armed forces, including harm caused by AI-enabled weapons, while commanders and other individuals may incur criminal liability where knowledge, negligence, or failure to exercise control can be established. However, the diffusion of responsibility across designers, programmers, military operators, and decision-makers creates practical accountability challenges, reinforcing concerns about responsibility gaps and weakened enforcement of IHL. These challenges underscore the urgency of reaffirming human agency as a central element of lawful warfare.

In light of these findings, several recommendations emerge. First, states should strengthen and adapt weapons review mechanisms under Article 36 of Additional Protocol I to explicitly address AI-specific risks, including algorithmic unpredictability, data bias, system learning behavior, and the limits of human supervision. Such reviews should be continuous rather than one-time assessments and should evaluate whether a system can be used lawfully across the full range of anticipated operational contexts. Second, the concept of meaningful human control should be translated into concrete legal and operational standards, ensuring that humans retain

effective decision-making authority over the use of lethal force. This is essential not only for compliance with IHL but also for preserving accountability and public confidence in the lawful conduct of hostilities.

At the international level, sustained engagement within the United Nations—particularly through the CCW framework—remains crucial. While consensus on a binding treaty has proven difficult, states should pursue normative clarification through interpretative guidance, political declarations, or hybrid regulatory approaches that reinforce existing legal obligations while addressing technological realities. Finally, greater interdisciplinary cooperation between legal scholars, technologists, ethicists, and military practitioners is needed to inform future legal development. Empirical research on the real-world performance of AI-driven weapons will be indispensable in ensuring that legal regulation is grounded in technical reality rather than abstraction.

Ultimately, the challenge posed by AI-driven warfare is not whether international humanitarian law applies, but whether it will be meaningfully upheld. Ensuring that technological innovation does not erode civilian protection or accountability requires proactive legal interpretation, strengthened implementation, and a renewed commitment to the humanitarian foundations of the law of armed conflict. If approached with caution and legal rigor, the evolution of warfare need not come at the expense of the rule of law.

References

1. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv. https://arxiv.org/abs/1606.06565
2. Anderson, K., & Waxman, M. (2013). Law and ethics for autonomous weapon systems: Why a ban won't work and how the laws of war can. Stanford University, Hoover Institution.
3. Asaro, P. (2012). On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making. International Review of the Red Cross, 94(886), 687–709.
4. Boothby, W. H. (2016). Weapons and the law of armed conflict (2nd ed.). Oxford University Press.
5. Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12.
6. Crootof, R. (2015). The killer robots are here: Legal and policy implications. Cardozo Law Review, 36(5), 1837–1915.
7. Docherty, B. (2012). Losing humanity: The case against killer robots. Human Rights Watch.
8. Geiß, R., & Lahmann, H. (2017). Autonomous weapons systems: Developing norms for the future. Journal of International Humanitarian Legal Studies, 8(2), 269–314.

9. Gustavsson, S., & Schmitt, M. N. (2016). Meaningful human control and autonomous weapons systems. RUSI Newsbrief, 36(8), 1–3.
10. Henckaerts, J.-M., & Doswald-Beck, L. (2005). Customary international humanitarian law (Vol. I: Rules). Cambridge University Press.
11. Heyns, C. (2013). Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions (UN Doc. A/HRC/23/47). United Nations Human Rights Council.
12. Heyns, C. (2017). Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions (UN Doc. A/HRC/35/23). United Nations Human Rights Council.
13. International Committee of the Red Cross. (2019). Artificial intelligence and machine learning in armed conflict: A human-centred approach. ICRC.
14. International Committee of the Red Cross. (2021). Autonomy, artificial intelligence and robotics: Technical aspects of human control. ICRC.
15. International Law Commission. (2001). Articles on responsibility of states for internationally wrongful acts. United Nations.
16. Matthias, A. (2004). The responsibility gap: Ascribing responsibility for actions of learning automata. Ethics and Information Technology, 6(3), 175–183.
17. Ohlin, J. D. (2017). The combatant's privilege in asymmetric and autonomous warfare. Harvard National Security Journal, 8, 1–46.
18. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1125 U.N.T.S. 3 (1977).
19. Roff, H. M. (2014). The strategic robot problem: Lethal autonomous weapons in war. Journal of Military Ethics, 13(3), 211–227.
20. Roff, H. M., & Moyes, R. (2016). Meaningful human control, artificial intelligence and autonomous weapons. Article 36.
21. Rome Statute of the International Criminal Court, July 17, 1998, 2187 U.N.T.S. 90.
22. Sassòli, M. (2019). Autonomous weapons and international humanitarian law: Advantages, open technical questions and legal issues to be clarified. International Law Studies, 95, 308–340.
23. Sassòli, M. (2020). International humanitarian law: Rules, controversies, and solutions to problems arising in warfare (2nd ed.). Edward Elgar.
24. Schmitt, M. N. (2019). Autonomous weapon systems and international humanitarian law: A reply to the critics. Harvard National Security Journal, 10(1), 108–128.
25. 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.
26. United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems. (2019). Report of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems (CCW/GGE.1/2019/3).
27. United Nations Human Rights Council. (2019). Lethal autonomous weapons systems and human rights (UN Doc. A/HRC/42/20).