Algorithmic Impunity and the Accountability Crisis in International Humanitarian Law Under AI-Driven Autonomous Weapons Systems

David Greacy Geovanie[1], Dewa Gede Sudika Mangku[2], Ni Putu Rai Yuliartini[3]

[1,2,3]Universitas Pendidikan Ganesha, Indonesia
✉ Corresponding email: david.geovanie24@gmail.com

History of Article
Submitted : June 30, 2026
Revised : July 22, 2026
Accepted : August 10, 2026
Published : September 22, 2026
DOI : https://doi.org/10.37253/conescintech.v6i1.12750

Copyright© 2026 by Author(s). This work is licensed under a Creative Commons Attribution-Non Commercial-Share Alike 4.0 International License. All writings published in this journal are personal views of the authors and do not represent the views of this journal and the author's affiliated institutions.

Abstract
This study aims to (1) construct the concept of algorithmic impunity as a new analytical category within International Humanitarian Law (IHL), addressing the legal accountability vacuum produced by the deployment of AI-driven Lethal Autonomous Weapons Systems (LAWS) in contemporary military operations. (2) Specifically, it examines the failure of international criminal accountability doctrines to establish mens rea for harms caused by autonomous AI systems,and evaluates the adequacy of UN General Assembly Resolution 79/239 (2024) as a normative response to this crisis. A juridical-normative approach is employed through doctrinal and comparative legal analysis. Primary materials include IHL instruments, CCW GGE LAWS documents, relevant UN resolutions (Res. 79/62 and 79/239), and applicable international jurisprudence. The analysis examines three core IHL principles, distinction, proportionality, and precaution, in the context of LAWS operating without meaningful human oversight. The research identifies a structural accountability crisis in the current IHL regime stemming from three conditions: (1) causal anonymity between algorithmic decisions and lethal consequences; (2) the disintegration of traditional command chains due to machine autonomy; (2) and the absence of a binding legal instrument specifically governing LAWS. The study concludes that adequate normative response requires developing an algorithmic impunity doctrine within an additional IHL protocol, moving beyond meaningful human control toward an AI life-cycle accountability framework.

Keywords: Algorithmic Impunity; International Humanitarian Law; Lethal Autonomous Weapons Systems (LAWS); Accountability Crisis.

Introduction

The transformation of contemporary armed conflict is increasingly defined by the integration of artificial intelligence (AI) into the targeting cycle of military operations. Lethal Autonomous Weapons Systems (LAWS), weapons capable of selecting and engaging targets without direct, real-time human intervention, have moved from theoretical possibility to operational reality, with loitering munitions, AI-enabled decision-support systems, and semi-autonomous targeting platforms already deployed in active theatres of conflict. Reports on the use of algorithmic targeting tools in recent conflicts illustrate how AI systems increasingly mediate, and in some cases substitute for, human judgment in decisions that determine who lives and who dies on the battlefield (Malik et al., 2025; Oringa et al., 2026). This shift is not merely technical; it strikes at the conceptual foundations of International Humanitarian Law (IHL), a body of law built on the assumption that the actors who plan, order, and execute attacks are human beings capable of intent, foresight, and moral agency (Garcia, 2023).

The urgency of this issue is compounded by the accelerating pace of state investment in military AI. Major military powers continue to develop increasingly autonomous targeting and decision-support systems even as multilateral negotiations under the UN Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts (GGE) on LAWS remain deadlocked over even a shared definition of autonomy (Sharikov, 2024; Jackson, 2023). This regulatory paralysis stands in sharp contrast to the operational tempo of AI deployment, generating what several scholars have termed a widening "accountability gap": a structural condition in which harmful outcomes produced by autonomous systems cannot be attributed, with legal certainty, to any human or state actor (Chengeta, 2016; Ahmad Khalil et al., 2024). Where such a gap persists unaddressed and normatively unresolved, this article argues, it hardens into something more corrosive than a mere gap, an entrenched condition of algorithmic impunity, in which the very sophistication of the technology becomes a shield against legal responsibility rather than an incidental byproduct of imperfect regulation.

Existing scholarship has extensively mapped the theoretical dimensions of the responsibility gap produced by autonomous weapons. Chengeta (2016) examined four modes of accountability, individual, command, corporate, and state responsibility and concluded that only the exercise of meaningful human control (MHC) can meaningfully close the gap. Sio and van den Hoven (2018) subsequently offered a philosophical account of MHC grounded in "tracking" and "tracing" conditions, while Amoroso and Tamburrini (2020, 2021) developed a differentiated, principled, and prudential normative model assigning humans the roles of fail-safe actor, accountability attractor, and moral agency enactor. Parallel to this literature, a growing body of international criminal law scholarship has interrogated whether existing doctrines of mens rea under the Rome Statute can accommodate harms produced by self-learning systems (Bo, 2021; Gaeta, 2024; Skuratova et al., 2026), while state-responsibility scholars have examined whether the law of attribution under the Articles on State Responsibility can absorb AI-related violations without conceptual strain (Boutin, 2022; Pacholska, 2022).

Despite this substantial body of work, three gaps remain insufficiently addressed in the literature, which this study seeks to fill. First, most existing studies treat the accountability gap as a static, singular phenomenon rather than disaggregating it into the distinct causal, institutional, and normative conditions that produce it. Second, comparatively little scholarship has systematically evaluated the adequacy of the most recent normative development in this field UN General Assembly Resolution 79/239 (2024) on artificial intelligence in the military domain against the specific accountability failures identified in the doctrinal literature. Third, the concept of "algorithmic impunity" itself has not yet been constructed as a distinct analytical category within IHL scholarship, despite its explanatory value in capturing how legal, institutional, and technical factors converge to produce not merely a temporary lacuna but a durable pattern of non-accountability. This article addresses these gaps by (1) constructing algorithmic impunity as an analytical category, (2) examining the mens rea failure specific to AI-mediated attacks, and (3) assessing the normative adequacy of Resolution 79/239 and the broader CCW-GGE process.

The novelty of this research lies in its integrative treatment of three IHL principles distinction, proportionality, and precaution as an analytical lens through which the structural sources of algorithmic impunity are diagnosed, rather than treating each principle in isolation as prior studies have tended to do (Bruun et al., 2025; Woodcock, 2023; Dorsey, 2025). By synthesising doctrinal analysis of positive IHL instruments with the emerging soft-law architecture of the CCW-GGE and the most recent UN General Assembly resolutions, this study offers a more complete diagnostic of why the current legal regime has failed to keep pace with battlefield autonomy and proposes a life-cycle-based accountability framework that reaches beyond the now-familiar but definitionally contested concept of meaningful human control.

Accordingly, this study pursues two specific objectives. First, it seeks to construct the concept of algorithmic impunity as a new analytical category within IHL, capable of describing the accountability vacuum generated by AI-driven LAWS in contemporary military operations. Second, it examines the doctrinal failure of international criminal accountability mechanisms to establish mens rea for AI-mediated harm and evaluates the adequacy of UNGA Resolution 79/239 (2024) as a normative response. Theoretically, the study contributes a refined vocabulary for describing accountability failure in military AI governance; practically, it is intended to inform ongoing treaty-drafting efforts within the CCW-GGE process and to support policymakers, military legal advisers, and international organisations engaged in the regulation of autonomous weapons. The scope of this research is confined to the international legal dimension of LAWS accountability under IHL and international criminal law; it does not extend to domestic criminal law regimes, arms-export control law, or the ethics of autonomous weapons as a freestanding philosophical inquiry, save where these bear directly on the legal analysis.

Research Method

This research employs a juridical-normative (doctrinal) approach, appropriate for a study whose central problem concerns the coherence, adequacy, and internal logic of

existing international legal norms rather than the empirical measurement of social behaviour (Gunawan et al., 2022). The research design is descriptive-analytical and comparative: it describes the current normative architecture governing LAWS under IHL and international criminal law and then compares that architecture against the accountability challenges generated by autonomous targeting to assess its adequacy. This design is well suited to the research problem because the central question whether existing doctrines of individual, command, and state responsibility can accommodate harms produced by algorithmic decision-making is fundamentally a question of legal interpretation and systemic coherence rather than of quantifiable outcomes.

The object of this research is the corpus of primary and secondary legal materials bearing on the regulation of LAWS. Primary legal materials consist of core IHL instruments the four Geneva Conventions of 1949 and their 1977 Additional Protocol I, particularly Articles 35, 36, 48, 51, 57, and 58 governing means and methods of warfare, distinction, proportionality, and precaution; the Rome Statute of the International Criminal Court, particularly its provisions on individual criminal responsibility (Article 25), command responsibility (Article 28), and mental elements (Article 30); the Convention on Certain Conventional Weapons (CCW) and the reports and Guiding Principles produced by its Group of Governmental Experts (GGE) on LAWS since 2014; and United Nations General Assembly Resolutions 79/62 and 79/239 (2024), together with the Draft Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA). Secondary legal materials comprise relevant international jurisprudence, doctrinal commentary, and peer-reviewed literature on autonomous weapons, meaningful human control, and AI-related state and individual responsibility, drawn from leading international law and international criminal law journals published between 2015 and 2026.

Data collection was conducted through systematic doctrinal document analysis. Legal instruments and GGE documents were obtained from official United Nations and International Committee of the Red Cross (ICRC) repositories, while peer-reviewed secondary sources were identified and retrieved through the Consensus academic search platform, which indexes peer-reviewed literature across international law, ethics of technology, and security studies. Search terms included combinations of "autonomous weapons," "accountability gap," "meaningful human control," "mens rea artificial intelligence," and "state responsibility artificial intelligence," cross-checked against citation counts and journal quality indicators to prioritise materials with demonstrated scholarly influence and topical currency. Sources were screened for direct relevance to the accountability dimension of LAWS regulation under IHL and international criminal law, and for methodological or doctrinal rigour, before being incorporated into the analytical corpus.

Data were analysed using doctrinal and comparative legal reasoning. First, a statutory and treaty-interpretation analysis was undertaken to establish the content and scope of the distinction, proportionality, and precaution principles under Additional Protocol I, applying the interpretive framework of the Vienna Convention on the Law of Treaties (ordinary meaning, context, object and purpose). Second, a comparative

doctrinal analysis examined how the three principal accountability regimes recognised in international law individual criminal responsibility, command responsibility, and state responsibility respond when confronted with autonomous decision-making, drawing on the analytical frameworks developed by Chengeta (2016), Boutin (2022), and Pacholska (2022). Third, a normative-adequacy analysis evaluated UNGA Resolution 79/239 and the CCW-GGE Guiding Principles against the specific accountability failures identified in the first two stages of analysis, to determine whether the existing normative architecture is capable of closing, rather than merely acknowledging, the accountability gap. This analytical sequence moving from doctrinal description to comparative diagnosis, to normative evaluation ensures that the study's conclusions are systematically derived from, and traceable to, the primary legal materials examined.

Results and Discussions

Causal Anonymity and the Erosion of Mens Rea in Autonomous Targeting

The first structural condition underlying algorithmic impunity is what this study terms causal anonymity: the inability to draw a clear, legally cognisable causal line between a specific algorithmic decision and its lethal consequence. Under Article 30 of the Rome Statute, individual criminal responsibility for war crimes generally requires that the material elements of the crime be committed with intent and knowledge; even where international criminal law accepts lowered mental thresholds such as dolus eventualis or recklessness, a prosecutable link between a human actor's mental state and the resulting harm remains indispensable. Bo (2021) demonstrates that this requirement becomes exceedingly difficult to satisfy in human-machine teaming scenarios: where a semi-autonomous or AI-supported system participates in target identification, the human operator's awareness of the specific risk that a given attack will strike civilians is diluted by the system's own probabilistic, opaque decision architecture. The operator may know only that the system carries a general error rate, not that a specific strike carries an elevated risk of unlawful harm a form of knowledge that dolus eventualis doctrine was not designed to capture.

This difficulty is intensified by the so-called "black box" phenomenon: the practical and, in the case of deep-learning systems, sometimes principled opacity of the reasoning by which an AI system reaches a targeting recommendation. Skuratova et al. (2026) observe that this opacity, combined with the "problem of many hands" arising from the distributed nature of AI development spanning data scientists, software engineers, systems integrators, military planners, and commanding officers creates a legal vacuum in which technical autonomy is frequently, and wrongly, treated as equivalent to legal autonomy. No individual actor in this distributed chain possesses both the requisite knowledge of the system's specific targeting logic and effective control over the specific engagement decision, which is precisely the combination that mens rea doctrine presupposes. Sudhakar et al. (2026) similarly characterise this condition as a "responsibility imbalance," in which multiple parties bear some causal relationship to the

harm, yet none satisfies the demanding evidentiary threshold required to establish individual culpability.

Empirical illustrations of this problem have emerged from recent conflicts in which AI-assisted targeting platforms were used to accelerate the identification and prioritisation of targets under significant time pressure. Analyses of such platforms indicate that human reviewers, operating under compressed decision cycles, tend to exhibit automation bias a tendency to defer to algorithmic recommendations rather than independently verifying them which further attenuates the meaningfulness of nominal human sign-off and, consequently, the mens rea attributable to that human act (Malik et al., 2025; Oringa et al., 2026). Suchman (2020) frames this dynamic as the paradox of "algorithmic accuracy": the promotion of automated targeting as a means of enhancing precision simultaneously produces operations that are, in practice, both more indiscriminate in outcome and less legally accountable in process, because the apparent objectivity of algorithmic outputs discourages the very scrutiny that IHL's principle of distinction demands.

The consequence of causal anonymity is not merely evidentiary difficulty but a doctrinal mismatch: existing international criminal law was built around a paradigm of individuated human agency that autonomous targeting structurally disrupts. Aprianti et al. (2026) argue that this mismatch calls for a reconstruction of mens rea doctrine along systemic and risk-based lines rather than its outright abandonment, a position with which this study concurs. Absent such reconstruction, however, causal anonymity operates as a de facto amnesty: harms occur, are documented, and are even publicly acknowledged, yet no individual is prosecutable because no individual can be shown, beyond reasonable doubt, to have possessed the required guilty mind in relation to the specific outcome. This is the first and, this study argues, most foundational pillar of algorithmic impunity.

Disintegration of the Command Chain and the Limits of Command Responsibility

The second structural condition is the disintegration of the traditional military command chain as a locus of accountability. Command responsibility under Article 28 of the Rome Statute and customary IHL rests on a doctrinal architecture calibrated to human subordination: a commander is held liable where they knew or should have known that subordinates were committing or about to commit crimes and failed to take necessary and reasonable measures to prevent or repress them. Gunawan et al. (2022) demonstrate that this doctrine assumes a relationship of hierarchical, intentional agency between commander and subordinate that simply does not exist between a human commander and an autonomous weapon acting on machine-learned targeting logic, particularly in the case of "human-out-of-the-loop" systems capable of selecting and engaging targets without any real-time human decision point.

Chengeta (2016) advances a related but analytically sharper argument: the very framing of a person who deploys an autonomous weapon as a "commander" is itself misleading, because it implies that the weapon occupies the subordinate role of a combatant capable of receiving and executing orders in a manner analogous to a human

soldier an analogy the study rejects as conceptually incoherent. An autonomous weapon does not "disobey" an order in the sense contemplated by command responsibility doctrine; it behaves in accordance with, or in departure from, statistical patterns learned during training, a behaviour that a human commander cannot meaningfully foresee, supervise, or correct in the way command responsibility presupposes. Gaeta (2024) offers a partial doctrinal rejoinder, arguing that the material act of an unlawful attack conducted through an "intelligent" AWS should, for the purposes of war-crimes attribution, be treated as the act of the human user who activates and deploys the system including in cases of unintended engagement precisely to prevent the technology from functioning as a liability-laundering device. Yet even this reconstruction does not resolve the mens rea difficulty identified in the preceding section, since attributing the material act (actus reus) to the user does nothing to establish the mental element required for individual criminal liability.

The disintegration of command responsibility is compounded by the phenomenon of the "many hands" problem operating vertically, not merely horizontally, across an organisation. Bo and colleagues, in their comprehensive report on retaining human responsibility for AWS-related IHL violations, emphasise that responsibility for a single autonomous engagement is now typically distributed across a temporal sequence spanning weapons design, testing, legal review, procurement, deployment authorisation, and real-time use each stage governed by different personnel operating with different levels of knowledge and control (Bo et al., 2022). No single point in this sequence corresponds neatly to the traditional commander-subordinate relationship targeted by Article 28, and the cumulative effect is a command structure that has been technically and organisationally disaggregated faster than the doctrine of command responsibility has evolved to track it.

At the level of state responsibility, Pacholska (2022) offers a more optimistic assessment, arguing that most challenges attributed to AI in fact stem from pre-existing systemic weaknesses in IHL's reverberating-effects doctrine rather than from anything unique to autonomy, and that state responsibility for AWS deployment is, in principle, always retained through the state's ultimate authorisation of the weapon's use. Boutin (2022) similarly maintains that a state may incur responsibility not only for the wrongful use of an AI-enabled weapon but also, prior to deployment, for a failure to ensure the system's compliance with IHL during development or acquisition, including through the Article 36 weapons-review obligation. While this study agrees that state responsibility remains conceptually available even where individual and command responsibility fail, it also observes consistent with Boutin's own caveat that state responsibility operates on a fundamentally different register from individual criminal accountability: it may yield reparations or a finding of an internationally wrongful act, but it does not supply the retributive, victim-facing accountability that criminal prosecution of responsible individuals is meant to deliver. The disintegration of the command chain therefore produces a second, distinct pillar of algorithmic impunity: even where state responsibility remains theoretically intact, the erosion of command responsibility leaves a persistent deficit in individualised accountability for battlefield harm.

Normative Inadequacy: UNGA Resolution 79/239 and the CCW-GGE Process

The third structural condition is the absence of a binding international instrument specifically governing LAWS, a gap that recent soft-law developments have thus far failed to close. UN General Assembly Resolution 79/239 (2024) adopted following years of preparatory work within the CCW-GGE represents the most significant normative development to date, calling on states to address the risks associated with autonomous weapons and to continue multilateral deliberation toward a possible legal framework. Yet an evaluation of the Resolution's substance against the accountability failures identified above reveals significant limitations. As with its predecessor Resolution 79/62, Resolution 79/239 is exhortatory rather than obligatory: it encourages continued dialogue and study rather than establishing binding obligations, verification mechanisms, or enforcement procedures, and it does not resolve the definitional disputes over "autonomy" and "human control" that have stalled the CCW-GGE process since 2014 (Sharikov, 2024; Windfeld, 2025).

The CCW-GGE's own output illustrates the depth of this normative inadequacy. The eleven Guiding Principles adopted by the GGE in 2019 affirm, in general terms, that IHL applies fully to LAWS and that human responsibility must be retained, yet Qerimi (2023) documents that state positions within the GGE remain fragmented across at least six distinct categories, ranging from support for a comprehensive prohibition to explicit opposition to any binding restriction with major military powers, including the United States, Russia, and China, resisting a legally binding instrument altogether. Begishev (2026) responds to this impasse by proposing a comprehensive draft convention prohibiting the development, production, and use of autonomous weapons systems lacking meaningful human control, complete with a verification regime and institutional enforcement mechanism; the very need for such a proposal, more than a decade into the GGE process, underscores how far current soft-law instruments remain from providing an operable accountability framework.

A further inadequacy concerns the substantive content of "meaningful human control" itself, which both the GGE Guiding Principles and Resolution 79/239 treat as the principal safeguard against unlawful autonomous engagements. Ekelhof (2019) demonstrates, through analysis of actual military operational practice, that meaningful human control is not a single, well-defined threshold but a contested and highly context-dependent concept whose meaning shifts across different weapons systems, targeting environments, and operational tempos. Zhu et al. (2025) reach a related conclusion through sourcedoctrinal analysis, finding that while meaningful human control can be inferred from a systemic reading of Articles 36 and 57 of Additional Protocol I at the level of treaty law, its status as customary international law remains unsettled precisely because the reservations of major military powers prevent the emergence of consistent state practice and opinio juris. The result is that the very concept intended to anchor accountability under the current normative architecture lacks the legal precision necessary to perform that function reliably across jurisdictions and weapons platforms.

The normative inadequacy is compounded by the way in which existing instruments address bias and epistemic reliability in AI-enabled targeting. Bruun et al. (2025) and Blanchard et al. (2024) show that systemic bias arising from skewed training data, flawed algorithm design, or biased patterns of use can degrade compliance with the principles of distinction and proportionality in ways that neither the CCW-GGE Guiding Principles nor Resolution 79/239 substantively address, since both instruments focus overwhelmingly on the question of human control at the point of engagement rather than on the integrity of the system across its entire development life cycle. Dorsey (2025) and Woodcock (2023) add that AI-enabled decision-support systems used for proportionality assessments increasingly displace the qualitative, context-sensitive judgment that Additional Protocol I's proportionality standard requires, replacing it with opaque, quantification-driven outputs that current legal review mechanisms under Article 36 were never designed to audit. Taken together, these findings demonstrate that the existing normative framework spanning both binding IHL instruments and the emerging soft law of the CCW-GGE and UNGA resolutions fails to close, and in some respects entrenches, the very accountability gap it purports to address. This normative inadequacy constitutes the third pillar of algorithmic impunity: a persistent absence of binding, technically calibrated, and life-cycle-sensitive regulation specifically tailored to autonomous weapons.

Toward an Algorithmic Impunity Doctrine: From Meaningful Human Control to Life-Cycle Accountability

The convergence of causal anonymity, command-chain disintegration, and normative inadequacy justifies treating algorithmic impunity not as a transient regulatory lag but as a structural feature of the current IHL regime one that requires a dedicated doctrinal response rather than incremental adjustment of existing categories. This study proposes that such a response should take the form of an algorithmic impunity doctrine, to be codified within a future additional protocol to IHL specifically addressing autonomous weapons systems. The doctrine would perform two functions that the current framework does not: first, it would formally recognise algorithmic impunity as a legally cognisable harm in itself a systemic failure of the accountability architecture thereby creating a basis for institutional and state responsibility even in the many cases where individual criminal responsibility cannot be established; second, it would anchor accountability obligations to the entire life cycle of a military AI system rather than to the single moment of human sign-off that meaningful human control doctrine currently privileges.

This life-cycle approach builds on, but moves beyond, the differentiated and prudential model of meaningful human control developed by Amoroso and Tamburrini (2021), which assigns humans the roles of fail-safe actor, accountability attractor, and moral agency enactor at the point of targeting decisions. While that model remains valuable for engagement-level safeguards, this study argues that it is insufficient on its own because it does not reach the upstream stages of design, data curation, testing, and legal review at which many of the sources of causal anonymity and bias identified above

originate (Blanchard et al., 2024; Bruun et al., 2025). A life-cycle accountability framework would instead impose enforceable obligations at each of five stages: (i) design and data governance, requiring documented bias-testing and provenance tracking for training data; (ii) legal review, strengthening Article 36 weapons-review obligations to include mandatory, independently auditable assessment of algorithmic decision logic, not merely of the weapon's physical effects; (iii) authorisation, requiring a documented, traceable human decision that satisfies a defined evidentiary threshold of contextual awareness before deployment in a specific operational environment; (iv) engagement, retaining meaningful human control requirements calibrated to the specific weapons system and environment, consistent with the differentiated model of Amoroso and Tamburrini (2021); and (v) post-incident review, mandating investigation protocols capable of reconstructing the algorithmic decision chain after an unlawful or contested engagement, addressing the black-box problem identified by Skuratova et al. (2026).

Institutionally, this doctrine would require the CCW-GGE process, or a successor diplomatic conference, to move beyond the exhortatory register of Resolution 79/239 toward a binding protocol that adopts three concrete obligations: first, a treaty-based definition of legally sufficient human control tied to the five-stage life-cycle framework rather than to an undefined, single-point notion of "meaningful" control, addressing the definitional instability documented by Ekelhof (2019) and Zhu et al. (2025); second, a presumption of state responsibility rebuttable only by evidence of full compliance with life-cycle obligations for internationally wrongful acts committed through AI-enabled weapons, building on Boutin's (2022) analysis of pre-deployment state responsibility and Gaeta's (2024) attribution of AWS acts to the deploying state; and third, a complementary reform of international criminal law recognising a distinct mode of liability analogous to, but doctrinally separate from, command responsibility for the knowing or reckless failure to discharge life-cycle obligations, addressing the doctrinal gap identified by Tong (2025) and Sudhakar et al. (2026) without requiring proof of intent regarding the specific lethal outcome that traditional mens rea doctrine demands.

This proposal is not without limitations, which future research should address. A life-cycle accountability framework requires verification and auditing capacities that many states, particularly in the Global South, may lack the technical infrastructure to implement or monitor, a concern consistent with Bode's (2019) documentation of Global South advocacy for a preventive ban precisely because verification-heavy compromise frameworks may favour technologically advanced states. It also requires resolving the interoperability of new life-cycle obligations with existing Article 36 review practice, which Malik et al. (2025) show is already weakly and inconsistently implemented even for non-autonomous decision-support tools. Notwithstanding these limitations, the analysis presented here demonstrates that closing the accountability gap in AI-driven warfare requires a doctrine that reaches beyond the engagement-level focus of meaningful human control to the full technical and institutional life cycle through which algorithmic impunity is produced.

Conclusion

This study set out to construct algorithmic impunity as an analytical category within International Humanitarian Law and to evaluate the adequacy of the international community's most recent normative response to AI-driven autonomous weapons. The findings demonstrate that the current accountability crisis is not incidental but structural, arising from the convergence of three distinct conditions: causal anonymity, which erodes the mens rea foundations of individual criminal responsibility (Bo, 2021; Skuratova et al., 2026); disintegration of the traditional command chain, which hollows out command responsibility as a viable accountability mechanism even as state responsibility remains theoretically intact (Chengeta, 2016; Pacholska, 2022; Boutin, 2022); and the normative inadequacy of existing instruments, including UNGA Resolution 79/239 and the CCW-GGE Guiding Principles, which remain exhortatory, definitionally unsettled, and inattentive to the full life cycle of military AI systems (Ekelhof, 2019; Zhu et al., 2025; Bruun et al., 2025).

The theoretical contribution of this study lies in reframing the widely discussed "responsibility gap" as algorithmic impunity a structural, self-reinforcing condition rather than a temporary lacuna awaiting incremental legal adjustment. Practically, the study recommends that the CCW-GGE process, states, and relevant international organisations pursue a binding additional protocol to IHL embedding a life-cycle accountability framework, encompassing design, legal review, authorisation, engagement, and post-incident investigation obligations, together with a rebuttable presumption of state responsibility and a distinct mode of individual liability tailored to life-cycle failures rather than to engagement-specific intent. Such a framework would move the international community beyond the currently indeterminate concept of meaningful human control toward an operable, auditable, and enforceable standard of accountability. This study is limited by its doctrinal-normative scope, which does not extend to empirical assessment of specific weapons platforms' technical performance or to comparative analysis of domestic legal responses to military AI. Future research should pursue empirical case studies of documented AI-assisted targeting incidents to test the life-cycle framework's practical applicability, as well as comparative analysis of how regional organisations and domestic legal systems might operationalise the accountability obligations proposed here pending the conclusion of a binding international instrument.

References

Ahmad Khalil, R., et al. (2024). Deployment of autonomous weapon systems in the warfare: Addressing accountability gaps and reformulating international criminal law. Balkan Social Science Review.

Amoroso, D., & Tamburrini, G. (2020). Autonomous weapons systems and meaningful human control: Ethical and legal issues. Current Robotics Reports, 1(4), 187–194.

Amoroso, D., & Tamburrini, G. (2021). Toward a normative model of meaningful human control over weapons systems. Ethics & International Affairs, 35(2), 245–272.

Aprianti, D., et al. (2026). Artificial intelligence and criminal liability: Reassessing mens rea in the Indonesian and global contexts. Jurisprudensi: Jurnal Ilmu Syariah, Perundang-Undangan dan Ekonomi Islam.

Begishev, I. (2026). United Nations Convention on the Prohibition of the Development, Production and Use of Autonomous Weapons Systems. Journal of Digital Technologies and Law.

Blanchard, A., et al. (2024). Bias in military artificial intelligence. SSRN Electronic Journal. https://consensus.app/papers/details/71a3e509d0d6592f8407384561bc997a/

Bo, M. (2021). Autonomous weapons and the responsibility gap in light of the mens rea of the war crime of attacking civilians in the ICC Statute. Journal of International Criminal Justice, 19(2), 275–299.

Bo, M., Dorsey, J., Scholz, J., Sharkey, N., Sparrow, R., & Lucas, G. (2022). Retaining human responsibility in the development and use of autonomous weapon systems: On accountability for violations of international humanitarian law involving AWS. Stockholm International Peace Research Institute.

Bode, I. (2019). Norm-making and the Global South: Attempts to regulate lethal autonomous weapons systems. Global Policy, 10(3), 359–364.

Boutin, B. (2022). State responsibility in relation to military applications of artificial intelligence. Leiden Journal of International Law, 36(1), 133–150.

Bruun, L., et al. (2025). Bias in military artificial intelligence and compliance with international humanitarian law. Stockholm International Peace Research Institute.

Chengeta, T. (2016). Accountability gap: Autonomous weapon systems and modes of responsibility in international law. Denver Journal of International Law & Policy, 45(1), 1–50. https://doi.org/10.2139/ssrn.2755211

Dorsey, J. (2025). The erosion of human(e) judgement in targeting? Quantification logics, AI-enabled decision support systems and proportionality assessments in IHL. International Review of the Red Cross.

Ekelhof, M. A. C. (2019). Moving beyond semantics on autonomous weapons: Meaningful human control in operation. Global Policy, 10(3), 343–348. https://consensus.app/papers/details/a5fbee81973a5172b9143d512d7d9c01/

Gaeta, P. (2024). Who acts when autonomous weapons strike? Journal of International Criminal Justice.

Garcia, D. (2023). Algorithms and decision-making in military artificial intelligence. Global Society, 37(4), 526–541.

Gunawan, Y., et al. (2022). Command responsibility of autonomous weapons under international humanitarian law. Cogent Social Sciences, 8(1).

Jackson, J. (2023). Mapping the lethal autonomous weapons debate: An introduction. Ethics & International Affairs, 37(4), 419–424. https://consensus.app/papers/details/914211435ca15cd895aa089ad865edac/

Kutovyi, O., et al. (2025). Autonomous weapon systems and artificial intelligence as a challenge to international humanitarian law and human rights. Uzhhorod National University Herald. Series: Law.

Malik, A., et al. (2025). From human-in-the-loop to accountability gaps: AI-enabled targeting tools and international humanitarian law. Physical Education, Health and Social Sciences.

Ojha, Y. (2025). Artificial intelligence in armed conflict: Perspectives from international humanitarian law. Unity Journal.

Oringa, J., et al. (2026). AI-assisted targeting and semi-autonomous weapons systems in the Israel-Gaza War: An international humanitarian law analysis. East African Journal of Law and Ethics.

Pacholska, M. (2022). Military artificial intelligence and the principle of distinction: A state responsibility perspective. Israel Law Review, 55(2), 181–204.

Qerimi, Q. (2023). Controlling lethal autonomous weapons systems: A typology of the position of states. Computer Law & Security Review, 49.

Sehrawat, V. (2025). Autonomous weapon systems and meaningful human control. International Journal of Law and Information Technology.

Sharikov, P. (2024). International regulations of lethal autonomous weapons systems: Transatlantic security dialogue. World Economy and International Relations.

Sio, F. S. de, & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. Frontiers in Robotics and AI, 5, 15.

Skuratova, A., et al. (2026). The problem of determining who is responsible under international criminal law for crimes committed by lethal autonomous weapons systems. Vestnik Permskogo Universiteta. Juridicheskie Nauki.

Suchman, L. (2020). Algorithmic warfare and the reinvention of accuracy. Critical Studies on Security, 8(2), 175–187.

Sudhakar, T. V. G. N. S., et al. (2026). Code to consequences: Assessing responsibility for international crimes through automated weapon systems. Asian Journal of Legal Education.

Tong, Z. (2025). Autonomous weapon systems and individual criminal liability under the Rome Statute. Scientific and Social Research, 7.

Windfeld, F. C. (2025). Struggles over epistemic capital: Complex governance objects and the making of lethal autonomous weapons systems. Contemporary Security Policy.

Woodcock, T. (2023). Human/machine(-learning) interactions, human agency and the international humanitarian law proportionality standard. Global Society, 37(4), 542–560.

Zakharchuk, I. (2025). Adapting international law to the challenges of new technologies: Artificial intelligence, autonomous weapon systems, and their regulation. Analytical and Comparative Jurisprudence.

Zhu, L., et al. (2025). The status of meaningful human control of lethal autonomous weapons system in international humanitarian law. Brawijaya Law Journal.

United Nations General Assembly. (2024). Resolution 79/239: Artificial intelligence in the military domain. UN Doc. A/RES/79/239.

United Nations General Assembly. (2024). Resolution 79/62: Lethal autonomous weapons systems. UN Doc. A/RES/79/62.

Protocol Additional to the Geneva Conventions of 12August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 UNTS 3.

Rome Statute of the International Criminal Court, 17 July 1998, 2187 UNTS 3.

International Law Commission. (2001). Draft Articles on Responsibility of States for Internationally Wrongful Acts. UN Doc. A/56/10.

Acknowledgments

The authors gratefully acknowledge the Faculty of Law, Universitas Pendidikan Ganesha, for institutional support during the preparation of this manuscript, and thank colleagues who provided constructive comments on earlier drafts.

Competing Interest

The authors state that there is no conflict of interest in the publication of this article.

Generative AI Statement

During the preparation of this manuscript, the author(s) used Claude (Anthropic) to assist with literature retrieval via the Consensus academic search integration, and to support language refinement and structural formatting in accordance with the journal template. Following this process, the author(s) critically reviewed, edited, and validated all content, arguments, and citations to ensure accuracy, doctrinal correctness, and scientific integrity. The author(s) maintain full accountability for the final content. AI was used solely as a supportive tool and is not credited with authorship. No original legal analysis or interpretation was performed by the AI without human review and validation.