BRIDGING THE RESPONSIBILITY GAP: A LIFECYCLE GOVERNANCE FRAMEWORK FOR AI-ENABLED AUTONOMOUS WEAPON SYSTEMS

Harman Hunjan*

ABSTRACT

The deployment of autonomous weapon systems (AWS) poses profound ethical and legal dilemmas. Just the thought of a drone, powered by opaque neural networks, which might target civilians in a crowded urban battlefield, raises accountability questions. A responsibility gap exists where no single human, including any operator, commander, or policy maker, can be fairly held accountable due to black box opacity and unpredictable emergent behaviour. IHL principles fail against self-learning systems. Lifecycle governance framework mandates human involvement throughout AWS implementation: Design, Testing, Deployment, Monitoring and Decommissioning. India considers the UN's CCW GGE the most suitable forum, chaired 2017-18, rejecting UNGA moratoriums. India's GGE-centric approach aligns innovation with humane accountability, before machines destroy modern warfare.

Keywords: Responsibility Gap, AWS, Lifecycle Governance, CCW GGE, IHL.

INTRODUCTION

Artificial intelligence has had the centre stage ever since its introduction and implementation in all spheres of life worldwide. With limitless usage, it has found its way into the military sectors of many countries. It is a matter of concern that the deployment of lethal autonomous weapon systems (LAWS) or the AI-enabled autonomous weapon systems poses profound ethical and legal dilemmas.[1] Just the thought of a drone, powered by opaque neural networks, which might target civilians in a crowded urban battlefield, is chilling. Its split-second choice raises the question of accountability and who to hold liable for the irreversible damage caused.

*BA LLB (HONS.), FIRST YEAR, G.S. FOUNDATION COLLEGE OF LAW, PANJAB UNIVERSITY.

[1] Alshammari AA, 'Artificial Intelligence in Defence: A Bioethical Examination of Societal Risks and Governance' (2026) 32 Acta Bioethics 41 https://doi.org/10.4067/s1726-569x2026000100041 accessed 19 March 2026.

This question on accountability is ideal for the Responsibility Gap. To define the responsibility gap, it is a loophole where no single human, including any operator, commander, or policy maker, can be fairly held accountable for the decisions taken by the autonomous weapons systems (AWS).[2] This exists owing to the black box opacity and the unpredictable emergent behaviour of the bots. This concept is clearly a manifestation of the violation of the international humanitarian law vision, as the risk is directly on the civilians, with no one held responsible for the same. Without resolution, human dignity and global IHL will be on the verge of erosion. Since the decisions made by AWS are machine-led, meaningful human control shines as the only path to stop the erosion.

The current remedies, which include the retroactive liability attribution and human in-loop safeguards, do address the symptoms of liability, but there still exists a gap in the Lifecycle governance framework and distributive responsibility to ensure ethical and fair compliance with IHL and AWS. This gap will be discussed in detail in this article.

LEGAL ADEQUACY OF INTERNATIONAL HUMANITARIAN LAWS (IHL) FOR THE AWS

When one considers the use of AI-enabled autonomous weapons systems, the traditional human-operated weapons regulation framework might not work effectively. IHL has based its structuring on major principles such as the separation of combatants from civilians and balancing out the military gain against harm, and to minimise the damage caused to other neighbouring sites. These rules are laid out in Articles 48, 51(5)(b) and 57 of the Additional Protocol I.[3] One such article, Article 36, deals with the new weapons and their prohibition in case they are identified as a threat. This section includes the AWS and the targeted algorithm since this section applies to both human and AI-driven systems. This section somehow fails to address the much more complex nature of the algorithm with its unpredictability and black box decision-making. The black box process is a challenge for the IHL's design, as AWS has the neural networks that can produce outcomes beyond the programmer's control. This causes a severe risk of a mass destruction attack. IHLs are still not well equipped to deal with the

[2] Sparrow R, 'Killer Robots Kill People: Issues in Shared Responsibility for Lethal Autonomous Weapon Systems' (2021) 34 Diplomacy & Security 155, 162.
[3] Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3, arts 48, 51(5)(b), 57 (Additional Protocol I).

systems that learn on their own and are self-emerging. This becomes the sole argument that human control must be considered as the outweighed option when it comes to lethal weapons.

There have been several debates and discussions stressing human responsibility for the weapons decision and attacks, but we have not reached a fixed set of rules. Without these rules, these weapons could cause irreversible damage with their decisions.[4] Accountability is yet another factor that adds to the gap. The rules and the scope are not clearly defined, and ultimately, who gets to be blamed is not decided.

TECHNICAL RISKS AND BLACK BOX OPACITY

AI-driven Lethal Autonomous Weapon Systems (LAWS) are trained and work on the machine learning data set and the Black box model.[5] The concern with the black box model is that it conceals the inner processing that was done by the system to reach the decision, which causes unreliability. If applied to practice, the AWS can simply optimise the narrow scope of the word "target hit" and might ignore civilian life, which is against the IHL's compliance. Another nail in the coffin is the degradation. Degradation occurs when the model must confront scenarios that it has never been trained for. This non-recognisable environment skyrockets the error rates and negative decision-making. These flaws are the major challenges for IHL's framework as it stresses upon foreseeability, but in practice, these systems hold a chance to fail in actual time despite being qualified in the pre-deployment tests.

THE GOVERNANCE AND RESPONSIBILITY GAPS

The autonomous weapon systems (AWS) create major cracks in the existing structures where technological independence and human command are not yet balanced out. These cracks represent legal accountability, moral considerations, ethical decision-making and demand an unbiased unified response from the machines.

Since AWS is an emerging system and makes decisions based on algorithms, some errors while training or situations not included in the pre-deployment training can cause irreversible damage that can include unintended deaths. Courts would eventually struggle in these situations to pin

[4] Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 'Report of the 2022 Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (UN Doc CCW/MSP/2022/4, 2022) para 12 https://dig.watch/processes/gge-laws accessed 10 March 2026.
[5] Encode Justice, 'Encode Justice Whitepaper: Technical Risks in Lethal Autonomous Weapons' (arXiv:2502.10174, 2025).

criminal negligence and civil damages. This is because the intent dissolves in the systems. This is similar to Sparrow's analysis, but with AI's evolution and self-learning nature, IHL's assumptions for traceable human agency fail.[6] AI algorithms hold the killing power that can work without command, and this will strip away the warfare and human solemnity. This will risk the normalisation of violence and weakening of restraints, which is a concern in human rights jurisprudence.[7] When we talk about AWS or L(AWS), these robots enabled warfare could unleash barrages like a wildfire. It will flare up reaction windows and further amplify the errors, leading to a global crisis.

LIFECYCLE-BASED HUMAN AGENCY MODEL

To deal with these gaps, there can be the implementation of a five-phase governance model can be implemented, which makes human involvement throughout the AWS implementation mandatory. This will transform the vague codified duties of the system managers and align the network with IHL's imperatives.

Design Phase: It is potent to start with risk management from the initial phase. There should be enforcement of an interpretable AI structure and networking with risk evaluations, which will certify developers with liability for foreseeable flaws. The main aim for this phase is to reduce the opacity.

Testing Phase: As discussed above, AWS can develop uncertainty or errors based on unrecognised situations. So, to prevent this, the systems must be trained and well tested with adversarial war game scenarios. The simulations must include the simulation breakdown and further evaluation to get accountability sanctions for testers and approvers.

Deployment Phase: The deployment must include graduated human intervention. There should be complete in-loop control for urban or civilian risks, and on-loop supervision for any remote operations. This will confirm real-time risk analysis and adherence to IHL rules.

[6] Sparrow, 'Robots and the Future of War: Identifying who is to Blame When an Autonomous Weapon System Causes Death or Damage' (2016) 24(1) International Journal of Human Rights 127.
[7] International Committee of the Red Cross, 'Why Multilateral Regulation of Autonomy in Weapons Systems is Needed Now' (International Review of the Red Cross, 2024) https://international-review.icrc.org/articles/stepping-back-from-brink-regulation-of-autonomous-weapons-systems-913 accessed 14 March 2026.

Monitoring Phase: This phase holds great significance in the accountability and liability criteria. Constant monitoring will help with post-event blame assessment and will hold the record for continuous performance of both systems and the supervising officers.

Decommissioning Phase: This phase will help to stabilise those AWS that hold the potential to go bad. It will work as a 'kill-switch', which can be used if the system remains unchecked for long or starts to develop errors during its operation.

This lifecycle structure provides a range of technical safeguards with legal duties to the developers and approvers, ensuring IHL principles like precautions permeate every step for enduring responsibility.[8]

COMPARATIVE ANALYSIS OF AI-DSS AND AWS GOVERNANCE

AI decision-support systems include all the domains of AI, such as machine learning, neural networks and deep learning, which influence the decision-making process. These systems are used in domains such as healthcare diagnostics and financial risk models. DSS frameworks mandate transparency audits and human vetoes throughout the Lifecycles. These also provide integrating interactive feedback, which is crucial for AWS adoption. In contrast with AWS, DSS enforces 'explainable AI', which is XAI. This feature is to unpack decision logic, unlike AWS, which uses black box shielding codes. Another relevant feature is the post-deployment logging in DSS. This enables traceability, which helps to pinpoint the error and liability clause. The feature which can be well praised is the regulatory tiers. These assign phase-specific duties in every step when the AI system implementation is used. It is used by the EU AI Act high-risk categories.[9] Borrowing DSS policies can make the use of AWS more accountable and within the scope of IHL, prioritising verification and safety over velocity. The real difference between the DSS and AWS frameworks is the military targeting cycle. AI-DSS works with the whole cycle, including target acquisition to damage control, while AWS only pulls the 'final trigger'.[10] Israel used AI-DSS in Gaza to suggest about forty thousand targets, which were approved by human command.

[8] Encode Justice, 'Encode Justice Whitepaper: Technical Risks in Lethal Autonomous Weapons' (arXiv:2502.10174, 2025) 15.
[9] European Commission, 'Proposal for a Regulation on Artificial Intelligence (AI Act)' COM (2021) 206 final (21 April 2021) art 52 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 accessed 15 March 2026.
[10] Blanchard & Bruun, *SIPRI Report* (2025).

PATHWAYS TO MULTILATERAL REGULATION AND TREATY DESIGN

The rules and regulations for AWS are supervised and come under the UN's CCW GGE group. This regulatory body monitors the successes and meaningful human control based on the norms and other morals.[11] To make sure that systems work efficiently, the Lifecycle compliance protocol could check the AWS at every stage of its life. It can start with the starting phase, which includes the building to testing and the point when the system goes bad. Another relevant tool that could be used is XAI Minima, as discussed above. This will promote transparency and help with drafting the minimum standards, which every AWS and L(AWS) must comply with, and lastly, verification regimes should be set up so that proper reporting can be achieved and inspection can be done through a peer board.

Indian Perspective: India considers the UN's CCW GGE, group of governmental experts, the most suitable forum related to AWS and, more significantly, the 1995 Protocol IV on blinding laser weapons, reached through consensus.[12] India, having chaired GGE (2017-18), during the genesis of guiding principles, prioritised IHL-neutral technology discussions and affirmed human responsibility across the AWS lifecycle.[13] In 2024, India's AI command centre in Bengaluru ran over the DSS structure to expand supervision over human control on drone strikes. In 2025, DRDO tested the prototypes of drones with limited autonomy in Rajasthan, which does not align with Article 36 of DRDO, related to weapon reviews. Hence, by adopting DRDO design phase certification, it will force transparency, and the monitoring phase logs will create accountability. As a global south leader, India aims towards a perfect blend of power with humanitarian protection, which is core for the lifecycle model.

CONCLUSION

AWS models have transformed warfare through the smart use of artificial intelligence, yet the gaps remain in responsibility and liability, opacity, and control under the IHL model. The lifecycle agency framework would work best, according to DSS, as it provides a more transparent and ethical solution. India's GGE-centric approach aligns with innovation and

[11] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons (adopted 10 October 1980, entered into force 2 December 1983) 1342 UNTS 137 (CCW).
[12] Convention on Certain Conventional Weapons (CCW) (adopted 10 October 1980) 1342 UNTS 137, Protocol IV.
[13] Ruchir Joshi and Abhijit Singh, 'India's Normative Stance on Lethal Autonomous Weapons Systems' (Carnegie Endowment for International Peace, 26 February 2024) https://carnegieindia.org/2024/02/26/india-s-normative-stance-on-lethal-autonomous-weapons-systems-pub-91796

stresses humane accountability. Before the machines destroy modern warfare, action related to supervision is the need of the hour to ensure AI weapons obey humanitarian laws.