

Autonomy in Weapon Systems: The Unpredictable and Unaccountable Confluence of Mechanization of Weapons and International Humanitarian Law

Ashlesha Pandey

A fourth-year law student at Hidayatullah National Law University, Raipur

Abstract

This essay delves into the complex intersection of autonomous weapon systems (AWS) and international humanitarian law (IHL), exploring the ethical, moral, and legal dilemmas associated with the mechanization of weapons. The essay scrutinizes their legality based on fundamental principles like the distinction and proportionality in the absence of clear IHL covenant governing AWS. The principle of distinction, crucial in safeguarding civilians during armed conflict, is challenged by the AWS's inability to precisely differentiate between combatants and civilians. The principle of proportionality presents challenges in the context of unpredictable AWS actions. The essay also examines the accountability vacuum in AWS scenarios, questioning the roles of system operators, military commanders, and programmers. The Martens Clause is highlighted as a crucial ethical consideration, emphasizing the importance of humanity and public dictate in the absence of specific AWS-related laws. Finally, the essay advocates for a careful balance between human agency and automated systems, emphasizing the imperative of international cooperation to mitigate the risks associated with AWS in armed conflicts.

Keywords

AWS, Humanitarian Law, Armed Conflict, International Law, AI

Introduction

Unpredictable, unreliable, unaccountable- the realm of autonomous weapon systems (“AWS”) is laced with the dilemmas of ethics of war, moral responsibilities and legal obligations. Currently, there is no prescribed covenant under international humanitarian law (“IHL”) that governs the usage of AWS- which leads to a derisive view regarding their legality. As of 2023, despite numerous attempts having been made to lay down the criteria for determining what qualifies as an AWS, there has been no concrete conclusion.¹ Therefore, even though the usage of such weaponry has been questioned innumerable times on accounts of morals and ethics, the following essay will examine the growing usage of such systems strictly in light of the underlying legal responsibilities of states under IHL, considering the unweighted merit of autonomous artificial intelligence (“AI”) and weaponry.

AWS can be defined as “Any weapon system with autonomy in its critical functions—that is, a weapon system that can select (search for, detect, identify, track or select) and attack (use force against, neutralise, damage or destroy) targets without human intervention.”² The world in itself is defining new limits for “autonomy”, especially with the increasing advent of artificial intelligence, thus it is pertinent to develop a few more definitions for the term, which encompass the limits set by the abovementioned definition.

However, by defining the broad subject matter of the debate over AWS without having to directly address any specific legal issues, the definition offers a helpful

¹ Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System Geneva, 6-10 March, and 15-19 May 2023 Item 5, Revised working paper Submitted by Austria, n.d.

² Neil Davison, ‘A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law’ (2018) 30 UNODA Occasional Paper 5.

foundation for a legal analysis. In that regard, the definition is not meant to make assumptions about the degree of autonomy in weapon systems that may or may not be considered legal but offers a level playing field in deciding the ambit of inclusion. Thus, it would form a strong foundation to develop further arguments, as has been attempted below.

The Initial Trepidations- What is Antithetical to IHL

The initial legal argument levied against AWS rests on the two primary concepts of IHL: the principle of distinction and the principle of proportionality.³ The crux of the arguments essentially prioritises the involvement of human agency in armed conflict due to the human nature of conflict, the requisite rules and conventions, and the underlying quantitative and qualitative judgement necessary in conflict.

The Principle of Distinction

The *principle of distinction* is an established norm⁴ under customary international law and is applicable in cases of both international and non-international armed conflict⁵, and forms the first safety net for civilians during armed conflict. There needs to be the utmost distinction between civilians and combatants- attacks on the former being in contravention of IHL, with the immunity of civilians from direct attack as one of the basic rules of international humanitarian law⁶. However, when AWS comes into the picture, the distinction aspect stands

³ Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), adopted June 8, 1977, 1125 UNTS 3, entered into force December 7, 1978, arts. 48 and 51(4-5).

⁴ Prosecutor v Katanga and Ngudjolo [2012] ICC Case No ICC-01/04-01/07, Judgment (International Criminal Court, Trial Chamber II, 21 November 2012)

⁵ 'Customary IHL - Rule 1. The Principle of Distinction between Civilians and Combatants' (ICRC.org2022) <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule1> accessed 4 November 2022.

⁶ **MILITARY PROSECUTOR v. OMAR MAHMUD KASSEM AND OTHERS, Israel, Military Court at Ramallah**

questioned. The distinction has blurred with the advent of AWS, for the ability of an AWS to differentiate between civilians and combatants when built purely to attack.

Initially, the question should be rephrased to a prospective, inventive aspect- is the creation of such AI-AWS possible where the distinction between civilians and combatants is made with precision? However, the prospective nature of the question and the futuristic approach are essentially hypothetical. Concepts and contingencies have never been the principles of war, and in the current picture, there exists no AWS mechanism known that is impeccable in determining the distinction between a civilian and a combatant- something even human soldiers can falter at. Further, gauging the needs of *hors de combat*⁷ and mediating the attacking nature and intensity is something that relies primarily on human understanding and judgement- something AWS are incapable of.⁸ Even if there were to be an AWS with a mechanism precise enough to distinguish military insignia/symbols/uniforms, there would be no protection from it being extended to ***hors de combat***- who bear the same symbols but have lost the ability to combat. Despite having the prerequisite of objective judgements of facts⁹ during the conflict as the premise for decision-making, the existence of case-to-case basis analysis by human soldiers stands vastly different from that of AWS.

Owing to the increasing involvement of civilian areas and residences as battlegrounds, the probability of the intersection of belligerents and civilians stands heightened. The recent years have seen increasing turmoil across the globe with indiscriminate involvement of

⁷ Human Rights Watch and IHRC, *Making the Case: The Dangers of Killer Robots and the Need for a Preemptive Ban* (December 2016), https://www.hrw.org/sites/default/files/report_pdf/arms1216_web.pdf

⁸ Afonso Seixas-Nunes, *The Legality and Accountability of Autonomous Weapon Systems* (2022).

⁹ Geneva, 'EXPERT MEETING AUTONOMOUS WEAPON SYSTEMS TECHNICAL, MILITARY, LEGAL and HUMANITARIAN ASPECTS' (2014).

civilians. For instance, the Israel-Palestine conflict witnessed the civilian death toll rising to 1,200 deaths in Israel and 21,507 deaths in Gaza.¹⁰ The evolution of war strategies has also increasingly used the civilian protection aspect in their favour, having soldiers and combatants dressed as civilians.¹¹ Despite violating IHL, the strategy minimises damage to the self while increasing harm to the other end. The conclusion then becomes simple- when humans have constantly been frazzling out in upholding the principle of distinction during the conflict, AWS- which inherently functions on the identification of set "characteristics" and lacks the qualitative, moment-mediated judgement- cannot act in consonance with the IHL and principle of distinction.

The Principle of Proportionality

Similarly, the ***principle of proportionality*** is another root of IHL, where the mandate is that any attack must be scaled against the civilian harm metric.¹² Thus, civilian harm cannot outweigh the military advantage during attacks.¹³ The basis of this analysis is subject to dynamic situations and infinite possibilities that may arise- where due care needs to be according to the battlefield's nature, location and demographics, the environmental impact assessment, and numerous other contingencies that may arise. These two principles- proportionality and distinction- in combination accrue the principle of precautions during the attack.

¹⁰ CNN, 'Title of the Article' (2023) [published on 7 November 2023] CNN <https://edition.cnn.com/2023/11/07/middleeast/palestinian-israeli-deaths-gaza-dg/index.html> accessed 17 November, 2023.

¹¹ Ido Rosenzweig, *Combatants Dressed as Civilians?* under International Law, https://en.idi.org.il/media/6191/combatants_dressed_as_civilians.pdf.

¹² Article 51(5)(b) of the 1977 Additional Protocol I, ICRC.

¹³ Bonnie Docherty, 'The Need for and Elements of a New Treaty on Fully Autonomous Weapons' (*Human Rights Watch* June 2020) <https://www.hrw.org/news/2020/06/01/need-and-elements-new-treaty-fully-autonomous-weapons#_ftn1> accessed 4 November 2022.

For AWS, the same judgement cannot be effectively mediated unless the attacking sector has designated weapons for infinite battlefields and situations- which is not just impossible to quantify and create but, at the very outset, is a massively improbable situation. Thus, where does AWS stand on the scale of preservation of these principles is gravely unclear.

The abovementioned principles thus create the liability of human agency. When the question of **accountability** strikes, IHL's mandate has always been to hold human actors accountable in cases of war crimes or grave breaches of the Geneva Conventions.¹⁴ In the case of AWS, there exists no mechanism to strike an accountability stroke at one person. Even if one position, as would later be expounded upon- is culled out to be prosecuted for the deployment of an AWS, no "real" accountability would exist as IHL mandates the attack be carried out in its entirety by the accused. Thus, accountability as a background stands vague and imprecise in the realm of allegedly precise machinery.

The Governing Principle- Martens Clause

However, in cases of the utmost unpredictability, is wherein the Martens Clause comes to the fore¹⁵. The Martens Clause is of extreme relevance to the issue considering there are no codified or specified laws on the AWS or their usage under the IHL. The clause simply prescribes the primacy of principles of humanity and public dictate in cases where there are no specific IHL principles governing a subject.¹⁶ Finding its origin at the

¹⁴ Prosecutor v Ongwen [2021] ICC Case No ICC-02/04-01/15, Judgment (International Criminal Court, Trial Chamber IX, 4 February 2021)

¹⁵ IHRC, *Heed the Call: A Moral and Legal Imperative to Ban Killer Robots* (August 2018), https://www.hrw.org/sites/default/files/report_pdf/arms0818_web.pdf accessed 5 November 2022

¹⁶ International Committee of the Red Cross, *THE PRINCIPLES OF HUMANITY AND NECESSITY*, (2023), https://www.icrc.org/sites/default/files/wysiwyg/war-and-law/02_humanity_and_necessity-0.pdf.

Hague Peace Conference of 1899, the Martens clause has been critical in defining the role of civilians in conflict, as it is the underscoring factor which ties principles of natural law with the positive norms of IHL.¹⁷ This is also the arena where ethics and morals are questioned at their highest-principles of humanity having their roots in the general concepts of moral rights and wrongs concerning humans.

Due to the absolute lack of qualitative judgements, any AWS or AI-powered machinery cannot possibly be held accountable for violating the Martens Clause. Such systems have one job, and that is victory. Examining “victory” and “success” through the metrics of human life is morally and ethically incorrect. Despite the dearth of statutory protection, a human soldier in unprecedented circumstances could exercise, if not morally correct judgements but at least act in self-preservation. The destruction that may follow the usage of AWS would be arbitrary, to say the least, and would undermine the right to life, remedy and dignity as well.¹⁸

A Hypothetical Scenario- And Everything That Can Go Wrong

The answer to the contentions raised above can be attempted with a hypothetical, which presumes infinite precision of the technology and AI involved. The world of AI is fast expanding and evolving, with ever-dynamic niches carving further specifications and pushing their boundaries. The inception of AWS, which is precise, reliable, and accurate, is not entirely impossible, if not in the near future, then somewhere further down the lane.

¹⁷ Rupert Ticehurst, ‘The Martens Clause and the Laws of Armed Conflict’ (1997) 37 International Review of the Red Cross 125.

¹⁸ Human Rights Watch and IHRC, *Shaking the Foundations: The Human Rights Implications of Killer Robots* (May 2014), <https://www.hrw.org/sites/default/files/reports/arms0514_ForUpload_0.pdf> accessed 5 November 2022

The entire debacle of AWS essentially revolves around existing weapons. However, for the sake of argument, let us assume that there exist hyper-precise mechanisms of distinction and proportion determination and that this technology is being actively used while manufacturing AWS. At the outset, the prospect seems enticing as it minimises human intervention on actual battlegrounds, minimises damages owing to the precise nature of the arsenal, and would reduce the costs massively due to specific targeting- among many other possible advantages. Now even if such weapons have been created, the question comes of state accountability and responsibility.

Expecting the Unlikely: Favouring AWS

Evidence supports the assertion that conflicting states have been consistently violating the IHL and have tendencies of twisting and disregarding principles of protection in lieu of victory and garb of self-defence.¹⁹ The past two years especially stand witness to it- with Russia blatantly invading Ukraine, to Israel's inhumane attacks in Palestine. Going back, the same instances have traced the history through Tigray, Nargono-Karabakh, Myanmar-amongst various others. Despite the inherent nature of the right of self-defence in international law, the right is only exercisable in case of an "apparent" threat²⁰, i.e., a tangible, calculated analysis is done before attacking-something far from reality. The current paradigm rests more on asserting oneself towards victory. Now in such cases, how far will the advent of the hypothetically infinitely precise technology be positively used in the development of weapons has an obvious answer. States

¹⁹ International Committee of the Red Cross, Violations of IHL, International Committee of the Red Cross (2014), <https://www.icrc.org/en/war-and-law/international-criminal-jurisdiction#:~:text=When%20violations%20of%20IHL%20occur.>

²⁰ United Nations Charter (adopted 26 June 1945, entered into force 24 October 1945) art 51. .

would still serve and devise such AWS, which eventually benefit them²¹- with all their precisions.

Programmed to Kill: Shrouding Immorality

Another popular contention raised in favour of AWS is that they might as well be programmed to suit IHL better. Due to the formulation of strict codes and programs, such weapons have a limited scope of functioning. They follow prescribed commands, which, experts presume, would be encoded in compliance with IHL.

However, if the previous situation is true, this assertion also falls flat here- considering that if the very programming disregards IHL, the machine becomes no less antithetical to IHL than a rogue soldier. A robot cannot hate²²- but the questions of morality would still be in the hands of the creator. Even if the creator in a so-called 'peaceful' environment creates a robot²³, the possibility of its ill-usage would still rest in the hands of other organs of conflict. This lack of a moral compass is argued in favour of AWS, considering they cannot decide adversely. However, this also brings forth the Martens Clause, as discussed previously. Further, considering the basis of Martens Clause to be public dictate, the situation worsens as opinion polls have hinted towards a disdain in the public towards the usage of AWS.²⁴

²¹ Marco Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified' *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified*, vol. 90 (US Naval War College 2014) <<https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1017&context=ils>> accessed 4 November 2022.

²² Ronald C. Arkin, *Ethical Robots in Warfare*, GEORGIA INSTITUTE OF TECHNOLOGY (Jan. 20, 2009), <http://www.cc.gatech.edu/ai/robot-lab/online-publications/arkin-rev.pdf>; Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Interim Report, ¶ 20 (by Philip Alston), transmitted by Note of the Secretary-General, UN Doc. A/65/321

²³ *Ibid.*, n. 8

²⁴ Executive Summary, International Committee of the Red Cross (ICRC), 'Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?', ¶ 4, ICRC Database (2018).

The situation can stand worsened in case of hyperactive development of such systems, wherein the situation would transcend into *weapon amassment and uncontrolled proliferation*- and the gravest scenario of all might be the falling of these weapons into the hands of non-state armed groups²⁵ which brings us to the next question- accountability.

Mechanical Questions: Who is Responsible?

Accountability can primarily be accorded to three positions in cases of attacks by AWS- the system operator, the military commander, and/or programmer. Nevertheless, there exist not just moral or ethical dilemmas to such liability accrual, but also the very legal principles of IHL stand questioned in this regime. Before delving deeper, it is crucial to study the regulatory principle, i.e., Rule 102²⁶ which states-

“No one may be convicted of an offence except on the basis of individual criminal responsibility.”

The above rule clarifies that any crime committed during an armed conflict asserts criminal responsibility solely on the doer. The basic principle of criminal law applies here, as has been upheld by the International Criminal Court as well²⁷ in case of individual liability, which includes “attempting to commit such crime, as well as assisting in, facilitating, aiding or abetting, the commission of a crime, along with planning and instigation” of the crime. When the same rule is read in the extension of Chapter 43 of the International Committee of

²⁵ “Autonomous Weapons: An Open Letter from AI & Robotics Researchers”; Human Rights Watch and IHRC, *Making the Case*, pp. 29-30.

²⁶ Volume II, Chapter 32, Section O., 'Customary IHL - Rule 102. Individual Criminal Responsibility (ICRC.org2022)<https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule102#:~:text=It%20is%20a%20basic%20principle,the%20commission%20of%20a%20crime.> accessed 5 November 2022.

²⁷ Art. 25, The Rome Statute of International Criminal Court.

the Red Cross' IHL database²⁸, the question of responsibility gives out a few issues.

Considering that an individual is himself responsible for any manifestly unlawful act as per Rule 151²⁹, and even for the carrying out of a manifestly unlawful order by a superior as per Rule 155³⁰, the direct responsibility would then fall on the system operator first. However, the question of duress and undue influence stands rampant in such a scenario at the very inception, considering the relationship between a superior and a junior. Further, the liability also stands, morally and legally diluted, considering the proximate foreseeability of the damage. In cases of AWS, there might be an unpredictable failure of machinery, collateral damage, misidentification, and unreliable killings- among other flaws that might come to the fore. In such a case, any damage or detriment would merely be the result of the failure of technology and not the human per se- as is evident with a few existing technologies (like Phalanx) where a limited timeframe is accorded to the system operator³¹.

In such a case, neither the system operator nor the military commander can be held liable- their defence would rest in the non-foreseeability of the damage that occurred. Instead of this, the liability might be accorded to the

²⁸ ., ICRC, Customary IHL Database, < https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_cha_chapter43_rule151> accessed 5 November 2022.

²⁹ Rule 151. "*Individuals are criminally responsible for war crimes they commit.*" Volume II, Chapter 43, Section A., ICRC, Customary IHL Database, < https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_cha_chapter43_rule151> accessed 5 November 2022.

³⁰ Rule 155. "*Obeying a superior order does not relieve a subordinate of criminal responsibility if the subordinate knew that the act ordered was unlawful or should have known because of the manifestly unlawful nature of the act ordered.*" Volume II, Chapter 43, Section E., ICRC, Customary IHL Database, < https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_cha_chapter43_rule151> accessed 5 November 2022.

³¹ Lieutenant Colonel Andre Haider, 'Autonomous Weapon Systems in International Humanitarian Law - Joint Air Power Competence Centre' (2018) 27 *The Journal of Joint Air Power Competence Centre* <<https://www.japcc.org/articles/autonomous-weapon-systems-in-international-humanitarian-law/#:~:text=Regardless%20of%20the%20presence%20or>> accessed 5 November 2022.

programmer for not creating an infallible system- another arbitrary accusation for the existence of human error is the very underscoring principle of human existence. Further, even if the program was 'impeccable', any last-minute situational dynamism cannot be predicted- so who is to be held liable in such a case remains to be seen.

Even if the two varying dictates of IHL are considered- one rooted in human action and the other focusing on minimising damages- the existence of liability still stands murky. The above discussion was of the school which accords responsibility to human agents.³² The latter line of thought where the minimising of unlawful damages is the case. This interpretation reflects a more utilitarian approach to the nature and character of IHL rules governing AWS, emphasising the net humanitarian effect with less focus on whether that effect was produced by reliance on human agency.³³ The middle ground between the two rests on the accountability of human agency unless the usage of machinery would minimise damages. Nonetheless, the abovementioned analysis makes it clear that unless there are AWS-specific regulatory statutes in existence, there will always remain accountability loopholes. Even if the eleven rules of Group of Governmental Experts, 2019 are adopted, the determination of “user” in cases of AWS³⁴ would be challenging and might lead to the prosecution of innocents. The debate about human agency and its extent remains unanswered; thus, determining responsibility would produce inaccurate results.

³² Vincent Boulanin, Laura Bruun and Netta Goussac, 'AUTONOMOUS WEAPON SYSTEMS and INTERNATIONAL HUMANITARIAN LAW Identifying Limits and the Required Type and Degree of Human-Machine Interaction' (Stockholm International Peace Research Institute 2021).

³³ CCW Convention, 'Israel considerations on the operationalisation of the eleven guiding principles adopted by the Group of Governmental Experts, Aug. 2020

³⁴ Guiding Principle (d) states that: 'Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system.', Guiding Principle (d) quoted from CCW Convention (note 2), Annex IV

Towards the Future

The dearth of AWS-specific legislation leads to extreme misuse and circumvention of liability. The absolute unpredictability during an attack furthers the admissibility of fault under IHL. Thus, as opposed to human combatants, AWS cannot be mandated, restricted, or accounted for during combat or attack. Their control should thus exist primarily at their very developmental stages. Declaring such weapons illegal would further the issue, amounting to double illegality of creation and usage- as their creation or usage will not stop any time in the near future owing to their efficiency and precision- being result oriented³⁵.

Updating Warfare: Drafting Statutes that Keep Up

The usage of AWS by irresponsible or despotic sovereigns or even NSAs poses an enormous responsibility. To keep a check on the same, the mandate needs to focus on creating warfare-specific AI statutorily bound to have integrated codes and mechanisms to comply with IHL. This needs to be the essential requirement for every AWS without exception. However, there also needs to be fault-specific provisions inculcated, considering the creation of such weapons might also be laced with human errors. Further, the distinction dimension needs to be refined and prioritised over the other aspects of machinery whilst compulsorily involving human agency at this stage. The suggestion is to include a two-step identification system- before the ultimate attack, the distinction needs to be verified by a human agent. Even though the suggestion would still lend errors in cases of hors de combat especially, it might be active in minimising civilian damage. The same would also make it comparatively more straightforward to accord human liability and aid in

³⁵ Ibid, n. 15, ¶ 6

resolving the moral and ethical dilemma which rests on lending the power of life and death to machines.

Need for Human Intervention

There needs to be a more active intersection of human agency, even in cases of AWS, especially in the deeper stages of combat. Further, a separate class of limited-automation weapon systems can exist, which can be used in cases where the battle proceeds in civil residential areas and the probability of civilian harm is higher. The proportion of human agency and decision involvement would also be higher in such cases. The existence of inherently and completely automated weapon systems would still be functional in areas lower in civilian populations.

The accretion of state liability would also rest in proportion determination. The existence of various weapons and the decision to deploy them during wartime is still in the hands of humans and might mitigate predictable proportional damage. The issue of unpredictability can be resolved only with highly advanced and precise AI- something that's development is contingent on future scientific developments. Until then, the state's responsibility would primarily rest with the developer, but the penalty should be designed with caution and consideration. The additional check and futuristic liability in cases of failure of such weapons might help reduce their development in fear of failure and prosecution. The human agency versus liability should exist based on internationally determined and agreed-upon metrics ensuring the difference between meaningful (to ensure the moral dilemma is resolved) and effective (to accord specific liability) is kept in mind.

Conclusion

The nature of the above arguments is purely suggestive, and their relevance is apt even in cases where a certain constant evolution is witnessed. The questions raised and the attempted answers offer a peek into what lies in the dark underbelly of the regime of AWS. Indeed, where there is infinite potential in AWS to redefine the quantum of damage, the possibility of ill-use is always higher.

The principles of proportionality and principles of distinction come to the fore in line with the above statement. In lieu of developing AWS so sophisticated that they are in line with these principles, there will always exist the logical parallel of the absolute disregard of the same as well.

However, none of these suggestions can be helpful if the conflicting states are unwilling to comply. War is the ultimate resort and the vilest form of human cruelty, for it sees no class or creed but damages beyond measure. As long as conflicting states are not complying with IHL, the question of AWS or human agency stands redundant. Thus, the first step would be to ensure that the option of war is the least exercised and that reliance on soft power is increased.