Can AI-powered weapons comply with international humanitarian laws?

Hager Radi Abdelwahed*
Applied AI Research Scientist
Mila Quebec AI Institute

Omnia Farrag Othman*
Human Rights Researcher

Hadeer El Ashhab
AI PhD Student

Artificial Intelligence (AI) has recorded unprecedented progress in the last few years, which sparked debates about AI safety. Concerns emerged that AI is advancing too fast without considering all safety issues, which led to calls for a slowdown in AI research due to its growing impact on everyone's lives. For example, AI leaders call for a 6-month pause on advanced AI development[1]. Similarly, the International Scientific Report on the Safety of Advanced AI raised several safety concerns, even for issues less critical than life-or-death decision, such as military use [2]. We believe the AI community has largely overlooked autonomous weapons systems (AWS), which is a concerning use case of AI and an alarming threat to human life. Recent work [3] is among the few research papers in AI explicitly addressing the risks of AWS, offering recommendations to mitigate these risks and their implications for AI research and global stability. We build upon this type of work as we question if AI in military can be safely regulated, in light of international humanitarian law (IHL).

AWS and IHL AI in military has been widely used in different conflicts around the world[4]. For example, South Korea's Samsung SGR-A1 sentry robot, deployed along the DMZ, autonomously detects, tracks, and engages intruders using advanced AI systems[5]. In the ongoing conflict between Russia and Ukraine, both sides deploy AI-equipped drones. Russia's use of AWS is exemplified by the deployment of the ZALA Lancet and Uran-9 systems. The ZALA Lancet, an exploding drone, has been upgraded with AI capabilities, enabling it to autonomously identify and engage high-value targets such as Ukrainian artillery or air defense systems [6]. Finally, Israel widely uses AI weaponery to identify targets in Gaza, resulting in a huge number of civilian atrocities [7] and [8]. A British Surgeon who volunteered in Gaza this year recently testified at the British Parliament recounting Palestinian children's testimonies about how drones would hover over the injured civilians lying down and shoot them after bombs hit an area [9]. According to the book [10], Israel uses Gaza and the occupied west bank to test their weaponry and surveillance technology on ground before exporting it around the world.

While there are no international treaties regulating the usage of AWS, international law experts and several states expressed concerns that their usage can cause violations of international law namely the charter of the United Nations, customary international law, IHL, international human rights law, international criminal law, and the law of state responsibility [11, 12, 13]. In a United Nations General Assembly (UNGA) session in July 2024, several states affirmed that any weapons, including AWS, that don't comply with IHL are "de facto already prohibited and must not be used" [11]. There was a focus on IHL where major concerns were raised by experts and states that AWS can lead to violations of IHL. Some of these violations can constitute war crimes as per the Rome Statute of the International Criminal Court [14] as further defined in Appendix 1.
IHL experts and several states expressed concerns that the usage of AWS can violate both customary and treaty IHL, especially the rules of distinction, proportionality, military necessity, and precautions in attack, along with Article 36 of the I Protocol to the Geneva Conventions and the Martens Clause (defined in Appendix 1) [11, 12, 13]. The Rule of Distinction requires the distinction between active combatants, non-active combatants, and civilians as the latter two are protected by IHL and intentionally killing them amounts to a war crime [12, 13, 14]. This is clearly violated in ongoing

*Equal Contribution.
Corresponding authors: hager.radi@mila.quebec & omniafarrag95@gmail.com

2024 Harms and Risks of AI in the Military (HRAIM 2024).

AWS-enabled conflicts where a very high number of civilian causalities especially women and children were reported in Gaza. A recent analysis by the UN Human Rights Office of verified victims in Gaza over a six-month period showed that 80% of the victims were killed in residential houses. It showed that 26% of the victims were women and 44% were children mainly aging between 5 to 9 years old [15].

The rule of proportionality prohibits attacks whose civilian harm outweighs their military advantage [12, 13]. When considering the proportionality of an attack, the civilian harm in question does not only mean death or injury of civilians but also the destruction of civilian objects as per Protocol I in [12]. Civilian objects include infrastructure, hospitals, places of worship, and historical monuments [16]. It is doubtful that AWS will be able to weigh the significance of such objects to determine whether an attack is proportionate. The concept of military necessity permits attacks only when they will lead to a winning at the war or defeating an adversary [12]. This is part of complex decision making that AI cannot handle. Finally, the rules of precautions in attack necessitate "[taking] constant care and/or to [taking] precautions to avoid or minimize incidental civilian losses," [13, 17]. Human judgment is necessary to abide by these four rules on battlefields which are by definition complex and constantly changing. Additionally, in the case of AWS, decisions already made cannot be reversed if the situation changes in a way that leads to a violation of these IHL rules.

The usage of AWS also raises concerns about accountability under IHL, whose subjects and enforcers are exclusively humans. During the aforementioned UNGA discussion of AWS, states stressed that humans bear the responsibility and accountability for the results of their actions during wars and "that accountability could not be transferred to machines" [11]. A real-world example demonstrated this concern when AWS targeted a significant number of civilians, raising questions about whether these incidents were the result of malfunctions within the AWS or decisions made by their human operators [18, 19].

Regulating AWS While there are no IHL treaties on the usage of AWS, article 36 of Protocol I [20] obliges treaty members who study, develop, acquire, or adopt new weapons or means of warfare to determine whether these weapons are in breach of the protocol or rules of IHL. The Protocol is signed by 59 states and has ratification/accession of 174 States; while Article 36 is considered by some experts as a customary international law, which binds all states [20, 12].[2] This means that even states developing and studying AWS without carrying out a legal review of its compliance with IHL fall short of their international law commitments, let alone those who use them. Outside the context of IHL, there is an increasing number of tools and frameworks for assessing and mitigating potential ethical and social risks of AI systems. The US Algorithmic Accountability Act [22] is a proposed law in the United States that would require companies to conduct impact assessments of their algorithms to identify and mitigate any potential biases or discriminatory effects. NTIA Artificial intelligence Accountability Policy Report [23] defines an AI accountability framework where one requites is to recognize potential risks and harm including intended harm to humans and human right violations. The report also mentions AI actors putting AI systems out in the world should ensure that systems perform as claimed in a trustworthy manner. In that case, it is not clear whether developers, governments or end users are to blame, while AWS does not even pass the requisites mentioned in the report. Practically, using AWS can make prosecution of war crimes impossible since most of the crimes stipulated in the Rome Statute [14] must be done with intent. How can human intent be proved if the attack was determined by a machine? As a result, current regulations and International laws are not effective for the use of AWS around the world.

We call on nation-states to domesticate international law by enforcing regulations on the development and use of AWS, ensuring compliance with the principles of distinction, proportionality, and accountability. Furthermore, the scientific community must recognize the ethical implications of their research in AI and autonomous technologies. Researchers developing AI systems need to consider how their innovations could be weaponized or lead to unintended civilian harm. Ethical oversight in AI research is essential to prevent the misuse of these technologies and to ensure that AI systems are designed with safety, transparency, and accountability in mind. In conclusion, we highlighted the gap between the current state of AI use in military applications and international laws, and showed the challenges to legalize AWS. In future work, we will analyze, more technically, how current AI systems fail to align with international humanitarian law and hence unfit for deployment in warfare.

[2] Some scholars are, however, contend that Article 36 of Protocol I is not a part of the customary international law [21].

References

[1] Yoshua Bengio. Yoshua bengio, major tech leaders call for six-month pause on advanced ai development in open letter, 2023.

[2] International Scientific Community. International scientific report on the safety of advanced ai, 2023.

[3] Riley Simmons-Edler, Ryan Badman, Shayne Longpre, and Kanaka Rajan. Ai-powered autonomous weapons risk geopolitical instability and threaten ai research. 2024.

[4] Stuart Russell. AI weapons: Russia's war in Ukraine shows why the world must enact a ban. *Nature*, 614(7949):620–623, February 2023. Bandiera_abtest: a Cg_type: Comment Publisher: Nature Publishing Group Subject_term: Machine learning, Engineering, Computer science, Technology, Policy.

[5] Samsung Techwin SGR-A1 Sentry Guard Robot.

[6] Russian Kamikaze Drone 'Lancet' Empowered by AI for Autonomous Targeting, October 2023.

[7] 972mag Editorial. Mass assassination factory: Israel's calculated bombing of gaza, 2023.

[8] 972mag Editorial. Lavender ai and the israeli army in gaza, 2023.

[9] BBC News. Gaza surgeon describes drones targeting children, November 2024. Accessed: 2024-11-20.

[10] Antony Loewenstein. The palestine laboratory: How israel exports the technology of occupation around the world, 2023.

[11] United Nations General Assembly. Report of the secretary-general on lethal autonomous weapons systems, 2024.

[12] Human Rights Watch and International Human Rights Clinic. Losing humanity: The case against killer robots, 2012.

[13] Neil Davison. Autonomous weapon systems under international humanitarian law. Technical report, International Committee of the Red Cross (ICRC), 2018.

[14] Rome statute of the international criminal court: Article 8. `https://www.law.cornell.edu/women-and-justice/resource/rome_statute_of_the_international_criminal_court_article_8`.

[15] BBC News. The impact of artificial intelligence on modern warfare. November 2024.

[16] Ihl database. rule 9. definition of civilian objects. `https://ihl-databases.icrc.org/en/customary-ihl/v1/rule9`.

[17] Military necessity (glossary entry). `https://casebook.icrc.org/a_to_z/glossary/military-necessity`.

[18] New Interactive Map Gives Ground-Level Look at High Civilian Toll of Israel's Gaza War | Common Dreams.

[19] Gaza civilian deaths test Israel's AI precision claims, March 2024. Section: Middle East.

[20] Protocol additional to the geneva conventions of 12 august 1949, and relating to the protection of victims of international armed conflicts (protocol i). `https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/state-parties`, 1977.

[21] U.S. Naval War College International Law Studies. Autonomous weapons and international law: A complex debate. 2024. Accessed on November 20, 2024.

[22] U.S. Congress. Algorithmic accountability act of 2023, 2023.

[23] National Telecommunications and Information Administration (NTIA). Ntia ai accountability policy report, 2024.

[24] Ihl database. rule 156. definition of war crimes. `https://ihl-databases.icrc.org/en/customary-ihl/v1/rule156`.

1 Appendix

Definition of War crimes: Drawing on article 8 of the Rome Statute, war crimes can be defined as a) "Grave breaches of the Geneva Conventions", b) "In the case of an armed conflict not of an international character, other serious violations of the laws and customs applicable in international armed conflict, within the established framework of international law," c) "Other serious violations of the laws and customs applicable in armed conflicts not of an international character, within the established framework of international law" [14]. It is worth noting that some states' legislation amount any violation of IHL to be a war crime [24].
Martens Clause: The Martens Clause, which is fundamental in IHL and adopted in Protocol I, stipulate that aspects of warfare should follow the laws of humanity and "the dictates of public conscience" [11, 12, 13]. AWS raises many concerns when it comes to The Martens Clause as machines will not be able to use concepts and laws of humanity [13].