Should autonomous weapons be allowed under international law?

Flore Dikkers (MLaw student, Université de Fribourg)

Les conflits armés font partie de l'existence humaine. Le droit international humanitaire est né dans la deuxième moitié du XIXe siècle. Depuis lors, de plus en plus d'instruments ont été créés afin de tenter d'établir un équilibre entre l'humanité et les principes militaires. Un certain nombre de principes clés ont été établis afin de veiller à ce que les effets des conflits armés ne soient pas trop dévastateurs. En parallèle, la technologie a progressé et les armes aussi. La possibilité de disposer d'armes entièrement autonomes est devenu une menace imminente pour les conflits dans le monde entier. Ces armes posent de nouveaux défis notamment en termes de proportionnalité et de responsabilité. Est-ce raisonnable d'autoriser de telles armes ? Ces questionnements s'élèvent la nécessité d'examiner la compatibilité de ces nouvelles technologies d'armement avec les principes de droit international humanitaire existant, ceci afin de déterminer si elles pourront s'y conformer et, par extension, comment elles pourront être utilisée dans le futur.

Introduction

The law on armed conflict, also known as International humanitarian law (IHL), accepts the fact that war is a part of human existence. Instead of outlawing conflict, it seeks to establish a balance between humanity and military principles[1]. The law on armed conflict evolves in response to experiences with new types of warfare. As weapons and other means of warfare continue to be developed so does the legislation[2]. A party is not free to choose whichever means of warfare they want[3]. IHL is largely laid down by the four 1949 Geneva Conventions and its two additional protocols[4], and the Law of The Hague[5], as well as by customary international law[6]. Customary international law is unwritten , it derives from widespread practices that become accepted as laws[7].

Over the years, IHL has prohibited or restricted the use of certain weapons[8], an example of such legislation is the Chemical Weapons Convention[9]. Should the same apply to autonomous weapons? In this paper we will review the definition of such weapons and explore whether or not they should be allowed in international law.

We will begin by defining autonomous weapons and differentiate them from other types of weapons (in-

[1] A. Henriksen, *International Law*, 3rd edn, Oxford 2021, p. 274 ; International Expert Meeting, *The Principle of Proportionality in the Rules Governing the Conduct of Hostilities Under International Humanitarian Law* (report edited by Laurent Gisel (ICRC)), Quebec 2016, p. 5.

[2] Henriksen, p. 274.

[3] S. Casey-Maslen, *Weapons*, in: Ben Saul and Dapo Akande (eds), *The Oxford Guide to International Humanitarian Law*, Oxford 2020, p. 263.

[4] Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 1949 (cited: GCI) ; Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, 1949(cited: GCII) ; Geneva Convention (III) relative to the Treatment of Prisoners of War, 1949 (cited: GCIII) ; Geneva Convention (IV) relative to the Protection of Civilian Persons in Time of War, 1949 (GCIV) ; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1977 (cited: AP I) ; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), 1977 (cited: AP II).

[5] Hague Convention (I) on Hospital Ships, 1904 ; Hague Convention (II) on the Laws and Customs of War on Land, 1899 (cited: HC II) ; Hague Convention (III) on the Opening of Hostilities, 1907 ; Hague Convention (IV) on War on Land and its Annexed Regulations, 1907 ; Hague Convention (V) on Neutral Powers in case of War on Land, 1907 ; Hague Convention (VI) on Enemy Merchant Ships, 1907 (cited : HC VI) ; Hague Convention (VII) on Conversion of Merchant Ships, 1907 ; Hague Convention (VIII) on Submarine Mines, 1907 ; Hague Convention (IX) on Bombardment by Naval Forces, 1907 ; Hague Convention (XI) on Restrictions of the Right of Capture, 1907 ; Hague Convention (XIII) on Neutral Powers in Naval War, 1907 ; Hague Convention for the Protection of Cultural Property, 1954.

[6] Casey-Maslen, p. 263.

[7] International Committee of The Red Cross, *Customary international Law*, 29 October 2010, in : <https://www.icrc.org/en/document/customary-international-humanitarian-law-0> (accessed 25.5.2024).

[8] *Idem*, p. 261-262.

[9] Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction, 1993 (cited: CWC).

fra I). Subsequently, we will review the existing legal framework surrounding the use of weapons (infra II) and finally, consider the problems stemming from the possibility of having and using autonomous weapons (infra III).

I. Defining autonomous weapons

Weapons are essential to the use of force and yet neither the Geneva Conventions nor the rules of the Hague provide a universally accepted definition of what a weapon under international law is. "Weapons" seems to be a rather broad term encompassing knives as well as guns or explosive devices[10]. Some instruments, such as the previously mentioned Chemical Weapons Convention, include articles that explicitely define the weapons they cover (art. 2 CWC).

With the progress of weapon technology, we have reached a stage where the person launching the weapon no longer needs to be physically present at the location of the deployment[11]. There are multiple levels of autonomy a weapon can have. The level of autonomy can be quite limited to where it simply involves the weapon's ability to return to its base if a malfunction occurs[12]. Higher levels of autonomy exist in remotely controlled systems such as armed drones, where the operator can remotely release a missile or projectile while being nowhere the location of the attack[13]. For example, the "Reaper", which was until recently considered the world's deadliest drone, with its ability to self-navigate, find, and attack targets using laser guided technology[14].

Currently we are entering a new era of technology: the development of fully autonomous weapons[15]. A fully autonomous weapon can be defined as a weapon that, once engaged, does not require any human intervention to select targets and attack[16]. An automatic weapon system (AWS) uses AI to analyze information from the surrounding environment received through sensors and utilizes a pre-defined target profile to carry out the required action. The key differentiation between this category of weaponry and others is that the operator does not choose or even know when or what the autonomous weapon will attack[17]. In 2021, a report by the United Nations Panel of Experts on Libya recorded the use of lethal AWS engaging retreating convoys. These were programmed to strike targets without needing data connectivity between the operator and the munition[18].

II. Legal Framework

In the Advisory Opinion on nuclear weapons, the International Court of Justice (ICJ) referred to two cardinal principles of IHL : the principle of distinction and the prohibition against unnecessary suffering[19]. Both of these refer to the basic principle of international law that states that the "right of belligerents to adopt means of injuring the enemy is not unlimited" (art. 22 HC IV).

The principle of distinction (arts. 48 and following AP I) obliges parties to distinguish between on one side, combatants, and military objectives, and on the other, civilians and civilian objectives[20]. The definition

[10] Casey-Maslen, p. 261.

[11] C. Heyns, *Autonomous weapons in armed conflict and the right to a dignified life: an African perspective*, South African Journal on Human Rights (2017) 33:1, p. 47.

[12] Human Rights Watch, *Losing Humanity : The Case against Killer Robots*, 2012, p. 6.

[13] Heyns, p. 47.

[14] H. Brollowski, *Military Robots and the Principle of Humanity: Distorting the Human Face of the Law?*, in: Mariëlle Matthee et al. (eds.), *Armed Conflict and International Law: In Search of the Human Face*, The Hague 2013, p. 62.

[15] Heyns, p. 47.

[16] K. Anderson / M.C. Waxman, Debating autonomous weapon systems, their ethics, and their regulation under international law, in: Roger Brownsword, Eloise Scotford and Karen Yeung (eds.), *The Oxford Handbook of Law, Regulation and Technology*, Oxford 2017, p. 1100 ; Henriksen, p. 288 ; Heyns, p. 46 ; L. Lema, *La guerre en Ukraine durcit les discussions sur les "robots tueurs*, *Le Temps,* Geneva, 2022.

[17] International Committee of the Red Cross, ICRC Position on Autonomous Weapons Systems, S/2021/229, 2021, <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems> (accessed 25.9.2023).

[18] UN Security Council, Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011), 8 March 2021, in :<https://documents.un.org/doc/undoc/gen/n21/037/72/pdf/n2103772.pdf?token=HHdl4VqXa4nraN6cyB&fe=true> (accessed 25.5.2024).

[19] Legal Consequences of the Use of Nuclear Weapons, Advisory Opinion, ICJ Reports (1996) 226, para 78.

[20] Boulanin et al., p. 4 ; E. Cannizzaro, Proportionality in the law of armed conflict, in: Andrew Clapham and Paola Gaeta (eds.), *The Oxford Handbook of International Law in Armed Conflict*, Oxford 2010, p. 335 ; Casey-Maslen, p. 263 ; Henriksen, p. 285.

of military objectives is found in art. 52(2) of AP I. If there is any doubt on the qualification of an object, it must be assumed that it is a civilian object[21]. Attacks on military objectives are permitted as long as the damage to civilian and/or civilian objects (collateral damage or incidental harm) is not excessive in comparison to the expected military advantage (art. 51 (3) AP I). It must be proportionate since indiscriminate and disproportionate attacks are prohibited[22]. Even if a weapon is targeted at a military objective the rule of proportionality must be respected[23]. The principle of proportionality is codified in art. 51(5)(b) AP I and complements the principle of distinction[24]. The assessment of the respect of this principle depends on value judgment at context specific times of attack[25]. This principle seeks to reduce the impact of armed attacks on protected persons and civilians[26].

The prohibition against unnecessary suffering forbids States to use weapons which cause unnecessary harm or uselessly worsen suffering[27]. This means parties do not have the right to use any means of warfare[28], as explicitly prohibited by art. 23 HC IV, as well as art. 35(2) AP I.

In accordance with both of these principles, weapons with effects that cause harm greater than the unavoidable are prohibited[29]. When a new weapon is deemed to cause excessive harm, states have come together to seek to restrict and/or prohibit these weapons using treaty negotiations[30]. States have an obligation to review new weapons and determine whether they are compliant with international law as codified by art. 36 AP . In order for a weapon to be allowed, it must be both legal and used in a legally compliant manner[31]. It is also important to note that a violation of IHL does not have to result from an obligation contained in an existing treaty or convention[32]. Experts agree that this clause, known as the Martens Clause[33], can be applied to weapon law as well[34]. The clause was first introduced in the preamble of the 1899 HC II. It was proposed by Fyodor Fyodorovich Martens to address the concerns of small states and avoid deadlock. Indeed, during the negotiations, small states objected to articles favoring occupying powers, as they were more likely to be occupied rather than occupiers[35]. The ICJ also expressed that the Martens Clause "has proved to be an effective mean of addressing the rapid evolution of military technology"[36].

International criminal law is a fairly recent aspect of international law. Its goal is to make sure that individuals responsible for heinous acts are held liable in front of national or international courts[37]. The primary sources of international criminal law are treaties establishing international courts, notably the 1998 Rome Statute[38]. Four crimes are widely recognized as binding on individuals under customary international law: genocide, crimes against humanity, certain war crimes and aggression[39]. Accountability serves several purposes notably deterring future harm and providing a sense of retribution to victims[40]. Indeed, the right to life is not only violated due to arbitrary

[21] Henriksen, p. 286.
[22] Boulanin et al., p. 4.
[23] Casey-Maslen, p. 264 ; International Expert Meeting, *The Principle of Proportionality in the Rules Governing the Conduct of Hostilities Under International Humanitarian Law* (report edited by Laurent Gisel (ICRC), Quebec 2016), p. 8.
[24] Y. Zerbe, *Autonomous Weapons Systems and International Law: Aspects of International Humanitarian Law, Individual Accountability, and State Responsibility*, 2019, Swiss Review of International and European Law 581, p. 588.
[25] Henriksen, p. 286 ; Heyns, p. 54 ; Losing Humanity, p. 332 ; Zerbe, p. 584.
[26] Heyns, p. 52.
[27] Nuclear Weapons, Advisory Opinion, ICJ, para 78.
[28] Casey-Maslen, p. 263.
[29] Nuclear Weapons, Advisory Opinion, ICJ, para 78 ; Henriksen, p. 287.
[30] Casey-Maslen, p. 265.
[31] Zerbe, p. 584.
[32] International Committee of the Red Cross , *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, Int'l Review of the Red Cross, 2008, p. 17 ; Boulanin et al., p. 11 ; Zerbe, p. 591.
[33] Boulanin et al., p. 11 ; Zerbe, p. 591.
[34] International Committee of the Red Cross, *A Guide to the Legal Review of New Weapons*, p. 17 ; T. Evans, *At War with the Robots: Autonomous Weapons Systems and the Martens Clause*, Hofstra Law Review, 2013, 700 ; Zerbe, p. 592.
[35] V. Koutroulis, Martens Clause, 2023, in : <https://www.oxfordbibliographies.com/display/document/obo-9780199796953/obo-97801997969530101.xml#:~:text=The%20clause%20was%20introduced%20for,the%201899%20International%20Peace%20Conference.> (accessed 26.5.2024).
[36] Brollowski, p. 70 ; Nuclear Weapons, Advisory Opinion, para 79 ; Zerbe, p. 592.
[37] Henriksen, p. 307.
[38] *Idem*, p. 308.
[39] Henriksen, p. 315.
[40] Losing Humanity, p. 42.

killings, but also when there is an absence of accountability[41].

III. Problems raised by autonomous weapons

The challenges created by autonomous weapons have been the focus of interstate discussions for close to a decade[42]. There seems to be an emerging consensus between states that autonomy in weapons should not be unlimited[43]. Some states, the majority of them being African or South American, are pleading for a total prohibition. Others, notably NATO member states, disagree with the need for a binding legal framework[44]. The International Committee of the Red Cross (ICRC) supports initiatives aimed at establishing international limits on AWS[45]. In October 2023, the Secretary General of the United Nations as well as the President of the ICRC called on political leaders to immediately creates new international rules on AWS. They argued that it was necessary for the protection of humanity as human control needs to be retained in life and death decisions. In their joint appeal, they advocate for clear restrictions to ensure that AWS comply with international law and ethical concerns, notably by limiting the location and the timing of their usage and ensuring the possibility of effective human oversight[46].

One key issue to address is the principle of distinction. How would an autonomous weapon distinguish between soldiers and civilians in complicated combat situations? The *Losing Humanity* Report cites the example of modern-day combat environments. In said environments, fighters do not necessarily look like soldiers and tend to blend with the civilian population and can only be identified by their "*direct participation in hostilities*" [47]. It is impossible to be certain that a fully autonomous would have the sensors and coding required to make those differences[48]. AWS lack the human emotions necessary to understand people and distinguish between specific groups. Another example, cited by the same report, is one of a mother running after her two children playing with toy guns next to soldiers. A human soldier would be able to understand the situation and hold fire, but the autonomous weapons might register it as a threat. Indeed, seeing a person running towards it followed by two armed individuals seems like a situation where force has to be used[49]. I believe those examples, already give too much credit to humans. Even for us, distinguishing the differences between specific groups can be challenging, This highlights how big of challenge it would be to code the ability to make said differences in AWS. Indeed, one might argue that AWS, unlike humans, do not have a sense of self-preservation. In a combat situation where a soldier is confronted with someone who cannot be immediately identified as combatant or civilian, the soldier may shoot immediately in self-defense, whereas an AWS could act in a more restrained manner and potentially save a life[50].

A second aspect to consider is AWS in relation to proportionality. Compliance with this principle depends on value judgements and estimates in specific contexts. Can autonomous weapon systems make these proportionality evaluations correctly[51]? These evaluations often rely on typically "human" qualities such as common sense, morality, or good faith. If those are not accurately translated into codes, how can the proportionality test done by autonomous weapons be accurate[52]? It is unlikely an AWS could be programmed to understand and handle the infinite number of different potential scenarios[53]. Indeed, the computer

[41] Heyns, p. 56.

[42] V. Boulanin et al., *Limits on Autonomy in weapon systems : Identifying Practical Elements of Human Control*, SIPRI 2020, p. 13.

[43] Boulanin et al., p. 13.

[44] L. Lema, Les Conventions de Genève, un rempart face à la barbarie ? « Nous n'avons rien d'autre », 12 August 2024, in: <https://www.letemps.ch/monde/les-conventions-de-geneve-un-rempart-face-a-la-barbarie-nous-n-avons-rien-d-autre> (accessed 26.8.2024).

[45] ICRC, Position on Autonomous Weapons Systems.

[46] United Nations, Note to Correspondents: Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems, 5 October 2023, in: <https://www.un.org/sg/en/content/sg/note-correspondents/2023-10-05/note-correspondents-joint-call-the-united-nations-secretary-general-and-the-president-of-the-international-committee-of-the-red-cross-for-states-establish-new> (accessed 25.5.2024).

[47] Losing Humanity, p. 30 : Zerbe, p. 586 ; see also : Brollowski, p. 79 n° 167.

[48] Heyns, p. 53 ; Losing Humanity, p. 31.

[49] Losing Humanity, p. 30-31.

[50] Heyns, p. 53 ; Zerbe, p. 587.

[51] Boulanin et al., p. 6 ; Heyns, p. 54 ; Losing Humanity, p. 32.

[52] Boulanin et al., p. 5 ; Heyns, p. 54 ; Zerbe, p. 588.

[53] Heyns, p. 48 ; Boulanin et al., p. 5 ; Losing Humanity, p. 32 ;

science professor Noel Sharkey has stated that programming an autonomous weapon with enough reactions to the infinite array of possibilities that could arise during a conflict is impossible. Similarly, if an autonomous car can be rendered useless by snow blocking its sensors, could the fog of war render an AWS unpredictable[54]. The test of proportionality is not just a simple matter of balancing data. This is why it is doubtful that AWS could replicate the human judgment needed to assess proportionality[55]. Akin to proportionality, military necessity requires the same subjective analysis that AWS are unlikely to be able to comprehend[56]. However, in this case AWS do have an advantage over human brains: their information processing and probabilities calculating capabilities. Them not having the capacity for emotions removes the risk of miscalculations induced by the stress of combat environments[57].

Another notable issue is coded bias. There is a growing recognition than an algorithm is not neutral and thus can exhibit bias when operating. These biases can stem from different reasons, notably interpretation bias or inappropriate training data[58]. Racism pervades our society and is thus articulated through what we make and then encode in AI models[59]. Indeed, there is an increasing amount of evidence that makes it clear that racism impacts the functioning of AI[60]. A notable and dangerous bias found in artificial intelligences (AI) is a racial bias. As example of this, a 2016 ProPublica study which examined predictive recidivism and analyzed the scores of 7000 individuals over two years. It revealed that the software was biased against African Americans, giving them a 45% higher reoffending risk compared to white offenders of the same gender, age and criminal record[61]. Using biased AI in AWS makes it evident how long-standing biases pose ethical and human rights threats, rendering some groups significantly more vulnerable than others. In this context, AWS would not only reinforce existing inequalities but could also worsen them which could lead to deadly consequences[62].

There is also a possibility that those weapons could be willingly programmed to look for a certain criterion before attack and thus become biased[63]. Indeed, there is a possibility that leaders might turn such weapons against their own people or use them to genocidal ends[64].

Lastly, we need to address the crucial issue of accountability. These weapons are, by definition, autonomous, so who is to be held legally and morally accountable when a mistake happens, and a war crime is potentially committed? It is futile to prosecute AWS as it is by no means a moral agent[65]. There are a several options for attributing responsibility, for example to the military commander, to the manufacturer, or to the programmer[66]. Some scholars have suggested that autonomous weapons and their algorithmic agents should be traced so that any reckless or negligent actions can be linked to the weapon's owner or controller[67]. The issue is that the autonomy creates a responsibility gap, and it seems unjust to hold people accountable for actions stemming from autonomous systems over which they have no complete control[68].

Zerbe, p. 588.

[54] F. Pasquale, *Machines set loose to slaughter: the dangerous rise of military AI*, The Guardian, London 2020.

[55] Brollowski, p. 79-80 ; Losing Humanity, p. 33.

[56] Losing Humanity, p. 33-34.

[57] Zerbe, p. 589.

[58] United Nations Institute For Disarmament Research, *Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies*, 2018, p. 2-3.

[59] J. McCrosky, *AI Weapons Could Risk Racist Decisions*, DataEthics, 2024, <https://dataethics.eu/ai-weapons-could-risk-racist-decisions/> (accessed 26.4.2024).

[60] H. Ramsay-Jones, *Racism and Fully Autonomous Weapons* (Submission to the UN Special Rapporteur regarding the thematic report on new information technologies), 2019, p. 1.

[61] J. Angwin / J. Larson / S. Mattu / L. Kirchner, Machine Bias, ProPublica, 2016, in: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed 26.5.2024).

[62] H. Ramsay-Jones, *Intersectionality and Racism*, Soka Gakkai International, 2020, p. 27, in: <https://www.stopkillerrobots.org/wp-content/uploads/2021/09/Intersectionality-and-Racism-Hayley-Ramsay-Jones.pdf> (accessed 26.5.2024) ; Ramsay-Jones, *Racism and Fully Autonomous Weapons*, p. 3.

[63] L. Mieno, *'Killer Robots' are coming, and U.N. is worried*, The Harvard Gazette, 2024, in : <https://news.harvard.edu/gazette/story/2024/01/killer-robots-are-coming-and-u-n-is-worried/> (accessed 26.5.2024)

[64] P. Asaro, *Autonomous Weapons and the Ethics of Artificial Intelligence*, in: S.M. Liao (ed.), *Ethics of Artificial Intelligence*, Oxford 2020, p. 212-236.

[65] Losing Humanity, p. 57.

[66] *Idem*, p. 42.

[67] Pasquale, The Guardian (n. 55).

[68] Losing Humanity, p. 42 ; A. Matthias, The responsibility

It is difficult to see how accountability could be attributed without actual human control[69]. Perhaps the simple solution that the person who orders the use of the AWS will be responsible for its actions.

Conclusion

The question as to whether autonomous weapons are allowed under international law is a complex issue raising notable legal (and ethical) challenges. To attempt to provide a satisfying answer to this question we began by attempting to define autonomous weapons (*supra* I), then we examined the existing legal framework (*supra* II) and finally we considered the dilemmas raised by applying those rules to AWS (*supra* III).

So, should AWS be allowed in international law? There is currently no evidence that they will be able to adhere to the principles outlined above. In view of actual developments and the current level of autonomy AWS possess, I would argue that they should not be permitted under international law because there are no concrete guarantees that they will be able to function according to the aforementioned IHL principles. Of course, the analysis of their compliance with international law is highly speculative[70]. If autonomy advances to a point where the aforementioned problems are resolved, AWS could become a valuable tool for protecting human lives in war. We do not know how, for example, how AI will evolve and if it will be a solution for the problems currently faced by AWS. Indeed, one might say that AWS with strong AI could be a solution to this, but it seems likely that militaries will introduce AWS before the required AI abilities are developed[71]. Besides, humans are not perfect in combat situations either, so how can we ensure that the behaviour we program into weapons will be appropriate. As long as we are not certain that AWS can hold up in combat, I strongly believe we should not take the risk. Additionally, until AWS benefits from a clear legal structure, it seems best to avoid employing it, especially regarding accountability. Without the rule of law, impunity reigns. People need to be held accountable for their crimes, whether they are committed by proxy or not. Without that guarantee, we should not move forward. It will be interesting to see if we do get a treaty on AWS in the next few years and to see how the legislators will formulate the rules for such weapons.

Gap: Ascribing Responsibility for the Actions of Learning Automata, *Ethics and Information Technology*, vol. 6, 2004, 176, p. 183.

[69] Heyns, p. 57.

[70] Zerbe, p. 595.

[71] Losing Humanity, p. 34.