Artificial intelligence as a new actor of international security? Assessing the perspective of "non-human autonomy" under international law.

Théo Antunes

The use of artificial intelligence in the context of armed conflicts or to guarantee security has always nurtured the human mind, from literatures to movies and video games contents, such perspective has been extensively pictured. What belonged to science fiction is turning into reality due to the recent developments in the artificial intelligence technology. However, the question lies on why such technology fascinates as much as it frightens. Why does artificial intelligence bring these many controversies? Out of its many unique features, its autonomy is the feature that bring dire legal challenges.

The capacity for artificial intelligence to be autonomous implies that, with a set of information (input), it is capable to produce a result (output) without any human intervention[1]. Such autonomy is even more enhanced by the incapacity for humans to access the reasoning the software in its most advanced shape (such as deep learning)[2]. The challenge that lies with such technology is to frame such autonomy, even in 2023, this question is not easily answered. Who should be held responsible, at what extent, what kind of responsibility and what legal framework to apply are common questions that surrounds the use of artificial intelligence.

If many artificial intelligent software is built to assist human decision-making; artificial intelligence in the context of international security is not limited to assisting decision making. In the context of armed conflict, applications of artificial intelligence software are directed toward devices that can directly take part in an armed conflict.[3] It does not aim at assisting human decisions; but to take one by itself based on its environment. The use of this technology takes place in a wider context of the use of unregulated entities under international law by states.

[1] Vladimir Nasteski, An overview of the supervised machine learning methods, (2017), Faculty of Information and Communication Technologies, Partizanska, 1-11, 4; Michael Wheeler, Autonomy, *The Oxford Handbook of Ethics of AI,* Oxford University Press, 719-736, 726

[2] Burrell 'How the machine 'thinks': Understanding opacity in machine learning algorithms' (2016), Big Data & Society, 3, 1, 1–12, 10.

[3] Gabriele Reitmeier, Artificial intelligence in weapon systems and new challenges for arms control, (2020), Friedrich Naumann Foundation for Freedom, Department of Global Topics / International Affairs, 1-22, 5.

Since the beginning of the XXI century, States have increasingly relied on irregular forces to evade their responsibility[4]. The latest developments demonstrate uses of private military companies to act as their extended arm while the mechanisms to attribute their conduct to states are challenging[5]. The use of artificial intelligence could be another mean for States to circumvent their responsibility by deploying "autonomous" agents on the battlefield and in the cyber-space. To attribute conducts to States, international law requires a strict application of legal mechanisms where autonomy lies at the core for non-attribution to States.

However, a thorough study of the intrinsic features of artificial intelligence questions the relevance of applying the term autonomy to such devices as well as the legal consequences such qualification bring. The concept of autonomy as it has emerged in international law had been erected for other purposes. To label artificial intelligence as "autonomous" could thus be misleading regarding its functioning and regarding the legal mechanisms that could apply. As such, this article will assess the extent of this "non-human autonomy" and its implications for attributing its conduct to states under international law and its impact on the law of armed conflicts.

Such perspective remains challenging under international law. This paper will firstly analyse the concept of autonomy and clarify its meaning, its features and purpose under international law (I). Secondly this article will analyse the impacts of the extent of autonomy for artificial intelligence (II). Finally, this paper will assess that this "non-human autonomy" limits the application of international law and further fragilize the legal mechanisms for attribution while proposing alternative pathways (III).

I. The concept of autonomy under the international law

Autonomy is a complex concept under international law. It is challenging to concretely grasp its meaning as it can vary depending on the entity for which it applies. Nevertheless, under the international law framework, the concept of autonomy is of great importance, notably in the context of international responsibility and the law of armed conflict.

Allegedly, artificial intelligence relies on a certain extent of autonomous decision making, its inner functioning allows it to receive input, interpret them and decide accordingly. It would thus be freed from human control as it can alone take a real-time decision. However, to fully grasp the stakes of artificial intelligence autonomy, one needs to step back and assess what autonomy encompasses under international law. As

[4] Marcus Rediker, 'Villains of All Nations: Atlantic Pirates in the Golden Age', (2004), Boston: Beacon press, 1- 248, 28; Wilhelmy, Jean-Pierre, 'Soldiers for Sale: German "Mercenaries" with the British in Canada during the American Revolution (1776-83)' (2012), Baraka Books, 296; Peter W Singer, 'Corporate Warriors: The Rise of the Privatized Military Industry',(2003) Cornell University Press, 1-360, 9.

[5] Théo Antunes, Challenge of Regulation of Private companies in international security operations, (2020), Research thesis, Université Catholique de Lille, 1-173, 133.

such one needs to look at the framing of the concept (A); and its rationale under international law (B).

A. Framing the concept of autonomy under international law

Autonomy encompasses a wide range of actors under international law as it can encompass territories, legal entities, and human conducts[6]. In international law, the mainstream use of the concept of autonomy refers to the right of states and its people to self-determination free from any other states intervention in their development[7]. Such concept also encompasses human conducts in the context of acts by individuals acting on behalf of others. In such approach, autonomy encompasses the capacity of individuals to act freely from external pressure or influence[8]. One can see that the common point between these two approaches of autonomy is the capacity of an entity, to freely decide by its own of its own path. In this context, autonomy is relevant in assessing the international responsibility of states for the wrongful conducts of non-state entities[9]. These entities can range from other states to individuals and armed groups and theoretically every non-state entity[10].

Under such framework, international law requires a link to the State for such conduct to fall under its scope as it does not have a direct effect on other entities except international organizations[11]. Enjoying autonomy from a State would ultimately mean that the link is severed, thus averting the possibility of applying international law. Hence, the more a non-state entity is enjoying autonomy from a State the less international law would be directly applicable[12]. As such, a territory that is not autonomous, meaning subjected to a state control, would mean that the controlling is responsible in cases of international wrongful acts[13]. Autonomy is a concept originally attached to states activities, as a mean to frame their conduct and to assess their

[6] Hurst Hannum and Richard B. Lillich, the concept of autonomy in international law, (1980), the American journal of international law, 74, 858-889, 860.
[7] Zubeida Mustafa, The Principle of Self-Determination in International Law (1971), International lawyer, 5, 3, 479-488, 484
[8] Conseil Consultatif des juges européens (CCJE), Magna Carta des juges, (2010), adoptée à sa 11ème réunion plénière le 17-19 novembre 2010, point 4 ; CJUE, *l'Associação Sindical dos Juízes Portugueses c. Tribunal de* Contas, arrêt du 27 février 2018, C-64/16, para 44.
[9] James Crawford, the international law commission's articles on state responsibility, (2005), third edition, Cambridge university press, 111.
[10] International Law Commission, Articles on the Responsibility of States for their International Wrongful Act, A/56/49(Vol. I)/Corr.4 Article 4, Article 5, Article 7, Article 8, Article 11.
[11] Reparation of Injuries Suffered in Service of the U.N., Advisory Opinion, 1949 I.C.J. 174 (Apr. 11), p9
[12] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America), Jurisdiction and Admissibility, Judgment, I.C.J. Reports 1984, p. 392 para 114; Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43, para 394
[13] International Law Commission, Articles on the Responsibility of States for their International Wrongful Act, A/56/49(Vol. I)/Corr.4 Article 17.

existence on the international scene. As international law is primarily the law of states, it was deemed logical for such concept to only encompass these entities[14].

Nevertheless, the evolution of international law witnessed the inclusion of other legal entities; especially the rise of private companies[15]. International law recognises a certain degree of autonomy of these entities because of their private status thus not automatically attributing their conduct to States[16]. However, in the context of their regulation, international law imposes positive obligations on states in order for private companies to adopt minimum standards of international law, mostly stemming from human rights[17]. One would assess that international law only applies indirectly to these companies pending the good will of States to enact legal frameworks in their domestic jurisdictions.

Autonomy from States thus prevents a direct application of the provisions of international responsibility and by extension, all international obligations. The conduct, no matter how wrongful would not bring the international responsibility of the State. A state might be found responsible for an imperfect framework; but not for the conduct of the legal entity itself.

B. The rationale behind autonomy in international law

The consideration of autonomy under international law is vital for assessing the responsibility framework. As such, the link between autonomy and responsibility must be deepened (1) whereas the concept itself, applied in the context of international responsibility, shares different conceptions (2).

1. Autonomy at the core of assessing international responsibility.

The concept of autonomy allows to determine whether international law can be extended to other entities than States and international organization[18]. Such concept becomes crucial in assessing their international responsibility. International law thus binds the concept of autonomy and attribution altogether under the international responsibility framework. One could assess what would be the extent of autonomy to bring the responsibility of the state under international law?

[14] Reparation of Injuries Suffered in Service of the U.N., Advisory Opinion, 1949 I.C.J. 174 (Apr. 11), p9; UN, 'Statute of the International Court of Justice', 18 April 1946, Article 34;
[15] Karsten Nowrot, 'Reconceptualising International Legal Personality of Influential Non-State Actors: towards a Rebuttable Presumption of Normative Responsibilities' (2005) Philippine Law Journal, Volume 80, 563-586, 579.
[16] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43, para 406
[17] Office of the High Commissioner, Guiding Principles on Business and Human Rights, (2011), UN New York and Geneva, Article 14; Swiss Government, 'International Code of Conduct for Private Security Service Providers' (2010), 5
[18] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America), Jurisdiction and Admissibility, Judgment, I.C.J. Reports 1984, p. 392 para 112.

International responsibility is primarily built upon the violation of an international obligation by state organs, whether *de jure* or *de facto*[19]. Agents of states are always considered as acting under its authority when they act in that capacity; thus, they are not considered autonomous under international law[20]. Such approach is also shared by the law of armed conflicts framework, where military officials and combatants' conducts are always attributable to the State[21]. Hence, public agents incorporated within the State architecture are not qualified as autonomous when they act in this capacity. The challenge of autonomy under the international responsibility framework lies in the context where non-state entities are used by States to perform activities in its service. The international responsibility framework is built on a *a contrario* basis. This means that non-state organs and agents are considered as autonomous until demonstrated otherwise[22]. Hence, it asserts that, in principle, the conduct of non-states entities will not be attributable to the State[23]. The exception lies where non-state entities are acting as state agents or are under their effective and direct control.

The attribution test set forth by the International Court of Justice (ICJ) entails two stakes: whether a non-state entity can be assimilated to a state organ; or whether it is under the effective and direct control of a State.

The ICJ indicated that "*Certain acts of individuals or entities which do not have the status of organs of the State may be attributed to the State in international law*"[24]. Such perspective is also shared in the law of armed conflict considering that "*The armed forces of a Party to a conflict consist of all organized armed forces, groups and units which are* ***under a command responsible to that Party*** *for the conduct of its subordinates*"[25]. The terminology used by such rules does not exclude non-state entities; but aims at encompassing private entities or individuals[26]. However, such attribution as a state organ is challenging to meet in practice, as it requires a total absence of autonomy from

[19] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43, para 405; International Law Commission, Articles on the Responsibility of States for their International Wrongful Act, A/56/49(Vol. I)/Corr.4 Article 4. HRC, General Comment No 36 on Article 6 of the International Covenant on Civil and Political rights, on the right to life, published on 30 October 2018, CCPR /C/GC/36, [27].

[20] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43, para 406.

[21] Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v. Uganda), Judgment, I.C.J. Reports 2005, p. 168, para 179

[22] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43, para 399;

[23] International Law Commission, Articles on the Responsibility of States for their International Wrongful Act, A/56/49(Vol. I)/Corr.4 Article 1

[24] UN, 'Materials on the Responsibility of States for Internationally Wrongful Acts' (2012), United Nations legislative series, ST/LEG/SER B/25,

[25] ICRC, Additional Protocol I to the Geneva Conventions of 12 August 1949, 8 June 1977, 1125 UNTS 3 Article 43(1)

[26] Such perspective is discussed later in this article.

the non-state entity[27]. The demonstration of such autonomy requires a total subjugation to the State, whether by being incorporated in the chain of command or in a state of total dependence[28]. Such agency test was never met in practice. The attribution can also be achieved with the requirement of a direct and effective control of a non-state entity by the State[29]. The ICJ indicated that "*The conduct of a person or group of persons shall be considered an act of a State under international law if the person or group of persons* ***is in fact*** *acting on the instructions of, or under the direction or control of, that State in carrying out the conduct*"[30]. The requirements of such a test are also very high as it requires a strict control from the State, casting aside a general control over an entity[31]. Thus, the control must be asserted over such an entity and over one instance that constitute the international wrongful act in order to bring the State responsibility.

The challenges with such test are their strictness as a non-state entity has never been found as acting under the effective control of the State. Nevertheless, these tests bring a further clarification on the concept of autonomy that would later be relevant for assessing artificial intelligence under international law.

2. Differentiating autonomies under international law.

The concept of autonomy stands as a counter point for attributing a conduct to the State. The more the autonomy, the less the conduct will be attributable to the State. Autonomy would be a mean for State to evade its international responsibility. The degree of control on the entity for an international wrongful act would require that it is acting as the extended arm of the State. Such control implies that the State is deliberately using the entity as a mean for committing violations of international law obligations. As such, even if the U.S.A had delivered the means for the contras to violate human and humanitarian law provisions, the conduct was not attributable for the U.S.A had not effectively directed the contras to commit such violations[32]. The main rationale was that they enjoyed a certain autonomy in the use of such means. Such autonomy of non-state entities thus does not permit to attribute their conduct to States.

In this perspective, one could see that the ICJ operates a distinction between two types of autonomy: the operational autonomy and the functional autonomy from states of non-state entities. The operational autonomy reflects the means and methods that are at the hand of these entities. It can encompass weaponry, information, or training for

[27] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14. paras 109-110.
[28] *Ibidem*
[29] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14, para 115.
[30] International Law Commission, Articles on the Responsibility of States for their International Wrongful Act, A/56/49(Vol. I)/Corr.4 Article 8
[31] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43, paras 405-407
[32] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14. Para 110.

instance. It represents the effective capacity of non-state entities to conduct operations. In the context of the Contras, they did not enjoy operational autonomy since their means and methods of warfare were dependent on the U.S.A deliveries.

The second type of autonomy is the functional one; meaning the freedom of using these means and methods freely from a state control[33]. In this perspective, if a non-state entity does not have such autonomy because it is obeying orders made by States agents; then the State will ultimately be responsible for such conduct. The ICJ tends to emphasise the functional autonomy as being the main nexus for attribution[34] whereas the lack of operational autonomy would bring the responsibility of states under the positive obligations framework to refrain of giving such means[35]. As such, while operational autonomy is not per se required to bring attribution, functional autonomy is the main criterion to assess the dependence to the State. This is the core component in determining an attribution of non-states conducts to States.

Consequently, the means delivered to a non-state entity are irrelevant for an attribution mechanism. Operational autonomy only entails a responsibility under positive international obligations. The distinction between operational and functional autonomy present opportunities and challenges for determining the status of artificial intelligence under international law.

II. The extent of the application of autonomy to artificial intelligence under international responsibility law.

Artificial intelligence represents a new challenge for international security. Indeed, devices deployed on battlefields or intended to do so are qualified as "autonomous" weapons. In this aspect, they can take many shapes and strike different targets, both in the physical and cyber realm[36]. However, one would need to assess the true extent of artificial intelligence autonomy under international law and what are the consequences on attribution? This question leads to assess artificial intelligence under the concept of autonomy as analysed by international law (A). This also leads to determine why the current terminology is misleading for the international legal framework (B).

A. Can the concept of autonomy apply to artificial intelligence?

International law witnessed the rise of new advanced weaponry and new means to target both military and civilian infrastructures, both in the physical and cyber-realm. As such, any offensives through these realms can damage key structures such as power

[33] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43, para 399; Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14, para 115.
[34] *Ibidem*.
[35] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14, para 228.
[36] United Nations Institute for Disarmament Research (UNIDIR), The Weaponization of Increasingly Autonomous Technologies: Autonomous Weapon Systems and Cyber Operations, 2017.

plant, nuclear plant, electricity grid etc... The cyber-realm offers a wider range as it reaches becomes virtually unlimited due to the internet. The use of drones also brings new perspective, as humans stand further from the field while remotely controlling the mean of warfare. Nevertheless, these new means, how advanced are they, have a common feature: the presence of a human being as the main controller of its course of action. These viruses are sent by humans and pre-programmed to attack a certain target, drones are controlled by humans and are dependent on its choices, irrelevant of the location of the human controller[37]. The ultimate purpose of these means reflects the ultimate purpose of the human being that is controlling it. They act as mere token and their extended arm. As such, these means, how advanced they might be, do not enjoy any autonomy in their conducts.

This is the where artificial intelligence stands out among all these means, by its capacity to take a decision without further human intervention once deployed on the field[38]. Such freedom of acting led to apply the terminology of "autonomous weapons" for such software[39]. They are defined as "a *weapon system that, once activated, can select and engage targets without further intervention by a human operator*"[40]. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control"[41]. The "autonomous" qualification stems from the fact that artificial intelligence does not need human intervention to fulfil its tasks once it is deployed. An armoured vehicle operating with an artificial intelligence software does not need humans to target and take actions in the course of battle.

Under the dual conception of autonomy, one could see that they enjoy a certain amount of operational autonomy, in the sense the data used in its developments and the weaponry it is equipped with gives it the capacity to take actions on the field. Artificial intelligence is developed from training data, it allows it to build correlations between these data and the outcome linked to it[42]. For instance, artificial intelligence program can recognize a person wearing a uniform and a weapon and qualify him as a combatant under international law thanks to the training data it is fed with. It can further correlate this information by using lethal force against such person. However, a person not wearing a uniform without carrying a weapon should be considered as a civilian and the

---

[37] *Ibidem*.

[38] ICRC, Artificial intelligence and machine learning in armed conflict: A human-centred approach, (2020), International Review of the Red Cross Digital technologies and war, 102, 463–479, 465

[39] ICRC website, What you need to know about autonomous weapons, , available at https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons [Accessed on 26th May 2023].

[40] Office of the Under Secretary of Defense for Policy, Autonomy in weapon systems, (2023), US Department of defense, 1-24, 21.

[41] The Development, Concepts and Doctrine Centre (DCDC) , Joint Doctrine Publication 0-30.2 Unmanned Aircraft Systems, (2017), 13.

[42] Ripley B.D, *Pattern Recognition and Neural Networks*, (1996), Cambridge : Cambridge University Press, p. 354

use of force thus prohibited. The training data enables artificial intelligence to apply such correlation in a live-time environment. Training data thus gives the capacity for artificial intelligence to recognize military objectives and act accordingly. One could assess that it enjoys a certain amount of operational autonomy. However, does artificial intelligence hold the freedom to use these means as understood under the concept of functional autonomy?

Artificial intelligence programming is intrinsically bound and confined to human choices, incorporated in its design and development[43]. They are developed for specific tasks and cannot effectively adapt to others, because they are not unable to do so. Such lack of adaptation earns such system the qualification of "weak AI" compared to "strong AI" that can adapt to other tasks like humans do[44].

An "autonomous" weapon is likely deployed to conduct offensive or defensive actions during an armed conflict and skirmishes. In this aspect, its main purpose is to use lethal force against enemies in a combat scenario[45]. Theoretically, it could also be programmed to make the person stand down until the threat he poses disappear by requiring a formal surrender. The programming can adopt a wide range of behaviour; nevertheless, this range of behaviour is virtually limited, predefined, and pre-determined by human decisions[46]. The choices on how to use the means at hand (such as the weapons installed on the artificial intelligence vehicle) is controlled by the programming, which is also a conduit of human choices and supervisions. The choice of the learning methods of the artificial intelligence model reflects human choices. For instance, to use a supervised training, an unsupervised training, whether to limit the software to some conducts or to consider others. What if an artificial intelligent software, equipped with weapons is confronted to a scenario where a combatant is using a human shield, but its training data did not consider such scenario? The outcome would risk being unsatisfactory or not adapted due to the lack of adaptation outside of the artificial intelligence programming.

The training data upon which the artificial intelligence is deemed to base its conduct on also reflects the human mind of the developers[47]. All the correlations artificial intelligence is making are intrinsically linked to remote human choices[48]. If developers

---

[43] Yavar Bathaee, the artificial intelligence black box and the failure of intent and causation (2018),Harvard Journal of Law & Technology Volume 31, n°2, pp 890-938; p933

[44] Ng Gee Wah and Leung Wang Chi, Strong Artĭcial Intelligence and Consciousness, (2020), Journal of Artĭcial Intelligence and Consciousness, 7, 1, 63-72, 65. However no such systems exist today.

[45] ICRC Report of the ICRC Expert Meeting on 'Autonomous weapon systems: technical, military, legal and humanitarian aspects', 26-28 March 2014, Geneva, published on 9 May 2014, p1.

[46] European Parliament, Artificial intelligence: How does it work, why does it matter and what can we do about it ? (2020), étude du service de recherche du Parlement Européen, p5

[47] Adrien Van den Branden, *Les robots à l'assaut de la Justice*, (2019), Bruylant, P107 ; Boris Barraud, Le droit en datas : comment l'intelligence artificielle redessine le monde juridique, (2019), Revue Lamy Droit de l'immatériel, volume 164, 1-22, 19.

[48] Jean de Codt, Juger avec un algorithme et juger l'algorithme, (2019), Juges augmentés ou justice diminuée, édition Larcier, 1-180, 26

decide to choose a training set that will ultimately lead the artificial intelligence to consider civilians as a potential threat in an armed conflict; it is their choices that would lead the machine to act. Artificial intelligence does not enjoy the freedom to act from the programmers as it is a slave to its own programming. Its conduct represents a mere repetition of the developer's intent imbued in its programming. The functioning of artificial intelligence is thus anchored in the frame the developers intended. This weak AI programming makes it an expert for one task, but not all the tasks that a human can accomplish in the course of an armed conflict. Hence, artificial intelligence cannot be qualified as autonomous in the sense that it cannot think outside of the box that it was originally programmed for. For instance, if autonomous weapons are not programmed to recognize combatants surrendering, it will not accept it can lead to a conduct in multiple violations of humanitarian law[49]. But the choice to target civilians can be linked to its developers. The functional autonomy of artificial intelligence is thus voided while it enjoys an extensive amount of operational autonomy. The repetition of a predetermined behaviour cannot characterize artificial intelligence as fully autonomous, but rather an automatic answer to a predefined set of situations: Automatic does not mean autonomous.

If artificial intelligence is enjoying operational autonomy by not requiring further input to act; the requirements of functional autonomy, the capacity to take a decision outside its programming are manifestly lacking. In this context, one should determine whether the current terminology of "autonomous weapon" is relevant.

B. Should artificial intelligence be qualified as autonomous?

Artificial intelligence only enjoys partial autonomy, the operational one. Thus, it cannot be qualified as fully autonomous under international law.

As mentioned before, the lack of functional autonomy allows for an attribution to a State for an international wrongful act. However, the artificial intelligence lack of autonomy is not due to a state order, but a developer's programming. They are often private companies and are intrinsically not responsible for an international wrongful act under international law. In order for such attribution to be considered, the State would have to either: exercise an effective control, consider such company as an agent, transferring governmental authority or if it approves its conduct ex-post[50]. It thus requires demonstrating that the State is holding such a link to the developers who themselves are holding a link to the programming of the artificial intelligence programming. It is unlikely that such links would be met in practice. Lack of afunctional autonomy of artificial intelligence would thus not directly make the State responsible for its conduct. This stretching of international law is thus very challenging under the

[49] Committee on Legal Affairs and Human Rights, Emergence of lethal autonomous weapons systems (LAWS) and their necessary apprehension through European human rights law, provisional version, (2020), Council of Europe, p1.

[50] International Law Commission, Articles on the Responsibility of States for their International Wrongful Act, A/56/49(Vol. I)/Corr.4 Article 4, 5, 7, 8 and 11

international responsibility framework. Hence, not only the terminology of autonomous systems is misleading considering the overall functioning, it also does not carry an automatic attribution to the State. The autonomous quality of such systems is thus only partially correct, but it is not enough to effectively consider an attribution to the State under the current international legal framework.

The illusion casted by the flagrant operational autonomy of artificial intelligence shrouds the lack of effective functional autonomy from its programming. The international responsibility framework is thus falling short before this "non-human autonomy". Nevertheless, since such technology is intended to be deployed on the field of battle; one needs to take a closer look to the legal framework applying to it and to assess whether the challenge of non-human autonomy can meet an answer.

III. Partial AI autonomy as an obstacle for the law of armed conflicts.

International humanitarian law contains several categories of participants in armed conflicts, distinction between combatants and civilians[51] but also a categorisation of means and methods of warfare. Hence not only the persons but also the tools of armed conflicts are categorised in a dual approach: authorised means and forbidden means[52].

As emphasised earlier, artificial intelligence is not as autonomous as it appeared. Under international law, such lack of functional autonomy could present opportunities for attributing artificial intelligence conduct to states according to international law of attribution. Nevertheless, it is unlikely that international law would stretch to this extent.

One would need to assess the extent of the law of armed conflicts application to artificial intelligence. This requires a proper qualification of artificial intelligence and to assess its status considering its lack of functional autonomy. The intrinsic feature of artificial intelligence reveals the difficulty to equate it under such the strict categorization provided by the current international legal framework. If artificial intelligence can take many shapes: such as viruses or machines; it should not hinder the nature of its lack of autonomy. Nevertheless, the different shapes artificial intelligence can take would result in a different approach under international law (A). Moreover, one needs to assess the role of general international law in the perspective of attributing conduct of artificial intelligence to states (B).

A. Assessing the qualification of artificial intelligence under international law

Artificial intelligence is unique under international law. It only enjoys one partial autonomy, the operational one. Its nature is also challenging; for it is often the product of a private company, whom the developers are part of. In this perspective, the

[51] ICRC, Customary international humanitarian law, Rule 1, Volume II, Chapter 1, section A;
[52] ICRC, Customary international humanitarian law, Rule 70, Volume II, Chapter 20, section A; "The use of means and methods of warfare which are of a nature to cause superfluous or unnecessary suffering is prohibited".

developers who intrinsically limit the artificial intelligence behaviour in its functional autonomy, are private companies and not public agents. Hence, if such technology is lacking autonomy, it does not automatically involve a state component in its inner working, limiting the nexus of international law application. States would only qualify as the users of such technology, at the very end of its life cycle[53]. The qualification of artificial intelligence under the current framework is thus challenging, both for the physical (1) and the still developing cyber-realm (2).

1. Autonomy for artificial intelligence used in the physical realm.

Artificial intelligence in the physical realm can take the shape of small, armoured vehicles, terrestrial or aerial. Deployed on the battlefields, they can conduct military operations aside, or as a replacement of, combatants. They operate following their algorithm and can act without further a human order[54]. This ranges from defensive artificial intelligence, to protect infrastructure and individuals, to offensive, used on the field to use lethal means against persons. The defensive ones are however already used, as a mean to rapidly detect threats that require a faster response. The offensive ones are the ones directly involved in dealing with lethal force, thus targeting humans and not missiles or torpedoes like the defensive ones do. Nevertheless, despite these functions, one should question whether it can qualify as a weapon (a); or to a full participant in the armed conflict (b).

a. The difficult equation of AI as a weapon ?

Can artificial intelligence be fully equated to a weapon? In the mainstream approach of such technology, it would qualify as such[55]. The underpinning question is whether the actual terminology and qualification can encompass artificial intelligence and its "*non-human*" partial autonomy. Weapons are synonymous with a mean to warfare[56]. Under this qualification, many bindings legal instruments prohibited the use of certain means of warfare. The prohibition is built upon the principles that weapons should not "*cause superfluous injury or unnecessary suffering*"[57]. Weapons tend to bring no challenges for attribution, for they always are activated by humans (Whether a rifle or a ballistic missile) and are under their control. Their functions are totally limited by the will of humans in their uses, and they are solely functioning because of such use thus, making

[53] Such definition stems from the EU AI Act: "'user' means any natural or legal person, public authority, agency or other body using an AI system under its authority"; European Parliament and Council, Proposal for a regulation 2021/106 on laying down harmonized rules on artificial intelligence (Artificial intelligence act) and emending certain union legislative acts, published on 21st April 2021, COM/2021/206 final, article 3 (4).
[54] Committee on Legal Affairs and Human Rights, Emergence of lethal autonomous weapons systems (LAWS) and their necessary apprehension through European human rights law, provisional version, (2020), Council of Europe, article 7; Nathan Gabriel Wood, Autonomous Weapons Systems: A Clarification, (2023), Journal of military ethics, 1-21, 6
[55] ICRC, Artificial intelligence and machine learning in armed conflict: A human-centered approach, (2020), International Review of the Red Cross Digital technologies and war, 102, 463–479, 464
[56] ICRC, Customary international humanitarian law, Rule 70 and 71, Volume II, Chapter 20, section A;
[57] *Ibidem.*

the attribution of the use of the weapon to the human logical and unchallenging. A rifle will not shoot until the human activates the trigger; humans are necessary for fulfilling the weapons purpose on the field.

However, the challenge of equating artificial intelligence to a mean of warfare is the lack of control of humans on the functioning of artificial intelligence. Humans cannot press the trigger of artificial intelligence; they merely deploy it, and it fulfils its task automatically. It could be considered as a mean of warfare, but the lack of effective control over its functioning makes it difficult to equate it to a weapon in the most traditional aspect. The partial autonomy it enjoys frees it from the humans that are using such technology and does not represent their extended arm; but rather the one of the developers. Thus, the qualification as autonomous weapon is misleading under these two aspects. It does qualify as a weapon in the legal sense, and its qualification as weapon is misleading in the sense that they do not share the same fundamental features. A pathway for a possible status would be to qualify it as an automatic mean of warfare. However, the partial autonomy it enjoys should be considered and new mechanisms should be set in place in order to assess the true responsible for its deployment. Questions arise whether such partial autonomy can lead to qualify it as a participant in an armed conflict.

b. AI as a participant in an armed conflict?

In this perspective, one could assess whether artificial intelligence should be qualified as a participant in an armed conflict, given its partial autonomous feature, for such operational autonomy frees it from any human control while it acts like a combatant[58]. The functional approach of this technology allows to affirm that it is developed for a direct participation in armed conflicts. In this perspective, and inasmuch as they share the same functions as combatants, one could assess this qualification under international law.

Two considerations arise: firstly, whether States decide to incorporate *de jure* as part of their armed forces, the attribution would not present a challenge[59]. Secondly, if they are not *de jure* incorporated in the armed forces, they can be by following four criteria: they are commanded by a person responsible for his subordinates, they wear distinctive sign recognizable at a distance, they carry arms openly, they conduct their operations in accordance with laws and customs of war[60]. Such qualification could enable a direct attribution scheme to the State by being a member of its armed forces.

[58] The capacity to automatically detect and engage targets equates its function with one of a combatant: ICRC Report of the ICRC Expert Meeting on 'Autonomous weapon systems: technical, military, legal and humanitarian aspects', 26-28 March 2014, Geneva, published on 9 May 2014: "'autonomous weapon systems' were defined as weapons that can independently select and attack targets, i.e. with autonomy in the 'critical functions' of acquiring, tracking, selecting and attacking targets".
[59] International Law Commission, Articles on the Responsibility of States for their International Wrongful Act, A/56/49(Vol. I)/Corr.4 Article 4.
[60] ICRC, Customary international humanitarian law, Rule 4, Volume II, Chapter 1, section D

The vehicle that would contain the artificial intelligence could easily wear distinctive signs and carrying weapons. The extent under which it must lead operations in accordance with laws and customs of war represents a challenging criterion. Such answer would be found in the programming of such machine and whether it is programmed to adopt certain conduct that would prevent any international law violations such as using lethal force on civilians. It is the extent of being commanded by a person responsible that brings challenges under international law.

International law considers the chain of command the main pathway to attribute the conduct of combatants under the law of armed conflicts[61]. Commanders are those persons that must exert a control over the combatants and the means of warfare in order for its responsibility to be effective[62]. All commanders are responsible for the conduct of their subordinates, unless they can demonstrate that they explicitly forbidden this conduct or that they did everything in their power to avoid such conduct[63].

The principle remains that commanders should assert a control over its subordinates as to avoid any wrongful conduct. However, at what extent can it be asserted that commanders hold any form of control on the choices artificial intelligence might take. It could be argued that this control is asserted for those systems requiring an activation to function[64]. But what about the choices that artificial is making on the field? These choices are predefined in the development of its algorithm by the developers. Military commanders merely deploy such means on the field and if they can operate under their general watch (such as striking or targeting order), they do not have a control on every acts it may perform, nor can they change its programming. As such, if military commander deploys the artificial intelligence expecting it to only attack combatants, but it kills both combatants and civilians, can it be asserted that this commander had a control over the artificial intelligence conduct? Artificial intelligence does not need any order to fulfil its programming; it just requires to be deployed in a specific context. The attribution test for such software is thus very unlikely to be relevant with international standards. Commanders could likely evade their responsibility by asserting that they do not have direct and effective control overs the choices made by the algorithm.

The extent under which such technology can be equated to participants under a personal responsible is challenging. The lack of control over the functional autonomy by commanders is the main obstacle for it. The specificity of artificial intelligence to be

[61] Jamie Allan Williamson, Some considerations on command responsibility and criminal liability, (2008), International review of the Red cross, 90, 870, 303 318, 306

[62] ICRC, Protocol additional to the Geneva Conventions of 12 August 1949, and relating t the protection of victims of international armed conflicts (Protocol I), adopted on 8 June 1977, article 87,

[63] ICTY Blaskic case; Jean-Marie Henckaertz and Louise Doswald-Beck, Customary international humanitarian law (2005), Cambridge University press, 563

[64] Committee on Legal Affairs and Human Rights, Emergence of lethal autonomous weapons systems (LAWS) and their necessary apprehension through European human rights law, provisional version, (2020), Council of Europe, article 7.3

able to make choices without or with a limited intervention of humans on the field is thus challenging.

Is artificial intelligence a weapon or a participant in an armed conflict? The mainstream approach qualifies such systems as weapons because of their non-human nature. However, their partial autonomy does not satisfy such a view and cannot be fully qualified as weapons. The challenge lies in the fact that the current international framework did not foresee the use of such technology on the field and is human-centric by nature. To adapt such rules to this technology risk to fragilize it further by losing consistency and relevance. These rules are even more challenging considering the perspective of artificial intelligence in the cyber-realm.

2. *Autonomy of artificial intelligence used in the cyber-realm.*

Artificial intelligence software used in the cyber-realm leads to consider three points. Firstly, its capacity to target a wide range of infrastructures, from civilian to military without any geographical limit. Its use is thus not confined to a theatre of operation nor to the actors involved in the conflicts[65]. Secondly its point of origin, it can either be a public or a private actor, theoretically every person having an internet access could be a potential participant in the cyber armed conflict[66]. Thirdly, these uses can occur outside of any declared armed conflict. These three features reveal a more universal use than artificial intelligence deployed in the physical realm. The use of the internet as a conduit to lead operations with artificial intelligence render it threatening for civilians and militaries alike.

The challenge of such artificial intelligence weaponization is its universality. Whether a mere civilian, a private company or a specialized armed forces of a State could develop and use such software. This contrasts with the use of physical artificial intelligence, where the developers are not the final users of the technology. The Tallinn manual on the international law applicable to cyber-operations encompassed such use of viruses as an attribution to the State if there is a contractual relationship between the person that created it and the Government of that State[67]. Whereas in the context of private companies outside of such relationship, the classic attribution test of direct and effective control would apply.

In the case of cyber-realm artificial intelligence, the real user of such technology is blurred by the lack of effective tracing throughout the internet. Artificial intelligence for cyber operations would share the same functions and the same methodology as classic viruses but are more effective and more difficult to neutralize. Whereas viruses

[65] U.S. Department of Defense, Dictionary of Military and Associated Terms, Joint Publication 1-02, Nov. 8, 2010, as amended through Feb. 15, 2012, available at http://www.dtic.mil/doctrine/dod_dictionary/. [accessed 30 May 2023]; Jeffrey T. G. Kelsey, 'Hacking into International Humanitarian Law: The Principles of Distinction and Neutrality in the Age of Cyber Warfare' [May 2008] Michigan Law Review 106 7, 1443.
[66] Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations [2017] Rule 81.
[67] Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations [2017], Rule 29.

can be a simple program, designated to target one infrastructure or a set of infrastructures, the autonomy of AI can bring unforeseen consequences.

The partial autonomy of artificial intelligence is less relevant in this context, for the developing framework of cyber international law is not as striking as the physical realm one. Moreover, the cyber-realm presents specificities rendering the attributions of acts difficult. The partial autonomy is not as striking as it is in the physical realm. However, it is a still slave of its programming. The programming would decide the full extent of the damages caused by artificial intelligence in the cyber-realm; whether to commit data theft, to neutralize civilian infrastructure or shut down core military and civilian infrastructures. The cyber nature of such artificial intelligence should not cast a doubt on its very nature; this means a program enjoying operational autonomy but none regarding the functional one.

Hence, what is artificial intelligence under international law? Such a question remains unanswered on many aspects, depending on its deployment, what realm it is used in, its relationship with human and at what degree can it be detached from humans. In the context of artificial intelligence responding to real-time orders, it makes no doubt that the issuer of such an order could be held responsible for the artificial intelligence acts. However, in most cases such systems are intended to act without such orders. Nevertheless, international law would remain a pathway to regulate such technologies, but one has to take a step back on such regulation.

B. General international law as the security valve for the regulation of artificial intelligence?

To qualify artificial intelligence to an already existing status under international law is challenging. Its partial autonomy makes it unique: more complex than a mere mean of warfare whereas it does not reach the threshold to be equated as a participant in an armed conflict. Moreover, the challenge lies in the regulation perspective as no explicit rules forbid the use of artificial intelligence in the context of warfare.

Moreover, some artificial intelligence regulation instruments have expressly rejected artificial intelligence in armed conflict out of their scope[68].

International law applied in the context of armed conflict, international security and the general framework is thus challenging to adapt for such technology. It leaves the regulation perspective in a decentralized approach leaving states to decide upon all the aspects of development, use and deployments. Nevertheless, the lack of a special framework applying to artificial intelligence in armed conflict does not mean a complete

[68] Committee on artificial intelligence (CAI), Revised zero draft [Framework] Convention on artificial intelligence, human rights, democracy and the Rule of Law, published on 6 january 2023, CAI (2023)01, article 4.3: "*This Convention shall not apply to design, development and application of artificial intelligence systems used for purposes related to national defence.*"

legal void. Indeed, international provisions provide for a security valve enabling a certain number of general principles and regulations mechanisms.

The international court of justice indicated that an inexistent attribution from a non-state entity conduct to a State does not free states of their general international law obligations[69]. As such, the delivery of means and methods of warfare to contras breached the principle of non-intervention of a state in another's affairs. Attribution is not the panacea for determining responsibility of states. The approach of the ICJ demonstrates that when attribution is not possible, one needs to look at the States' international obligations. In the context of artificial intelligence however, one needs to identify such legal obligations as it brings new challenges. Since there are no international binding instruments focusing on artificial intelligence, one needs to take a step back and assess the general international law obligations of States.

What would be left then? During armed conflicts, one needs to look at the fundamental principles of human rights and humanitarian law. In this perspective, the distinction between civilians and combatants, the prohibition to attack those *hors de combat*, the prohibition to inflict unnecessary suffering, the principle of necessity and the principle of proportionality frame all theatre of operations. These principles are not applicable specially to armed forces or means and methods of warfare; but to "parties to the conflict" hence having a wider scope for application[70]. These principles are thus applicable irrelevant of any categorisation under international law. States hold a certain responsibility in ensuring that every deployment they are performing would have to be compatible with the above-mentioned principles[71]. The attribution regarding such technology under the current framework should not focus on the use, but the deployment phase.

The deployment of such a technology, even if not directly attributable to the state and without a special framework, can be encompassed in such obligations. It would provide for a general guidance for states to consider using such technology and could also provide a pathway for developers to construct software compatible with international standards. The enactment of general principles as a mean to answer the autonomy of artificial intelligence provides a safer path than the special regulative perspective in the sense that these principles will always frame such conflicts. The search for a special framework applying to artificial intelligence might not be the most relevant path in a context where such technology evolves rapidly and where such regulation might meet the reluctance of states.

---

[69] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14, para 228.

[70] ICRC, Customary international humanitarian law, Rule 1, Rule 7, , Volume II, Chapter 1, section D

[71] *Ibidem*. The absence of any further qualification other than "Parties to the conflict" aims at widening the scope beyond the distinction of civilian/combatant and means and methods of warfare. This would allow to evade the question of qualification of artificial intelligence under such framework. The challenge remains however that no attribution can be made under this framework.

The already established principles allows more flexibility and clarity in that sense. Such principles would thus frame the conduct of artificial intelligence and at what extent can states be responsible for its mere deployment. Moreover, the rising awareness surrounding the artificial intelligence development pinpoints stages where artificial intelligence could bring states responsibility for failure to act accordingly to these principles[72]. Instance where states deploy means and methods that they are aware of their capacity to commit violations of humanitarian law or human rights, by not upholding the principle of differentiating civilians from combatants, would constitute such a violation. States would likely in the development and the testing of such technologies to ensure its adequacy to international standards. The reliance on such a general framework for artificial intelligence allows to go beyond the question of its autonomy and the challenges for attributing the conduct to the State.

IV. Redefining autonomy for artificial intelligence under international law?

The general framework represents a safety valve in the regulation of artificial intelligence in international law. However, such framework needs to be bolstered by a specific one directly addressing the challenges brought by artificial intelligence.

In this context, the concept of autonomy as currently interpreted in international law is not adapted for such purposes. As demonstrated, the autonomy of artificial intelligence is misleading; it is confined to and by its programming and is lacking any sort of adaptability outside of it. Since it qualifies as a "weak-ai" it cannot adapt to other scenarios in a war-like environment. Artificial intelligence devices only focus on the tasks considered in its programming. It does not display the flexibility that full autonomy brings. The functional autonomy of artificial intelligence is void and is ultimately dependent on its programming, leaving only a display of operational autonomy, providing an illusion of full autonomy. The consequence of this non-human autonomy reveals challenges in adapting a legal framework that was thought and erected for human conducts.

The risk of deploying artificial intelligence with the same understanding of autonomy might lead to evade totally or partially the attribution of artificial intelligence to states or adapted agents. State could evade the attribution by assessing that they had no control over the programming while artificial intelligence does not fully qualify as a weapon nor a combatant. Since autonomy leads to state irresponsibility, artificial intelligence should not be recognised as autonomous under international law. Artificial intelligence would thus become a new actor in international law, but the current legal status does not fully encompass the challenges brought by artificial intelligence.

In this perspective, users, mainly military commanders, do not have a control over the artificial intelligence behaviour since it is programmed upstream of its use. Such

[72] ICRC, Artificial intelligence and machine learning in armed conflict: A human-centred approach, (2020), International Review of the Red Cross, 102, 913, 463-479, 471.

programming totally escapes from its hand and can cause international wrongful acts without the possibility of acting beforehand. The real stakes in the one occurring in the developing made by private companies. However, at what extent such companies would hold an international liability under international law for such developing? It is unlikely that the actual framework would allow such responsibility from occurring. The development of means of warfare would not present the delegation of "governmental authority" per se; not does it encompass a direct and effective control from State on companies.

The assimilation of artificial intelligence to an existing status of international humanitarian law is also challenging due to the human-centric approach of such framework. This uneasiness brought by the partial autonomy of such systems need to be addressed. The use of mechanisms applying to humans and legal entities cannot fully be adapted to artificial intelligence. As such, the conceptualization of autonomy as a mean of attributing the conduct of a non-state entity must evolve. The lack of functional autonomy of the artificial intelligence device bars any possibility for applying such attribution.

Non-human autonomy should be given an explicit recognition at national and international level for a recognition of artificial intelligence as a new actor of international security. Such recognition would give the pathway to an adapted legal regime and a framework for attributing its conduct to humans in order to provide for a safe and reliable framework.