20

Artificial Intelligence and Armed Conflicts

Katerina Yordanova[*]

20.1 INTRODUCTION

War has been an integral part of human history since the dawn of time. It evolved together with human society and influenced its development in a tremendous manner. Military historians and political scientists established numerous systems in an attempt to better classify and analyze military endeavors, taking into account global political trends, the evolution of the means and tactics of war, and local geographical and cultural specifics.[1] Nevertheless, armed conflicts of all kinds have been persistently disrupting human lives, leaving individuals with little to no remedy against various infringements of their human rights.[2] While rules such as the distinction between combatants and civilians have been around for centuries, their application was dependent on factors such as the personal honor of the individual soldier,[3] or whether the enemy was considered as belonging to a "civilized nation."[4]

The modern development of international humanitarian law (IHL), which started during the second half of the nineteenth century, initially did not change much vis-à-vis the means of response and remedy individuals had over violation of their rights during armed conflicts. A treaty-based framework was established, largely based on customary rules. Despite its wide scope, the framework was

[*] The present research was conducted in the framework of Cybersecurity Initiative Flanders – Strategic Research Program and Imec.

[1] See, for example, Max Boot, *Savage Wars of Peace: Small Wars and the Rise of American Power* (Basic Books, 2014); Mary Kaldor, *New and Old Wars: Organized Violence in a Global Era, 3rd ed* (Stanford University Press, 2013).

[2] The term here is used in the broadest sense, without prejudice to the modern-day framework of human rights developed post-Second World War.

[3] Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law. Volume I. Rules* (Cambridge University Press, 2009), p. xxxi.

[4] Sven Lindqvist, *"Exterminate All the Brutes": One Man's Odyssey into the Heart of Darkness and the Origins of European Genocide* (The New Press, 1997); James Sloan, "Civilized Nations" *Max Planck Encyclopedia of Public International Law* (April 2011), https://opil-ouplaw-com.kuleuven.e-bronnen.be/display/10.1093/law:epil/9780199231690/law-9780199231690-e1748, accessed December 28, 2022.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

founded on the deep humanist ideals of just one man to relieve the unnecessary suffering of fellow human beings elevating their protection as a main consideration for all parties involved in an armed conflict. International humanitarian law evolved dramatically during the twentieth century in response to the numerous wars that caused huge casualties and unimaginable destruction all over the world, directly related to the rapid change of military strategies and the utilization of new weapons. This tendency of IHL "lagging behind"[5] the contemporary challenges of the day is often characterized as one of its biggest weaknesses. However, this "weakness" permeates the entire legal system, as it became particularly evident in the last decades with the booming development of new technologies, and especially disruptive ones[6] such as blockchain, the internet of things (IoT), and artificial intelligence (AI).

Examples such as the drastic changes in the regulation of technologies used in the financial sector after the Financial Crisis of 2008 and the general abandonment of the laissez-faire approach toward them come to show that law in general struggles to adapt to technological progress and that the traditional reactive approach[7] is not always sufficient to ensure legal certainty and fairness in society. In the context of IHL, this problem becomes even more pressing due to its focus on the preservation of human life and the prevention of unnecessary suffering.

AI is largely recognized as a disruptive technology with the biggest current and potential future impact on armed conflicts. In this context, the media often uses AI and lethal autonomous weapons systems (LAWS) as synonyms, but in reality, LAWS is just one application in which AI is utilized for military and paramilitary purposes.

The purpose of the present chapter is to provide the reader with a brief overview of the main uses of AI in armed conflicts with a specific focus on LAWS, and the main societal concerns this raises. For this purpose, the chapter will first provide an overview of IHL, so as to supply the reader with general knowledge about its principles and development. This, in turn, will hopefully allow for a better understanding of the legal and ethical challenges that society currently faces regarding the use of AI in armed conflicts. Finally, the chapter also attempts to pose some provocative questions on AI's use in this context, in light of contemporary events and global policy development in this area.

[5] James D. Fry, "Contextualized Legal Reviews for the Methods and Means of Warfare: Cave Combat and International Humanitarian Law" (2006) *Columbia Journal of Transnational Law*, 44(453): 466.

[6] Disruptive technologies do not have a commonly accepted definition, but they are often characterized by their refinement, ground-breaking nature, and ability to create new industries. The term was first used by Clayton M. Christensen and later explored in Clayton M. Christensen, "The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail" (1997) Harvard Business Review Press.

[7] A "reactive approach" of the law implies that, traditionally, legal norms are adopted and amended as a response or reaction toward new relations in society or significant changes in already existing ones.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

20.2 INTERNATIONAL HUMANITARIAN LAW

20.2.1 *Brief Introduction to the History of IHL*

The term *armed conflict*, which stands central to IHL, has a distinctive meaning in international law. The most commonly referred definition[8] is the one provided by the International Criminal Tribunal for the Former Yugoslavia which describes it as existing "whenever there is a resort to armed force between States or protracted armed violence between governmental authorities and organized armed groups or between such groups within a state."[9] The definition itself shows the armed conflicts are typically divided into those having an international and a non-international dimension.[10] For the sake of simplifying the scope of the present chapter, it will focus on the usage of AI solely in international armed conflicts. As a matter of fact, international armed conflicts were the only subject of regulation before the Second World War[11] which once again demonstrates the reactive nature of IHL.

IHL is considered to be born in the second half of the nineteenth century after the infamous battle of Solferino on June 24, 1859.[12] The battle resulted in the victory of the allied French Army of Napoleon III and the Piedmont-Sardinian Army commanded by Victor Emmanuel II over the Austrian Army under Emperor Franz Joseph I. It was documented that over 300,000 soldiers participated in the fifteen-hour massacre from which 6,000 died and more than 35,000 were wounded or went missing.[13] A detailed testimonial regarding the fallout of the battle and the suffering of both combatants and civilians was given by the Swiss national Henry Dunant, who happened to be in the area and witnessed the gruesome picture of the battlefield, later described in his book named *A Memory of Solferino*.[14] Henry Dunant dedicated his life and work to the creation of an impartial organization tasked with caring for the wounded in armed conflicts and providing humanitarian relief. It became a reality on February 17, 1863, when the International Committee of the Red Cross (ICRC) was established in Geneva. This event,

[8] See cases such as Prosecutor v. Charles Ghankay Taylor, SCSL-03-1-T, Special Court for Sierra Leone, May 18, 2012, para. 506, The Prosecutor v. Thomas Lubanga Dyilo, ICC-01/04-01/06-2842, International Criminal Court, April 05, 2012.

[9] *Prosecutor* v. *Dusko Tadic Appeal* Judgement (1999) International Criminal Tribunal for the former Yugoslavia (ICTY), IT-94-1-A.

[10] Non-international armed conflicts are usually perceived as every armed conflict that is not qualified as international.

[11] Yoram Dinstein, *Non-International Armed Conflicts in International Law 2nd Edition* (Cambridge University Press, 2021), 3.

[12] ICRC, "What are the origins of International Humanitarian Law?" (August 2017) *ICRC Blog*, https://blogs.icrc.org/ilot/2017/08/07/origins-international-humanitarian-law/ accessed June 5, 2023.

[13] International Committee of the Red Cross, "Solferino and the International Committee of the Red Cross" (June 2010) *ICRC Feature*, www.icrc.org/en/doc/resources/documents/feature/2010/solferino-feature-240609.htm, accessed December 28, 2022.

[14] Original title: *Un souvenir de Solferino*, self-published in 1962.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

while significant, was just the beginning of the rapid development of modern IHL. In 1864, the Geneva Convention for the Amelioration of the Condition of the Wounded in Armies in the Field was adopted, becoming the first international treaty regulating armed conflicts in a universal manner, opened for all States to join.[15] That instrument, and the following Geneva conventions, established three key obligations for the parties: (1) providing medical assistance to the wounded regardless of their nationality, (2) respecting the neutrality of the medical personnel and establishments, and (3) recognizing and respecting the sign of the Red Cross on white background.

These first steps predefine the characteristics of modern IHL as part of the body of international law that governs the relations between states. Its subsequent development also broadens its scope to protecting "persons who are not or are no longer taking part in hostilities, the sick and wounded, prisoners and civilians, and to define the rights and obligations of the parties to a conflict in the conduct of hostilities."[16] In addition, IHL evolved to serve as a guarantee for preserving humanity even on the battlefield by attempting to ease the suffering of the combatants. This development in the role of IHL occurred early on due to the rapid uptake of new weapons. In particular, the invention of the *dum-dum bullet*[17] in 1863 inspired the adoption of the first international treaty concerning weaponry (namely the St. Petersburg Declaration Renouncing the Use, in Time of War, of Explosive Projectiles under 400 gm weight, in 1868).[18] It is a pivotal instrument for IHL not only because it was the first of its kind but also because it recognized the customary character of the rule according to which using arms, projectiles, and materials that cause unnecessary suffering is prohibited.[19] This line of work was continued through the Hague Conventions from 1899 and 1907, governing the "laws of war" as opposed to the Geneva Conventions that focus on the right to receive relief.

The First World War indicated the end of this period of development of IHL, bringing forward the concept of total war, new weapons (including weapons of

[15] Bilateral treaties protecting war victims existed before 1864, but they were often breached and circumvented. Moreover, they were concluded and enforced only regarding a specific armed conflict.

[16] International Committee of the Red Cross, "War and international humanitarian law" (October 2010) *ICRC Overview*, www.icrc.org/en/doc/war-and-law/overview-war-and-law.htm, accessed December 28, 2022.

[17] The expanding bullet or the dum-dum bullet was invented by the Russian military and as the name shows it expands when it comes into contact with a hard surface. Even though it was initially created in order to deal more effectively with ammunition wagons, it was soon modified to expand when coming into contact with soft services as well, thus causing serious wounds to human beings Edward M. Spiers, "The use of the Dum Dum bullet in colonial warfare" The Journal of Imperial and Commonwealth History, vol. 4, issue 1, p. 3.

[18] Orlin Borisov, *Public International Law* (2008), Nova Zvezda, 712.

[19] Luc Reydams and Jan Wouters, "A la guerre comme à la guerre: patterns of armed conflict, humanitarian law responses and new challenges" in Jan Wouters, Philip De Man, and Nele Verlinden (eds), *Armed Conflicts and the Law* (Intersentia, 2016), 6.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

mass destruction such as poison gas), and the anonymization of combat.[20] These novel technologies and the effects they had on people, as well as the outcome of the campaigns and individual battles, naturally resulted in the adoption of further legal instruments in response of the new treats to the already established law of war.[21] In addition, the Geneva Convention was overhauled by the 1929 Convention for the Amelioration of the Condition of the Wounded and Sick in Armies in the Field which further reflected the influence of the new technologies establishing, for example, protection of medical aircraft.

The Second World War brought another set of challenges for IHL besides the tremendously high percentage of civil causalities compared to the First World War.[22] The concept of total war which transforms the economies of the states into war economies fogged the distinction between civilians and combatants and also between civilian and military objects, which is one of the key principles of IHL. Other contributing factors were the civilian groups targeted by the Nazi ideology and the coercive warfare used by the Allied powers.[23]

The immediate response to the horrors of the biggest war in human history was the establishment of the United Nations and the International Military Tribunals of Nuremberg and Tokyo. In addition, four new conventions amended and reinforced the IHL framework. The 1949 Geneva Conventions on the sick and wounded on land; on the wounded, sick, and shipwrecked members of the armed forces at sea; on prisoners of war; and on civilian victims (complemented by the Additional Protocols from 1977) codified and cemented the core principles of the modern-day IHL, the way we know it at present days.

The second half of the twentieth century, and in particular the Cold War period and the Decolonization, contributed mostly to the development of the IHL rules regarding non-international armed conflicts. Nevertheless, the tendency of adopting treaties in response to technological advancement continued. Certain core legal instruments on arms control were adopted during that time, such as the Treaty on the Non-Proliferation of Nuclear Weapons from 1968,[24] the Biological Weapons Convention from 1972,[25] and the Chemical Weapons

[20] Ibid., 7–8.

[21] See, for example, The Protocol for the Prohibition of the Use of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare from 1925, Geneva Conference for the Supervision of the International Traffic in Arms.

[22] The percentage of civil casualties directly cause by the Second World War is close to 50 percent compared to 5 percent during the First World War, see Orlin Borisov, Public International Law, p. 670.

[23] Reydams and Wouters, *A la guerre comme à la guerre*, p. 11.

[24] Unlike biological and chemical weapons, nuclear weapons have not been formally banned, but regulatory efforts were concentrated on their nonproliferation. The Treaty on the Prohibition of Nuclear Weapons from 2017 is unfortunately lacking effectiveness due to the absence of the states that actually own nuclear weapons among the signatory states. Therefore, one can hardly talk about a ban on nuclear weapons in practice, despite the existence of the instrument in theory.

[25] The Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (1972) United Nations General Assembly.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

Convention from 1993.[26] Another important legal treaty, directly connected to the regulation of new technologies used as weapons, concerns the Convention on Certain Conventional Weapons (CCW) from 1980.[27] This Convention will be further discussed in Section 20.3.

The global tendencies of the twenty-first century signaled the decreasing political will of nation-states to enter into binding multilateral agreements beyond trade and finance due to their lack of effectiveness.[28] This is extremely worrisome not only because of its effect on the international legal order, but also due to the consequences it has on the ambition of making law anticipatory rather than reactionary, especially in the context of new weapons such as LAWS. Therefore, the principles of IHL which have already been established during the last century and a half, need to be taken into account when interpreting the existing body of law applicable vis-à-vis technologies such as AI, regardless of the capacity it is used for in military context. The next section offers an exposition of these principles which should provide the reader with a better understanding of the IHL challenges created by AI.

20.2.2 *Principles of IHL*

To understand the effect of AI on IHL, six core principles of IHL need to be unpacked, as they serve both as an interpretative and guiding tool for the technology's use in this area. This section will briefly discuss each principle in turn.

20.2.2.1 Distinction between Civilians and Combatants

The principle that a distinction should be made between civilians and combatants is extremely important in armed conflicts, as it could mean the difference between life and death for an individual. In essence, the principle requires belligerents to distinguish at all times between people who can be attacked lawfully and people who cannot be attacked and should instead be protected.[29] This principle reflects the idea that armed conflicts represent limited conflicts between the armed forces of certain

[26] Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (1992) Conference of Disarmament and General Assembly of United Nations.

[27] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (1981) United Nations Conference on Prohibitions or Restrictions of the Use of Certain Conventional Weapons Which May Be Deemed Excessively Injurious or to Have Indiscriminate Effects and the General Assembly of United Nations.

[28] Steven J. Hoffman, Prativa Baral et al., "International treaties have mostly failed to produce their intended effects" (2022) *Proceedings of the National Academy of Sciences (PNAS)*, 119(32).

[29] Nils Melzer, "The principle of distinction between civilians and combatants" in Andrew Clapham and Paola Gaeta (eds), *The Oxford Handbook of International Law in Armed Conflict* (Oxford University Press, 2014), 297.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

States and not between their populations. The only legitimate goal is hence to weaken the military forces of the opposing State.[30] The importance of the principle of distinction as a cornerstone of IHL was reaffirmed by the International Law Commission[31] which argued it should be considered as a rule of *jus cogens*. This term is used to describe peremptory norms of international law from which no derogation is allowed. While the scope of *jus cogens* norms is subject to a continuing debate,[32] the fact that the principle of distinction is regarded as such a norm has a lot of merit.[33]

While the principle of distinction between civilians and combatants sounds very clear and easy to follow, in reality, it is not always easy to apply. On the one hand, the changing nature of armed conflicts blurs the differences between the two categories and shifts the traditional battlefield into urban areas. On the other hand, many functions previously carried out by military personnel are currently being outsourced to private contractors and to government personnel sometimes located in different locations. Involving technologies such as AI either in attacking (e.g., LAWS) or defensive (e.g., in cybersecurity) capability further complicates the distinction between civilians and combatants[34] and brings uncertainty, potentially increasing the risk of civilians being targeted erroneously or arbitrarily.

20.2.2.2 Prohibition to Attack Those Hors De Combat

The principle of prohibition to attack those *hors de combat* shows some similarities with the principle of distinction between civilians and combatants, mainly because one needs to be able to properly identify those *hors de combat*. The term originates from the French language and literally means "out of combat." Article 41 from Additional Protocol I to the Geneva Conventions from 1949 stipulates that a person "who is recognized or who, in the circumstances, should be recognized to be hors de combat shall not be made the object of attack." Paragraph 2 provides additional defining criteria, including the person to be in the power of an adverse Party, in case that person clearly expresses an intention to surrender. This also covers persons who

[30] Jean Simon Pictet, *Development and Principles of International Humanitarian Law* (Martinus Nijho Publishers, 1985), 62.

[31] International Law Commission, "Report on State Responsibility" (Yearbook of the International Law Commission, 2001) vol II, Part Two, 113.

[32] Ulf Linderfalk, "The effect of jus cogens norms: whoever opened pandora's box, did you ever think about the consequences?" (2007) *European Journal of International Law*, 18(5): 853–871. Andrea Bianchi, "Human rights and the magic of jus cogens" (2008) *European Journal of International Law*, 19(3): 491–508.

[33] International Court of Justice, "Legality of the threat or use of nuclear weapons opinion" (July 8, 1996) *Advisory Opinion*: § 78.

[34] The distinction between civilians and combatants could be complicated by AI systems due to a number of factors including, for instance, bias in the data used to train the systems; the unclear status of the people who develop, deploy, and maintain such systems who are usually contractors and employees of private companies; and the subjective psychological element which often compromises mechanisms such as human oversight and authority due to situations resembling the Milgram experiment.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

have been rendered unconscious or who are otherwise incapacitated by wounds or sickness, and, therefore, are incapable of defending themselves, as long as they do not conduct any hostile acts or try to escape. These additional criteria are important because they expand the scope of the protection of the principle not only to individuals who are explicitly recognized as *hors de combat* in all situations but also to those who should be recognized as *hors de combat* in a given moment of time based on the specific circumstances.[35]

While the principle was historically easy to apply nowadays, it raises several issues. First and foremost, the changing nature of the various armed conflicts makes determining the status of *hors de combat* dependent on the context. Steven Umbrello and Nathan Gabriel Wood provide an interesting example involving poorly armed adversary soldiers who do not have any means to meaningfully engage a tank but do not surrender or fall under another condition described in Additional Protocol I. The authors, however, argue that the customary understanding of the principle involves "powerless" as well as "defenseless" as characteristics that define an individual as being *hors de combat*.[36]

Another possible issue stems from the utilization of autonomous and semi-autonomous weapons. The contextual dependency mentioned earlier makes the application of the principle complicated from a technical point of view. For example, the dominating approach to machine learning e.g., in language modeling, relies on calculating statistical probabilities.[37] State-of-the-art models have no contextual or semantic understanding, and this makes them susceptible to so-called "hallucinations."[38] Reliance on such models to assess whether a person is conducting any hostile acts or is trying to escape is therefore a risk, and an unreliable undertaking. In addition, despite the advances in mitigating the risks of adversarial attacks, computer vision applications remain vulnerable to evasion attacks with adversarial examples that do not require sophisticated skills on the part of the attacker.[39] In addition, any kind of identification is susceptible to unfair bias that could be extremely hard to overcome in military context due to the lack of availability of datasets that are domain specific and reflecting physical traits of combatants (e.g., skin color and uniforms).[40]

[35] Steven Umbrello and Nathan Gabriel Wood, "Autonomous Weapons Systems and the Contextual Nature of Hors de Combat Status" (2021) *Information*, 12(5): 216.

[36] Ibid.

[37] Daniel Jurafsky and James H. Martin, *Speech and Language Processing* (2022) Third Edition draft, 31 https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf, accessed June 5, 2023.

[38] Open AI, "GPT-4 Technical Report" (2023), arxiv Cornell University, 10 https://arxiv.org/abs/2303.08774, accessed June 5, 2023.

[39] Naveed Akhtar, Ajmal Mian, Navid Kardan, and Mubarak Shah, "Advances in adversarial attacks and defenses in computer vision: a survey" (2021) arxiv Cornell University: 24–25 https://arxiv.org/abs/2108.00401v2

[40] Marcus Comiter, "Attacking artificial intelligence. AI's security vulnerability and what policymakers can do about it" (August 2019) *Harvard Kennedy School for Science and International Affairs, Belfer Center*, 37, www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf accessed June 5, 2023.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

20.2.2.3 Prohibition to Inflict Unnecessary Suffering

The principle of prohibition to inflict unnecessary suffering was one of the cornerstones of IHL, as demonstrated by the fact that it was implemented in one of the first IHL legal instruments adopted internationally, namely the St. Petersburg Declaration.[41] This principle is also the key rationale behind important treaties such as the CCW that bans and limits a number of weapons inflicting unnecessary suffering, such as laser weapons and incendiary weapons.[42]

The principle has a customary character,[43] but it is nevertheless codified in written law, primarily in Article 35, paragraph 2 of Additional Protocol I to the Geneva Conventions. This article contains a general prohibition for States to employ "weapons, projectiles, and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering." This formulation of the rule is derived from the principle of humanity which is explained in Section 20.2.2.6.[44] While States agree that the only legitimate military goal is weakening the military of the adversary state, this goal needs to be achieved without unnecessary suffering. In other words, combatants are not prohibited of being killed during an armed conflict, but if it is to be done, it needs to be done "humanely." This notion, however, could be subject to discussion, due to the questionable coexistence of killing and humanely in one sentence in military context. The meaning of the terms "unnecessary" and "superfluous" is also problematic from the standpoint of employing LAWS which need to be designed, created, and used in accordance with this principle, as well as the rest of the rules of IHL. This is so because concepts such as "unnecessary" and "superfluous" suffering are not susceptible to mathematical formalization.

20.2.2.4 Military Necessity

The principle of military necessity[45] was already briefly touched upon in the previous sections due to its relation to the other principles of IHL. The Hague

[41] St. Petersburg Declaration Renouncing the Use, in Time of War, of Explosive Projectiles under 400 Grammes Weight, consideration 4.

[42] Protocols III and IV to the CCW.

[43] Henri Meyrowitz, "The principle of superfluous injury or unnecessary suffering: from the Declaration of St. Petersburg of 1868 to Additional Protocol 1 of 1977" (1994) *International Review of the Red Cross (1961–1997)*, 34(299): 101.

[44] Theodor Meron, "The Martens Clause, principles of humanity, and dictates of public conscience" (2000) *The American Journal of International Law*, 94(1): p. 87.

[45] It is important to mention that military necessity as principle of IHL should not be confused with the state of necessity usually related to state responsibility, as a circumstance precluding the wrongfulness of an act that would otherwise be considered wrongful under international law, see Article 25 of "Articles on Responsibility of States for Internationally Wrongful Acts" in "Report of the International Law Commission on the Work of its 53rd Session" (April 23–June 1, and July 2–August 10, 2001) UN Doc A/56/10.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

Regulations from 1899 and 1907[46] refer to the principle by proclaiming that "[t]he right of belligerents to adopt means of injuring the enemy is not unlimited."[47] This rule shows the collision between military necessity and humanitarian considerations, which results in limiting the former, sometimes even through the creation of new norms, for example the prohibition of destruction of cultural property.[48] As a result of this collision, the principle covers the range of justified and thus allowed use of armed force and violence by a State in order to achieve specific legitimate military objectives, as long as it stays within the limits of the principle of proportionality.

This principle is probably technically the hardest one to abide by when deploying autonomous technologies in military context. This could be explained by the fact that the military necessity is a justification for disregarding the principle of distinction under the circumstances provided in Articles 51 and 52 from Additional Protocol I. Both articles concern defining and protecting civilians in armed conflicts except in the cases when they take direct part in the hostilities. Article 52 also prohibits attacking objectives that are not military objectives, although it acknowledges that civilian objects could become military "when, 'by their nature, location, purpose or use,' such objects 'make an effective contribution to military action' and their total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[49] Evidently, performing such assessments is of paramount importance and in the author's opinion requires meaningful human oversight in order to avoid fatal mistakes.

20.2.2.5 Proportionality

The principle of proportionality in IHL prohibits attacks that may lead to "incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[50] The main issues that doctrine and practice have been facing regarding this principle are related to defining military advantage,[51]

[46] Convention Respecting the Laws and Customs of War on Land annex Article 22 and Convention with Respect to the Laws and Customs of War on Land annex Article 22.

[47] Regulations concerning the Laws and Customs of War on Land annexed to Hague Convention IV (adopted October 18, 1907, entered into force January 26, 1910) (1907) 205 CTS 227 (Hague Regulations), Art. 22.

[48] Nobuo Hayashi, "Hayashi, Nobuo, Requirements of Military Necessity in International Humanitarian Law and International Criminal Law" (2010) *Boston University International Law Journal*, 28(1): 48.

[49] Michael N. Schmitt, "Military necessity and humanity in International Humanitarian Law: preserving the delicate balance" in *Essays on Law and War at the Fault Lines* (Springer, 2012), 96.

[50] Article 51(5)(b) from Additional Protocol I to the Geneva Conventions.

[51] Robin Geiss, "The Principle of Proportionality: 'Force Protection' as a military advantage" (2012) *Israel Law Review*, 45(1): 71–89.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

incidental harm,[52] and excessiveness[53] – all three of these terms being heavily context reliant.

Furthermore, the comparison between the "military advantage" on the one hand and the "excessiveness" on the other hand when performed by a machine learning algorithm (for example by an autonomous military drone such as the STM Kargu-2)[54], requires weighting exercise represented in a computational format. This means that certain values are to be assigned to the categories, which is extremely difficult not only from a technical but also from a legal and ethical perspective.[55] For instance, using autonomous drones designed to destroy military equipment is considered a military advantage. However, doing so in densely populated areas could lead to significant collateral damage, including loss of human life. In this scenario, it is extremely hard if not impossible to assign numerical value and assess whether destroying one piece of military equipment constitutes such a military advantage that it is proportionate to killing two civilians in the blast.

20.2.2.6 The Principle of Humanity

Finally, the principle of humanity underlies every other principle of IHL. It could be defined as a prohibition of inflicting "all suffering, injury or destruction not necessary for achieving the legitimate purpose of a conflict"[56] that aims to protect combatants from unnecessary suffering. It also protects those *hors de combat*.[57] Customary by nature, the principle was first codified in the St. Petersburg Declaration, pointing out that "laws of humanity" are the reason behind the prohibition of arms which "uselessly aggravate the sufferings of disabled men, or render

[52] Eliav Lieblich, "Beyond life and limb: exploring incidental mental harm under International Humanitarian Law" in Derek Jinks, Jackson Maogoto, and Solon Solomon (eds) *Applying International Humanitarian Law in Judicial and Quasi-Judicial Bodies* (T.M.C. Asser Press, 2014), 159–187.

[53] Jason Wright, "'Excessive' ambiguity: analysing and refining the proportionality standard" (2012) *International Review of the Red Cross*, 94(886): 819–854.

[54] For the technical characteristics of the STM Kargu-2 and its capabilities, see www.stm.com.tr/en/kargu-autonomous-tactical-multi-rotor-attack-uav; Hitoshi Nasu, "The Kargu-2 autonomous attack drone: legal & ethical dimensions" (June 2021) Lieber Institute Blog, https://lieber.westpoint.edu/kargu-2-autonomous-attack-drone-legal-ethical/ accessed June 5, 2023.

[55] Tomasz Zurek, Taylor Woodcock, Magdalena Pacholska, and Tom van Engers, "Computational modelling of the proportionality analysis under International Humanitarian Law for military decision-support systems" (January 14, 2022) *SSRN* https://ssrn.com/abstract=4008946 accessed January 6, 2023.

[56] The definition is proposed by the ICRC in ICRC, "What is IHL?" (September 2015) *ICRC blog*, www.icrc.org/en/document/what-ihl, accessed June 7, 2023. However, the discussion on the content and the applicability of the principle is still ongoing, see Kjetil Larsen, Camilla Guldahl Cooper, & Gro Nystuen (Eds.), "Searching for a 'Principle of Humanity' in International Humanitarian Law" (2012) Cambridge University Press.

[57] Geoffrey S Corn, "Humanity, Principle of" Max Planck Encyclopedia of Public International Law (July 2013) https://opil-ouplaw-com.kuleuven.e-bronnen.be/display/10.1093/law:epil/9780199231690/law-9780199231690-e1810, accessed January 6, 2023.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

their death inevitable."[58] The principle of humanity also further balances military necessity, which is reflected in most IHL instruments. This balancing function is best described by the so-called Martens Clause in the Preamble to the Hague Conventions on the Laws and Customs of War on Land from 1899. The idea of the clause named after the diplomat Friedrich Martens is very simple. In essence, it attempts to fill a possible gap in the legislation by providing that, in the situation of an armed conflict in the absence of a legal norm, belligerents are still bound by the laws of humanity and the requirements of public conscience.

This principle is rather vague and open to interpretation, evident by the attempts to apply it in numerous contexts, from the deployment of LAWS to nuclear weapons.[59] Furthermore, it is one of the most challenging principles from an ethical and philosophical point of view. The principle of humanity requires an almost Schrödinger-like state of mind in which the enemy on the battlefield which is to be killed is also regarded as a fellow human being. This paradox makes the practical application of the principle a very hard task not only by humans but even more so by autonomous systems used in a military context, which more often than not reflect human bias.

20.3 USING AI IN ARMED CONFLICTS

The present chapter already sporadically touched upon several legal, ethical, and technical problems raised by applying the principles of IHL to AI systems used for military purposes. This next section is going to concentrate on the different applications for which AI systems are used, which includes LAWS or so-called "killer-robots" without being limited thereto. It also discusses the possible regulatory framework and challenges before the full-scale adoption of such systems.

20.3.1 *Lethal Autonomous Weapons Systems*

LAWS have been in the spotlight for several years already due to the blooming of the tech industry and in particular the huge investments and reliance on AI systems. Naturally, realizing the capabilities of AI in everyday life poses the question of how it can be used in military setting. Combining technological development with justified ethical concerns and the catchy phrase "killer robots" served as a great source of inspiration for media and entertainment, with the unfortunate result of shifting the focus of the discussion from people and their behavior to the regulation of inanimate objects.

[58] St. Petersburg Declaration Renouncing the Use, in Time of War, of Explosive Projectiles under 400 Grammes Weight, consideration 5.

[59] Theodor Mero, "The Martens Clause, principles of humanity, and dictates of public conscience" (2017) *American Journal of International Law*, 94(1): 78–89.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

There are a number of problematic points related to the discussion about LAWS that prevent policymakers and other relevant stakeholders to move away from speculation and concentrate instead on the real dangers and immediate issues caused by using AI in armed conflicts.

First, despite the concept of LAWS being around for quite a while, we still lack a universal definition of what constitutes a lethal autonomous weapon.[60] A definition is extremely important because "autonomy" is a scaled feature. Therefore, the answer to the question of whether a system is truly autonomous depends on where on the curve the definition positions LAWS. A similar issue appeared during the ongoing process of creating legislation on AI in several jurisdictions,[61] such as the AI Act in the EU, discussed in Chapter 12 of this Book.

Defining LAWS has been a key task for the High Contracting Parties and Signatories of the CCW, through the Group of Governmental Experts on emerging technologies in the area of LAWS (GGE on LAWS). However, their efforts to adopt a universal definition currently remain fruitless.[62] Hence, the lack of a universal definition is filled by a plethora of more regional definitions serving various purposes. For example, the United States Department of Defense characterizes fully autonomous weapons systems as systems that "once activated, can select and engage targets without further intervention by a human operator."[63]

The European Parliament, by comparison, adopted another definition through a resolution referring to LAWS as "weapon systems without meaningful human control over the critical functions of selecting and attacking individual targets."[64] The two definitions, although similar, reveal some important differences. The definition provided by the European Parliament is based on the absence of *meaningful* human oversight, while the US definition only speaks about control. Additionally, Directive 3000.09[65] also defines semi-autonomous systems, highlighting the scale-based nature of autonomy, while the resolution of the European Parliament does

[60] UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches" (2017) *UNIDIR Resources*, https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-concerns-characteristics-and, accessed January 3, 2023.

[61] Jonas Schuett, "Defining the scope of AI regulations" (August 22, 2021) *Law, Innovation and Technology, Legal Priorities Project Working Paper Series No. 9*; Katerina Yordanova, "The EU AI Act – Balancing Human Rights and Innovation Through Regulatory Sandboxes and Standardization" (2022) *Competition Policy International*, www.competitionpolicyinternational.com/the-eu-ai-act-balancing-human-rights-and-innovation-through-regulatory-sandboxes-and-standardization/, accessed January 7, 2023.

[62] Ann-Katrien Oimann, "The responsibility gap and LAWS: a critical mapping of the debate" (2023) *Philosophy & Technology*, 36(3): 4.

[63] US Department of Defense, "Directive 3000.09 Autonomy in Weapon Systems" (November 21, 2012) www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf, accessed January 7, 2023.

[64] European Parliament resolution of September 12, 2018 on autonomous weapon systems (2018/2752(RSP)), www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.html, accessed January 7, 2023.

[65] USA Department of Defense Directive 3000.09, "Autonomy in Weapon Systems," November 21, 2012.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

not mention them at all. The discrepancy between just these two examples of LAWS definitions demonstrates that a number of weapons systems might fall in or out of the scope of norms that should regulate LAWS based on the criteria applied by a certain State, which leads to uncertainty.

A second problematic point in the LAWS discussion is the speculative nature of their capabilities and deployment. While the general consensus is that the deployment of killer robots is a highly undesirable prospect for both ethical and legal reasons, such as the alleged breach of the principle of humanity[66] and compromised human dignity,[67] those arguments are based on what we imagine LAWS to be. Military technology has always been surrounded by a very high level of confidentiality, and this is also the case when it comes to LAWS. Until recently, their nature could only be speculated on based on more general knowledge about AI advances and robotic applications outside their military application. The first officially recognized use of LAWS was established by the UN Security Council's Panel of Experts on Libya regarding an incident from March 2020. During that incident, a Turkish STM Kargu-2 drone, referred to as a lethal autonomous weapon, "hunted down and remotely engaged" logistic convoys and retreating forces in Libya.[68] The drone was described as being "programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true 'fire, forget and find' capability."[69] There is little to no information, however, regarding the measure taken to ensure compatibility of the drone with the norms and principles of international law.

The UN's recognition of the use of autonomous drones in Libya reignited the public debate on the international rules applicable to LAWS. It also strengthened the calls for a total ban on killer robots, unfortunately without the participation of countries like the USA, Russia, Turkey, and others having the necessary resources for the research and development of autonomous weapons.

Nevertheless, relevant stakeholders such as states, NGOs, companies, and even individuals with high standing in society[70] did take some significant steps in two directions. First, they launched campaigns promoting a total ban on LAWS[71] and

[66] Elvira Rosert and Frank Sauer, "Prohibiting autonomous weapons: put human dignity first" (2019) *Global Policy*, 10(3): 373.

[67] Amanda Sharkey, "Autonomous weapons systems, killer robots and human dignity" (2019) *Ethics and Information Technologies*, 21(2): 87.

[68] UN Security Council's Panel of Experts on Libya, "Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011)" (March 8, 2021) S/2021/229.

[69] Ibid.

[70] Ian Sample, "Thousands of leading AI researchers sign pledge against killer robots" (2018) *The Guardian*, www.theguardian.com/science/2018/jul/18/thousands-of-scientists-pledge-not-to-help-build-killer-ai-robots, accessed January 13, 2023.

[71] Human Rights Watch, "Stopping killer robots. Country positions on banning fully autonomous weapons and retaining human control" (2020) *Human Rights Watch Website*, www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and, accessed January 13, 2023.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

second, they started developing and applying rules that govern the existing LAWS in accordance with IHL.[72]

The first initiative, even though remaining popular in civil society, does not currently show any significant development on a global scale.[73] The second one has a better success rate, being built around the application of the CCW, which was deemed to be the most suitable instrument to attempt to regulate LAWS. It is so because of its purpose to ban or restrict the use of specific types of weapons that are considered to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately. In addition, the structure of the Convention allows flexibility in dealing with new types of weapons, such as for example blinding laser weapons.[74]

Therefore, this particular forum was considered as best placed to discuss the technological development of LAWS and the legal and customary rules applicable to them,[75] as well as to codify the results of this discussion. The GGE on LAWS adopted the eleven guiding principles in the 2019 Meeting of the High Contracting Parties to the CCW.[76] The eleven guiding principles first and foremost address the need for LAWS to comply with IHL (guiding principle 1), as well as the human responsibility for the use of LAWS (guiding principle 2). This was important not only because of the ongoing trend of personification of machines[77] but also because of the existing debate regarding the so-called "responsibility gap" created by LAWS.[78] This is further supported by guiding principle 7, which forbids anthropomorphizing LAWS. The guiding principles also elaborate on the need for human–machine interaction, accountability mechanisms, and a sound balance between military necessity and humanitarian considerations.

[72] An example of that is the established GGE, under the mandate of the CCW Meeting of High Contracting Parties and their ongoing work, including the eleven guiding principles on LAWS.

[73] James Dawes, "UN fails to agree on 'killer robot' ban as nations pour billions into autonomous weapons research" (December 2021) *The Conversation*, https://theconversation.com/un-fails-to-agree-on-killer-robot-ban-as-nations-pour-billions-into-autonomous-weapons-research-173616, accessed January 13, 2023.

[74] Protocol IV of CCW adopted on October 13, 1995 during the First Review Conference of the States parties to the convention, entered into force July 30, 1998.

[75] In particular, see Additional Protocol I to the Geneva Conventions, Article 36 providing that "[i]n the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party."

[76] Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, "Final Report" (December 13, 2019) CCW/MSP/2019/9, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/343/64/PDF/G1934364.pdf?OpenElement, accessed January 13, 2023.

[77] Ugo Pagallo, "Vital, Sophia, and Co., The quest for the legal personhood of robots" (2018) *Information*, 9: 230.

[78] Oimann, *The Responsibility Gap and LAWS: A Critical Mapping of the Debate*, p. 7.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

Despite the guiding principles showing some progress and common ground among States Parties to the CCW on the subject of LAWS, they are not a binding legal instrument and are not a lot to show after eight years of work on the side of the GGE. Indeed, the formulated principles could become useful by demonstrating convergence on elements of customary international law such as *opinion juris*[79] or serving as a starting point for a new protocol to the CCW. Currently, however, neither of those appear to be on the agenda.

20.3.2 *Other Applications of Autonomous Systems in Armed Conflicts*

As discussed previously, the precise extent of the use and development of LAWS remain mostly confined to speculations. While weapons with a certain degree of autonomy such as unmanned aerial vehicles are currently used in many armed conflicts, truly autonomous weapons remain a rarity. However, outside the context of LAWS, the more general use of AI systems in armed conflicts and for military purposes is becoming the new normal.

AI remains a great tool for supporting decision-making in the military, assisting personnel with going through large quantities of data and analyzing it, and making the "right" judgment regarding transport, logistics, communications, and others.

Furthermore, AI has many applications in data gathering, including as a surveillance tool, although such AI products may fall under the category of dual-use items.[80] Other possible utilizations of AI systems are applications for predictive maintenance of military equipment,[81] unarmed vehicles such as ambulances, supply trucks or drones,[82] and medical aid.[83] While for some of these purposes of IHL might still be a consideration, such as for example the special regime of medical vehicles, other uses might raise more fundamental concerns when aspiring conformity with human rights law. A typical example is dual-use technologies used for

[79] In public international law, *opinio juris sive necessitates* refers to the second element needed for the establishment of a binding customary rule, namely the belief of a State that it is *bound* by a certain rule or obligation.

[80] Dual-use items are goods designed for civilian use, which also have military utilization. Usually, such goods are subject to a special export control regime. In EU, this regime is established by Regulation (EU) 2021/821. For more information, see Machiko Kanetake, "Dual-Use Export Control: Security and Human Rights Challenges to Multilateralism" in Marc Bungenberg, Markus Krajewski, Christian J. Tams, Jörg Philipp Terhechte, and Andreas R. Ziegler (eds), *European Yearbook of International Economic Law 2020* (Springer, 2021).

[81] For example, ALIS is an AI tool used to predictively maintain F-35 fighter jets.

[82] Millicent Abadicio, "Artificial intelligence for military logistics – current applications" (2019) *EMERJ website* https://emerj.com/ai-sector-overviews/artificial-intelligence-military-logistics/ accessed January 13, 2023.

[83] Ajinkya Jadhav, "Developing AI in combat healthcare" (2021) *CLAWS* www.claws.in/developing-ai-in-combat-healthcare/ accessed January 13, 2023.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

surveillance of individuals, which violate their right to privacy and often other rights too, without a plausible justification based on military necessity.[84]

Finally, AI is increasingly used in cyberwarfare, both for its attack and defense capabilities.[85] Typically, this involves the spreading of viruses and malware, data theft, distributed denial-of-service attacks, but also disinformation campaigns.[86] AI systems can enhance these activities, as well as help to circumvent defenses on the way. AI is, however, not only used for defense purposes beyond military infrastructure but also to protect civilian objects and infrastructure that is more susceptible to cyberattacks due to being part of IoT.[87]

While cyberwarfare has not been explicitly mentioned by IHL instruments, this does not mean that it falls outside its scope. On the contrary, the potential (and often the intent) of cyberattacks to be indiscriminate is a considerable reason for applying IHL to cyberwarfare in full force, especially when AI systems are involved given their ability to further increase the scale of the attack.

20.4 CONCLUSION

All around the world, regulators are thinking about the changing power of AI in every aspect of our lives. Although some are calling for the adoption of stricter rules that would not allow something to be done just because we have the technology to do it, others support a more innovation-friendly, liberal approach to AI regulation, which would assist the industry and allow a more rapid development of AI in the tech sector. While both positions have their merit, when it comes to using AI in the military context, the stakes change dramatically.

The potential of AI in armed conflicts goes in two directions. It can save resources and lives more efficiently by supporting humans to make the "right" decisions or avoid unnecessary causalities or it could assist in killing people more efficiently. During a war, those might be the two sides of the same coin. Therefore, we need to ensure that IHL is embedded in the design, development, and use of AI systems to the best extent possible.

[84] For a more in-depth discussion on dual-use AI, see Ilaria Carrozza, Nicholas Marsh & Gregory M. Reichberg, "Dual-use AI technology in China, the US and the EU: strategic implications for the balance of power" (2022) *PRIO Paper*. Oslo: PRIO, www.prio.org/publications/13150, accessed June 10, 2023.

[85] For a more in-depth discussion, see Maria Taddeo, Tom McCutcheon, and Luciano Floridi, "Trusting artificial intelligence in cybersecurity is a double-edged sword" (2019) *Nature Machine Intelligence*, 12(1): 557–560.

[86] Rod Thornton and Marina Miron, "Towards the 'third revolution in military affairs'. The Russian military's use of AI-enabled cyber warfare" (2020) *The 165 RUSI Journal*, 165(3): 12–21.

[87] Petri Vähäkainu, Martti Lehto, Antti Kariluoto, and Anniina Ojalainen, "Artificial intelligence in protecting smart building's cloud service infrastructure from cyberattacks" in Hamid Jahankhani, Stefan Kendzierskyj, Nishan Chelvachandran, and Jaime Ibarra (eds) (Springer, 2020) *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*, 289–319.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press

Even if killer robots would be successfully banned, that is not a guarantee that they would not be used anyway. Furthermore, as demonstrated, AI has many more applications on the battlefield, all of which need to be in accordance with the basic considerations of humanity during an armed conflict. Bringing military AI systems under the scope of IHL would give us at least some hope that, while we cannot stop AI from being used in wars, maybe we can change the ratio of its use as a weapon for killing, toward its use as a shield for protection.

https://doi.org/10.1017/9781009367783.024 Published online by Cambridge University Press