CROSSING BOUNDARIES: AUTONOMOUS WEAPON SYSTEMS AND THE CHALLENGE OF IHL COMPLIANCE

Maria Gevorgyan*

Abstract: The introduction of autonomous weapon systems (AWS) marks a transformative juncture in the modern landscape of warfare. Promising operational efficiency, enhanced soldier safety, cost reduction, and workforce minimization, these systems have ignited a global discourse concerning their compliance with international law and the necessity of comprehensive regulation. This study delves into the multifaceted challenges associated with the deployment of AWS within the context of International Humanitarian Law (IHL) while evaluating their alignment with the principles of IHL, especially in terms of direct participation in hostilities (DPH). The research begins by establishing a fundamental understanding of autonomy, delineating the criteria that define AWS, and addressing their legal categorization—whether they constitute a method of warfare or serve as a replacement for combatants. By employing a diverse range of research methodologies encompassing system analysis, comparative legal analysis, synthesis, comparison, analogy, deduction, classification, interviews, and case studies, this study provides a comprehensive examination of the intricate AWS-IHL relationship. Further depth is added to the theoretical analysis through real-world case studies, including the STM Kargu-2 and the United Nations (UN) expert group's involvement in Libya, offering practical insights into the challenges posed by AWS in armed conflicts. Additionally, consideration is given to the SGR-A1, an autonomous system employed for border safeguarding, further illuminating the complexities of AWS in practice. This research aims to provide a nuanced and insightful understanding of the pressing regulatory and challenges arising from the utilization of AWS within the contemporary framework of IHL.

Keywords: Autonomous Weapon Systems; International Humanitarian Law; IHL Principles; Modern Warfare

* Center for Truth and Justice, United Kingdom.

Table of Contents

INTRODUCTION

The development and deployment of autonomous weapons systems (AWS) represent a significant paradigm shift in the landscape of modern warfare. As militarily developed states continue to invest in the creation of increasingly autonomous weapons, the implications for both the means and methods of warfare, as well as compliance with the law of armed conflict, have become subjects of intense debate and scrutiny. While it is true that certain existing weapons systems have exhibited limited autonomous capabilities, the current trajectory of development suggests a substantial expansion of these capabilities in the near future. It is increasingly unlikely that the deployment of AWS can be halted, and this reality has ignited a profound global conversation.

This prospect has ignited a passionate debate. Human rights organizations are calling for a preemptive ban on the use of autonomous weapons systems, while numerous states are voicing their concerns within international forums, such as the UN. Despite these debates, there exists no specific international legal framework governing AWS. Nevertheless, the motivation for increasing the level of autonomy of weapon systems is compelling, driven by the promise of greater operational efficiency, enhanced safety for one's own soldiers, reduced personnel requirements, and significant cost savings. In June 2022, representatives of the US Defense Ministry underscored the pivotal role of digital transformation and artificial intelligence (AI) in maintaining a competitive edge on the battlefield.[1] Similar efforts in the field of AI and AWS are underway in the armed forces of other nations, including Israel[2] and China.[3] Russia, likewise, has not remained passive, and approximately a year before its another invasion of Ukraine in 2022, it became evident that Russia was expanding its arsenal of weapons equipped with AI capabilities.[4] However, these developments occur without internationally agreed guarantees, including legal ones. Despite advocating for the use of AI in armed conflicts as early as 2021, as reflected in the 2019 Report on AI, the ICRC has expressed[5] profound concerns regarding AWS. Furthermore, the UN Secretary-General articulated in March 2019 that AWS are politically unacceptable, morally repugnant, and should be prohibited by

---

[1] Dave Vergun, *'Digital Transformation, AI Important in Keeping Battlefield Edge, Leaders Say'* (2022) US DEPARTMENT OF DEFENSE'S NEWS, https://www.defense.gov/News/News-Stories/Article/Article/3058028/digital-transformation-ai-important-in-keeping-battlefield-edge-leaders-say/ (last visited 15 September 2023).

[2] S. Biddle, *'Documents reveal advanced AI tools Google is selling to Israel'* (2022) THE INTERCEPT https://theintercept.com/2022/07/24/google-israel-artificial-intelligence-project-nimbus/ (last visited 15 April 2023).

[3] PEOPLE'S REPUBLIC OF CHINA, 'POSITION PAPER ON REGULATING MILITARY APPLICATIONS OF ARTIFICIAL INTELLIGENCE (AI)' (2021) https://www.fmprc.gov.cn/mfa_eng/wjdt_665385/wjzcs/202112/t20211214_10469512.html (last visited 15 September 2023).

[4] Lieutenant General Michael Groen, the Director of the Pentagon's Joint Artificial Intelligence Center, who has been involved in implementing artificial intelligence within the US Department of Defense since 2018, stated, 'The Russian Armed Forces are striving to become a leader in artificial intelligence technologies.' Additionally, CNA, a research organization based in Arlington, Virginia, was commissioned to examine the Russian market. In a report titled 'Artificial Intelligence and Autonomy in Russia,' it is noted that there are over 150 military systems with artificial intelligence in various stages of development. Groen explained that the country aims to utilize AI for electronic warfare, intelligence, surveillance, reconnaissance, and strategic decision-making processes as its leaders seek information dominance on the battlefield. Yasmin Tajde, '*Algorithmic Warfare: Russia Expands its Fleet of Weapons with Artificial Intelligence Support*', NATIONAL DEFENSE (2021) https://www.nationaldefensemagazine.org/articles/2021/7/20/russia-expanding-fleet-of-ai-enabled-weapons (last visited 15 September 2023).

[5] 'ICRC POSITION ON AWS' (2021) https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems (last visited 15 April 2023).

international law.[6] This divergence in viewpoints underscores the urgency of addressing the regulatory and ethical challenges posed by AWS.

Member States of the Convention on Certain Conventional Weapons (CCW) have been engaged in discussions related to these systems for over seven years, and while these deliberations hold the potential for future regulation, concrete steps in that direction have yet to be taken. In the context of deploying AWS, a highly significant yet intricate matter emerges. The core concern revolves around the potential fallibility of machine-based decision-making, which may arise from the inability to account for intricate nuances specific to a particular situation. Human agents are presumed to possess superior capabilities in this regard. Consider, for instance, a scenario in which a drone, armed with explosive ordnance, is dispatched to eliminate an enemy military target positioned in close proximity to a residential building housing civilians. In such a situation, the act of bombing the military facility carries a substantial risk of collateral damage to civilians. When AI governs the drone's operations, numerous pressing questions arise. How effectively can AI navigate the complexities of decision-making? Can the drone ensure adherence to the norms of International Humanitarian Law (IHL) in a manner that strikes a delicate balance between the principles of humanity and military necessity? These represent merely a fraction of the weighty issues inherent in the deployment of AWS.

The primary objective of this study is to explore the multifaceted issues that emerge when deploying AWS within the context of IHL. Moreover, this research will assess the compliance of AWS with the foundational principles of IHL. With a foundational understanding of autonomy, legal classification of AWS according to IHL, and meticulous case studies involving modern AWS, this study aims to provide a comprehensive analysis of the complex relationship between AWS and IHL.

I. METHODOLOGY

To address these complex and multifaceted issues, this research relies on a diverse array of methodological approaches. These include system analysis, comparative legal analysis, synthesis, comparison, analogy, deduction, classification, interviews, monitoring, and case studies. Through this comprehensive approach, we aim to provide a nuanced and insightful understanding of the challenges posed by AWS within the framework of IHL.

II. DISCUSSION

A. Understanding an Autonomy

The initial step in evaluating the legal implications of heightened autonomy in weapon systems is to establish a clear understanding of the technological characteristics underpinning these changes. Only then will the legal significance of these developments become apparent. The technical discussions in this chapter will serve as the foundation for the legal analysis throughout the entire study. In the context of legal analysis, there is no need to delve excessively into technical details. Instead, the primary focus should be on the general possibilities for limiting and utilizing AWS.

[6] ANTÓNIO GUTERRES, 'REMARKS AT WEB SUMMIT' (2018) https://www.un.org/sg/en/content/sg/speeches/2018-11-05/remarks-web-summit (last visited 13 May 2023).

The concept of "autonomy" varies depending on the areas of study and the debates surrounding the use of AWS, which poses a significant challenge. The absence of a unified definition hampers rational discussions of their legal significance. It is crucial to understand that a weapon system, whether in its simplest form, or in a complex form, can be misleading, especially from a legal perspective. The lack of interaction between the machine and the operator during operation does not necessarily mean that the machine's behavior is not determined by a human. Rather, it indicates that the intended behavior was predetermined before the machine's activation and is executed by some part of it, typically through computerized control systems. This control system monitors the weapon system's operation and issues commands as necessary to achieve the desired programmed behavior.

The critical point to emphasize at this stage is that when a "manual" system is replaced by a system capable of a certain degree of autonomy, it may appear as though the control system is effectively supplanting the human operator. This is partially true because the operator's understanding of how to operate the machine is programmed into the control system: the physical means through which the operator manipulates the machine are transformed into a set of actuators that can be activated by the control system itself. Moreover, additional sensors can be incorporated, allowing the machine to process available information based on relatively accurate and dependable environmental perception, in order to develop a meaningful plan and utilize its actuators to set that plan into motion.[7]

Indeed, to an observer, these may seem like highly intricate programs; however, they are merely a collection of predefined instructions, and the machine executes instructions that were pre-written, rather than acting independently. For greater clarity, let's turn to the following practical example. A hypothetical Unmanned Aerial Vehicle (UAV) used for counterterrorism purposes may be equipped with cameras and image recognition software that matches images captured by the cameras to images of known terrorist locations. An instruction such as "if the camera image matches those of known terrorists, behave as if terrorists are present" or some other set of rules by which the UAV can compare the images captured by its cameras does not mean the UAV is making a determination in any real sense as to whether a person is a terrorist; it is still inaccurate to portray the UAV as making independent judgments. Instead, the UAV is simply following instructions provided to it in advance.

Such a lengthy description of technical capabilities aims to eliminate misleading formulations that exist in legal debates, suggesting that AWS have the ability to make real choices during operations. At this stage of development, no computer is capable of independently choosing to execute or not execute a specific instruction in a program.[8] Any such appearance of choice may result from other encoded instructions in the software. Thus, autonomy is the ability of a system to behave as desired and achieve the goals *provided by its operator* without the need for constant external instructions.

B. Defining AWS and Their Classification

In the context of the research, an interview was conducted with Dr. Alex Leveringhaus, a Ph.D. holder in the field of public administration, a research fellow at the Institute of Ethics,

[7] OFFICE OF THE SECRETARY OF DEFENSE, 'UNMANNED AIRCRAFT SYSTEMS ROADMAP, 2005– 2030', US DEPARTMENT OF DEFENSE (2005), https://fas.org/irp/program/collect/uav_roadmap2005.pdf (last visited: 10 October, 2023).

[8] T. MCFARLAND, AWS AND THE LAW OF ARMED CONFLICT, IN AWS AND THE LAW OF ARMED CONFLICT: COMPATIBILITY WITH INTERNATIONAL HUMANITARIAN LAW, 36, (CAMBRIDGE UNIVERSITY PRESS, 2020).

Law, and Armed Conflict at the University of Oxford, and the coordinator of the Special Group on Ethics and AI. Through this interview, it was possible to establish that one of the most challenging questions in nearly all debates concerning AWS is the issue of defining these systems.[9] Various definitions of AWS exist; however, there is no consensus on this issue.[10] A more technical approach to autonomy considers the actual ability of a system to control its behavior and deal with uncertainties.[11] According to this approach, an AWS is a system capable of, based on its perception of the environment, taking the necessary actions to achieve a desired goal. Machines that can adapt to changes in the environment and exercise control over their actions can be characterized as automated or autonomous. Here arises the question: what is the essential difference between these two mentioned systems? Some experts see the difference in the degree of self-governance, considering AWS as more complex, intellectually advanced forms of automated systems.[12] However, there is no definitive answer. There is also an approach that focuses on the command-administrative relationship between humans and AWS.

In the context of this research, the focus has deliberately shifted away from treating autonomy as a general attribute of AWS. This is because such a broad approach can be misleading and lead to intricate debates concerning the threshold for considering a weapon system as autonomous. Instead, the contention is that issues, particularly those with legal implications, should be addressed by considering the specific functions or tasks for which autonomy is employed. Autonomy is most effectively analyzed by categorizing it according to the functions performed at the level of an AWS.[13] A primary advantage of adopting this approach lies in its flexibility for investigating issues related to AWS. For instance, the functional approach allows for acknowledging that the extent of human interaction, including operator control, varies between different functions within AWS. Some functions may necessitate a higher degree of autonomy, while human control may be retained for others or relinquished entirely. Furthermore, the level of human operator involvement can fluctuate based on the mission at hand. Consequently, this approach implies that the concept of "AWS" is a comprehensive term encompassing a wide array of weaponry with autonomy integrated into their critical functions. This includes weapons capable of autonomously selecting (searching, identifying, tracking) and engaging (applying force to) targets without continuous human intervention.[14]

Thus, an AWS is a weapon that, once activated, can select and engage targets without further *constant* intervention by a human operator.[15] Furthermore, within this approach, the following types of AWS can be identified.

---

[9] Dr. Alex Leveringhaus, a Ph.D. holder in the field of public administration, a research fellow at the Institute of Ethics, Law, and Armed Conflict at the University of Oxford, the coordinator of the Special Group on Ethics and AI, personal interview 02.07.23.

[10] P. Scharre, 'Autonomous Weapons and Operational Risk', 16, Center for a New American Security, Ethical Autonomy Project, (2016).

[11] S. Thrun, 'Toward a Framework for Human-Robot Interaction', 9-24 (2004) 19(1–2) Human–Computer Interaction.

[12] D. Mindell, *Our Robots, Ourselves: Robotics and the Myths of Autonomy*, 12 (2015)Viking: New York,.

[13] United Nations Institute for Disarmament Research (UNIDIR), Framing Discussions on the Weaponization of Increasingly Autonomous Technologies, UNIDIR Resources No. 1 (UNIDIR: Geneva, 2014).

[14] 'AWS: Is it Morally Acceptable for a Machine to Make Life and Death Decisions?', 13-15, ICRC, (CCW Meeting of Experts on Lethal AWS, (Geneva, April 2015).

[15] P. Scharre, 'Where Does the Human Belong in the Loop?', 4, CCW Meeting of Experts on LAWS: Technical Issues, (May 2014).

1. Semi-autonomous system - human in the loop. In this configuration, the system is programmed to await input from a human operator before taking action.

2. Supervised autonomous system - human on the loop. In this case, the program allows for human intervention but does not require real-time and mandatory human involvement, as is the case with the aforementioned autonomous systems.

3. Fully autonomous system - human out of the loop. In this scenario, the system is programmed in a way that does not permit real-time human intervention.

C. Legal Categorisation of AWS

The subject of debates has also revolved around the question of the legal category to which AWS belong. Some participants in the discussions argue that AWS occupy a position between weaponry and combatants, which in turn raises questions about their ability to adhere to the norms of IHL. The central idea behind this approach is as follows: when a significant portion of the "targeting" process is encoded within the weapon system, the AWS assumes the responsibilities of a soldier, acting as a kind of delegate of the combatant or an artificial surrogate for the combatant. The concept of "AWS" often appears in formulations that seemingly position these systems as bearers of obligations under IHL.[16] In the context of this research, arguments of this nature, which imply the replacement of humans by AWS in operational and possibly legal terms, do not consider AWS as tools used by humans. However, within the scope of this study, autonomy implies a form of control rather than its absence, which is why the aforementioned positions do not appear sufficiently substantiated. Let us attempt to analyze the legal validity of the position presented above and address the question: is it appropriate to classify AWS as a means of warfare for legal purposes, as envisaged by IHL? The argument that the ability of AWS to perform tasks traditionally assigned to combatants justifies categorizing them into a legal category distinct from the means of warfare does not appear well-founded.

The provisions of Protocol Additional I to the Geneva Conventions 1949 (API) "Believing it necessary nevertheless to reaffirm and develop the provisions protecting the victims of armed conflicts and to supplement measures intended to reinforce their application," as well as the provisions related to weaponry, imply that terms should be interpreted expansively to ensure such protection. This is also confirmed in the Commentary of the ICRC to Article 35 of AP I, the provisions of which affirm the principle that the right of parties to the conflict to choose methods or means of warfare is not unlimited, and that means of warfare encompass weapons in the widest sense.[17] Article 36 of API refers to the scope of application of the weapons review mechanism, thereby indicating the need to adhere to a broad concept of weaponry, which is supported by state practice.

---

[16] HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS (2012) 30; C. HEYNS, REPORT OF THE SPECIAL RAPPORTEUR ON EXTRAJUDICIAL, SUMMARY OR ARBITRARY EXECUTIONS, HUMAN RIGHTS COUNCIL, 23RD SESS, 5-6 [28], AGENDA ITEM 3, UN DOC A/HRC/23/47 (9 April 2013).

[17] COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, (1987), 130, INTERNATIONAL REVIEW OF THE RED CROSS (1961 - 1997).

In accordance with the concept of a broad understanding of weaponry, substantiated by state practices, the Working Group of the U.S. Department of Defense on Military Law defines "weaponry" as encompassing all armaments, munitions, material components, mechanisms, or devices intended to have the presumed effect of causing injuries, damage, destruction, or incapacitation of personnel or equipment. Furthermore, the same Working Group defines a "weapon system" as the weapon itself and the components required for its operation, including new advanced or emerging technologies that may lead to the development of weapons or weapon systems with significant legal and political implications.[18] In turn, the Australian Department of Defense specifies that the concept of "weaponry" encompasses weapon systems, ammunition, submunitions, guidance devices, and other destructive mechanisms.[19]

These definitions and characteristics apply to the components, functions, and effects typically possessed by AWS, thereby allowing them to be considered as means of warfare, especially given the need for a broad interpretation. It should also be acknowledged that, despite the presented technical anthropomorphic designs, the physical composition and form of AWS do not significantly differ from other types of weaponry. Indeed, the software distinguishes itself with a range of capabilities, but it remains software developed by a human programmer, with a design fundamentally similar to that used in other armaments.

The examination of the functional aspects of AWS should not influence a change in the legal categorization of AWS as a means of warfare. Regarding input data, as previously noted, their role remains to receive a command from a human source and execute it, with the nature of commands varying as the autonomy of the weapon system changes. It is crucial to understand that the enhanced capabilities of AWS may lead to significant changes in the conduct of military operations, but they do not alter the legal category. In other words, there is no causal relationship between the function performed by an AWS and its legal categorization under IHL.

Consider another argument put forth by proponents of the concept that AWS are something more than just weapons. It is argued that the ability of a system to gather data about the surrounding world and use it to formulate high-level commands, which were not explicitly given, serves as yet another example of the common human tendency towards anthropomorphism.[20]Anthropomorphic concepts of autonomy bring to the forefront one or more human-like qualities or models of behavior that autonomous systems can exhibit. Definitions related to operational independence of weapon systems are among the most common. In technical terms, "independence" implies the absence of a need for explicit instructions from an operator, as previously explained; these instructions are pre-encoded in the weapon system's control software. However, this factor is often distorted to support legal arguments and claims that it signifies the independence of actions of AWS from humans.[21] In the legal context, arguments equating the role of a soldier with that of an AWS ignore a crucial fact: weapons and soldiers are distinct legal categories, despite potential functional overlap, which the author of this study does not acknowledge. Unlike weapons, computers, software,

---

[18] DEPARTMENT OF DEFENSE DICTIONARY OF MILITARY AND ASSOCIATED TERMS (12 April 2001), https://irp.fas.org/doddir/dod/jp1_02-april2010.pdf (last visited: 10 October 2023).

[19] 'LEGAL REVIEW OF NEW WEAPONS' (DEFENCE INSTRUCTION (GENERAL) OPS 44-1, AUSTRALIAN DEPARTMENT OF DEFENCE, (2 June 2005) sub-s 3(a).

[20] S. FUSSELL, 'HOW PEOPLE ANTHROPOMORPHIZE ROBOTS', 145, (2008) 3RD ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN ROBOT INTERACTION PROCEEDINGS.

[21] *'Mind the Gap: The Lack of Accountability for Killer Robots',* HUMAN RIGHTS WATCH, (April 2015), https://www.hrw.org/sites/default/files/report_pdf/arms0415_summary_mindthegap.pdf (last visited: 10 October 2023).

'combatant' is a legal category encompassing only humans, who, in turn, possess rights and responsibilities that underlie IHL. Positioning AWS as combatants in a legal sense would grant them the role of exercising primary control over decisions and committing acts of violence. However, the analysis of autonomous weapon technology presented above reveals that any such assertion of independence from human control is an illusion. AWS, like other forms of weaponry, differ from combatants, and drawing a legal analogy between these two categories appears fundamentally untenable and irrational. Neither existing legislation nor political considerations support the assumption that legal personhood should extend to artifacts, such as AWS.[22] In legal terms, AWS should be regarded as means of warfare, with humans and the states deploying and operating them being the bearers of legal obligations.

The question also necessitates an answer: can AWS be legally regarded as a method of warfare? Given the inherently fluid nature of machine autonomy and the myriad of ways it can be employed, a sufficiently expansive interpretation of the latter concept will be required to classify AWS as methods of warfare. A more plausible proposition is that the specific behavior exhibited by a particular AWS might qualify as a method for conducting military operations. Nevertheless, even in such cases, it is imperative to acknowledge that any weapon possesses distinct behavior patterns that should be construed as integral components of the means employed in the conduct of military operations rather than constituting distinct methods. For instance, the behavior of a mine's detonation mechanism represents a distinguishing feature of that particular means for conducting military operations. The software underpinning AWS, despite its considerably enhanced complexity, fundamentally adheres to this principle. Consequently, it is paramount to recognize that the concept of a method of warfare encompasses the manner in which weaponry is deployed (i.e., the interactions between a human operator and the weapon), rather than the intrinsic behavior of the weapon itself. Therefore, even if the pre-programmed behavior embedded within an AWS emulates actions that a human could undertake, it remains an integral aspect of the means employed in the conduct of military operations. However, it is essential to note that the aforementioned does not imply that the utilization of all categories of AWS inherently constitutes a means for conducting military operations.

D. Challenges in International Regulation of AWS

The participating states of the CCW are currently engaged in a series of discussions concerning issues related to AWS. One possible outcome of this process could be the establishment of a specialized international legal framework for AWS. However, at present, no concrete steps have been taken in this direction. Due to the absence of specialized international legal regulations for AWS, their design and use are governed by general conventional and customary norms of IHL that regulate weapons not subject to separate control.

Nevertheless, the legal implications of introducing specific legal regulations for AWS, which may contain restrictive or prohibitive provisions, are of particular interest. In most cases, rules that restrict or prohibit certain types of weapons define the regulated weaponry based on its inherent nature or the effect it has on the objects against which it is used, including the equipment used in conjunction with the weapons, such as delivery mechanisms. Thus, the introduction of autonomy is unlikely to significantly impact the applicable law.

---

[22] M. SASSÒLI, 'AUTONOMOUS WEAPONS AND INTERNATIONAL HUMANITARIAN LAW: ADVANTAGES, OPEN TECHNICAL QUESTIONS AND LEGAL ISSUES TO BE CLARIFIED', 308, 323, (2014) 90 INTERNATIONAL LAW STUDIES.

In examining the approach used in The Biological Weapons Convention (BWC), participants commit to never, under any circumstances, develop, produce, stockpile, acquire, or otherwise retain microbiological or other biological agents or toxins or weapons, equipment, or means of delivery intended for the use of such agents or toxins.[23] These prohibitions are not related to any properties that could be altered through the combination of biological weapons with autonomous control. Therefore, at first glance, the utilization of autonomy as a means of control over biological weapons would similarly be prohibited by the BWC. Comparable instances can be found by referencing the Chemical Weapons Convention (CWC) and the Ottawa Convention. Thus, the applicable law concerning AWS at present and in the near future is confined to general conventional norms, as well as customary IHL relating to weaponry and its use.

E. Examining Autonomous Weapons Systems AWS Through the Lens of IHL Principles

In the absence of any international legal regulation for AWS, this study has chosen to examine the application of AWS through the lens of IHL principles. The optimal approach is to peel back the layers of laws and assess the compliance of AWS with the fundamental principles of IHL, as these principles play a foundational role and can serve as a guide in addressing complex issues. It should be noted that the principles of IHL are not fixed in a single specific source, so the first step is to identify them. The International Court lists four core principles of IHL: the principle of distinction,[24] the principle of military necessity,[25] the principle of avoiding unnecessary suffering,[26] and the principle of proportionality.[27] However, scholars propose various sets of principles. For instance, C. Droege identifies among the fundamental principles of IHL: the principle of distinction, the principle of proportionality, and the principle of taking precautionary measures.[28]

1. Principle of Distinction

The principle of distinction consists of two components: parties in an armed conflict must distinguish between civilians and combatants, as well as between civilian and military objects. The target of an attack may only be military objects and combatants.[29] This principle is codified in Article 48 of Additional Protocol I (API) For states that have not ratified API, this principle is a customary rule of IHL. Additionally, the International Court has ruled that this norm is a cardinal principle of IHL.[30]

At first glance, the principle seems to be a relatively straightforward black-and-white rule: a potential target is either military or it is not. However, difficulties arise because whether a target is qualified as civilian or military can depend on the context. Indeed, the analysis

[23] Biological Weapons Convention, opened for signature 10 April 1972, 1015 UNTS 163 (entered into force 26 March 1975) art 1.
[24] Nuclear Weapons Advisory Opinion, 1996, ICJ Rep 226 paras 78, 92 and 95.
[25] ibd, paras 30, 32, 43 and 48.
[26] ibd, paras 77, 78 and 95.
[27] ibd, paras 41, 44 and 46.
[28] C. Droege, *'Get Off My Cloud: Cyber Warfare, International Humanitarian Law, and the Protection of Civilians'*, 533-553 (2012) 94(886) INTERNATIONAL REVIEW OF THE RED CROSS.
[29] Protocol I Additional to the Geneva Conventions, Arts 48 and 51, especially 51(4). ICRC Study on Customary International Humanitarian Law, Volume II, Chapter 24, Section B.
[30] Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion), 76, ICJ Reports (1996).

typically required to adhere to the principle of distinction is highly complex and highly contextual. Granting weaponry, the ability to choose its own targets means that part of the weapon targeting process is removed from the hands of a human operator and encoded in the control system of the AWS. It should be noted that this principle implies a simple dichotomy: a weapon is either capable of being directed or not directed against a specific military target in the circumstances of a particular attack. Thus, the legal assessment of AWS requires a broader understanding of "precision," which includes the elements of the system's ability to select lawful targets in addition to factors that typically characterize the "precision" of a weapon. This approach may seem illogical, especially to proponents of an anthropomorphic approach (see above). Nonetheless, this approach aligns with the paradigm chosen in this study, where autonomy is considered a form of control. The principle requires that the combination of the operator's actions and the behavior of the weapon system result in the identification of a lawful target and the realization of conditions under which it can be legally attacked. The weapon system must then behave in a way that ensures the attack is directed at the chosen target, which can be legally attacked. When a significant portion of the target selection process is encoded in the weapon control system, the distribution of tasks between the operator and the weapon system changes.

Consider the measure of "precision" in the specific case of an anti-missile defense system. The weapon operator applies a "time and place" restriction to the set of potential targets that can be set. Once this external limit is established, control shifts to the weapon system itself, which then aims and fires. Thus, the overall precision of the weapon system has two components: the precision with which it identifies that a potential target falls within the established set of targets it is programmed for and the precision with which it can actually engage that target. The same logic can be applied to more AWS. The fact that an AWS has greater operational freedom, such as increased range or adaptability, does not preclude its actions from remaining limited and tied to the point of activation; the system will only select targets within its operational range. The commitment of persons conducting an attack to target only military objectives remains unchanged. However, the share of the AWS in the fulfillment of this commitment increases. As long as it is possible for the operator to restrict the set of targets available to the AWS in such a way that there is sufficient certainty that only lawful targets will be engaged, the AWS will satisfy the threshold requirement of Article 51(4)(b) of the API. Other questions that require answers include: How effectively can the targeting system identify lawful targets under attack conditions? How well can the targeting system distinguish combatants from civilians? How accurately can the chosen target be engaged? Given the achievement of a sufficient level of accuracy in target identification and the operation of the AWS, it will not have uncontrolled effects, and thus legal requirements will be met. The perspective that certain obligations related to the "law of targeting," such as the requirement to take all feasible precautions to verify that targets are lawful, might be absorbed by the functions of the AWS may intuitively appear attractive. However, this concept is flawed because it implies that the software should do everything that is practically possible in the circumstances of the attack.[31] These assertions are components of an anthropomorphic concept and are fundamentally incorrect, as legal obligations cannot automatically transfer from humans to weapons. Obligations related to precautionary measures remain exclusively with those who plan or make decisions about an attack. It is they who must do everything possible to ensure

[31] C. PILLOUD, J. PICTET, 'PROTOCOL I - ARTICLE 57 - PRECAUTIONS IN ATTACK' IN YVES SANDOZ, (C. Swinarski, B. Zimmermann eds), Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949 (Martinus Nijhoff, 1987) n 43 in Chapter 2.

that the targeting functions of AWS operate appropriately. Any targeting functions encoded in AWS must conform to the requirements of the principle of distinction.

Opponents of AWS often cite (hypothetical) concerns related to attacks on protected persons, describing such incidents as indiscriminate attacks.[32] For example, it is difficult to distinguish a farmer tilling the soil from a member of an armed group laying improvised explosive devices. Disagreements persist regarding the precise circumstances under which deadly force can lawfully be used against civilians who are in some way connected to an armed conflict. Indeed, such a problem related to the principle of distinction exists. The opposing side, in turn, may use civilian objects enjoying special protection under IHL, such as hospitals or mosques and churches, for cover. But does the firing of an AWS, for instance, at a civilian necessarily equate to a violation of the principle of distinction? There are several factors that could lead AWS to cause harm to individuals or objects enjoying protection under IHL. It should be noted that some of these factors are common to other weapon systems, while others are specific to the characteristics of AWS. The legal character of harm inflicted on persons or objects enjoying protection under IHL depends on the grounds. Possible grounds can be classified by examining the components of an attack by an AWS. Firstly, there is the human operator who activates the weapon system. Secondly, the guidance system, which is part of the control system of the AWS. Thirdly, the weapon itself. Incidents related to the actions of the operator or the behavior of the weapon have the same legal character as the use of "manual" weapons as well as AWS. Therefore, it is not necessary to consider situations in which the operator intentionally conducts an attack on an unlawful target, which would violate Article 51(4)(a) of the API, or activates the weapon system in circumstances for which it is not intended, or the AWS simply misses and fails to hit the target due to limited accuracy, which could

One of the primary concerns related to AWS is the possibility of system malfunctions. An AWS, experiencing accidental software failure, may attack unlawful targets. It is not evident that such a failure would have any legal consequences beyond those associated with an equivalent software or hardware malfunction in "manual" weapon systems. Another issue is the potential for the "intentional" opening of fire by an AWS on an unlawful target. In this case, the AWS conducts an attack on an unlawful target, but unlike the previous example, in this case, the system's targeting code executes without errors, albeit not in line with the operator's intent. Presumably, there are two possible reasons for such an outcome: intentional and unintentional misconfiguration. In the first case, the AWS may be maliciously programmed to target civilians. The legal character of such a scenario does not differ from a similar act committed using any weapon relying on a targeting system. Depending on the responsible party, these actions may be classified as acts of sabotage or indiscriminate attacks under Article 51(4)(a) of the API. In the case of unintentional misconfiguration, the developer incorrectly configured the AWS in a way that its targeting system identifies a civilian as a legitimate target. Such errors are not inherent to AWS but are human errors.[33] Failure can be attributed to an individual, the programmed weapon system, or individuals who failed to detect an error during the weapon's testing process. The executable code for determining a lawful target ultimately reflects the decision-making process of a human. Any failure leading to an AWS opening fire on unlawful targets is equivalent to a failure that can occur with any other type of weaponry, so AWS are not an exception. At a technical level, there are no failure modes that are unique to AWS or cannot be regulated by the law applicable to AWS. Malevolent actions by a weapon

[32] ibd, 684.
[33] R. McLaughlin, *'Unmanned Naval Vehicles at Sea: USVs, UUVs, and the Adequacy of the Law'*, 105, (2011) 21(2) JOURNAL OF LAW, INFORMATION AND SCIENCE.

developer, in turn, do not alter their legal character simply because they are carried out using an AWS rather than another type of weaponry.

The second category includes inherently unlawful weapons, which are those types of weaponry whose consequences cannot be controlled well enough to limit their use solely for military purposes. API describes such weapons as those that employ methods of warfare whose effects cannot be confined within the requirements of the Protocol. While the first category of unlawful weapons is prohibited due to the inability to precisely direct their effects towards legitimate targets, this type of weaponry is prohibited because it is impossible to subsequently restrict the impact of these weapons to lawful targets. The reservation 'as required by the Protocol' deserves special attention: this wording indicates that the mentioned effects do not necessarily result from the direct impact of the weapon but can also include any effects that raise concerns in accordance with other articles of Additional Protocol I. These include Articles 35(3) and 55, which prohibit 'widespread, long-term, and severe damage to the natural environment,' suggesting that it pertains not only to the direct effects of the weapon.

It is worth noting that the prohibition of weapons whose effects cannot be confined is only loosely related to autonomy and is not a subject of discussion in the context of AWS as a class of weaponry. This prohibition is more closely associated with the physical means by which the weapon component of the system inflicts harm, such as the type of ammunition.[34] As such, AWS must possess advanced skills in observation and recognition, as well as the developed ability to make judgments because people make distinctions, taking into account various factors, such as raising hands in the air or signs of an enemy becoming incapacitated.

2. Direct Participation in Hostilities in the Context of AWS

As demonstrated by the analysis presented above, the ability of AWS to adhere to the principle of distinction is highly questionable. "Attacks shall be limited strictly to military objectives. In so far as objects are concerned, military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage.[35]"

With regard to individual persons, context can be important for the following reasons. In the context of modern warfare (as mentioned above), an individual may transition from being a combatant, who is a legitimate target, to a person protected under IHL, and vice versa. For example, an individual may be dressed in camouflage attire with various distinctive markings, carrying a rifle in the midst of combat. Based solely on observation and recognition capabilities, such an individual might be identified as a participant in hostilities. However, upon further analysis of the entire situation, taking the context into account, it may become evident that this person is not a legitimate target, for instance, having been incapacitated due to injury or illness.[36] Indeed, not only clothing and distinctive markings but also contextual factors are of significant importance in determining an individual's status under IHL.

---

[34] I. HENDERSON, 'THE CONTEMPORARY LAW OF TARGETING (2009), 237, ch 1, para 14, C. PILLOUD, J. PICTET, 'PROTOCOL I – ARTICLE 57 – PRECAUTIONS IN ATTACK' IN YVES SANDOZ, C. SWINARSKI, B. ZIMMERMANN (EDS), COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, 686, (1987) n 43 in ch 2.

[35] Additional Protocol I to the Geneva Conventions, Art 52.

[36] Additional Protocol I to the Geneva Conventions, Art 41(2).

Measures to protect individuals who are not or are no longer participating in hostilities rely on combatants' capacity to make reasoned judgments about contextual factors. For example, recognizing the raising of hands as a sign of surrender or identifying an enemy in a state of unconsciousness or disorientation, indicating their incapacity to continue fighting. Conversely, a situation can also arise where an individual transitions from being a person protected by IHL to a combatant, simultaneously becoming a legitimate target. IHL generally defines civilians negatively, meaning anyone who is not a combatant is considered a civilian.[37] Furthermore, IHL provides an additional safeguard for the protection of civilians: in cases of doubt, an individual is considered a civilian.[38] Under certain circumstances, civilians may lose the protection afforded to them by international humanitarian law, including the Geneva Conventions. For example, in cases of "levee en masse," the status of a prisoner of war, and hence a participant in the conflict, is extended to residents of unoccupied territory who voluntarily take up arms to resist invading forces without having had time to form regular armed forces. Today, the notion of "direct participation in hostilities" (DPH) is actively used.

It would be overly optimistic to think that there is a rule, however complex, that can definitively categorize every individual in terms of IHL. Let's demonstrate this with the example of determining a "civilian taking direct part in hostilities." The ICRC has developed a carefully thought-out guide defining what constitutes an act of direct participation in hostilities, based on which civilians do not receive the protection typically afforded to civilians under IHL. For a civilian to become a lawful target, the following criteria must be met: 1) the threshold of harm; 2) a causal link between the act and the harm; 3) affiliation with a party to the conflict.

To meet the threshold of harm necessary for an act to qualify as direct participation in hostilities, it must, with a high degree of probability, have a negative impact on military operations or the military capabilities of the opposing party in an armed conflict. In the absence of military harm, the harm threshold can also be reached when the act may lead to the death, destruction, or injury of individuals or objects protected against direct attack. In both cases, actions that meet the required harm threshold can be considered as direct participation in hostilities only if they additionally satisfy the criteria of a causal link between the act and the harm, as well as an affiliation with a party to the conflict (as mentioned above). For the causal link requirement to be met, it must be possible to establish that either the specific military operation or the specific action, as part of which the harm occurred, is directly causally linked to the harm inflicted, which reaches the required threshold. However, actions meeting the causality and harm thresholds can be considered as direct participation in hostilities only if they also satisfy the criterion of affiliation with a party to the conflict. To meet this requirement, the action must be specifically designed to cause the required harm in support of one of the parties to the armed conflict to the detriment of the other party. Note that certain actions, while causing harm that reaches the required threshold, may lack this specific affiliation. For example, harm caused in individual self-defense, protecting others from violence prohibited under IHL, exercising authority over individuals and territory, and addressing civil disturbances and internal violence may lack this specific affiliation.

When all these criteria are applied in combination, it allows us to distinguish between actions that are considered equivalent to direct participation in hostilities and actions that are not part of the behavior of armed forces, even though they occur within the context of armed

[37] Additional Protocol I to the Geneva Conventions, Art 50(1).
[38] Third Geneva Convention, Art 4A(6).

conflict. However, even in cases where a specific action is the equivalent of direct participation in hostilities, the type and degree of force used in response must comply with the norms and principles of IHL.[39] These guiding principles are an attempt to provide the means to determine whether an individual in specific circumstances meets each of these requirements. It requires a thorough understanding of the complex situation, including the strategic consequences of potential harm, the status of individuals facing potential harm, sociocultural and psychological indicators in which the intentions and actions of this individual are classified as military rather than, for instance, personal self-defense (see above). In other words, AWS must be able to perceive all the necessary information while taking into account the contextual element. For example, distinguishing a civilian holding a large piece of metal from an armed combatant with a rifle in civilian clothing.

The International Criminal Tribunal for the former Yugoslavia (ICTY), in its decision,[40] provided a definition of what behavior can be qualified as direct participation in hostilities. While this list raises more questions than it provides answers, it is worthy of attention. For example, examples of direct participation include not only carrying or using weapons during involvement in armed conflict or operations but also engaging in attacks on the personnel, property, or equipment of the opposing armed forces, transmitting military information for immediate use, transporting weapons in close proximity to combat operations, conducting reconnaissance, and observing on behalf of armed forces.[41]

These examples illustrate the complexity and nuances involved in determining what constitutes direct participation in hostilities. Such determination often depends on the specific context and factual circumstances of each case. AWS, if involved in making such determinations, must be capable of assessing these complex situations, considering both actions and intent, and evaluating them within the broader framework of IHL. The interpretation of the rules established by the ICTY also requires a higher-level judgment. For example, judgment will be needed to determine whether the military information was transmitted for "immediate use" (qualifying as DPH) or merely for use in the more distant future (which does not qualify as DPH). Similarly, judgment will be required to decide whether weapons were transported "in the immediate vicinity of hostilities" or further away from the theater of operations, which also affects the classification of actions: actions will be classified as direct participation in hostilities only in the first case.

In order to comply with the principle of distinction using AWS, it is necessary for these systems to possess cognitive abilities inherent to humans, which allow them to assess and analyze contextual factors, as mentioned earlier. AI capable of such abilities currently exists only in theory and is referred to as General AI. At the moment, only narrow AI is available, which is also used in AWS. Narrow AI is capable of automatically performing very specific, point-specific functions or tasks at a human level or even exceeding it in some areas.

It is essential not to equate narrow AI with general AI since the process of task-solving is fundamentally different. General AI is capable of not only self-learning but also making its conclusions based on received information and environmental conditions, considering context, and even having a form of self-awareness and self-improvement. General AI can solve complex

[39] INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION IN HOSTILITIES UNDER INTERNATIONAL HUMANITARIAN LAW, ICRC, (Geneva, 2009).
[40] Prosecutor v Pavle Strugar (Appeal Judgment), International Criminal Tribunal for the Former Yugoslavia, Case No ICTY-01-42, 17 July 2008.
[41] ibid, para 77.

composite tasks, including recognizing human emotions. This can be useful in field conditions, for instance, to determine whether a person has been incapacitated or whether they are surrendering. To make a definitive judgment, general AI incorporated into autonomous weapons systems would need to process data on facial expressions, posture, spoken words, and so on.

Despite the fact that General AI is yet to become a reality, it should be noted that this field is advancing rapidly. However, there are significant discrepancies in the scientific community regarding predictions and timelines for the realization of general AI. According to experts in artificial intelligence, it's unlikely that General AI will emerge in the near future[42]. This raises the question of whether AWS can provide the necessary level of judgment for the principle of distinction at the present moment, especially in the context of the concept of direct participation in hostilities. The absence of AGI does not necessarily preclude this capability. While the development of full AGI may take a while, the artificial intelligence currently in use (Narrow AI) in AWS may still be capable of making distinctions. However, this capability is limited, and such systems should ideally be employed in specific settings, as mentioned earlier, such as exclusively for defensive operations and in areas where the presence of civilians is highly unlikely.

3. The Principle of Proportionality

Despite the fact that the principles of IHL are primarily oriented towards the protection of civilian populations, it is essential to maintain a realistic perspective on armed conflicts. In modern conditions, there are no effective means to completely eliminate the possibility of civilian casualties and injuries. This principle is established both in Article 51 of AP I and in customary IHL.

The Principle of Proportionality entails that in cases where harm is inflicted on civilian individuals, such harm should be commensurate with the military advantage gained. In essence, an attack becomes unlawful when the incidental harm to "civilian" persons is excessive as a result of the attack. IHL dictates that the use of force and the means employed must always be proportionate to the military advantage sought.[43] In the context of military advantage, it should be specific and direct, rather than abstract. Specific and direct nature of the advantage indicates that it should be significant and relatively immediate, and military advantages that may arise in the long term should be ignored.[44]

In the context of proportionality, it requires a contextual balancing of two factors: the potential harm to civilians and civilian objects on one hand, and the expected military advantage on the other. The ICTY established criminal liability for disproportionate attacks, depending on whether a person was well-informed, used available information reasonably, and expected excessive harm to civilians and objects as a result of the attack.[45] It is essential to note that the concept that there is no formula for balancing these factors is not endorsed in this study. Proportionality is not a vague notion but rather a clear directive, setting a "fixed standard" on the constraints of what commanders and soldiers can do, eliminating undesirable freedom of

---

[42] V. MULLER, N. BOSTROM, 'FUTURE PROGRESS IN ARTIFICIAL INTELLIGENCE: A SURVEY OF EXPERT OPINION' IN VINCENT (C. Muller ed), FUNDAMENTAL ISSUES OF ARTIFICIAL INTELLIGENCE (Springer, 2016).
[43] Article 51(5)(b) Additional Protocol I to the Geneva Conventions. ICRC Study on Customary International Humanitarian Law, Rule 14.
[44] J. HENCKAERTS, ICRC STUDY ON CUSTOMARY INTERNATIONAL HUMANITARIAN LAW, 49, (1996).
[45] Prosecutor v Galic (Judgment), (International Criminal Tribunal for the Former Yugoslavia, Case No ICTY-98-29, 5 December 2003) Para 808.

action. Since assessing proportionality involves weighing competing interests, AWS must be able to anticipate the consequences of all potential decisions and the potential number of civilian casualties. They must also be responsive to changing circumstances and, subsequently, be able to calculate the military advantage and determine whether the incidental harm is acceptable.[46]

Here we will address the issue of the "frame": in order to calculate the potential collateral damage resulting from an attack by an AWS, it is necessary either to calculate the consequences of every possible action, which would take an infinite amount of time, or to make assumptions, which could potentially lead to a disproportionate attack and, consequently, a violation of IHL. Determining collateral damage is associated with assumptions because, in the conditions of armed conflict, exact certainty is rarely achievable. In cases where AWS are used in open civilian areas, due care must be taken to ensure that the information on which the assumptions necessary to determine collateral damage are based is sufficient and reliable.

Despite the prevailing scientific position[47] that the implementation of the principle of proportionality requires more than just balancing quantitative data and that "a robot cannot be programmed to replicate the psychological processes in human judgment necessary for assessing proportionality," this study does not share that view. The principle of proportionality demands a quantitative calculation, albeit a rather complex one. In this study, proportionality is examined through the lens of utilitarianism. Utilitarianism allows for the translation of ethical aspects into a specific observable dimension by replacing normative categories with observable outcomes and recognizing the moral significance of actions' consequences based on the criteria of increasing so-called "happiness." Just as the principle of utilitarianism is satisfied when an action leads to more "happiness" than "unhappiness," proportionality is satisfied when an attack results in greater military advantage than collateral damage. Utilitarianism allows for the translation of ethical aspects into a specific observable context, achieved by substituting the categories of the morally required with observable objectives. It also recognizes the moral significance of the consequences of actions, based on criteria aimed at increasing what is commonly referred to as "happiness." Therefore, just as the principle of utilitarianism is adhered to when an action leads to greater "happiness" than "unhappiness," proportionality is satisfied when an attack results in greater military advantage than accompanying harm.[48]

While maintaining the principle of proportionality, AWS must also have the capability to weigh military advantage against collateral damage. A proposed starting point for this evaluation is the methodology for assessing collateral damage estimates (CDEM[49]). A five-stage analytical system is used for assessing collateral damage, based on factors such as the area of effect of different weapon types, demographic data in the anticipated strike area, the timing of the attack and its potential impact on the likely level of civilian casualties.[50]

---

[46] M. WAGNER, 'TAKING HUMANS OUT OF THE LOOP', 159-162, (2011).
[47] J. PETMAN, 'AWS AND IHL: "OUT OF THE LOOP?"', 39, (2017).
[48] E. Winter, *'Autonomous Weapons in Humanitarian Law: Understanding the Technology, Its Compliance with the Principle of Proportionality and the Role of Utilitarianism', 183,194,* (2018) 6(1) GRONINGEN JOURNAL OF INTERNATIONAL LAW.
[49] CDEM - Collateral Damage Estimation Methodology, is a methodology developed by the United States. It is used to identify any "collateral concerns" (i.e., civilians, civilian objects, or other protected entities) within the radius of action of a weapon under consideration for use in an attack.
[50] E. Winter, *'The Compatibility of the Use of Autonomous Weapons with the Principle of Precaution in the Law of Armed Conflict'*, 240, 262, (2020) 58(2) MILITARY LAW AND THE LAW OF WAR REVIEW.

Through this methodology, utilitarianism, as originally defined by Bentham, can be applied to real values that can be used in proportionality calculations. In this approach, proportionality is assessed through a utilitarian framework, which translates ethical aspects into observable terms, focusing on the consequences of actions, based on criteria that increase "happiness." Just as utilitarianism is adhered to when an action leads to greater "happiness" than "unhappiness," proportionality is met when an attack leads to a greater military advantage than collateral damage.[51]

The key challenge here is to weigh factors that may seem incomparable. The study suggests quantifying collateral damage in terms of lives lost or injuries sustained (possibly utilizing CDEM, as mentioned earlier) and then calculating military advantage in terms of lives saved or injuries prevented (possibly using a CDEM equivalent). By comparing these two values, a reliable proportionality assessment can be made. This approach simplifies the task by making the values for autonomous systems comparable.[52]

However, this is a complex process that requires a high level of AI, which might not be readily available at the time of this study. Nevertheless, it is suggested that with time, AWS will become capable of adhering to the principle of proportionality.[53]

Furthermore, the compliance of AWS with the principle of proportionality depends on the degree of firepower they control and how much firepower is used simultaneously. For example, if an AWS uses nuclear weapons, the collateral damage could be extensive. However, if it employs a high-precision micro-projectile or a laser beam, the potential collateral damage in case of a miss would be relatively low. Therefore, it is recommended to keep the level of firepower controlled by AWS relatively low to avoid causing disproportionate damage in case of malfunctions or software errors. Additionally, since machines lack the right to self-defense, they should not have the same authority to use force and firepower as a human in the field.

4. The Precautionary Principle

This principle was first enshrined in Article 2(3) of the Hague Convention (IX) concerning Bombardment by Naval Forces in Time of War of 1907, and state practice establishes this norm as customary international law.[54] The principle of taking precautionary measures requires the adoption of all possible measures to avoid or minimize incidental harm to the civilian population. Firstly, there is an obligation to provide timely warning to the civilian population of an attack on military objects if such an attack may cause them harm, except when this is impossible due to tactical circumstances. Secondly, the principle of precaution also extends to the choice of means and methods of warfare. For example, if military advantage necessitates an attack on a military object located near a soccer stadium, the commander, in adhering to precautionary measures, must time the strike so that there is no civilian population on the soccer stadium at that moment.

It is worth noting that the requirement for taking precautionary measures applies throughout the entire process of armed deployment planning and concerns all individuals

---

[51] US Department of Defense, 'No-Strike and the Collateral Damage Estimation Methodology', (2012) CJCSI 3160.01A D-A-7.
[52] M. Newton and L. May, Proportionality in International Law, 285, (OUP 2014).
[53] M. Guetlein, 'Lethal Autonomous Weapons: Ethical and Doctrinal Implications', 126, (2005) Research Report Maxwell, AFB: Air War College, https://apps.dtic.mil/sti/pdfs/ADA464896.pdf (Last visited: 23 October 2023).
[54] ICRC Study on Customary International Humanitarian Law, 44, (1996).

involved in preparation. This includes not only commanders but potentially also weapon system manufacturers and programmers of AWS.[55] Consequently, considering the possibility of various unforeseen situations and challenges arising during the deployment of AWS and the execution of combat tasks, there is a basis for the obligation to always have a human "in the loop." In other words, an operator who will be responsible for monitoring and responding to various new situations as they develop.[56]

Adhering to caution implies measures that are highly context-dependent and subject to rapid and unpredictable changes. They involve continuous target assessment, weapon selection, timing, and method of attack. An attack must be halted if it becomes evident that it will have disproportionate consequences or if the target is no longer (or no longer remains) lawful. However, it is essential to always remember the need for ongoing reassessment, which raises the question: can AWS perform the required assessment without human intervention? The necessity for constant reassessment of circumstances, in conjunction with the existing capabilities of AI in AWS, suggests that compliance with this principle by autonomous systems may not be feasible in the near future, at least for several years.

Therefore, it appears prudent to deploy autonomous functions primarily against military targets, such as military aircraft, ships, or in situations where the risk to civilian populations is virtually non-existent. While the deployment of defensive weaponry with autonomous capabilities against enemy projectiles or missiles does not pose significant compliance issues with respect to IHL, an intriguing question arises regarding whether the deployment of fully AWS will be restricted to situations where the encounter with the civilian population is ruled out from the outset. This is because, given the fact that conflicts are increasingly becoming non-international, with no clearly defined geographical frontlines, military objectives primarily located in civilian areas, and combatants who intentionally do not clearly distinguish themselves from non-combatants, the ability of AWS to exercise caution is currently not feasible.

5. The Principle of Unnecessary Suffering

This principle is enshrined in Article 35(2) of API and essentially reiterates the customary principle of IHL, which prohibits the use of weapons, projectiles, materials, and methods of warfare that are capable of causing excessive damage or unnecessary suffering. The primary purpose of this principle is to ensure that the injuries and suffering inflicted on combatants do not exceed the necessary level required to render them ineffective. Therefore, it is forbidden to use any means or methods that cause excessive damage or inflict unnecessary suffering unless such effects are required to render the enemy combatants ineffective in any case.[57]

---

[55] W. Boothby, *'Conflict Law: The Influence of New Weapons Technology'*, 115, (2014), T.M.C. Asser Press: Den Haag.

[56] M. Trascasas, N. Weizmann, AWS under International Law, Academy Briefing No. 8, N. Weizmann, Geneva Academy of International Humanitarian Law and Human Rights, 4, (November 2014).

[57] C. Pilloud, J. Pictet, 'Protocol I – Article 57 – Precautions in Attack' in Yves Sandoz', C. Swinarski and B. Zimmermann, Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, 686, (1987).

F. Case study: Current Challenges in the Deployment of AWS in Armed Conflicts (Autonomy in Existing Weapon Systems)

1. United Nations Expert Group on Libya: Existing Lethal Autonomous Weapon Systems - A Line Crossed?

To raise public awareness of the issue, in 2017, the Future of Life Institute released a viral video called "Slaughter Bots." [58] At that time, many experts considered it overly fantastical, believing that the emergence of such systems should not be expected before the mid-21st century. However, a few years later, the potential use of such systems is already being discussed within the UN. In March 2021, a UN expert group on Libya, where a civil war has been ongoing for several years, reported the possible use of lethal AWS, such as STM Kargu-2. The official report[59], consisting of over 500 pages, provides a detailed account of events from October 2019 to January 2021, with information regarding lethal drones presented within the context of the March 2020 battles. The report indicates that on March 27, 2020, the Prime Minister of Libya, Fayez al-Sarraj, issued an order for the "Peace Storm" operation, during which drones were used against forces associated with Haftar. According to the report, Libyan authorities employed Turkish-made STM Kargu-2 drones to strike a column of the Libyan National Army forces retreating from Tripoli. Subsequently, logistical columns and retreating armed forces were tracked and remotely engaged using combat drones or lethal AWS, such as STM Kargu-2 and other loitering munitions. The augmented capacity for operational reconnaissance included Turkish electronic warfare assets and reconnaissance, surveillance, and reconnaissance capabilities provided by Bayraktar TB-2 drones and, likely, TAI Anka-S. This, in turn, enabled the deployment of an attrition warfare strategy.[60] But this was unlike previous drone strikes. According to the incident description, the STM Kargu-2 drone could track and engage targets remotely. According to the report, “the lethal autonomous weapons systems were programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true “fire, forget and find” capability.”[61]

The report does not provide information on whether there were casualties or fatalities associated with the attack. However, it does note that the drones "very effectively" assisted in causing "significant losses" to enemy anti-aircraft missile systems, which generated a flurry of sensational headlines in the media about the first-ever fully autonomous drone attack on humans in history.[62] Furthermore, the report lacks technical details, and even with the presence of such details, it is presumable that the situation would not necessarily be clarified. Despite the report containing numerous pieces of information about violations of IHL and international human rights law (IHRL), it does not mention the unlawful use of Kargu-2. Therefore, the following legal aspects that could be relevant to the deployment of STM Kargu-2 will be considered. However, for an assessment of the legality of using STM Kargu-2 in the specific context of the armed conflict in Libya, insufficient information is available.

The analysis of the report's content has revealed that UN experts do not differentiate between lethal AWS and kamikaze drones (loitering munitions). The latter, while waiting for

---

[58] "Slaughter Bots" (DUST), https://www.youtube.com/watch?v=O-2tpwW0kmU (last visited: 9 September 2023).
[59] LETTER DATED 8 MARCH 2021 FROM THE PANEL OF EXPERTS ON LIBYA ESTABLISHED PURSUANT TO RESOLUTION 1973 (2011) ADDRESSED TO THE PRESIDENT OF THE UN SECURITY COUNCIL (S/2021/229).
[60] ibid, para 60.
[61] ibid, para 63.
[62] *'Turkish Drones STM Kargu-2 Autonomously Attacked Humans*', ROBOCRAFT, (31 May 2021), https://robocraft.ru/news/4177 (last visited: 31 September 2023).

a specific target and signal to strike, patrol a particular area. For example, the Israeli IAI Harpy, developed as far back as 1980, has predecessors that are still produced, including in Azerbaijan under the name "Zarba," and they were actively used during the Nagorno-Karabakh armed conflict.[63] Yes, the deployment of loitering munitions in Libya, in this case, does not introduce any elements of novelty. Moreover, the manufacturers of these loitering munitions themselves classify them as "all-weather autonomous weapons."[64] However, the wordings featured in the UN report still indicate that there is something new at play.

Kargu-2 is a UAV created by the Turkish company STM. It utilizes machine learning algorithms integrated into the platform, enabling it to operate autonomously and be manually controlled. Unlike the Bayraktar TB2 or the Israeli loitering munition Harpy (as mentioned above), Kargu-2 is a weapon (tactical quadcopter) capable of selecting human targets and engaging them based on object classification through machine learning.[65] Although various ammunition options are available, the Kargu drone detonates the explosive charge near the target, minimizing the likelihood of collateral damage. It's worth noting that despite not having particularly impressive flight characteristics, Kargu is capable of orienting itself using visual data and reference points in the landscape, which makes it less susceptible to interference with GPS signal jamming systems.[66] The manufacturer, STM, advertises the anti-personnel capabilities of the Kargu drone in a grim video, showcasing a model of the Kargu diving towards a target amidst a group of mannequins representing people. According to the manufacturer's claims, the systems employ various facial recognition tools to identify and track potential targets, and in case they cannot fulfill their mission, the systems return to base. Of particular note in this context is the existence of separate loitering munitions (drones) in STM's product line, such as Alpagu, which presumably indicates Kargu-2's capability for fully autonomous tracking and targeting. STM's CEO, Murat Ikinci, emphasized that a squadron of 30 Kargu drones is powerful enough to destroy a military unit and a warship. Each Kargu has a specific mission. If one of the drones in the team is attacked or disabled during an operation, others replace it and carry out the assigned task, confirming the presence of AI and facial recognition systems in the drones. However, as of now, there is no confirmed case of AWS attacking humans, including by STM and Turkish authorities.

Returning to the events of 2020 in Libya, it is highly likely that drones possessed some capability to identify moving objects in videos, potentially including the ability to distinguish between people and other objects such as cars and buildings. However, they lacked certain other functions typically associated with full autonomy, such as the ability to prioritize targets, dynamically execute complex tactics, or make decisions in accordance with the laws of armed conflict.[67] In such scenarios, image processing may prove insufficient for detecting and identifying individuals who are unconscious or suffering from internal injuries or illnesses. On the other hand, leaving one's weapon may be programmed into an autonomous weapons system as a sign of surrender. However, there are situations where individuals may be unable to

---

[63] J. Antal, *'7 Seconds to Die: A Military Analysis of the Second Nagorno-Karabakh War and the Future of Warfighting'* (2022), https://www.airuniversity.af.edu/Aether/Book-Reviews/Article-Display/Article/3218364/7-seconds-to-die-a-mi litary-analysis-of-the-second-nagorno-karabakh-war-and-the/ (Last visited: 17 September 2023).

[64] Autonomous weapon for all Weather, https://www.iai.co.il/p/harpy (Last visited: 31 September 2023).

[65] Kargu-2 Specifications, https://www.stm.com.tr/en/kargu-autonomous-tactical-multi-rotor-attack-uav (Last visited: 31 September 2023).

[66] R. Fishman, *'Everything You Need to Know About Killer Drones from Turkey*,' TECHINSODER, https://www.techinsider.ru/weapon/712883-vse-chto-nado-znat-o-dronah-ubiycah-iz-turcii/ (Last visited: 1 September 2023).

[67] Footnote 60, para 63.

relinquish their weapons, potentially leading to misinterpretation by the machine. The use of autonomous weapons systems like Kargu-2 without the capacity to detect and provide assistance to incapacitated individuals raises questions regarding these systems' compliance with IHL. Indeed, employing such a system while aware of its inability to meet these legal requirements may be akin to an order not to spare anyone, i.e., to show no mercy or leniency and take a life in exchange for surrender, which is prohibited under customary IHL.[68]

2. GR-A1 Autonomous Security Robot: Border Safeguard or Severe IHL Violation on the Korean Peninsula

As previously noted, an operator can trust the machine and not monitor its operation properly, practicing "blind trust in the machine". It is also important to consider that in the case of the "human on the loop" concept, an AWS is capable of functioning without operator intervention and can be controlled not constantly, but only as needed, for example, as in the case of the all-weather autonomous security robot SGR-A1, which can support troops in the demilitarized zone separating North and South Korea, guarding key military facilities. The developers of SGR-A1 report that the system was created to replace people who may suffer from bad weather or fatigue. SGR-A1 has three low-light cameras,[69] thermal detectors, motion detectors, as well as image recognition software, [70]allowing it to detect targets up to two miles during the day and one mile at night.[71] Upon detecting an intruder, SGR-A1 issues verbal warnings, followed by rubber bullets, and subsequently, live metal bullets from the onboard machine gun of the AWS. SGR-A1 can operate as a "man on the loop" system. This means that the system is capable of autonomously selecting and engaging targets, but if necessary, a human operator can intervene to deactivate the system. This is fundamentally different from the concept of "human in the loop" (see above), where the weapon system awaits commands from operators to engage a target.

Due to the unstable situation in the region and the criticism directed at the developers of SGR-A1 and the authorities of South Korea for using "killer robots" to guard their borders, it is impossible to assert with certainty the presence of fully autonomous functions in this weapon system, just as it is impossible to determine the specific number of "sentries" and unsuccessful incidents that have occurred or continue to occur in the demilitarized zone of South Korea, as this data is classified from the public. Based solely on the available information at the moment, an analysis of the application of SGR-A1 reveals that the ability to address immediate threats is indeed valuable in the demilitarized zone since South Korean defense is stretched over 250 kilometers. This means that forces must respond to any threats as quickly as possible to hold their positions until reinforcements arrive and prevent an invasion.

Therefore, the "man on the loop" system is suitable for addressing this problem. Firstly, it is difficult for operators to manage multiple systems simultaneously, even in non-combat situations. In the rapidly changing battlefield scenario, the capabilities of system operators,

---

[68] ICRC Study on Customary International Humanitarian Law, Volume II, Chapter 15, Section A. Rule 46, D. Hambling, '*Israel used world's first AI-guided combat drone swarm in Gaza attacks*' New Scientist (2021), https://www.newscientist.com/article/2282656-israel-used-worlds-first-ai-guided-combat-drone-swarm-in-gaza-attacks/ (Last visited: 4 September 2023).

[69] J. Kumagai, 'A Robotic Sentry For Korea's Demilitarized Zone' (2007) IEEE Spectrum https://spectrum.ieee.org/a-robotic-sentry-for-koreas-demilitarized-zone (Last visited 4 September 2023).

[70] S. Weinberger, *'Next Generation Military Robots Have Minds of Their Own'*, BBC, (2014) https://www.bbc.com/future/article/20120928-battle-bots-think-for-themselves (Last visited 4 September 2023).

[71] Samsung Techwin SGR-A1 Sentry Guard Robot https://www.globalsecurity.org/military/world/rok/sgr-a1.htm (Last visited: 9 September 2023).

"human in the loop," to make timely decisions will be limited. Attacks or synchronized incursions in different parts of the battlefield can overwhelm operators, depriving them of the ability to track multiple threats and allocate defensive resources appropriately. This, in turn, increases the likelihood that the defensive line may be breached. Furthermore, the "human in the loop" system relies on human input before engaging a target, creating an opportunity for the target to escape. This can be costly and cause irreparable harm to military interests. For instance, in a "counter-sniper" situation, the fraction of a second needed to obtain operator approval to open fire can hinder the neutralization of the enemy.

An interesting aspect from a compliance standpoint with the laws of armed conflict is the claimed ability of the system to recognize signs of surrender.[72] For instance, upon detecting an intruder, SGR-A1, in addition to issuing verbal warnings, can recognize behavior and movements indicative of surrender.[73] However, details about how SGR-A1 recognizes surrender signs are lacking. From the manufacturer's claims in a promotional video, which suggests that this weapon system can be used on military bases and deployed on tanks, it follows that SGR-A1 could be deployed in the context of an armed conflict. This raises numerous questions, as mere capability for recognizing surrender is insufficient to comply with the laws of armed conflict during an armed conflict. For example, how does the system determine what constitutes surrender? Presumably, programming "raising hands" or "dropping weapons" as surrender signs would not be enough to enable the system to qualify certain behaviors as surrender and, therefore, to cease the use of force and adhere to the laws of armed conflict. Can the system determine if a soldier is incapacitated (which also requires the cessation of the use of force in accordance with the laws of armed conflict) and, for example, needs assistance?

Furthermore, the issue of determining a legitimate target remains relevant in this context, considering the concept of direct participation in hostilities (as mentioned above) required to comply with the principle of distinction. As repeatedly noted, modern machine learning-based systems are unlikely to make decisions considering contextual factors. For example, an autonomous system cannot distinguish a farmer in camouflage from a soldier. Adequate classification of a vehicle is also challenging since a variety of contextual and other factors may hinder this. Research[74] has shown that systems cannot recognize partially obscured objects: due to poor visibility of the wheels and front windows of a bus, the system identified the bus as a bicycle. In a similar vein, the system can easily distinguish a tank in open terrain with good lighting from a bus, but it would struggle with the same task if key distinguishing features of the tank were obstructed, for example, by trees or buildings. Weather conditions also play a significant role: studies confirm that in foggy weather, the accuracy of the AI system used to detect obstacles on roads drops to 58% compared to 92% in clear weather, which is also typical for humans.[75] Moreover, there is a high probability of deceiving AWS on the

[72] SAMSUNG SGR-1 - SECURITY GUARD, https://www.youtube.com/watch?v=xtE9hpwrDg4 (last visited: 4 September 2023).
[73] SAMSUNG TECHWIN SGR-A1 SENTRY GUARD ROBOT, https://www.globalsecurity.org/military/world/rok/sgr-a1.htm (Last visited: 4 September 2023).
[74] A. KORTYLEWSKI, Q. LIU, H. WANG, Z. ZHANG, A. YUILLE, 'COMBINING COMPOSITIONAL MODELS AND DEEP NETWORKS FOR ROBUST OBJECT CLASSIFICATION UNDER OCCLUSION', (2020), https://arxiv.org/abs/1905.11826 (last visited: 10 October 2023).
[75] Z. LIU, Y. HE, C. WANG, AND R. SONG, 'ANALYSIS OF THE INFLUENCE OF FOGGY WEATHER ENVIRONMENT ON THE DETECTION EFFECT OF MACHINE VISION OBSTACLES' (2020) https://www.mdpi.com/1424-8220/20/2/349 (Last visited: 4 September 2023).

battlefield, for instance, by placing a simulator image of a school bus on a tank, as confirmed by a similar example, which has been verified in practice.

The stated problems necessitate the need for operator intervention in the machine's operations. Since SGR-A1 functions as a "human on the loop" system, the possibility of eliminating errors remains. This means that if SGR-A1 mistakenly targets non-hostile entities, the operator can deactivate it using the "soft" or "hard kill" option. The "soft kill" option is based on a wired or wireless communication link between the remote position and the robot; if something goes wrong, the operator sends a kill signal that stops the robot's activities. The "hard kill" option is a hardware access point on the machine itself that the operator can use to manually shut it down. In the event that a system like SGR-A1 mistakenly targets civilians, the operator can disable it. However, a problem arises here: this may not save the lives of the initial civilians attacked by the machine since the operator is unlikely to foresee the wrongful targeting, as is the case with the "human in the loop" system. Nevertheless, this would prevent mass wrongful casualties as the operator could disable the machine after the initial wrongful engagement. To avoid the tragedy of losing even a few innocent lives due to incorrect targeting by the "man on the loop" system, it is necessary to deploy the weapon system exclusively in an environment where the presence of the civilian population is minimal.

Today, the demilitarized zone is so heavily fortified that there are no civilians in it, and most North Korean defectors have to travel through China, Laos, and Thailand to bypass it.[76] This makes the demilitarized zone a controlled environment, where anyone who can physically enter SGR-A1's targeting area is reasonably considered a combatant. This is why Samsung engineers programmed SGR-A1 to identify anyone in the demilitarized zone as an enemy. Undoubtedly, such a software solution cannot be used on all borders since not all borders are equally controlled environments. For example, deploying SGR-A1 would do more harm than good on the US-Mexico border, where the vast majority of border violators do not pose a military threat. However, along borders where civilians do not move or can be restricted from moving, where a controlled environment can be reasonably established, "man on the loop" systems can be used to deter invasions.

Furthermore, to prevent errors and accidents, these systems should only be deployed in defensive operations, which also minimizes the threat to civilians. However, it is essential to prevent the use of these systems for offensive purposes, which requires the commitment of states using SGR-A1 and other AWS to establish "defensive intent." Such intent can be expressed, for example, by setting a specific firing range to prevent the possibility of striking deep into the territory of neighboring countries, ensuring the deployment of systems in areas with minimal civilian presence. In this regard, an AWS deployed in densely populated cities in the context of new wars (as mentioned above) will face a high likelihood of failing to comply with the laws of armed conflict, especially the principle of distinction.

From the examples mentioned above, it becomes clear that despite the lack of technical details, the information regarding violations of IHL by these AWS, a key question still remains. To what extent can AWS adjust and regulate their own behavior after activation to the point

[76] Debra Kamin, *'How to escape from North Korea*' (2014), THE TIMES OF ISRAEL, https://www.timesofisrael.com/what-does-it-take-to-escape-north-korea/ (Last visited: 04 October 2023).

where their behavior becomes unpredictable in terms of their ability to comply with IHL? This was also confirmed during the interview with Alex Leveringhaus.[77]

III. RESULTS

Based on the adopted approach, the term "AWS" encompasses a broad category of systems capable of independently selecting, tracking, and engaging targets without constant human intervention. This approach facilitated the categorization of these systems into three distinct types: semi-autonomous, supervised autonomous, and fully autonomous systems.

In the course of research into the legal classification of AWS within IHL, a number of challenges were encountered. While there are viewpoints characterizing these systems as artificial combatants, it was discerned that their functional aspects do not influence their legal categorization. Instead, they should be considered as instruments for executing military operations, following instructions pre-programmed by humans.

When examining compliance with IHL principles, it became evident that adherence to these principles goes beyond target selection. It necessitates considering various contextual factors, particularly for the principle of distinction, which requires evaluating contextual elements, including socio-cultural factors. This level of analysis surpasses the capabilities of AI in the foreseeable future. In addressing compliance with the principle of proportionality, we employed utilitarian principles, translating ethical considerations into measurable terms, such as collateral damage and military necessity. It was concluded that AWS may excel at making these calculations more efficiently than humans. However, the need for constant contextual reassessment poses a challenge to the principle of precautionary measures.

CONCLUSION

In conclusion, the impact of AWS on armed conflicts and their compliance with IHL remains uncertain due to limited empirical data. Nevertheless, there is an imperative for the international community to proactively address these concerns. Fully AWS, in the current state of AI development, may potentially violate IHL norms if deployed without careful consideration. The risks of these systems falling into non-state actor hands are substantial, emphasizing the necessity for comprehensive regulation. Interested states, international organizations, the UN, and global civil society should advocate for consensus on specific "rules" governing the development and deployment of AWS. These rules should encompass their goals, geographical coverage, deployment scenarios, operator interaction requirements, and weaponry specifications while these systems are still in their early stages. Failing to do so could exacerbate the gap between modern technologies and the law, potentially leading to the development of uncontrolled weapons, as was the case with nuclear weapons.

[77] Dr. Alex Leveringhaus, a Ph.D. holder in the field of public administration, a research fellow at the Institute of Ethics, Law, and Armed Conflict at the University of Oxford, the coordinator of the Special Group on Ethics and Artificial Intelligence, personal interview 02.07.23.