7-2021

Killer Robots on Trial: Autonomous Weapons Systems in the Context of International Law

Mikaela Heck
Western University

UNIVERSITY OF WESTERN ONTARIO

Killer Robots on Trial: Autonomous Weapons Systems in the Context of International Law

by

Mikaela Heck

MAJOR RESEARCH PAPER
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF ARTS

SUPERVISOR: PROFESSOR DAN BOUSFIELD

GRADUATE PROGRAM IN POLITICAL SCIENCE

LONDON, ONTARIO

JULY 2021

TABLE OF CONTENTS

Killer robots are no longer a facet of science fiction, but rather an imminent reality. The development of autonomous weapons systems (AWS) has been something states and military operations have been working towards to build their arsenal and change the landscape of conflict. With this changing landscape, these AWS fit within public international law in a unique way, existing somewhere in between a weapon and a combatant. With increased autonomy and diminished human control over their behaviour, AWS present an interesting dilemma to existing international legal structures, as they are typically written in a fashion designed to be adhered by humans, not machines. In order to better understand and solidify the place of AWS within these structures, this paper will analyze legal scholars' works regarding AWS in armed conflict. Within the defined boundaries set forth in the international regulatory legal structures, this paper will provide analysis situated in context in order to provide a more grounded interpretation of AWS within these structures. literature review seeks to draw conclusions from these authors and their work, and how they contribute to finding a place for AWS within the existing international legal structures.

This literature review will look at these existing legal structures, specifically the 1977 Additional Protocol I (API) to the 1949 Geneva Conventions, as it broadly and holistically covers the norms and principles that must be adhered to in conflict. The norms and principles of international law set forth in the API are analyzed in terms of how AWS fit within these structures, identifying aspects and characteristics of the systems that would allow them to perform certain tasks, such as their ability to perform lawful attacks. Possible avenues for which AWS can be included in the clauses and definitions presented in these Conventions and principles are also provided, identifying that these weapons can be interpreted into these structures based on their autonomous system and weapons features, such as munition and

sensors. Specifically, the principles of discretion and proportionality are explored and the capabilities of AWS are analyzed as to whether they are actually able to comply with these principles. Components such as sensors and decision-making software are used to identify targets and perform proportionality calculations in order to adhere with these principles, though with varying levels of compliance confidence.

This literature review delves into the two different avenues of attributing responsibility for the actions of AWS. One option is the developers of the AWS, who designed and programmed the system with set parameters and code that determines its decision-making and action capabilities. More specifically, these developers ultimately determine how the AWS will perform its tasks, and therefore vicariously contribute to the acts it commits. However, these developers are also too far removed from the acts and their creation, and some aspects of the autonomous features diminish the influence of the developers over the AWS's behaviour. Another option for attributing responsibility is the military commander who deployed the weapon and has command responsibility over the AWS. Since the decision is deliberate on the part of the military commanders to ultimately deploy the AWS, they must be responsible for their actions on the battlefield, and should take all reasonable steps to ensure the AWS is in the best state to perform its military objectives.

WHERE DO THEY FIT? EXPLORING AWS'S PLACE IN INTERNATIONAL LAW

It is important to understand the current landscape of international law, which provides the foundation for interpreting how AWS fit within these existing structures. International law exists in four branches: the law of state responsibility, the law on the use of force, international humanitarian law (IHL) and human rights law.[1] This section of this paper will focus primarily on

[1] Denise Garcia, "Killer Robots: Why the US should Lead the Ban," *Global Policy* Volume 6, Issue 1 (2015): 60.

the law on the use of force, or law of armed conflict (LOAC), and IHL. International law concerning conduct in war focuses on human soldiers and their use of weapons and other force.[2] In other words, much of the language coded in the current international legal structures focuses on characteristics such as reasonability and common sense, which are traditionally associated as innately human characteristics and therefore create a difficulty in translating them for an AWS. Currently, there are no laws or treaties that explicitly provide governance on AWS, but rather regarding use of force and weapons as a whole.[3] Aspects of AWS are covered within existing legislation that deal with the projection of force, specific technologies and practices, and interpretations of IHL and principles of the LOAC.[4] For example, the principle of distinction – which will be discussed in length later in this section – requires that combatants distinguish between the civilian population and combatants, and between civilian objects and military objectives, and thus only attack military objectives.[5] Therefore, a weapon such the US Air Force's Low Cost Autonomous Attack system, which is capable of searching for and identifying targets, is capable of complying with this principle as it is capable of identifying a military target.[6] Thus, although AWS are not explicitly named in this principle, there is still the possibility of interpreting the law to match with the capabilities of an AWS.

Another aspect of international law is the concept of good governance. Good governance has been described by Gary Marchant et al. – who are members of the Autonomous Robotics group of the Consortium on Emerging Technologies, Military Operations, and National Security, and individually conduct research in legal and robotic ethics – as realistic, holistic, inclusive,

---

[2] Gary E. Marchant, Braden Allenby, Ronald Arkin and Edward T. Barrett, "International Governance of Autonomous Military Robots," *Columbia Science and Technology Law Review* 12 (2011): 289.
[3] Ibid, 289.
[4] Ibid, 289.
[5] Michael N. Schmitt and Jeffrey S. Thurnher, "Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict," *Harvard National Security Journal* 4, no. 2 (2013): 251.
[6] Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy*, Vol. 24, No. 1 (2007): 63.

feasible and malleable norms that can be accomplished under existing policy and legal constraints, and therefore can be assessed and improved upon when necessary.[7] In other words, good governance is not codified, but rather is a set of agreed upon norms that states can use when conducting operations, and it is understood that other states will adhere to these norms as well. Since they are not codified, states are able to re-evaluate and change these norms where necessary, for example the introduction of new technologies or new conflict environments. Good governance is a self-regulatory concept in which states hold each other accountable to a certain level of behaviour. Bode Ingvild and Henrick Huelss specialize in international politics and relations, and define norms as "standards of appropriateness for specific practices," allowing for broad guiding principles for states to adhere to.[8] In other words, the fundamental and procedural norms set out in existing principles of international law shape how states will act and provide standards for "appropriate" warfare.[9] As broad principles, they allow for a wider range of interpretations and scenarios in which they can be deemed applicable. Therefore, the malleability and self-regulatory nature of good governance allows for states to adapt to new technologies, such as AWS, and create new standards that are more inclusive of these developments as they happen.

IHL presents key legal steps when conducting an attack, which provides states and military commanders with a standardized approach to warfare. Alan Blackstrom and Ian Henderson summarize these steps as: collective information about a target; analyzing said information in order to determine the lawfulness of the target at the time of the attack; understanding the potential incidental effects of the weapon and taking precautions to minimize

---

[7] Marchant et al., 2011: 291.
[8] Ingvild Bode and Hendrik Huelss, "Autonomous weapons systems and changing norms in international relations," *Review of International Studies*, Vol. 44 (2018): 407.
[9] Ibid, 407-408.

them; assessing the proportionality of those effects against the anticipated military advantage of the attack; firing the weapon at the directed target; and monitoring the situation in order to cancel or suspend the attack if necessary.[10] In other words, in order for an attack to be considered lawful, it must adhere to the principles of distinction and proportionality, have collected all necessary information to have an exhaustive and conclusive understanding of the environment in which the attack will be made, and employ mitigation strategies in case an attack goes array. This list of relatively simple tasks becomes far more complicated with AWS, as the more complex a weapon is, the greater the potential is for discrimination to be affected by design errors or manufacturing errors.[11] This creates issues in anticipating the actions of these AWS and ensuring their compliance with these norms. If we are unable to anticipate the actions of an AWS, we are unable to reasonably have all the information necessary to ensure a lawful attack, as well as be confident that the AWS will be able to accurately calculate the proportionality of the attack before firing, or suspend or cancel their attack if necessary.

In order to ensure greater confidence that AWS will be able to conduct a lawful attack, it may be necessary to have safeguards in place to intervene and ensure effective compliance if the AWS itself cannot confidently do so. It is important, therefore, to have what Geoffrey Corn calls LOAC compliance enablers.[12] These can include military commanders and programmers who can ensure that the weapon has the capability to follow those key legal steps.[13] This approach has been used when issuing orders to subordinate units; for example, a unit of soldiers is told by their

---

[10] Alan Blackstrom and Ian Henderson, "New capabilities in warfare: an overview of contemporary technological developments and the associated legal and engineering issues in Article 36 weapons reviews," *International Review of the Red Cross* 94, no. 886 (2012): 485.
[11] Ibid, 486.
[12] Geoffrey S. Corn, "Autonomous weapons systems: managing the inevitability of 'taking the man out of the loop,'" In *Autonomous Weapons Systems: Law, Ethics, Policy*, ed. Nehal C. Bhuta et al. (Cambridge: Cambridge, 2016): 224.
[13] Ibid, 224.

commanding officer what their task is and the purpose of that task.[14] Therefore, Geoffrey Corn translates this idea to AWS, demonstrating how maximizing the articulation of this "task and purpose" within the intended tactical function of the AWS will facilitate compliance with the LOAC.[15] In other words, Corn argues that by having the developer program the AWS with certain functions capable of objectives while adhering to principles such as distinction and proportionality, and by having the military commander deploy the AWS within specific parameters into attack and monitoring its actions, this will create more confidence in the AWS's ability to comply with the LOAC. Corn additionally argues that this increased oversight of the development of AWS allows for better assessments of the potential risk of LOAC violation, which in turn will help define the weapon's purpose and intended use; therefore, the development phase of AWS is decisive in establishing LOAC compliance confidence.[16] By being more involved in the development of the AWS, officials are able to determine the capabilities of the AWS and therefore determine how and where it should be used to maximize confidence of LOAC compliance.

In terms of concrete codified law, the 1977 Additional Protocol I (API) to the 1949 Geneva Conventions provides a comprehensive outline for weapons law and LOAC. Article 36 of the API has required that a High Contracting Party is under obligation to determine whether a new weapon, means or method of warfare that it is developing or acquiring may be prohibited by the Protocol or any other rule of international law that the High Contracting Party is party to.[17] This links to Corn's argument, that it is important for States to have oversight over the development of a new weapon, such as an AWS, to ensure that its functionality is capable of

[14] Ibid, 224-225.
[15] Ibid, 225.
[16] Ibid, 225-228.
[17] Ibid, 227.

complying with the LOAC. Additionally, under this article, it is required of states that are party to this API to conduct legal reviews of all weapons being developed to ensure they meet the requirements laid out in the API.[18] These reviews are self-regulatory, rooted mainly in customary international law and norms. Article 38 is considered to embody the customary law of obligation regarding weapons and therefore acts as a starting point for understanding weapons law.[19] It includes a wide range of weapons and how they should be used, which can be considered unlawful in itself or unlawful in the way that it is used in certain circumstances only.[20] This provides a very broad approach to understanding and governing weapons. Article 51(4)(c) of the API also details that weapons systems that have uncontrollable effects, despite being able to strike their targets accurately, are not allowed.[21] Article 50(1) deals with more "human" aspects of conflict, specifically looking at doubt and its role during an attack.[22] Article 50(1) explains that "doubt as to status of a person must be resolved in favour of treating that individual as a civilian" during an attack.[23] This threshold for doubt is, however, framed in terms of human reasonableness, and therefore complicates translation for AWS.[24] This is not to say that it cannot be done; algorithms that can precisely measure doubt and reliability of target identification can mitigate this issue, providing these systems with all the information necessarily to act as if they were a reasonable human attacker.[25]

---

[18] Ibid, 28.
[19] Hin-Yan Liu, "Categorization and legality of autonomous and remote weapons systems," *International Review of the Red Cross* 94, no. 886 (Summer 2012): 638.
[20] Ibid, 639.
[21] Michael N. Schmitt, "Autonomous weapon systems and international humanitarian law: a reply to the critics," *Harvard National Security Journal*, 4 (2013): 14.
[22] Schmitt and Thurnher, 2013: 262.
[23] Ibid, 262.
[24] Ibid, 263.
[25] Ibid, 263.

Ultimately, a main concern when determining whether AWS will ever be able to conduct themselves in the nuanced way necessary to comply with international law rests in the two key principles of international LOAC: distinction and proportionality. The two substantive rules within the API are codified in Article 54(b)(4) and Article 35(2).[26] These two rules determine the lawfulness of the weapons themselves; the first stating that a weapon is deemed indiscriminate if it cannot be aimed at a specific target, and the second stating that a weapon is deemed disproportionate if its nature is to cause unnecessary suffering or superfluous injury to combatants.[27] It is important, therefore, to delve further into these principles and examine whether AWS have the capabilities to comply with them in order to better understand their place within the legal landscape.

Distinction

Distinction, also referred to as discrimination, is a crucial element of international law which protects civilians and civilian objects. This principle is codified in law through Article 48 of the API, stating that: "In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."[28] This means that states have an obligation to distinguish combatants from non-combatants, and military targets from civilian objects, in order to protect civilians and ensure that they are only attacking targets that fulfill their military objective. Civilian objects are those that are "indispensable to the survival of the civilian population," as well as the natural environment, historic monuments, places of worship

[26] Kenneth Anderson and Matthew Waxman, "Law and ethics for autonomous weapon systems: why a ban won't work and how the laws of war can," *Hoover Institution, Stanford University* (2013): 10.
[27] Ibid, 10.
[28] Schmitt and Thurnher, 2013: 251.

and works of art.[29] Additionally, distinction requires that attacks be limited to military objectives, and defines a military objective as "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."[30] There is little difference between the application of these rules governing attacks on individuals and objects, but ultimately AWS that lack the capability to distinguish between lawful and unlawful targets may be considered illegal under the Protocol.[31] There are AWS that have been developed with this capability; for example, the South Korean military developed a stationary sentry robot that is capable of detecting and selecting targets, as well as respond with lethal or non-lethal force depending on the circumstances at the time, without human input.[32] Additionally, the Phalanx Close In Weapons Systems for Aegis class cruisers in the US Navy are capable of autonomously conducting their own searches, detection, evaluation, tracking and killing of targets.[33] Therefore, in order for an AWS to be considered discriminatory in nature, it must have the ability to distinguish its target from non-targets, and aim in a way that only attacks that target. It must also have the ability to actively survey its area and make distinctions throughout the course of its military operation.

Distinction is not always black and white in terms of identifying civilian and military objects. Distinguishing between military and non-military objects becomes more difficult with the labelling of targets as "suspected terrorists;" persons in this category in armed conflict are

---

[29] Markus Wagner, "Autonomy in the Battlespace: Independently Operating Weapon Systems and the Law of Armed Conflict," In *International Humanitarian Law and the Changing Technology of War vol. 41.* Edited by Dan Saxon (Leiden, Brill: 2013): 110.
[30] David Akerson, "The Illegality of Offensive Lethal Autonomy," In *International Humanitarian Law and the Changing Technology of War vol. 41*, ed. Dan Saxon (Leiden, Brill: 2013): 77-78.
[31] Schmitt, 2013: 18.
[32] James Foy, "Autonomous Weapons Systems: Taking the Human out of International Humanitarian Law," *Dalhousie Journal of Legal Studies* 23 (2014): 50.
[33] Ibid, 51.

prima facie, or at first appearance, civilians and are protected as such until they participate directly in the hostilities.[34] Determining whether a civilian is actually directly participating in the conflict requires analyzing whether that participation is "direct or indirect, continuous or sporadic, and caused a sufficient level of harm of a military nature."[35] This, therefore, creates a highly subjective scenario for a soldier, or AWS, to decipher in order to apply the principle of distinction, and the question remains whether AWS have the capability to perform such tasks.[36] Some difficulties can also arise for the attacker in cases where they must distinguish between civilian and military objects when said object can be classified as both civilian and military in purpose.[37] An example of this could be a bridge an army uses to get supplies; the bridge can be considered to serve a civilian purpose as it was designed for civilian commuting, but can also be considered to serve a military purpose as the army is using it to transport its supplies, and therefore aids the army in gaining supplies and building their attack. [38] Thus, the attacker must be able to make the decision as to whether or not it serves more of a civilian or military purpose at the time of the attack. The uncertainties surrounding the ability of AWS to discriminate between seemingly undetermined individuals to determine which are legitimate military targets and which are not raise some serious concerns regarding discrimination.[39] This malleable identity can cause issues for AWS, as their software would have to be capable of constantly re-evaluating and re-defining their targets depending on their interactions, as well as being able to identify when an individual is in fact participating in the conflict. Thus, there are a myriad of considerations to

[34] Liu, 2012: 645.
[35] Akerson, 2013: 77.
[36] Ibid, 77.
[37] Markus Wagner, "Taking Humans out of the Loop: Implications for International Humanitarian Law," *Journal of Law, Information and Science 21*, no. 2 (2011/2012): 160.
[38] Schmitt and Thurnher, 2013: 160.
[39] Liu, 2012: 645.

make when designing an AWS to ensure that it has the capabilities to accurately comply with the principle of distinction.

This principle not only details rules pertaining to the attacker, but also outlines guidelines for the weapons themselves. Article 51(4)(a) states that attacks that are not directed at a specific target and but rather strikes to lawful targets without discrimination are banned.[40] This article is particularly important for AWS, as it is different than the ban on indiscriminate weapons because this ban also involves weapons that have the capability to aim at a lawful target, but do not do so.[41] Therefore, under this Article, AWS must have and use sensors to enhance their ability to distinguish between lawful and unlawful targets.[42] However, this becomes far more difficult the more complex a mission becomes. An AWS can only compute a given procedure under the confines in which its code was written, and therefore may not be able to sufficiently operate within those confines in order to properly identify what is and is not a target.[43] For example, using the example of the "suspected terrorist" laid out above, if an AWS is not coded to be able to properly identify when an individual becomes directly involved in a conflict, it is unable to comply with the principle of distinction. This also can produce some difficulties, as the principle of distinction not only requires the proper distinction of legitimate and illegitimate targets, but also requires that an attack be carried out by weapons that have the capability to prosecute the attack in a discriminatory way.[44] For example, an AWS must be able to attack a target without attacking other objects in the process; if it is the case where an AWS is in a position where it is unable to discriminately attack its target, it must be able to abort in order to adhere to the

[40] Schmitt and Thurnher, 2013, 253.
[41] Ibid, 253.
[42] Ibid, 253.
[43] Noel E. Sharkey, "The inevitability of autonomous robot warfare," *International Review of the Red Cross* 94, no. 886 (2012): 789.
[44] Wagner, 2011/2012, 161.

principle of distinction. Therefore, their underlying software would have to be sophisticated enough to determine whether a target it civilian or military in nature, and would have to be coded to take into account uncertainty and abort the attack if needed.[45]

Civilians are a central focus of this principle, specifically regarding a duty of constant care to ensure that civilians and civilian objects are spared during conflict. This duty of constant care, in terms of AWS in attacks, would include the procurement of said AWS and its preparation for deployment, as well as the deployment and its operation in the battlefield.[46] Article 57(2)(a)(i) amplifies this by requiring that combatants "… do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives within the meaning of paragraph 2 of Article 52 and that it is not prohibited by the provisions of this Protocol of attack them accordingly."[47] In other words, the attacker must collect all relevant information to ensure the validity of their target's identity, and that the target is in fact a military objective and not a civilian or civilian object. This includes preparing the mission, programming the autonomous software, reviewing the available information, prescribing the areas to be searched and when, and setting the target identification criteria for the weapon in order to ensure that it is able to appropriately adhere to the obligations under this Article.[48] However, feasibility is very subjective, and invokes human judgement and discretion in order to make these decisions.[49] This feasibility, therefore, is ultimately an issue of reasonableness, and under the LOAC, it would require an attacker to assume greater risk to avoid damage if a reasonable attacker in the same or

[45] Ibid, 161.
[46] Ibid, 87.
[47] Ibid, 80.
[48] Bill Boothby, "Weapons Law, Weapon Reviews and New Technologies," In *Routledge Handbook of War, Law and Technology*, ed. by James Gow et al. (London and New York: Routledge Taylor & Francis Group, 2019): 34.
[49] Akerson, 2013: 80.

similar situation would do the same. [50] With the subjectivity of this aspect of distinction, an AWS alone may not be able to comply, as it is not able to operate in a way that allows for reason in its current definition, which is inherently human-focused.

Reasonableness and feasibility are both understood in an innately human and subjective way, and therefore present an obstacle for defining AWS within this principle. For example, the Geneva Convention cites "common sense" as a requirement for being a combatant, and thus this would render AWS as potentially unlawful, as machines are unable to conduct this type of reasoning due to their programming. [51] Noel Sharkey, a specialist in the ethics of robotics, argues that common sense is still necessarily for reasoning and making discrimination decisions, and therefore this lack of "battlefield awareness" renders AWS unable to have the independent facility to operate on the principles of distinction with the limited constrains in which they are coded.[52] As common sense is grounded in rationality and reason that require a broader scope of decision-making, this argument follows that the constraints that are programmed into AWS render it incapable of having common sense and operating within this space. It would logically follow that in order for this "common sense" requirement to be met, an AWS must be supervised by a superior human operator in order to be considered compliant with the principle of distinction in this sense.

Proportionality

Proportionality works in tandem with distinction, focusing on the scale of the attack and its effects on civilians. This principle is codified in Articles 51(5)(b) and 57(2)(a)(iii) of the API, prohibiting: "an attack which may be expected to cause incidental loss of civilian life, injury to

[50] Schmitt and Thurnher, 2013: 261.
[51] Sharkey, 2012: 789.
[52] Ibid, 789.

civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[53] Therefore, proportionality requires a combatant, before attacking, to weigh the potential loss of civilian lives against the military advantage, and determine whether that loss would be proportional to the advantage, meaning that the effects would balance each other out. This proportional weighing of potential harm to civilians and civilian objects and the potential military advantage of the attack requires contextual and discretionary decision-making with emphasis on reasoning.[54] However, as seen with distinction, reasonableness is vague and subjective, and therefore makes it difficult to uniformly enforce and determine, especially in terms of AWS and their capabilities.

Regarding human combatants, the principle is considered one of the most complex and misunderstood norms in the law of armed conflict in terms of interpretation and application.[55] Michael Schmitt and Jeffrey Thurnher, both experts in international and military law, argue that this occurs because core notion of proportionality lies within the idea of "excessiveness."[56] With no accepted definition of excessive in the law of armed conflict, determining excessiveness is done on a case-by-case basis, evaluating it in terms of reasonableness within the given circumstances of the action.[57] It therefore follows that excessiveness can be looked at through the idea that the greater the reasonably anticipated military advantage that might occur from the attack, the more the law will tolerate the expected collateral damage of said attack.[58] IHL presents that proportionality rests on the notion that "belligerents must exercise restraint in the face of highly uncertain environments," and decisions regarding proportionality should be

[53] Schmitt and Thurnher, 2013: 253.
[54] Ibid, 56.
[55] Schmitt and Thurnher, 2013: 254.
[56] Ibid, 254.
[57] Ibid, 254.
[58] Ibid, 254.

weighed against dynamic environments through highly qualitative and subjective knowledge.[59] However, this becomes difficult for AWS, as their programming is constrained in terms of its decision-making process and is unable to process information in a qualitative and subjective way, and thus may not be able to make the calculations necessary to appropriately comply with proportionality.

There are some ways that AWS can be designed to ameliorate their decision-making abilities to better comply with the principle of proportionality. Not all AWS are capable of proportionality; for example, landmines can be considered AWS, as the decision to detonate is made by the machine, but they are simply programmed to detect the intended scenario and detonate.[60] This programming is very simple, consisting of a mechanical spring or arrangement that is unable to calculate proportionality.[61] AWS, however, can be programmed to have pre-determined parameters for when they can and cannot attack to adhere to proportionality.[62] This can be done through programming "doubt values" to increase reasonableness in decision making, and ensure there are adequate sensors to ensure the correct identification of targets and other civilian individuals and objects.[63] In this sense, Schmitt and Thurnher interpret reasonableness in terms of a calculation based in "doubt," and the amount of doubt there is regarding a target or the attack's effect. Therefore, reason becomes quantified and tangible for AWS to process and input into their actions. These parameters can also be set by programming relative judgement into an AWS, which will allow it to measure anticipated civilian harm and military advantage, subtract and measure the balance against a standard of "excessiveness," and, if excessive, not attack.[64]

---

[59] Garcia, 2015: 59.
[60] Thrishantha Nanayakkara, "Autonomy of Humans and Robots," in *Routledge Handbook of War, Law and Technology*, ed. by James Gow et al. (London and New York: Routledge Taylor & Francis Group, 2019): 136.
[61] Ibid, 136.
[62] Schmitt and Thurnher, 2013: 264.
[63] Ibid, 264.
[64] Anderson and Waxman, 2012: 10.

Kenneth Anderson and Matthew Waxman – a scholar of international and technology law, and scholar of international law and LOAC, respectively – thus turn this relative judgement into a concrete calculation that is weighted against a set coded standard, allowing the AWS to be able to make these decisions. Although this technically fails to holistically approach proportionality in the way that international law has traditionally expected, it provides a method for which AWS to adhere to the principle in their own way. Though the type of reason that has traditionally been linked to proportionality has relied mainly on human judgement and subjectivity, looking at proportionality as Schmitt and Thurnher, and Anderson and Waxman, have, makes it possible to understand reason in a more quantitative way, and therefore allow for an interpretation of the principle that fits with AWS.

Proportionality and the concept of excess can also be viewed through the concepts of "superfluous injury [and] unnecessary suffering."[65] Schmitt and Thurnher note that Article 35(2), the article that addresses this concept, only addresses the effect of weapons systems on target individuals, and not the actual manner of engagement, and therefore AWS would not automatically violate this principle.[66] This, therefore, states that AWS are able to comply with proportionality and causes harm that determines whether or not the effect was proportional or not. William Boothby, a leading authority in new weapons technologies and the development of international law, argues that the autonomous nature of AWS would not likely directly contribute to the degree of suffering or injury, but rather the munition that is being delivered to the target.[67] In other words, he argues that AWS are capable of complying with this principle so long as the human operator in charge of loading the weapon or choosing its munition would be responsible

[65] Corn, 244.
[66] Schmitt and Thurnher, 2013: 244.
[67] Bill Boothby, "Autonomous Attack - Opportunity or Spectre?," In *Yearbook of International Humanitarian Law Volume 16*, ed. Terry D. Gill. (Cambridge: Springer, 2013): 75.

for ensuring proportionality. However, Markus Wagner, a scholar specializing in IHL and the technology of war, argues that AWS can be able to determine what type of effect and munition to use to produce an attack in any given circumstance, and understand the weighting of the effect on the military objective and civilian population.[68] This would thus eliminate the need for human supervision, but would require the AWS to have the capability to calculate the proportionality of its attack before choosing its method. This, therefore, adds another element to the proportionality calculations that the AWS would need to be capable of making for itself.

The question regarding AWS therefore is whether they are capable of performing these proportionality calculations properly. The decision of proportionality rests on assessing and processing complex data that can sometimes be based on contradictory signals if they are measured against a preprogrammed set of action criteria that tend to be characteristic of AWS.[69] Thus, proportionality calculations must subjectively determine the "value" of the anticipated military advantage gained from the attack and weigh it against the harm expected to civilians and civilian objects, and take precautions or ultimately forfeit the attack if there is too high of a degree of doubt.[70] However, there still remains a level of vagueness when it comes to these calculations, even though they rely on pre-set parameters which make the decision-making process of the AWS much more stringent. Therefore, AWS do not have the capability to properly calculate proportionality, as their programming limits them from doing so.

These calculations are simply not holistic or malleable enough to ensure adequate compliance with the principle of proportionality. Pablo Kalmanovitz argues that simply setting threshold values for proportionality assessments within narrow settings does not get rid of human

[68] Wagner, 2013, 121.
[69] Bode and Huelss, 2018: 402.
[70] Schmitt and Thurnher, 2013: 266.

judgement, but rather makes it indispensable.[71] AWS themselves are incapable of making the type of judgements that proportionality currently requires, even with human-made algorithms and choices.[72] AWS also lack the reasonableness to balance the two sides of the proportionality calculation in a meaningful way.[73] Thus, as Kalmanovitz argues, that the test of proportionality will ultimately rest on the human decision to deploy the AWS, assess "in good faith" and according to "common sense" whether the weapon is able to act proportionally given the algorithm and action parameters, and in the specific conditions of its deployment.[74] In other words, a human must make the decision to deploy an AWS based on its judgement, but the AWS itself is still capable of acting proportionally. Therefore, though there is still a need for human supervision, an AWS does have the capability to calculate proportionality if the parameters of the military operation allow for it to do so and is not too complicated as to render the decision out of the scope of the programming's decision-making capabilities.

ROBOTS ON TRIAL: EXPLORING RESPONSIBILITY AND LIABILITY

This section of the paper it will examine the different options for attributing responsibility for the actions of AWS. Before exploring this topic further, and exploring the different roles and responsibilities of the developer and military commander with regards to AWS, it is important to understand the definitions of responsibility and liability. Andreas Matthias specializes in the ethics of new technologies, and presents that an agent can only be held responsible if they know the particular facts that surround their action, they are able to freely form a decision to act, and are able to select one of the suitable available alternative

[71] Kalmanovitz, 2016: 151.
[72] Ibid, 151.
[73] Ibid, 151.
[74] Ibid, 151.

actions based on the facts of the given situation.[75] For an AWS, as their behaviours and ability to make decisions are based on pre-determined code and parameters, they are unable to freely make decisions and do not have full control over their behaviour, which makes them unable to be held responsible under this definition.

This definition of responsibility reflects a more human-focused idea of responsibility, and therefore responsibility of an AWS or other technology must be defined differently. Giovanni Sartor and Andrea Omicini are scholars in law and artificial intelligence, and computer science and autonomous systems, respectively. They present three notions of responsibility concerning technology: functional responsibility, blameworthiness and legal liabilities for harm.[76] Functional responsibility assumes that "the harm would not have resulted had the responsible component correctly exercised the function attributed to it."[77] In other words, if the machine had been functioning correctly, specifically the component of the AWS that was caused harm to the object of individual, said harm would not have occurred; essentially, it is harm caused by a malfunction or error. This means that any component or subcomponent of the system could fail to exercise its expected function and therefore have harmful consequences for which the malfunctioning component may be considered responsible.[78] For example, failure in the system's sensor causes the weapon to be incapable of adhering to the principle of distinction and harms an innocent civilian; therefore, the failure of the sensor can be considered responsible for the AWS not complying with the principle of distinction and the harm caused. Blameworthiness continues this idea by stating that the fact that the failure that caused the harm involves a fault.[79] Faulty

---

[75] Andreas Matthias, "The responsibility gap: Ascribing responsibility for the actions of learning automata," *Ethics and Information Technology* 6 (2004): 175.
[76] Giovanni Sartor, and Andrea Omicini, "The autonomy of technological systems and responsibilities for their use," In *Autonomous Weapons Systems: Law, Ethics, Policy*, ed. Nehal C. Bhuta et al. (Cambridge: Cambridge, 2016): 62.
[77] Ibid, 62.
[78] Ibid, 62.
[79] Ibid, 62.

design of a weapon system could result in its inability to exercise the function that is attributed to it; for example, if a computer supporting an autopilot in a drone burns out due to a design failure, and an accident occurs, then the fault should be allocated to the developers of the computer.[80] The third notion of legal liabilities for harm can be related to a number of forms of liability, such as strict liability, vicarious liability, product liability and negligence.[81] Under these two notions, responsibility can also be attributed to the humans who work on and with the AWS, such as the developers who design the weapons and military commanders who deploy them, and not just the functional aspects of the weapon itself.

Criminal and international law recognizes this notion of shared responsibility, and allows agents associated with the performance and actions of an AWS to be considered responsible for its actions. Neha Jain is a legal scholar who specializes in public international law, and explores how criminal law recognizes that there are instances in which responsibility can be shared, as the immediate agent who is most directly related to the offence is "autonomous" in the material sense, but another agent can still be held responsible for their conduct.[82] In other words, even though the agent who committed the act did so of their own ability, another agent could still be responsible for those actions based on their relationship with the immediate agent, such as the one who commanded the attack. She explains that criminal and civil law systems recognize that there are various categories of perpetration and principal responsibility.[83] Most consistently in these systems, a person is considered a principal through personal fulfillment of both the action (*actus reus*) and the intent (*mens rea*), but some systems recognize that principal responsibility

[80] Ibid, 62.
[81] Ibid, 64.
[82] Neha Jain, "Autonomous weapons systems: new frameworks for individual responsibility," in *Autonomous Weapons Systems: Law, Ethics, Policy*, ed. Nehal C. Bhuta et al. (Cambridge: Cambridge, 2016): 307-308.
[83] Ibid, 308.

can come even when the accused acts "through" another.[84] For example, a military commander acting "through" his soldiers, since he was the one to deploy the troops and give them the military operation.

There are two concepts that help give guidance for how to attribute responsibility. One concept is the "innocent agent", who is an individual whose actions are not deemed "free, informed or voluntary" due to factors such as ignorance, insanity or minority, which in turn can be regarded as having been "caused" by the conduct of another person.[85] AWS can, thus, be considered an innocent agent, as their actions are not free or voluntary due to the nature of their behaviour being pre-determined by another person. They also present the idea of the "semi-innocent" agent, which applies when perpetrators actions are considered not fully voluntary, but not to the extent that it would absolve them of criminal responsibility; they can be characterized as having "caused" to the extent that they did not possess the complete knowledge necessary to fully comprehend the nature or circumstances of their conduct.[86] The *mens rea* of the direct perpetrator therefore must be judged in terms of the secondary party's mental state, and will require intent or knowledge.[87] This can also apply to AWS, as their code gives them the ability to perform some decision-making capabilities, and therefore be able to comprehend certain elements of their actions. However, ultimately, their actions are limited by a human agent, who sets parameters for how they are able to act. Therefore, responsibility can be shared by both the AWS and another human counterpart who is involved in its behaviours and actions.

Responsibility does not only occur when the act happens, but rather can be attributed at any time throughout the operation. According to specialist in the ethics of artificial intelligence

[84] Ibid, 308.
[85] Ibid, 308.
[86] Ibid, 308.
[87] Ibid, 310.

and autonomous systems Johannes Himmelrich, responsibility can be forward-looking and backward-looking.[88] Forward-looking responsibility refers to "obligations to manage risks, perform certain actions, or produce certain outcomes."[89] This encompasses the obligations under IHL and the LOAC laid out above, in that actors are held responsible for complying with these principles and norms. For example, under this concept of forward-thinking responsibility, a military commander is responsible for ensuring that all systems are functioning properly in an AWS before deploying it, and programming it with the parameters necessary to ensure the intended military objective is met. Backward-looking responsibility refers to "what an agent acquires because of what she has done or brought about that grounds permission of other agents to react to this agent in certain ways."[90] In other words, the actor can be held responsible for their the results of their actions after they have been committed. For example, if an AWS harms an innocent civilian during their mission, then it and/or someone connected to the machine's actions must be held responsible for the harm done to that civilian.

This backward-looking responsibility is the most common way of thinking about responsibility, and presents a more holistic approach to attributing responsibility for an unintended result. Himmelreich explains that when an agent is responsible in this backward-looking sense, then others are justified in holding them responsible in terms of attributability, accountability and answerability.[91] In terms of attributability, responsibility is determined by express or constitute judgement of a person's action and of the person themselves.[92] In other words, responsibility is attributed to someone based on the judgement of an official legal body.

---

[88] Johannes Himmelreich, "Responsibility for Killer Robots," *Ethical Theory and Moral Practice* 22 (2019): 733.
[89] Ibid, 733.
[90] Ibid, 733.
[91] Ibid, 733.
[92] Ibid, 733.

Thus, attribution of responsibility is determined in an ununiform and case-by-case basis depending on the facts of the case and the judgement of those proceeding over the case. In terms of accountability, responsibility is determined by justifying taking a "certain stance towards this person and forming evaluative or emotive attitudes, such as blame, praise, or resentment, as a part of this stance."[93] This means that accountability is determined through both qualitative analysis and emotional response; for example, to hold someone accountable for the harm caused by an AWS, it must be determined that the person was involved in process of that harm occurring, such as programming the system or determining the operation parameters, and can attribute blame to that person's fault. Himmelreich, however, argues that responsibility need not always involve accountability and blameworthiness, but rather involve that an agent by answerable for their actions and apologize for the harms that ensued.[94] Answerability can be determined by assessing and questioning the reasons the agent took in justifying their actions in order to justify the attitudinal stances taken.[95] This aspect of responsibility becomes more muddied with AWS, as with their complex systems, it is hard to accurately determine its decision-making process and therefore make it difficult, if not impossible, to assess the reasoning or justification for its actions. However, determining answerability for AWS can take a different approach by assessing the initial code written by the developer, or the reasoning for deploying the AWS in the way that it was with the guidance or military objectives it received. Although this may not be able to specifically determine the reasons for the actions taken by the AWS, it will give a better idea of the decision-making process and reasoning by understanding the pre-determined parameters within which it was operating.

---

[93] Ibid, 733.
[94] Ibid, 733.
[95] Ibid, 733.

To bring it back to forward-looking responsibility, liability can be attributed based on whether an actor failed to take precautions in order to ensure compliance with IHL and the LOAC. Boothby presents that there is no liability for "the damage lawfully done to military objectives, for the death or injury lawfully caused to members of the opposing armed forces, for expected death, injury or damage to civilians or civilian objects which is not excessive in relation to the anticipated concrete and direct military advantage, or for the death or injury of civilians or damage to civilian objects caused by mistaken or erroneous attacks caused, for example, by the malfunction of military equipment."[96] In other words, since these actions are considered lawful under the principles of distinction and proportionality, as well as other principles codified in law, the agent who caused these effects will not be held liable for them. Liability, therefore, rests in damage caused by the failure to take all feasible precautions in relation to the attack operation which results in disobeying the law.[97] For example, excessive harm to civilians caused by an AWS that failed to complete all necessary proportionality calculations is liable for that harm, since it did not take all feasible precautions by not completing all of the calculations. This could occur because of a developer not programming all the possible calculations into the AWS's software, or the military commander not providing the appropriate parameters or information needed to perform the tasks.

With AWS, there exists that problem of responsibility gaps in determining liability and responsibility, as it becomes difficult to attribute individual responsibility to the actions of an AWS due to the nature of its programming and operation. Himmelreich explains that responsibility gaps occur when an AWS harms someone but there is no one responsible for that

[96] William Boothby, "Some legal challenges posed by remote attack," *International Review of the Red Cross* 94, no. 886 (2012): 591.
[97] Ibid, 591.

harm.[98] He goes on to explain that these responsibility gaps seem to only occur when minimal agents that have intentional agency but not moral agency are used in conflicts; they have intentional agency in the sense that they can form beliefs, decision and actions but they cannot be responsible for those actions without that moral agency.[99] An agent has moral agency if their actions originate within themselves and reflect their end, which must come from their capacity to reason on the basis of their past experiences, and this end must have been chosen by themselves.[100] Therefore, an AWS does not have moral agency because even though their actions do originate within themselves from their coding and software, their software was written and determined by someone else, and their end was, to an extent, chosen by someone else who decided their military objectives. Sartor and Omicini build on this by articulating that the deployment of AWS could determine responsibility and liability gaps due to their impossibility to attribute moral responsibility and legal liabilities to anyone based on harms caused by the AWS's autonomous operation.[101] Since their autonomous operation causes a diminishing influence of their human operators and developers, this creates a responsibility gap in that these humans can be considered too far removed from the actions of the AWS to be held responsible for those actions.

The idea of attributing responsibility to AWS can been compared to the use of child soldiers in war. Sparrow explains that while children can be argued to lack full moral authority, they are autonomous and are capable of acting and making decisions on their own, just like AWS.[102] He argues that they are not appropriate objects of punishment, as they are not capable

[98] Himmelreich, 2019: 731.
[99] Ibid, 734.
[100] Sparrow, 2007: 65.
[101] Sartor and Omicini, 2016: 68.
[102] Sparrow, 2007: 73.

of fully understanding the moral dimensions of what they are doing and therefore understanding their crime and punishment.[103] However, the limited autonomy that these child soldiers do have is enough to ensure that those who order them into doing those actions do not control them, which makes attributing responsibility problematic.[104] Sparrow explores a space where these children are sufficiently autonomous to make it difficult to attribute responsibility to appropriate adults, but not autonomous enough to be responsible for their own actions, and argues that it should be the person who placed them in the position where they played the causal role who should be held responsible.[105] Analogous with this example, AWS have the ability to act on their own without immediate control, but these actions are pre-determined based on their coding and the military objectives delivered to them. Therefore, the person who is held responsible can be the person who developed the weapon and pre-determined its capabilities, or the commander who deployed it and delivered its military objectives, as they were the ones involved in guiding their actions.

The developers

Responsibility for the actions of an AWS can lie with the developer of that weapon, as the design of the weapon contributes to its ability to perform tasks, and therefore the developer plays a crucial role in determining the actions of an AWS. Although it is rarely an individual who is solely in charge of developing, designing and programming the weapon, but rather more likely a team who works together for a technological corporation, for the purposes of this paper, developer will be used as a shorthand in explaining this responsibility attribution. The phases in t the lifetime of a new weapon include concept, assessment, development, manufacture, in-service

[103] Ibid, 73.
[104] Ibid, 73.
[105] Ibid, 74.

and disposal.[106] This means that for a significant portion of the lifecycle of the AWS, the developers are involved and in charge of determining how it is programmed, designed and developed. Michael Schmitt argues that since a human, or rather team of humans, must decide how to program a system, they would be accountable for programming it to engage in actions that could amount to war crimes.[107] Though he does argue that it is (hopefully) improbable that the developers would design an AWS to commit war crimes, he explains that it would be much more likely that a system that has not been programmed to do so is used in a manner that constitutes a war crime.[108] In this case, it could be argued that the developers be held responsible for that war crime if the AWS is unable to discriminate between a combatant and non-combatant due to the programming of its sensors. Sartor and Omicini also contribute to this debate, stating that developers will be considered blameworthy when they "negligently or intentionally contribute to delivering a device that either (a) would not achieve the intended function, or (b) would achieve the intended function, but this function necessarily entails unacceptable consequences."[109] Therefore, responsibility can be placed on developers that did not equip the AWS with, for example, sensors and software that is capable of discriminating at the level required for the intended task for harm caused to innocent civilians due to this error.[110] The developers are directly involved in determining the capabilities of the AWS, and therefore have a role in their actions and decision-making.

---

[106] Tony Gillespie, "Humanity and Lethal Robots: An engineering perspective," in *Routledge Handbook of War, Law and Technology*, ed. by James Gow et al. (London and New York: Routledge Taylor & Francis Group, 2019): 194
[107] Ibid, 33.
[108] Ibid, 34.
[109] Sartor and Omicini, 2013: 63.
[110] Ibid, 63-64.

Developers play a role in the actions of AWS, and therefore can be responsible for them, as their capacity to cause harm is a direct result of its design.[111] Tim McFarland and Tim McCormack, authors of "Mind the Gap: Can Developers of Autonomous Weapons Systems Be Liable for War Crimes," argue that the developers exert greater control over the range of actions of the AWS and the specific actions it performs after its been deployed than the analogous relationship between combatant and commander, and therefore responsibility for the proscribed acts committed by an AWS can be more easily ascribed to the former relationship than the latter.[112] This is because these actions are pre-determined by the programming and design of the AWS that the developers created. However, McFarland and McCormack go on to argue that to the extent that weapon developers may be considered instigators of an action by an AWS, they do so through control software, and the degree of control exercised by the developers depends on the degree of autonomy of the AWS with respect to the action.[113] In other words, as the degree of autonomy increases – as in the less control an operator or commander has – the greater the share of control the developers have over the behaviour of the AWS. Therefore, as long as the operator of the AWS remains fully or partially connected to the operation of the weapon, the control is shared between the developers and the operator in terms of the actions conducted.[114] There may be points during a conflict in which developers occupy control over the actions of a system such that soldiers and commanders may be excluded and are unable to instigate or intervene in the actions of the AWS.[115] Thus, the developers are considered the primary determinants of the AWS's actions and behaviour, as they are the ones who program those capabilities.

---

[111] Tim McFarland and Tim McCormack, "Mind the Gap: Can Developers of Autonomous Weapons Systems Be Liable for War Crimes," *International Law Studies: U.S. Naval War College* 90 (2014): 366.
[112] Ibid, 366.
[113] Ibid, 375.
[114] Ibid, 376.
[115] Ibid, 376.

Due to the nature of the relationship between developers and AWS, the developers can be held liable for the actions of their weapons as they determine what actions they are able to conduct. McFarland and McCormack look at how *mens rea* requirements could be ground for developers to be held liable, particularly within the definition given above regarding aiding and abetting.[116] They bring up the concept that in order for a person or persons to be held criminally responsible, they must have committed the act in question with intent and knowledge, on the basis that intent exists "when a person means to engage in conduct or cause a consequence, and knowledge refers to awareness that a circumstance exists or a consequence will occur in the ordinary course of events."[117] This means that the person or persons must have had full knowledge of the circumstances under which the action happened, and intended for the action to happen. However, McFarland and McCormack use the example of a judgement made in the Trial Chamber of the International Criminal Tribunal of the Former Yugoslavia (ICTY) in the trial of Anto Furundžija.[118] The judgement noted that "it is not necessary that the aider and abettor should know the precise crime that was intended and which in the event was committed. If he is aware that one of a number of crimes will probably be committed, and one of those crimes is in fact committed, he has intended to facilitate the commission of that crime, and is guilty as an aider and abettor."[119] In other words, developers do not need to have been involved in the specific plans to commit the crime or intend for that specific crime to occur, but rather have programmed behaviours into the AWS that would be capable of committing criminal acts.[120] For example, developers that knowingly programmed sensors that were not entirely accurate in

[116] Ibid, 378.
[117] Ibid, 378.
[118] Ibid, 379.
[119] Ibid, 379.
[120] Ibid, 379.

detecting a target, or had limitations in certain circumstances, would be held responsible for an AWS killing an innocent civilian because they were unable to distinguish between targets. However, a prosecutor would need to demonstrate that these developers understood that the weapon was capable of behaving in an illegal manner in order to be held liable.[121] For example, if an AWS was placed in an environment that its developers had not anticipated it to operate in, and therefore its programming is not equipped to confidently and compliantly operate within that environment, and there is a crime committed, this could absolve developers of liability as they did not understand that the weapon could behave illegally as they did not understand how the AWS would operate in this foreign environment.

There are different ways that these system limitations can come about that can produce different levels of responsibility ascribed to developers. Robert Sparrow argues that fault could lie with the person or persons who designed and/or programmed the AWS if the fault was a result of negligence on the part of the design and/or programming team.[122] In other words, fault could be attributed to the developers if did not properly take all steps necessary to ensure that the AWS was designed and/or programmed in a way that allowed it to function the way it was intended without undue errors. He does argue, however, that this need not necessarily be the case. He explores the possibility that the machine may attack the wrong targets due to an acknowledged limitation of the system, and therefore if the manufacturers of the weapon have made it clear to those who purchased or deployed the AWS, then they cannot be held responsible.[123] Using the example presented earlier, developers who know that an AWS is unable to reliably detect targets in certain environments must tell the purchaser or deployer of the

[121] Ibid, 380-381.
[122] Sparrow, 2007: 69.
[123] Ibid, 69.

weapon that it should not be operated in those environments, as it has the possibility of distinction error. Additionally, the connection between the programmers and the results of the system which would attribute responsibility to them is broken as the system becomes more autonomous.[124] This is because as an AWS becomes more autonomous, the possibility that the AWS will make a choice other than those predicted by its programmers and designers increases as well.[125] In other words, the developers become less responsible for the actions of an AWS as they become more autonomous, as they are less likely to accurately predict its actions with greater autonomy. Sparrow likens this to the relationship between a parent and child, making the analogy that just as a parent is not responsible for the actions of their child after they have left their care, programmers should not be responsible for the actions of their AWS after they have "left".[126] The responsibility for the actions of an AWS attributed to developers becomes more distant as those actions become further outside of the control of the developers if they are deployed to operate outside of their operational limits, and as their degree of autonomy increases, as both of these instances have the possibility to cause unpredictable effects.

Degrees of autonomy will influence the role of developers in terms of attributing responsibility to them. Matthias builds on the idea of growing autonomy shifting the role of developers in terms of responsibility for their "creation".[127] With increased techniques of artificial intelligence programming being developed for these AWS, the developers' role becomes more distant as their code becomes more ambiguous, making it more difficult to isolate and identify errors in the decision-making process of the AWS.[128] Since this programming uses

[124] Ibid, 70.
[125] Ibid, 70.
[126] Ibid, 70.
[127] Matthias, 2004: 181.
[128] Ibid, 181.

predicate logic that is not executed in the same linear fashion, but rather runs deductions though inference rules, the flow of control in these systems is more difficult to describe.[129] However, as long as there is symbolic representation of the facts and rules involved in these deductions, there are methods of checking the stored information to ensure that it is correct.[130] Therefore, as long as there is a way to assess how the decision-making process was made, and how the information was gathered and interpreted by the AWS, developers can be held responsible for those actions if they are found to be at fault in terms of how they programmed the system.

In AWS with greater levels of autonomy, the representation of information becomes even more abstract and difficult to isolate and interpreted. The symbolic representation of this information and flow control disappears and is replaced with a matrix of synaptic weights, which cannot be interpreted directly and, therefore, any information stored in this network can only be inferred indirectly through experimentation.[131] This makes it incredibly difficult to ensure predictability, as the network is constantly changing.[132] Thus, this constantly changing network makes it impossible for the developers of the system to eliminate errors, but rather forces them to permit these errors so the system can learn and improve its operational performance.[133] Even with developers defining the operational parameters of the system, as well as define the alphabet used and the semantics of its symbols, the system they create programs itself through this genetic programming inherent in this autonomous network.[134] This makes the AWS work essentially outside of the observation of the developers, who is rendered unable to intervene manually.[135] Due to the nature of these increasingly autonomous networks, it becomes harder to attribute

[129] Ibid, 181.
[130] Ibid, 181.
[131] Ibid, 181.
[132] Ibid, 182.
[133] Ibid, 182.
[134] Ibid, 182.
[135] Ibid, 182.

responsibility to the developers, as they could not predict the actions of the AWS even though they were the ones who programmed them. With this increased ability to self-learn, AWS become more independent from their "creators", and therefore can be considered acting on their own outside of the realm of control of their developers.

Although responsibility can be diminished, it cannot be entirely removed from the developers. Hin-Yan Liu is a legal scholar focusing in artificial intelligence and its legal disruption, who explains how the complexity of the software can lead to the diminishing influence of the developers on the AWS's behaviour and conduct.[136] He explains that though the developers are able to program constraints to AWS behaviour, these constraints are likely to be quite broad and abstract, and therefore will allow the AWS to act in a way that could be difficult to predict given the range and complexity of the programming.[137] He argues that because of this, the obstacles of attributing responsibility regarding AWS are issues of control, predictability and foreseeability, which ultimately work together to determine the manner in which the AWS are developed and deployed.[138] Therefore, responsibility is determined by the circumstances under which the AWS is used, and determining the distribution of this responsibility is based on the level of control, predictability and foreseeability each actor has in the situation.[139] Liu illustrates this disconnection in responsibility through exploring the individual roles of a developer or commander of an AWS.[140] The individual or individuals are responsible for fulfilling their expectations and obligations that attach to their function of the AWS, and when they have

---

[136] Hin-Yan Liu, "Refining responsibility: differentiating two types of responsibility issues raised by autonomous weapons systems," in *Autonomous Weapons Systems: Law, Ethics, Policy*, ed. Nehal C. Bhuta et al. (Cambridge: Cambridge, 2016): 330.
[137] Ibid, 330-331.
[138] Ibid, 335.
[139] Ibid, 335.
[140] Ibid, 337.

fulfilled these obligations, they are deemed to have acted responsibly.[141] For example, if the developers have fulfilled their obligations to program the AWS to the best of their ability with the appropriate constraints and notify the commander of the limitations of its designs, they are deemed to have acted responsibly. However, Liu points out that AWS behaviour also depends on the commander who deploys it, and therefore this facilitates the displacement of blame from the programmer to the commander.[142] This is because the programmer can argue that they had discharged their obligations by implementing the general parameters, and therefore the unlawful system behaviour is the responsibility of the commander who failed to complement those parameters with more specific constraints in the AWS's deployment.[143] In other words, if the developers outline the parameters and limitations of the AWS to the commander, but they fail to work within them causing harm to occur, then the developers cannot be held responsible for those actions, as it was the commander who failed to deploy the AWS in the right circumstances. Thus, it is important to explore the role and responsibility of military commanders regarding AWS.

The military commander

The relationship between AWS and their military commander can be looked at as analogous to the relationship between commander and soldier. Heather Roff, a specialist in the law, policy and ethics of emerging technologies, explores this relationship through the doctrine of command responsibility.[144] This doctrine allows for the claim that commanders can be held morally and legally responsible for the actions of their subordinates, and these commanders

[141] Ibid, 337.
[142] Ibid, 331.
[143] Ibid, 331.
[144] Heather M. Roff, "Killing In War: Responsibility, liability, and lethal autonomous robots," in: *Routlege Handbook of Ethics and War: Just War Theory in the 21st Century*, ed. Fritz Allhoff et al. (New York & London: Routlege, 2013): 357.

"should have known" what would happen.[145] In other words, a commander should have known the possible outcomes of a military operation or commands given to the soldiers. Therefore, if a commander gives their soldiers an unlawful order, then they should have known unlawful outcomes would occur and thus be held responsible for those outcomes. Command responsibility is therefore premised on the fact that a superior-subordinate relationship does exist, and that the test to see whether a person is a superior is one of "effective control."[146] Effective control exculpates superiors from prosecution where a "person who is formally a superior in the line of command may be excluded from criminal liability if that superior does not exercise actual control."[147] Looking at AWS, Roff argues that effective control would exculpate commanders from legal responsibility as they are unable to control the machines if they have a heightened amount of autonomy.[148] Similarly as with developers, as the level of autonomy increases in an AWS, the level of control decreases for military commanders as the actions of the machine are less predictable and foreseeable. Additionally, she states that AWS are "impossible" to control "by a human in real-time due to its processing speed and the multitude of operational variables involved," and therefore there is no way for a commander to prevent or punish a violation of *jus in bello* by said AWS.[149] In other words, since the AWS is a self-regulatory system, it becomes difficult for commanders to exercise control over them and therefore cannot be responsible for the actions of the AWS. She continues by explaining that there can be no prevention of action as there can be no foresight of the actions the AWS will take, and therefore can be no punishment ascribed to the military commander.[150] Though she does acknowledge that some responsibility

[145] Ibid, 357.
[146] Ibid, 357.
[147] Ibid, 357.
[148] Ibid, 357.
[149] Ibid, 357.
[150] Ibid, 357.

can be attributed to commanders, as they made the decision to deploy the AWS, she argues that this level of moral responsibility is relatively low.[151] Therefore, the level of responsibility for military commanders for the actions of an AWS becomes quite limited as their level autonomy increases and diminishes the commander's control over those actions.

Command responsibility and control can occur to varying degrees between a superior and their subordinates. Himmelreich builds on the idea of control, and the differing degrees of control that can exist between a commander and a subordinate.[152] One way of understanding control that he proposes is Robust Tracking Control, which outlines that an agent has control over an outcome if that agent gave an order such that the outcome would occur in all relatively similar circumstances.[153] Conversely, if that agent did not give that order, the outcome would not occur in all relatively similar circumstances.[154] Robust tracking control also takes into account errors and that control need not be perfect; it allows for risky actions in that the outcomes can represent disjunctive descriptions and therefore the outcomes can include consequences that are unintended.[155] Therefore, this type of control can be attributed to commanders and the AWS they deploy, in that if a commander deploys an AWS with the order to kill a specific target, it will do so if all circumstances and parameters are appropriate for that outcome to occur, and it will occur in similar situations. Since the AWS is unable to act in any way until they are activated, this becomes the central decision of the commander and therefore places responsibility on the commander to engage the AWS in any action.[156] This means that the commander is responsible

---

[151] Ibid, 357.
[152] Himmelreich, 2019: 736.
[153] Ibid, 736.
[154] Ibid, 736.
[155] Ibid, 736.
[156] James Foy, "Autonomous Weapons Systems: Taking the Human out of International Humanitarian Law," *Dalhousie Journal of Legal Studies* 23 (2014): 49.

for the outcomes of the AWS as those outcomes would not have occurred had the order not been given.

To illustrate this concept, Himmelreich provides an example regarding AWS and their commanders. He provides a scenario in which a commander orders an AWS to patrol a large region and engage legitimate targets.[157] During the mission, communication is not maintained, and the AWS identifies a potential target, which can only be engaged immediately.[158] The AWS takes the target to be legitimate and engages it, and the target turns out to be a legitimate target.[159] In this example, the commander does not have control over the actions of the AWS. However, this is not because this particular target would be bombed if the commander were to give the order, but rather because there is plausibility that in similar situations where the commander gave the order, the AWS decides against the bombing of the particular target in favour of bombing another.[160] In this case, the commander is not responsible because this particular bombing does not track control order, but at the least, the commander still has control over whether or not they give an order.[161] In other words, even though the commander has no control over any of the particular bombings, they do have control over whether some targets might be bombed.[162] In this scenario, the commander would be held liable, but not directly responsible, for the actions of the AWS. Himmelreich, therefore, presents two outcomes that may arise from the deployment of this AWS: Outcome A, which states that this particular target is bombed; and Outcome B, which states that some target is bombed or no target is bombed.[163] He explains that outcome A represents a possible world where the AWS is deployed and the

[157] Himmelreich, 2019: 738.
[158] Ibid, 738.
[159] Ibid, 738.
[160] Ibid, 738.
[161] Ibid, 738.
[162] Ibid, 738.
[163] Ibid, 738.

particular target is bombed; this represents the particular outcome.[164] Outcome B, on the other hand, is a probabilistic outcome, which represents a possible world where the AWS is deployed and some target is bombed.[165] In this world, if the commander were to give the order, then some targets might be bombed; if the commander were not to give the order, no targets would be bombed.[166] Himmelreich argues that, within this context and with regards to the condition that responsibility requires control, outcome B would render the commander responsible.[167] Since the commander has a certain level of control in deploying the AWS, knowing that there is the possibility that a target other than the particular target may be bombed, that commander is responsible for the actions of the AWS because if the order had not been given, no bombings would occur at all. The decision to deploy the weapon, therefore, constitutes control and ascribes responsibility to the commander.

The decision to deploy an AWS into the battlefield will ascribe responsibility for the actions after deployment to the military commander that ordered it. Sparrow argues that by making the decision to send the AWS into the battlefield, the commander is accepting the risk that it might go awry.[168] In other words, if is the case that the autonomy of the AWS rests on the fact that its actions are not always reliably predictable and thus may cause unwanted deaths, then the commander who deployed the weapon is held responsible for those unwanted deaths as they understood that the weapon had the potential to kill people other than the intended target.[169] However, he acknowledges that AWS have the capacity to choose their own targets, and therefore with greater autonomy becomes less confidence in their reliability to attack the

[164] Ibid, 739.
[165] Ibid, 739.
[166] Ibid, 739.
[167] Ibid, 739.
[168] Sparrow, 2007: 70.
[169] Ibid, 70.

intended target.[170] Thus, using the unpredictability of a system as the only aspect of autonomy is incomplete; the use of AWS involves a risk that military personnel will be half responsible for, even though they did not control its decisions, as autonomy refers to the machine's ability to determine its own actions.[171] In other words, the military personnel would be responsible for acknowledging and accepting the risk of deploying the AWS, as their actions can be unpredictable, but are not directly responsible for their actions, as AWS have the capability to determine their own actions. In this relationship, it is therefore necessary that there is a sort of cooperation between the commander and the AWS, as the commander has the cognitive understanding of the machine's capabilities and monitor its progress towards the military objective, similarly to a "human-like" sense of teamwork.[172] Sparrow, therefore, accepts the analogy of child soldiers (which was explored earlier) to better understand how to determine responsibility. As with child soldiers, the possible solution to the responsibility gap identified in this relationship is assigning responsibility to the military commander who issued the attack order, but acknowledges that this solution may hold the commander responsible for things out of their control and therefore leaves open the possibility that they be punished unfairly.[173] Under this analogy, a military commander could be held responsible for all actions of the AWS after its deployment, which could ascribe more blame and fault than reasonable to the military commander, as the actions of the AWS are not directly controlled by the commander but rather merely influenced by the order to attack.

---

[170] Ibid, 70.
[171] Ibid, 71.
[172] Amitai Etzioni, and Oren Etzioni, "Pros and Cons of Autonomous Weapons Systems," *Military Review* (2017): 74.
[173] Sparrow, 2007:74.

Continuing with the analogy of the superior-subordinate relationship between soldiers and their commanders, the relationship between AWS and their commanders can be analyzed in a similar way. Jack McDonald, author of "Autonomous Agents and Command Responsibility," argues that military organizations have a "top-down" compliance structure.[174] This means that this structural compliance places constraints on individual autonomy and enables them to act based on information passed on to them by others who are higher ranked.[175] In terms of AWS, this means that the actions that the weapons take are directly influenced and partially determined by the commanders who set the parameters of their military operation. Aiden Warren and Alek Hillas, authors of "Friend or frenemy? The role of trust in human-machine teaming and lethal autonomous weapons systems," follow this line of reasoning by arguing that the successful adoption of AWS will depend on the direction of military commanders and will be reliant on effective human-machine teaming to be reliable and perform correctly.[176] They argue that this relationship between humans and machines is necessary for the proper employment of AWS, and therefore they are interconnected in terms of the actions that occur.[177] In other words, these humans will be held responsible for the actions of the AWS they operate because they are directly in charge of ensuring the proper and reliable operation of the AWS before deployment and during its mission. However, Sharkey argues that this human commander must have "full contextual and situational awareness of the target area at the time of a specific attack and be able to perceive and react to any change or anticipated situations that may have arisen since planning

[174] Jack McDonald, "Autonomous Agents and Command Responsibility," in *Routledge Handbook of War, Law and Technology*, ed. James Gow et al. (London and New York: Routledge Taylor & Francis Group, 2019): 143.
[175] Ibid, 143.
[176] Aiden Warren and Alek Hillas, "Friend or frenemy? The role of trust in human-machine teaming and lethal autonomous weapons systems," *Small Wars & Insurgencies*, 31:4 (2020): 823.
[177] Ibid, 825.

the attack."[178] In other words, the commander must be entirely aware of the circumstances of the operation during the attack to the extent that they are able to intervene if need be. Therefore, the military commander must always be monitoring the AWS and have the capability to intervene during its mission in order to effectively be responsible for its actions. Kalmanovitz continues this idea by arguing that military commanders must be confident that they have taken all reasonable steps to adhere to the principles of distinction and proportionality, for which they have a legal duty to do.[179] He argues that if, for example, a commander is unable to anticipate the range of action and corresponding risk with sufficient confidence, then it would be wrong to field the weapon and possibly a case of criminal negligence, and therefore the commander who fielded the weapon would be held criminally responsible for any mistakes that the AWS makes.[180] Military commanders are accountable for the deployment of the AWS as it is a deliberate decision, and therefore any actions taken by the AWS after deployment directly stem from that decision.[181] Military commanders can therefore be considered linked to the AWS that they deploy, making them responsible for their actions under their control and supervision.

Autonomy plays a large role in defining the superior-subordinate relationship in terms of attributing responsibility. Corn elaborates on this relationship by linking command, LOAC compliance and lawful combatant status together, explaining that under the law only individuals who are capable of autonomous reasoning that are incorporated into the military organization capable of managing that reasoning should be allowed to engage in hostilities.[182] He explains that the law establishes a high degree of confidence that an "autonomous human" will not use the

[178] Noel Sharkey, "Staying in the loop: human supervisory control of weapons," in *Autonomous Weapons Systems: Law, Ethics, Policy*, ed. Nehal C. Bhuta et al. (Cambridge: Cambridge, 2016): 35.
[179] Kalmanovitz, 2016: 155-156.
[180] Ibid, 156.
[181] Ibid, 156.
[182] Corn, 2016: 222.

power entrusted in them in an unconstrained manner, but rather exercise that autonomy within the boundaries that their superior imposed on them to ensure legal compliance.[183] In terms of AWS, those constraints are pre-determined for them through their programming, and therefore commanders are able to more confidently ensure legal compliance as these constraints are more strictly binding to machines than human soldiers. However, this superior-subordinate relationship obviously is not entirely analogous for AWS. For the soldier, training prior to being put into the battlefield lays a foundation for the ongoing process of framing the exercise of cognitive reasoning and independent judgement, and the military commander builds upon this foundation by exercising their responsibility to develop the soldier through continued training.[184] Therefore, the development of an AWS becomes more crucial in determining its behaviour on the battlefield, and creates an analogous counterpart to this training stage. However, Corn argues that military commanders will not have a meaningful opportunity to influence the judgement and reasoning of truly autonomous weapons, but rather deploy the weapons when a situation allows that the capabilities of the particular AWS will produce the desired result.[185] Therefore, the military commander the military commander will have to have faith that the weapon has the capacity to exercise the necessary cognitive judgement to comply with the LOAC.[186] Thus, this initial development phase is the decisive point in establishing parameters to ensure that the cognitive functions of the AWS only are exercised within these parameters in order to comply with the LOAC, as well as the military objectives and interests of the force.[187] In other words, the military commander must be able to work within the parameters of the AWS's capabilities and

[183] Ibid, 222.
[184] Ibid, 222.
[185] Ibid, 223.
[186] Ibid, 224.
[187] Ibid, 224.

deploy the weapon when they are confident that the circumstances under which it is deployed will allow it to comply with the LOAC and produce the intended military objective. They are, therefore, responsible for the actions of the AWS even though they are not directly involved in determining its capabilities and judgement.

Command responsibility can be explained through two legal theories of criminal liability: traditional accomplice liability and "should-have-known" command responsibility theory.[188] Traditional accomplice liability is attributed when a commander shares the criminal intent and acts in a way that contributes to or facilitates a violation of the LOAC through a subordinate.[189] In other words, the commander knowingly deployed the AWS into a situation it was not adequately equipped to operate within with the intent to cause undue harm to civilians, and therefore would be held responsible for that harm. The "should-have-known" theory attributes liability to a commander for foreseeable LOAC violations that are committed by their subordinates, even when there is no proof that the commander shared the intent with the subordinate, because the commanders are responsible for violations that they "should have known" would occur.[190] For example, the commander deploys an AWS in a situation that it was not adequately equipped to operate within, but did not intend for undue harm to happen, but should have known that this undue harm was a possibility in the circumstances under which the AWS was deployed. Not all technology is neutral, and some have properties that make some tasks easier to do than others; for example, some AWS may have constraints that make it easier for them to operate in some environments over others.[191] Therefore, even though the commander

[188] Ibid, 233.
[189] Ibid, 233.
[190] Ibid, 233.
[191] Malachy Eaton, "Lethal Autonomous Weapons Systems," In *Computers, People, and Thought: From Data Mining to Evolutionary Robotics* (Limerick: Springer, 2020): 194.

did not intend to cause the harm, they are still responsible for the actions that occurred as they "should have known" the AWS was unable to operate accurately in these environments and should have foreseen the possible LOAC violations. This theory provides that it is the responsibility of the military commanders to ensure that any AWS that is deployed is reliable and performs its capabilities in the ways expected of it.[192] Corn sums up the idea of command responsibility by stating that the concept of "mission command" is central to the planning and execution of military operations, and is premised on the expectation that subordinates advance the commander's intent.[193] In other words, the orders of the military commander directly dictate the subordinates' actions and therefore they follow through with the commander's intent to perform actions in order to achieve the military objective. For AWS, this means that the military commander's intent is advanced by the AWS as the commander gave the AWS the military objectives knowing its capabilities, and the AWS performed actions in order to follow through with those objectives.

Command responsibility can be understood as direct or indirect responsibility. Jain explains that command responsibility is combination of direct and indirect responsibility in that the commander is held directly responsible for their own failure to supervise or interevent, and indirectly responsible for the criminal acts of their subordinates.[194] She explains that the doctrine includes three common characteristics: the existence of a superior-subordinate relationship exists, there is a requisite mental element in that the superior knew or had reason to know of their subordinates' crimes, and that superior had failed to control, prevent or punish the subordinates for those crimes.[195] In other words, the commander should have had effective control over the

[192] Ibid, 233.
[193] Ibid, 239.
[194] Jain, 2016: 310.
[195] Ibid, 310.

subordinate at the time of the act, and that commander should have had the ability to prevent that act.[196] In terms of AWS, this doctrine would require that the military commander had control over the AWS, through deployment and monitoring, during the criminal act, and had the ability to intervene or abort the mission to prevent the act from happening. Regarding the mental element, Article 28 of the Rome Statute states that a "should have known" or negligence standard suffices for criminal liability.[197] Article 28(1) also requires a causal connection between the crimes committed by subordinates and the superior's culpability, and therefore the superior's omission could have facilitated or encouraged the crimes, or increased the risk of crime.[198] In other words, if a commander deployed an AWS that was unable to appropriately function within the parameters and environment in which it was deployed, the commander could be held liable as they acted negligently. The duty to prevent exists when an offence is going to, or is about to, occur and could materialize due to a commander failing to account for factors in their subordinates that could give rise to the crime being committed.[199] For AWS, this means that the military commander must be aware of any offence that may occur due to the design of the AWS, and by failing to account for the limitations in its d esign that caused the offence to occur, they are held responsible for that offence. Therefore, a military commander must understand, acknowledge and take into account the limitations of an AWS before deploying it to ensure that it is limiting the risk of potential criminal activity, and is responsible for the actions of that AWS that may occur due to these limitations.

There are, however, some issues that arise when attributing responsibility to the commander of an AWS. Liu argues that because the commander acts at a later stage than the

[196] Ibid, 311.
[197] Ibid, 311.
[198] Ibid, 311.
[199] Ibid, 312.

developer, the commander's ability to set constraints for the AWS is limited by and contingent upon the constraints already implemented by the developer, and therefore this narrows the commander's control over the AWS and limits their predictability over the system's behaviour.[200] Additionally, the replacement of a direct human operator of a weapons system by an artificial counterpart could disrupt the superior-subordinate relationship required, because the relationship has historically been an interpersonal one.[201] Because of this, Liu argues that it would be impossible to use the doctrine of command responsibility for an AWS due to a lack of superior-subordinate relationship.[202] Liu also explores the requirement of "effective control," arguing that the powers that the commander has to influence, suppress or prevent behaviours of an AWS may be severely limited due to the technical parameters of the system, be contingent on the technical knowledge and capabilities of the commander, and be impractical due to the inability to meaningfully punish a machine.[203] These considerations present obstacles in presenting effective control over an AWS, and therefore cause issues in attributing responsibility to the commander in charge of the AWS.[204] Therefore, the relationship between AWS and their commanders becomes more distant in that the commander's ability to effectively control and determine the AWS's actions is limited by the design already pre-determined by the developer, and thus the responsibility would be shared between the two in terms of being in control of the AWS's behaviours and actions.

CONCLUSION

[200] Liu, 2016: 331-332.
[201] Ibid, 332.
[202] Ibid, 333.
[203] Ibid, 333.
[204] Ibid, 333.

With the rapid advancement of technology and the introduction of AWS into the battlefield, it has become increasingly more important to explore how these robots will fit within the international legal landscape. The current legal landscape, though not explicitly including AWS within their rules and regulations, provides clear guidelines with which to contextualize how AWS fit; through exploring their capabilities and matching them with principles of IHL and LOAC, we are able to identify how AWS are able to comply with these rules and consequently be deemed to have broken the law. For example, through the use of sensors, AWS are able to adhere to the principle of distinction by being able to identify between targets and non-targets. However, attributing responsibility and liability present a separate issue, as the requirements for each are rooted in common sense and reasonableness, which are coded as inherently human characteristics that require more abstract judgement that AWS are incapable of processing. Therefore, we explore vicarious liability that can be attributed to the AWS's developers and their military commanders. Developers can be held responsible for the actions of AWS due to their role in programming the weapons capabilities and parameters. Although the role of the developer becomes more distant and diminished as the AWS operates in a more autonomous fashion, the capabilities of the AWS are still pre-determined by the software and programming that they were designed with, and therefore the developers are responsible for the actions made due to the information that was programmed into the system. Military commanders can be held responsible for the actions of AWS because they made the decision to deploy the weapon and determine its military objectives. Though they are working within the confines and constraints of the programming of the AWS, military commanders are still held responsible for the actions of the AWS as they made the orders to the machine knowing its capabilities and limitations, predicting the actions that could be taken to complete the military objectives, and therefore be held

responsible for those actions. However, one cannot be responsible and the other not; since both the developers and commanders are simultaneously involved with the behaviours and actions of the AWS, they can be held jointly responsible for the actions of the AWS. Both play a role in determining how the AWS will act and react in any given circumstance, and therefore both play a role in the final effects of the AWS. Though the autonomous nature of AWS provides the weapon itself with the ability to make its own decisions and take its own actions, it is ultimately the influence of the developers and commanders that dictate those actions and, consequently, their effects.

BIBLIOGRAPHY

Akerson, David. "The Illegality of Offensive Lethal Autonomy." In *International Humanitarian Law and the Changing Technology of War vol. 41*. Edited by Dan Saxon (Leiden, Brill: 2013).

Anderson, Kenneth, and Matthew Waxman. "Law and ethics for autonomous weapon systems: why a ban won't work and how the laws of war can." *Hoover Institution, Stanford University* (2013).

Blackstrom, Alan and Ian Henderson. "New capabilities in warfare: an overview of contemporary technological developments and the associated legal and engineering issues in Article 36 weapons reviews." *International Review of the Red Cross* 94, no. 886 (2012).

Bode, Ingvild, and Hendrik Huelss. "Autonomous weapons systems and changing norms in international relations." *Review of International Studies*, Vol. 44 (2018).

Boothby, Bill. "Autonomous Attack - Opportunity or Spectre?" In *Yearbook of International Humanitarian Law Volume 16*, ed. Terry D. Gill. (Cambridge: Springer) (2013)

Boothby, William. "Some legal challenges posed by remote attack." *International Review of the Red Cross 94*, no. 886 (2012).

Boothby, Bill. "Weapons Law, Weapon Reviews and New Technologies." In Routledge Handbook of War, Law and Technology. Edited by James Gow, Ernst Dijxhoorn, Rachel Kerr and Guglielmo Verdirame. (London and New York: Routledge Taylor & Francis Group, 2019).

Corn, Geoffrey S.. "Autonomous weapons systems: managing the inevitability of 'taking the man out of the loop.'" In *Autonomous Weapons Systems: Law, Ethics, Policy*. Edited by Nehal C. Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Claus Kreß. (Cambridge: Cambridge, 2016).

Eaton, Malachy. "Lethal Autonomous Weapons Systems." In Computers, People, and Thought: From Data Mining to Evolutionary Robotics (Limerick: Springer, 2020)

Etzioni, Amitai, and Oren Etzioni. "Pros and Cons of Autonomous Weapons Systems." *Military Review* (2017)

Foy, James. "Autonomous Weapons Systems: Taking the Human out of International Humanitarian Law." *Dalhousie Journal of Legal Studies 23* (2014)

Garcia, Denise. "Killer Robots: Why the US should Lead the Ban." *Global Policy Volume 6*, Issue 1 (2015).

Gillespie, Tony. "Humanity and Lethal Robots: An engineering perspective." In Routledge Handbook of War, Law and Technology. Edited by James Gow, Ernst Dijxhoorn, Rachel Kerr and Guglielmo Verdirame. (London and New York: Routledge Taylor & Francis Group, 2019).

Himmelreich, Johannes. "Responsibility for Killer Robots." *Ethical Theory and Moral Practice* 22 (2019).

Jain, Neha. "Autonomous weapons systems: new frameworks for individual responsibility." In *Autonomous Weapons Systems: Law, Ethics, Policy*. Edited by Nehal C. Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Claus Kreß. (Cambridge: Cambridge, 2016).

Kalmanovitz, Pablo. "Judgement, liability and the risks of riskless warfare." In *Autonomous Weapons Systems: Law, Ethics, Policy*. Edited by Nehal C. Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Claus Kreß. (Cambridge: Cambridge, 2016).

Liu, Hin-Yan. "Categorization and legality of autonomous and remote weapons systems." *International Review of the Red Cross* 94, no. 886 (2012).

Liu, Hin-Yan. "Refining responsibility: differentiating two types of responsibility issues raised by autonomous weapons systems." In *Autonomous Weapons Systems: Law, Ethics, Policy*. Edited by Nehal C. Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Claus Kreß. (Cambridge: Cambridge, 2016).

Marchant, Gary E., Braden Allenby, Ronald Arkin, and Edward T. Barrett. "International Governance of Autonomous Military Robots." *Columbia Science and Technology Law Review* 12 (2011).

Matthias, Andreas. "The responsibility gap: Ascribing responsibility for the actions of learning automata." *Ethics and Information Technology* 6 (2004).

McDonald, Jack. "Autonomous Agents and Command Responsibility." In *Routledge Handbook of War, Law and Technology*. Edited by James Gow, Ernst Dijxhoorn, Rachel Kerr and Guglielmo Verdirame. (London and New York: Routledge Taylor & Francis Group, 2019).

McFarland, Tim, and Tim McCormack. "Mind the Gap: Can Developers of Autonomous Weapons Systems Be Liable for War Crimes." *International Law Studies: U.S. Naval War College* 90 (2014).

Nanayakkara, Thrishantha. "Autonomy of Humans and Robots." In *Routledge Handbook of War, Law and Technology*. Edited by James Gow, Ernst Dijxhoorn, Rachel Kerr and Guglielmo Verdirame. (London and New York: Routledge Taylor & Francis Group, 2019).

Roff, Heather M.. "Killing In War: Responsibility, liability, and lethal autonomous robots." In: *Routlege Handbook of Ethics and War: Just War Theory in the 21st Century*. Edited by Fritz Allhoff, Nicholas G. Evans, Adam Henschke (New York & London: Routlege, 2013).

Sartor, Giovanni and Andrea Omicini. "The autonomy of technological systems and responsibilities for their use." In *Autonomous Weapons Systems: Law, Ethics, Policy*. Edited by Nehal C. Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Claus Kreß. (Cambridge: Cambridge, 2016).

Schmitt, Michael N. and Jeffrey S. Thurnher. "Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict." *Harvard National Security Journal* 4, no. 2 (2013)

Schmitt, Michael. "Autonomous weapon systems and international humanitarian law: a reply to the critics." *Harvard National Security Journal*, 4 (2013).

Sharkey, Noel E.. "The evitability of autonomous robot warfare." *International Review of the Red Cross 94*, no. 886 (2012).

Sharkey, Noel. "Staying in the loop: human supervisory control of weapons." In *Autonomous Weapons Systems: Law, Ethics, Policy*. Edited by Nehal C. Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Claus Kreß. (Cambridge: Cambridge, 2016).

Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy*, Vol. 24, No. 1 (2007).

Wagner, Markus. "Autonomy in the Battlespace: Independently Operating Weapon Systems and the Law of Armed Conflict." In *International Humanitarian Law and the Changing Technology of War vol. 41*. Edited by Dan Saxon (Leiden, Brill: 2013).

Wagner, Markus. "Taking Humans out of the Loop: Implications for International Humanitarian Law." *Journal of Law, Information and Science* 21, no. 2 (2011/2012).

Warren, Aiden and Alek Hillas. "Friend or frenemy? The role of trust in human-machine teaming and lethal autonomous weapons systems." *Small Wars & Insurgencies*, 31:4 (2020).