The application of command responsibility to Lethal Autonomous Weapon Systems (LAWS) – who is responsible?

Thesis Submitted in partial fulfillment for the degree of:

Master of Laws in International Operational Law

Author: Emma Gonzales-Puell

Supervisor: Jann K. Kleffner

Word Count: 12.570 words (ex. Bibliography)
Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor Jann K. Kleffner for his invaluable support and guidance during the writing process. His knowledge and expertise allowed me to write a piece I can be proud of.

I am grateful for the Swedish Defence University and Nobuo Hayashi for tailoring such an interesting program in International Operational Law and for providing us with unique support and materials during our studies.

I am also thankful to my friends and family who have given me support in the most desperate times. I am especially thankful for Cornelia who has served as sounding board during my legal analyses. And I am thankful for my parents without whom I wouldn’t have had the opportunity to study in Stockholm.

I would also like to give special thanks to Nobuo Hayashi, Geoffrey Corn and Stuart Maslen who have given me the confidence to write on a topic on which much has not be written before.
# TABLE OF CONTENTS

1. **Introduction and factual background** ................................................. 4  
   1.1 Decision making by Lethal Autonomous Weapons ........................... 7  
      1.1.1 The OODA Loop ................................................................ 8  
      1.1.1.1 Human-in-the-loop ...................................................... 9  
      1.1.1.2 Human-on-the-Loop ................................................... 9  
      1.1.1.3 Human-out-of-the-loop .............................................. 11  
2. **Defining Lethal Autonomous Weapon Systems** .................................. 12  
   2.1 Intricacies of state’s definitions of Lethal Autonomous Weapons ........ 12  
      2.1.1 Defining an Autonomous Weapon System in Relation to Human Interaction 13  
      2.1.2 Defining an Autonomous Weapon System in Relation to Technology 15  
      2.1.3 Defining Autonomous Weapon Systems in Relation to the Task ......... 16  
   2.2 How to define Lethal Autonomous Weapon Systems? ......................... 18  
   2.3 The legal status of Lethal Autonomous Weapon Systems .................. 18  
3. **Command Responsibility** .................................................................. 21  
   3.1 The doctrine of command responsibility ......................................... 22  
      3.1.1 Superior/Subordinate relationship ....................................... 24  
      3.1.2 The knowledge requirement ............................................... 25  
      3.1.3 Failure to take measures .................................................... 26  
      3.1.4 Causation ..................................................................... 26  
   3.2 The mental element in laws ................................................................. 27  
   3.3 Command Responsibility and Malfunctions, accidents and system errors of LAWS .............................................................. 28  
      3.3.1 Should have known: Command responsibility and accidents ........ 30  
      3.3.2 Should have known: Command responsibility and errors/malfunctions 31  
   3.4 Ethical Concerns ........................................................................... 32  
4. **Other sources of responsibility** .............................................................. 34  
   4.1 Responsibility of manufacturers, designers, engineers and programmers .......................................................... 34  
   4.2 State responsibility .......................................................................... 38  
      4.2.1 Legal review of new weapons under IHL .................................. 40  
      4.2.2 Challenges of LAWS weapons review .................................... 43  
Bibliography ................................................................................................. 47  
   Books: ................................................................................................. 47  
   Journal Articles: ................................................................................... 47  
   Government Publications: ................................................................. 48  
   Websites: ............................................................................................ 49  
   Reports: ............................................................................................... 50  
   Cases: ................................................................................................. 50
1. INTRODUCTION AND FACTUAL BACKGROUND

The development of advanced artificial intelligence (AI) and its use in the military domain has brought new discussions on the use of weapons using AI on the battlefield. Countries have been working towards weapon systems which, through AI, would be able to act autonomously – meaning without human intervention. While some welcome the use of weapons being able to select and engage military targets autonomously as a safer, faster and more efficient manner\(^1\) to conduct war, others find that such systems which are able to make their own decisions could create legal issues of their own as they lead to the ‘dehumanization of war’. Some authors view LAWS favorably, while they acknowledge the inherent limitations of these systems, as they may lack the necessary discernment and nuanced human judgement that complex battlefields demand.\(^2\)

Thus far, there are no weapon systems which are entirely autonomous. Currently there are autonomous systems such as cruise missiles, torpedoes, submersibles, robots for urban reconnaissance, Uninhabited Aerial Vehicles (UAVs) and Uninhabited Combat Aerial Vehicles (UCAVs).\(^3\) These weapon systems are however subject to some level of human control. The most widely used robots in the military sphere are UAVs, these are remotely piloted drones which are able to fly, do reconnaissance and come back to base with minimal human intervention. They are nonetheless not exempted from human intervention, or human control.

---


These types of weapons are referenced to as ‘fire and forget’ system. Indeed, these weapon systems are engaged by a human operator, the system is then able to determine its trajectory and follow its target for a limited time. They are nonetheless engaged by a human controller which does not make the weapon fully autonomous.

Much of the debate surrounding Lethal Autonomous Weapons revolves around the definition that the United Kingdom (UK) and the United States (US) have given of autonomous weapon systems. The US defines it as “weapon systems that operate without human intervention”. The UK defines it as “weapon systems that operate without human control”. These two varying definition have given rise to the much heated debate of what are the factors that are to be taken into account when defining autonomous weapon systems.

The increasing debates surrounding what constitute an autonomous weapons system is of great importance, however most troubling of all debates is the responsibility for autonomous weapon systems. Such weapon systems which are entirely autonomous would be able to gather and evaluate data and to autonomously engage their targets without human control/intervention. Therefore, it is important to bridge the gap of responsibility for violations of international humanitarian law (IHL) which might ensue from the use of such systems. After all, international humanitarian law is made by humans, to protect humans, from humans so how far can it apply to machines? Does it at all?

---

4 United States, Department of Defence Directive 3000.09 (2012)
5 United Kingdom, Joint Doctrine Publication 0-30.2 Unmanned Aircraft Systems (2017)
In this debate there are Autonomous Weapon Systems (AWS), and there are Lethal Autonomous Weapon Systems (LAWS). The next chapter of the thesis will show that the debate surrounding LAWS and AWS is very similar, LAWS are in the end AWS with lethal capabilities.\(^6\)

The question thus arises, who is responsible for violation of IHL by LAWS? The first one is the programmer, since the weapon must be designed not to act in violation of IHL the programmer should bear the responsibility for violations. The second one is the commander, since the commander has ordered the deployment and use of the weapon, he should bear responsibility for violations of the law. The third one is the machine itself; it was programmed to conduct a task and it has failed to do so. The last one is the state, since the state is responsible for ensuring that the weapon systems which are used on the battlefield comply with IHL.

However, there are various problems which ensue from each of these propositions. The programmers have designed the weapon but limitations to any weapons is inherent to them, and these limitations shall be communicated to the state acquiring them. The problem which arises from commanders being held responsible is that they would be responsible for the actions of autonomous weapons, which they are not able to control. The machine itself is not a moral agent, and it would be hard to conceive a justice system in which a machine would be held responsible for the commission of war crimes.

The focus of this thesis is the following question: is command responsibility a viable basis for responsibility in case of violation of IHL by LAWS? If not, what other basis can one rely on? In order to construct a clear argument, the thesis will first define LAWS and provide main differences of the debates surrounding the difficulties in defining LAWS. The second part will focus on command responsibility and its limitations to ascribe responsibility in case of mistakes,

\(^6\) This will be further developed in the next chapter.
errors and malfunctions. Finally, the third part will focus on other sources of responsibility, in which individual criminal responsibility of developers will be discussed as well as the responsibility of states. Prior to delving into these debates, it is important to understand the decision-making process of LAWS.

1.1 DECISION MAKING BY LETHAL AUTONOMOUS WEAPONS

Decision making by Lethal Autonomous Weapon Systems (LAWS) is complex, these systems are designed to collect data from their environment using sensors (cameras, radar or other type of sensors). The data collected by the sensors is then processed by the system’s algorithm to identify potential targets and determine the appropriate course of action. The algorithms and machine learning models that are used by the LAWS are trained to identify and engage specific types of targets. LAWS uses the information it has acquired to determine whether the target meets the criteria for engagement – factors such as size, location, movement and shape are taken into account. Moreover, the models can be designed to operate within specific parameters established by the human operator – e.g., Rules of Engagement.

LAWS have therefore the technological capacity of working independently without any human oversight. Indeed, LAWS are designed to make decisions quickly and efficiently, based on their programming and the information which they have gathered from their sensors. Nonetheless, LAWS decision-making may not be able to process the complexities of real-world situation, such as the presence of civilians or a dirty emblem. It is therefore important to ensure that there is some level of human oversight at some level of the decision-making process.
1.1.1 THE OODA LOOP

Engineers and roboticists have embraced the OODA Loop mode to analyze the decision-making abilities of machines. The model was first introduced by U.S Air Force Colonel John Boyd. OODA stands for “observe, orient, decide, and act” and the loop refer to the continuous cycle of decision-making that occurs during an operation.

The first step in the OODA loop is to observe the situation or environment in which an operation is taking place. This involves gathering information about the conditions, potential threats and other factors that may affect the success of the operations. Once observations have been made, the next step is to orient oneself to the situation by analyzing and synthesizing the information gathered during the observation phase. This involves assessing the situation, identifying potential courses of action and anticipating likely outcomes. The third step is to make a decision based on the observations and orientation. This involves selecting a course of action that is likely to achieve the desired outcome while minimizing risk and uncertainty. Finally, the last step in the OODA loop is to act on the decision made in the previous step. This involves executing the chosen course of action, monitoring the results, and adjusting plans as necessary based on ongoing observation and orientation.

In order to determine the extent of human intervention that is necessary in that loop, there are three models which are currently being used. The models of human oversight are the following: human-in-the-loop, human-on-the-loop, human-out-of-the-loop.

---


8 “The Ooda Loop” (*The Decision Lab*) [https://thedecisionlab.com/reference-guide/computer-science/the-ooda-loop]
1.1.1.1 Human-in-the-Loop

System control exists on a continuum ranging from manual control to full system autonomy. A control mode in which humans retain selected key functions and make all or most decisions is referred to as in-the-loop control.

Human operators are an integral part of the system’s control loop.\(^9\)

Human in the loop refers to a system where the human operator is involved in the decision-making process. The system thus requires human intervention and oversight before making decisions. The human would be reviewing and approving certain decision made by the system, or it could take the form of presenting multiple options to the human operator who then selects the most appropriate one. The Human-in-the-loop model ensure that the weapon system is used only in situations where a human operator has determined that the use of lethal force is necessary and proportionate. This allows for human control to be kept over the weapon system and thus preventing possible illegal outcomes. In essence, the human-in-the-loop model means that a human operator must instigate the action of the weapon, the weapon is therefore not autonomous, but may present characteristics of automated weapon systems.

1.1.1.2 Human-on-the-Loop

Another term for supervisory control. The operator sets goals, monitors system actions, and intervenes when necessary.\(^10\)

Human on the loop model is where a human operator is involved in the decision-making process and has the ability to intervene and take control of the system if he/she deems it necessary.


\(^10\) Ibid.
These systems are able to identify, select and engage targets without human intervention. This model is a step forward autonomy of the weapon system as the human operator will not be actively controlling the system, but rather overseeing and monitoring the system. This means that while the weapon system is autonomous and may engage in lethal force without a human operator deciding to do so, the human operator can override the system’s decision.

These types of systems are referred to as “human supervised autonomous weapon systems” by DoD 3000.09 and included in the category of automated systems according to JDN 2/11 and JDP o-30.2. Examples of such ‘human supervised autonomous weapon systems’ include: AEGIS Combat System: American integrated naval weapons system. Goalkeeper Close-in Weapon System (CIWS): Dutch close-in weapon system. MK Phalanx Close-in Weapon System (CIWS): American made gun-based close-in weapon system. Patriot

---


12 Ibid.


14 Used by the United States Navy, Japan Maritime Self-Defense Force, Spanish Navy, Royal Norwegian Navy, Republic of Korea Navy and Royal Australian Army


16 Used by Belgian Navy, Chilean Navy, Royal Netherlands Navy, Peruvian Navy, Portuguese Navy, Qatar Armed Forces, Republic of South Korea Navy.


18 used by Australia, Bahrain, Canada, Chile, Greece, Ecuador, Egypt, India, Israel, Mexico, New Zealand, Pakistan, Poland, Portugal, Saudi Arabia, South Korea, Thailand, Turkey, Taiwan, United Kingdom and the United States.
**Missile**\(^{19}\): American made surface-to-air missile\(^{20}\). **SeaRam**\(^{21}\) Anti-Ship Missile Defence System: Developed by General Dynamics under agreement with Denmark and West Germany, Denmark dropped out, but The United States joined.\(^{22}\)

As can be seen from the above-mentioned systems, human-on-the-loop weapon systems are widely used by a plethora of states for air-defense. The systems are able to autonomously engage incoming threats, unless overridden by the human operator.

1.1.1.3 **HUMAN-OUT-OF-THE-LOOP**

“Robots that are capable of selecting targets and delivering force without any human input or interaction”\(^{23}\)

Human out of the loop refers to the model of human involvement in operating lethal autonomous weapon systems where the system operates without any human intervention or control. This means that the decision to engage a target and use lethal force is entirely automated without any human oversight or intervention. It is nonetheless important to note that the decision to use the weapon still lies in the human operator, but once the weapon has been chosen it would not need further human oversight. Essentially, the human-out-of-the-loop model means that the

---


\(^{22}\) Used by Egypt, Germany, Greece, Japan, Mexico, Qatar, South Korea, Saudi Arabia, Turkey, United Arab Emirates and the United States

human operator would be out of the decision-making process, but still hold some human control over the system.\textsuperscript{24}

2. DEFINING LETHAL AUTONOMOUS WEAPON SYSTEMS

Lethal Autonomous Weapon Systems are relatively new to the conversation, states have not yet been able to agree on a common definition. This is a major issue for the regulation and attribution of responsibility for violations of IHL by LAWS. This section aims to provide an overview of the definitional trends which have come up surrounding the debate on LAWS. The final part of the section will provide a definition which encompasses various attributes of states definition.

2.1 INTRICACIES OF STATE’S DEFINITIONS OF LETHAL AUTONOMOUS WEAPONS

In defining autonomous weapons, we can see three distinct trends of definition. All three trends put importance or define in relation to one aspect of the weapon system. While states struggle to find common grounds in defining autonomous weapon systems, some states\textsuperscript{25} have still showed support for a preemptive ban on autonomous weapon systems. In contradiction to this lack of consensus around a definition of AWS, there is the acceptance of all the states calling for a pre-emptive ban to accept that AWS are those weapon systems which can select and engage targets without meaningful human control. There are multiple definitions of autonomous weapon

\textsuperscript{24} Afonso Seixas-Nunes, The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective (Cambridge University Press 2022), 95

\textsuperscript{25} Since 2013, 30 countries have called for a ban on fully autonomous weapon systems or weapons systems that do not have meaningful human control. These include Algeria, Argentina, Austria, Bolivia, Brazil, Chile, China, Colombia, Costa Rica, Cuba, Djibouti, Ecuador, Egypt, El Salvador, Ghana, Guatemala, Holy See, Iraq, Jordan, Mexico, Morocco, Namibia, Nicaragua, Pakistan, Panama, Peru, State of Palestine, Uganda, Venezuela, and Zimbabwe.
systems that have been proposed and used by states as well as international actors, in fact, the National Institute of Health of the United States government has identified 12 definitions.26

States have thus far defined Lethal Autonomous Weapon Systems (LAWs) as an AWS having lethal capabilities. Therefore, in the following sections the definition of AWS will be discussed. Important to note that all LAWs are AWS, but not all AWS are LAWs, therefore the main difference between the two is the capability of LAWs to be lethal but it remains and AWS. 27 In essence, the definition proposed by states for AWS apply to all LAWs insofar as they are equipped with the necessary means to be able to lethal. At the end of the section a final working definition of LAWs will be provided and will be the reference for the remainder of the analysis.

2.1.1 Defining an Autonomous Weapon System in Relation to Human Interaction

The first definitional trend defines the machine (or the weapon system) in relation to the human. This is the case of the United States who defines an autonomous system as a “weapon system that operates without human control”28. It can be seen here that the weapon is not defined by stating its characteristics of what it is capable to do, but rather to what extend a human can interact or counteract the machine. Similarly, Human Rights Watch categorizes autonomous weapon in relation to the level of human involvement in the operation of the weapon.29 For Human Rights Watch the level of involvement depends on the whether the man is in the loop, on the loop or out

26 Mariarosaria Taddeo, Alexander Bachard, A Comparative Analysis of the Definition of Autonomous Weapons Systems. Sci ENg Ethics, 23 August 2022, 36
28 United States, Department of Defence Directive 3000.09 (2012)
of the loop. Central to defining an autonomous weapon system in relation to the human interaction is ensuring that humanitarian principles will be abided by.

The positives of defining an AWS in relation to the human is that governments, international organizations and/or legal scholars do not need to possess extensive knowledge of technology in order to be able to define them. Nonetheless, by putting such emphasis on the human, the definition lacks sufficient understanding of how the machine might change the relationship between humans, as introducing a non-human actor into the equation might lead to regular human-to-human interaction to change. Indeed, humans and machines are not able to carry out similar tasks as the same speed. For example, a machine is able to spot through various data sets what it believes to be a F-18 while the human would only be able to carry out that tasks if it sees the F-18. Of course, humans use machines in order to help them carry out the task more rapidly, but the decision to engage or not remains at the discretion of the human operator.

This can be seen from the failures of human-machine interaction of the automated system ‘Patriot’. The Air Defense Artillery fired a Patriot missile at what the system had mistakenly reported as a ballistic missile which was in fact an F-18 (which latter turned out to be a radar electromagnetic interference). The Tactical direction ordered the launcher to be in tactical ballistic missile automatic mode, which resulted in automatic engagement of the first available launcher – thus once a launcher was made available the F-18 was engaged.\(^\text{30}\) From this incident it transpires that it is not only essential to define the weapon system in relation to the human-machine interaction, but also to the human-human interaction. As a result of this incident, it was noted that there was a lack of understanding of the machine by the human therefore a lack of human-human communication on how the machine was to be interacted with.

2.1.2 Defining an Autonomous Weapon System in Relation to Technology

Some states prefer to define AWS in relation to their technological capabilities. This has typically been the approach of the GGE CCW (Group of Governmental Experts on Certain Conventional Weapons), who has defined in relation to “specific aspect of the weapon such as technical specifications, range, payload and intended operating environment”\(^{31}\). This is for example the case of the United Kingdom who defined an autonomous system as: “Capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be.”\(^{32}\)

This type of definition allows for the distinction of different degrees of autonomy. It distances itself from definitions of automatic systems and automated systems. However, while it tries to distance itself from other types of systems, it provides with a definition that could be applied to various weapon systems. Indeed, defining AWS as “capable of higher-level intent” provides an unusually high threshold for “autonomous” and therefore has the consequence of providing an umbrella definition for AWS rather than providing a specific definition which could limit AWS.


The definition refers to the cognitive abilities of the Artificial Intelligence of AWS but highlights that the AWS they refer to “do not currently exist”. Adding such a statement to the definition allows for future developments to encompassed by the definition. However, while this may seem like a positive outcome of the definition of the UK, it may have the effect of not focusing on current legal problems which are posed by AWS. It might also have the detrimental effect of not allowing current AWS to be defined under the current definition if they do not have the sufficient level of “understanding” and “higher-level intent”.

2.1.3 Defining Autonomous Weapon Systems in Relation to the Task

Yet other states find it more fitting to define AWS in relation to the tasks or functions that they are able to do autonomously. For example, the International Committee of the Red Cross (ICRC) has proposed that AWS are those “with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention.”34 The definition of the ICRC proposes an avenue in which the AWS would be defined in relation to the different tasks it is able to carry out, while ensuring that there is an element of the lack of human intervention – as seen in the definition in relation to human interaction. The proposed definition allows for the AWS definition to encompass both current AWS as well as future AWS developments.

In contrast to the high threshold proposed by the United Kingdom’s definition, the ICRC’s takes into consideration the fact that AWS developed post-UK definition could fall into a ‘grey area’ where they will not be able to be regulated correctly. In fact, it is behind this rationale that the ICRC proposed the above definition: the definition “encompasses some existing weapon systems, [and so] enable real-world consideration of weapons technology to assess what may make certain existing weapon systems acceptable – legally and ethically – and which emerging technology developments may raise concerns under international humanitarian law (IHL) and under the principles of humanity and the dictates of public conscience.”

Similarly, Switzerland has proposed the following definition of AWS: “weapon systems that are capable of carrying out tasks governed by international humanitarian law in partial or full replacement of a human in the use of force, notably in the targeting cycle” The definition proposed by Switzerland proposes an interesting avenue to consider when defining AWS. The definition takes upon elements of autonomy from pre-existings system rather than providing a brand-new definition specifically designed for AWS. This allows – similarly to the ICRC’s definition – to encompass a wider range of pre-existing AWS and future AWS in the definition. Nonetheless, Switzerland concedes, however, that its working definition “could and probably should evolve to become more specific and purposeful.”

35 Ibid.
2.2 How to define Lethal Autonomous Weapon Systems?

For the purpose of this thesis, a definition based on various elements discussed above will be presented. The aim of the proposed definition is to be able to encompass all the attractive elements of certain definitions while ensuring that the shortcomings of some of the definitions are dealt with. Furthermore, the aim will also be to prepare for the following chapter on command responsibility in which a definition of LAWS is essential to conclude whether command responsibility is the correct avenue to pursue or not.

A lethal autonomous weapon system (LAWs) is a type of military technology that uses artificial intelligence and machine learning to autonomously make decisions about selecting and engaging targets without the need for human intervention. These systems can include a wide range of weapons, from unmanned aerial vehicles to ground-based robots and naval vessels. A LAWS is able to make decisions about the use of lethal force without direct human control. This means that once a LAWS has been deployed it will choose its targets, select weapons and engage in combat without further human input.

2.3 The Legal Status of Lethal Autonomous Weapon Systems

Some scholars have argued that autonomous weapon systems blur the line between a combatant and a weapon. Therefore, making it difficult to understand which legal status they have under international law – in fact they have been referred to as ‘new agents’. Crootof and Hin-Yan Liu are authors who are raising the questions surrounding this new agent on the battlefield and the difficulties in evaluating these entities. Crootof has raised the issue that generally Lethal Autonomous Weapon Systems have been referred to as either weapons or combatant, but referring

---

to them as either of these is inexact.\textsuperscript{39} If referred to as weapons, they would be a weapon which is capable to take independent and [sometime] unpredictable action, which is not the case of other weapons who are fully operated by human actors and are therefore not capable of independent action. If they were to be referred to as combatants, they will lack human morality and motivation that a combatant has as they would be operated by algorithms.

This particular difficulty in understanding what legal status LAWS have is what makes regulation so difficult. For Crootof, weapons \textit{are} lawful or unlawful, and combatants \textit{act} lawfully or unlawfully.\textsuperscript{40} Therefore, the question that is central is: how can a lawful weapon who acts autonomously act unlawfully. This question is the one that should be studied in depth when conducting the weapons review under Art. 36 AP I. Yet, subjecting LAWS to weapons review may not be sufficient to ensure that they comply with IHL, this stems from the fact that IHL is based on human-decision making and therefore weapons are to be used in relation to human decision-making. Consequently, introducing an agent which does not act either as a [human] combatant or as a weapon blurs the pre-existing lines to ensure IHL compliance.

International Humanitarian Law has established regulation regarding the nature and manufacturing of weapons to ensure proposer use under IHL. According to Article 36 of AP 1, the state responsible for manufacturing a weapon is obligated to investigate its compliance with IHL. A weapon must therefore be able to distinguish between combatants and civilians to be considered legal\textsuperscript{41}. Article 51 (4)(b) and (c) prohibits the use of weapons that are unable to make


\textsuperscript{40} Ibid.

\textsuperscript{41} See, Legality of the Threat or Use of Nuclear Weapons Advisory Opinion, [1996] ICJ Rep 226
this distinction and those that cannot be directed at a specific target without controlled damage. It stems from this that the law governing weapons is concerned with the physical design and capabilities of the weapon.

The law regulating combatants requires that combatants distinguish between military objectives and civilians, and that they only attack military objectives. Combatants must also take precautions to avoid or minimize harms to civilians and civilian objects. They must therefore not use weapons or tactics that cause unnecessary suffering or whose effects are indiscriminate. Combatants who violate these rules may be subject to prosecution and punishment for war crimes or other serious violations of international law. The law regulating combatants therefore seeks to regulate their behavior through training and accountability measures.

The difficulty of assessing their legal status is in the fact that under the law of weaponry they would not be able to be classified as lawful or unlawful and they would not be able to classify as human combatants and therefore would not fall under the law of combatants. This poses a challenge to the traditional assumption underlying these legal frameworks, as LAWS do not fit into either category of dependent weapons or independent human beings.  

3. **Command Responsibility**

“Willingness to accept mortal risk in pursuit of important objectives is, of course, a core ethos of a professional military.”\(^{43}\) Military commands bear the hard task and responsibility to give orders in order to achieve these said important objectives, knowing the subordinates are entering in potentially life-threatening situations. The relationship between superior and subordinate is crucial to the adherence of the Law of Armed Conflict (LOAC), as it depends on the commander to provide LOAC training to subordinates and ensure a high level of compliance.\(^ {44}\) This relationship is also built on the disciple of subordinates and the respond for military command, which forms the foundation of military organizations. Violations of LOAC may occur without this relationship.

Under the principle of command responsibility, commander can be held liable for war crimes, crimes against humanity, or other serious violation of international law committed by their subordinates. The principle is based on the idea that commanders have a duty to exercise effective control over their subordinates, including taking measures to prevent and punish any violations of international law. It is also intended to ensure that those with the greatest power and authority are held accountable for the actions of their subordinates. The principle of command responsibility is an important component of international humanitarian law and is recognized by various international legal instruments and tribunals, including the Geneva Conventions and the International Criminal Court (ICC) – as well as in the domestic law of various countries.


The doctrine of command responsibility saw the light especially with the case of Yamashita, a commander of the Japanese forces in the Philippines during 1944 and 1945. Due to his inability to control the actions of the persons under his command, who has committed war crimes, he was found guilty of the murders committed by them. While Yamashita had not ordered the commission of the war crimes by his subordinates, but the court noted that his liability was not direct complicity but his failure to prevent them from committing the crimes, essentially not ensuring adequate control over his subordinates. The court finally noted that:

“Hence the law of war presupposes that its violation is to be avoided through the control of the operations of war by commanders who are to some extent responsible for their subordinates. This is recognized by the Annex to the Fourth Hague Convention of 1907, respecting the laws and customs of war on land. Article 1 lays down as a condition which an armed force must fulfill in order to be accorded the rights of lawful belligerents, that it must be ‘commanded by a person responsible for his subordinates.’”

3.1 THE DOCTRINE OF COMMAND RESPONSIBILITY

Command/superior responsibility is an inculpatory doctrine specific to ICL, which consists in a broad form of liability justified by the privileges, honours and responsibilities that command entails. We will focus on the criminal responsibility of a commander/superior for offences committed by his or her subordinates. As mentioned before the origin is the trial of the General Yamashita by an American military tribunal. Command responsibility was included in

\[45\] “The law of war imposes on an army commander a duty to take such appropriate measures as are within his power to control the troops under his command for the prevention of acts which are violations of the law of war and which are likely to attend the occupation of hostile territory by an uncontrolled soldiery; and he may be charged with personal responsibility for his failure to take such measures when violations result” In Supreme Court of the United States 327 U.S 1, Yamashita, 14-16

\[46\] Supreme Court of the United States 327 U.S 1, Yamashita, 15
military manuals after WWII, but made its first clear appearance in a treaty in Articles 86 and 87 of AP I, followed by Article 7(3) International Criminal Tribunal for the former Yugoslavia (ICTY) Statute and a more detailed version in Article 28 ICC Statute. The Trial Chamber in Celebici elaborated the requirements of command responsibility under customary law:

- A superior/subordinate relationship;
- The ‘mental element’;
- Failure to take reasonable measures to prevent or punish violations of ICL.

This trio has been adopted by the ad hoc Tribunals. To those, with which the ICC has concurred, the Rome Statute has added the requirement of causation. Under the ICC statute, article 28 provides that military commanders may be held criminally responsible for crimes committed by his or her subordinates:

“In addition to other grounds of criminal responsibility under this Statute for crimes within the jurisdiction of the Court:

(a) “A military commander or person effectively acting as a military commander shall be criminally responsible for crimes within the jurisdiction of the Court committed by forces under his or her effective command and control, or effective authority and control as the case may be, as a result of his or her failure to exercise control properly over such forces, where:

i. That military commander or person either knew or, owing to the circumstances at the time, should have known that the forces were committing or about to commit such crimes; and

ii. That military commander or person failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission or to submit the matter to the competent authorities for investigation and prosecution.

(b) With respect to superior and subordinate relationships not described in paragraph (a), a superior shall be criminally responsible for crimes within the jurisdiction of the Court committed by subordinates under his
or her effective authority and control, as a result of his or her failure to exercise control properly over such subordinates, where:

i. The superior either knew, or consciously disregarded information which clearly indicated, that the subordinates were committing or about to commit such crimes;

ii. The crimes concerned activities that were within the effective responsibility and control of the superior; and

iii. The superior failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission or to submit the matter to the competent authorities for investigation and prosecution.47

The subsequent analysis will be based on article 28.

3.1.1 SUPERIOR/SUBORDINATE RELATIONSHIP

There are the clear formal chains of command that characterize modern well-disciplined armies. However, modern conflicts are not always fought on this basis and by such forces. Therefore, the Appeals Chamber in Celebici based on a test of ‘effective control’ and the material ability to prevent or punish criminal conduct. The de jure position of the superior is not determinative, it is the largely factual ability that counts. A de jure position may be evidence of effective control, so does the ability to require people to engage or withdraw from hostilities. Issuance of orders may also be good evidence but, if they are not obeyed, this will count the other way (case-by-case basis, the fact that force is necessary to enforce authority does not automatically mean that a person does not have effective control over subordinates).

3.1.2 THE KNOWLEDGE REQUIREMENT

The mental element of command responsibility is one of its most controversial aspects. The superior knew, or had reason to know (was information available), that the subordinates were about to commit or were committing crimes. It is accepted that actual knowledge can be determined by a direct proof, or with reference to circumstantial evidence (such as type and scope of illegal acts or the time during which the illegal acts occurred). What the superior knew or had reason to know must be crimes, and the type of crimes committed (or that might be about to be committed) by their subordinates.\(^{48}\) These include actions such as the giving of orders or instructions, the widespread nature of the offenses, public report and media coverage, and the number of subordinates engaged in acts of disobedience. Article 82 of AP 1 even suggests that expert advice is available to the commander.\(^{49}\)

Article 28 ICC Statute sets a different standard for military and non-military superiors\(^{50}\), the standard for the former being that the superior ‘knew or, owing to the circumstances at the time, should have known that the forces were committing or about to commit such crimes’. For civilians, the civilian superior ‘knew, or consciously disregarded information which clearly indicated, that the subordinates were committing or about to commit such crimes’. This higher standard is in line with the Yamashita trial and Additional Protocol 1.

\(^{48}\) The Prosecutor v. Radovan Karadzic (Judgement), IT-95-5/18-T, 24 March 2016, 586

\(^{49}\) Geoffrey Corn, Ken Watkin, Jaimie Williamson, ‘The Law in War: A Concise Overview’ (2nd Ed.), Routledge (2023), 357

\(^{50}\) Geoffrey Corn, Ken Watkin, Jaimie Williamson, ‘The Law in War: A Concise Overview’ (2nd Ed.), Routledge (2023), 360
3.1.3 FAILURE TO TAKE MEASURES

The final link in the chain of liability under customary law is the failure or refusal to take necessary and reasonable measures to prevent or punish the offences the superior knew or culpably ought to have known of. crimes.\textsuperscript{51} There is no necessity that a person knew or should have known of the offences before they occurred for failure to punish liability to arise. The measures that can be expected to be taken depend on the precise nature of the control exercised by the superior.\textsuperscript{52}

There are certain circumstances in which the duty to punish may be fulfilled by disciplinary sanctions rather than criminal prosecutions, but, for international crimes, these will be rare. The commander cannot just later punish its subordinates. In the Bemba case, the Trial Chamber found that Bemba did not take all necessary and reasonable measures to prevent and repress the commission of crimes within his material ability. However, in the opinion of the Appeals Chamber, the Trial Chamber should have indicated \textit{in concreto} the measures that were available to Bemba at the time, while taking into account that he operated from another country, far removed from his forces.

3.1.4 CAUSATION

For superior responsibility to arise, crimes must be committed \textbf{whilst the superior had effective control} over the offenders since the offences must occur as a result of a superior’s failure to exercise control properly over such forces. The rapid processing capabilities of fully autonomous weapons, combined with unforeseen situations (communication disruptions, system errors, mechanical failures) could hinder commanders from aborting an attack. Additionally, a

\textsuperscript{51} International Committee of the Red Cross (ICRC), \textit{Protocol Additional to the Geneva Conventions of 12 August, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)}, 8 June 1977, 1125 UNTS 3, 86(2)

\textsuperscript{52} The Prosecutor v. Jean-Pierre Bemba (Judgment), ICC-01/05-01/08, 21 March 2016, 198
commander's formal accountability for a subordinate or, in this context, a fully autonomous weapon, would not automatically entail responsibility for the weapon's unlawful actions unless there is actual control. Therefore, if the criminal act is taking place during a time where the commander was not exercising effective control, command responsibility would not apply.

3.2 THE MENTAL ELEMENT IN LAWS

For command responsibility to arise, the subordinate must commit a chargeable offence, and satisfy all the elements of the crime. There are two important notions to these offences. First, the *actus reus*, this is the criminal act, and that act must be done with a certain mental state or *mens rea*. LAWS are able to commit criminal acts, since they are able to direct attacks, these attacks could be directed against civilians, therefore satisfying the *actus reus*. The issue with LAWS is satisfying the *mens rea*, since LAWS are not humans, they do not have moral agency they would therefore lack the independent intentionality to accompany the commission of the crime.53

The issue of finding accountability for LAWS which are unable to satisfy the intent part of the crime is where the problem lies. LAWS are able to select and engage targets which has in the past been a role held by human means that these machines cannot be held directly responsible for the criminal acts which they perpetrate. The issue further lies in the fact that these machines are prone to errors, accidents and malfunctions.

3.3 Command Responsibility and Malfunctions, Accidents and System Errors of LAWS

Malfunctions, accidents, and errors are potential risks associated with the use of LAWS. These risks arise from the fact that LAWS rely on complex software and hardware systems that can fail or produce unexpected results. All of these potential risks can lead to unintended or harmful consequences, such as civilian casualties, damage to infrastructure, or other negative impacts.

Malfunctions are technical failures or glitches that can occur in hardware of LAWS. These malfunctions can cause the system to behave unpredictably potentially leading to unintended consequences. Malfunctions can happen despite humans being on the loop. An example of hardware malfunction in the military domain is a failure of the guidance system of a missile. Guidance systems are responsible for controlling the trajectory of the missile, ensuring that it reaches its intended target accurately. If a malfunction occurs in the guidance system, it can cause the missile to steer off-course or to fail to hit its target. For example, on February 25th 1991, during the Gulf War and American Patriot missile experienced a hardware malfunction in its guidance system. Due to a software error the system’s clock became inaccurate over time, causing the missiles’ calculations to become progressively less precise. The result was a failure to intercept an incoming Scud missile fired by Iraq, leading to the death of 28 American soldiers and injury of around 100 others.

An accident refers to unintentional events that result in harm or other damage, such as a weapon firing when it is not intended to. Accidents can occur due to factors such as faulty sensors,

---


55 The Patriot Missile Failure <https://www-users.cse.umn.edu/~arnold/disasters/patriot.html>

56 Ibid.
incorrect programming, or lack of maintenance. An accident is therefore traceable to a human fault. For example, in 2016, a Tesla Model S car autopilot system resulted in the death of the driver. The driver was using the auto pilot system and the car’s sensors failed to detect a truck turning in front of it, resulting in the car crashing into the truck.57

Errors refer to mistakes or incorrect decisions made by the LAWS. These errors can result from factors such as an incorrect programming, incomplete or inaccurate data or misinterpretation of sensor information.58 Errors associated with the software of the decision-making process of the autonomous system. For example, in 2018 an Uber self-driving car hit and killed a pedestrian in Arizona. The car’s sensors had detected the pedestrian, but the software system failed to properly identify her as a human.59

In order to test whether command responsibility is an adequate basis for responsibility in cases of breach of IHL, we must look at probably the most important element of command responsibility: the superior/subordinate relationship. This is one of the constitutive elements of this mode of liability. As mentioned before this relationship is based on the superior being able to prevent or punish the crimes of his or her forces, who are the subordinates the superior-subordinate relationship is paramount to the doctrine of command responsibility. This relationship is crucial as it allows the superior to prevent or punish the offenses committed by their subordinates. It is evident that this relationship is based on personal interaction between two individuals who are human. As mentioned earlier, it was clarified in the preceding chapter that


LAWS are merely machines or weapons and lack the capacity to act as moral agents on the battlefield. Additionally, superiors who lack actual control over their subordinates are not included in this relationship.\(^{60}\) Therefore, if a superior lacks effective control over their subordinates, they will be exempted from criminal responsibility.

### 3.3.1 *Should Have Known: Command Responsibility and Accidents*

The ICC Statute, by imposing the general requirement for liability that the crimes occur as a result of a failure to supervise subordinates, excludes liability where there is no form of causation (contrary to the Tribunals), even in the expanded sense that a failure to prevent may facilitate commission. In the Bemba Gombo Case, the court stated: “The Chamber therefore considers that the chapeau of article 28(a) of the Statute includes an element of causality between a superior’s dereliction of duty and the underlying crimes. This interpretation is consistent with the principle of strict construction mirrored in article 22(2) of the Statute which, as part of the principle *nullum crimen sine lege*, compels the Chamber to interpret this provision strictly.”\(^{61}\)

It can therefore be inferred from this excerpt that commanders will not be held liable in cases in which there is no causal link, or where they are not able to exercise control. If we go back to our examples LAWS failures, out of the three types of failures which have been mentioned, there is only one which is traceable to a human: accidents. For errors and malfunctions it is a ‘glitch’ in the system which cannot be traced back to a human and therefore no military commander could be held liable for such failures of the system. This is linked to the fact that a commander is the able to prevent and punish machines, their fast-learning capabilities make it

---

60 The Prosecutor v. Mucic et al (Celebeci)(Trial Chamber), 1998 IT-96-21-A, 303

61 The Prosecutor v. Jean-Pierre Bemba (Pre-Trial Chamber), ICC-01/05-01/08, 15 June 2009, 423
impossible to conceive that a commander could prevent a machine from committing the criminal act, a human is not able to process information as fast as the machine can. It would also be complicated for the commander to punish the machine, owing to the lack of moral agency, or more generally lack of human nature of the machine, how could such a machine be punished?

3.3.2 SHOULD HAVE KNOWN: COMMAND RESPONSIBILITY AND ERRORS/MALFUNCTIONS

Let us now consider the issue of malfunctions and errors as previously posed. These malfunctions and errors are essentially autonomous errors made by LAWS. It would be challenging for a criminal tribunal to determine whether a military commander has any control over the autonomous malfunctions and errors of the weapon, which can be IHL violations. Additionally, LAWS operate on complex algorithms which a military commander would not be expected to understand in its entirety. As a result, if a commander were to be held responsible, they must have access to sufficient information to reasonably expect or be aware of any violation of IHL. Furthermore, to hold a commander accountable for these errors and malfunctions, it is imperative to demonstrate that the commander did not take the necessary steps to prevent or repress such errors. However, this would be a daunting task for the court to accomplish. The relationship between the commander and the LAWS is non-human in nature. The LAWS remains inactive until the parameters for its deployment are established. Once set, malfunction may arise, of which the commander may have no knowledge. Therefore, this is no avenue for the commander to take any measures to prevent the commission of a crime. Nonetheless, it has been posited that a commander has a proactive duty to seek and scrutinize information concerning attacks with LAWS. In a SIPRI report it was mentioned that this could take the form of keeping up to date with technological developments, otherwise they could be considered negligent if they knew that

---

a software update which was available but no implemented and this resulted in IHL violations.\textsuperscript{63} Regardless, without the superior-subordinate relationship being established, commanders wouldn’t be held responsible under this mode of liability.

3.4 Ethical Concerns

There are nonetheless ethical concerns which are bound to be raised when discussing responsibility of LAWS. They do not have human agency, intent, moral responsibility, and human dignity, yet they are being designed to take human lives, these poses great concerns for the use of such human-less machines to be able to decide whether to take human life. According to the ICRC, responsibility and accountability for decisions to use force cannot be transferred to a machine or a computer program.\textsuperscript{64} Ensuring the principles of IHL in war primarily rest on the assumption that any decision to take human life is inherently a human one.

Moreover, as we have mentioned before LAWS are based on complex algorithms and they are able to learn from experiences and surroundings. This machine learning process is drastically different from a human learning process. It is therefore impossible for human to understand the decision-making process of such a complex machine unless the machine is able to present a thought process at the end of the targeting. Nonetheless, even if the machine would be able to present that thought process, it would be unfair to hold humans responsible for the errors and malfunctions of these machines.

\textsuperscript{63} Marta Bo, Laura Bruun, Vincent Boulanin, Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On accountability for Violation of International Humanitarian Law Involving AWS, SIPRI, October 2022, 37

\textsuperscript{64} ICRC, Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control, 3 April 2018
4. OTHER SOURCES OF RESPONSIBILITY

As previously discussed in the preceding chapter, command responsibility may not be a sufficient framework for assigning accountability in the event of violations of IHL by LAWS. Nonetheless, it is untenable for a weapon capable of breaching the law to be deployed without an appropriate framework for establishing responsibility. This chapter aims to explore alternative avenues for assigning responsibility in cases of IHL violations. The chapter will focus on the responsibility of developers and state responsibility.

4.1 RESPONSIBILITY OF MANUFACTURERS, DESIGNERS, ENGINEERS AND PROGRAMMERS

In the development stage of LAWS, many persons are involved at different stages of the process. The code which is written for LAWS is not written by a singular programmer, but by many. The number of designers and engineers involved in the development is also considerable. For the purpose of this section, I will refer to developers more generally. Developers are essentially the ones creating these weapon systems, they are the ones which are writing the codes, which will lead the LAWS to take a certain course of action rather than the other based on their surroundings. Therefore, to some extent the developers are the ones which give the weapon system the ability to act the way that they will. However, to what extent are they responsible for violations of IHL resulting from failures, errors, and mistakes?

In domestic legal systems, various states have the possibility to file civil lawsuits for failures resulting in harm to persons. For example, the United States has national product liability and safety laws which allow civil suits to be brought against corporations in case negligence in manufacturing, designing and programming the machine. It has been suggested that this sort of liability could be applied in the case of LAWS. Sehrawat believes that US tort liability litigation

---

would be able to provide an adequate avenue for prosecution in the case of failures, errors or malfunctions of LAWS. He explains that in United States v. Kick, the all-civilian court stressed the necessity to “criminalize behavior that breached the relatively low standard of simple negligence in the military”\(^\text{66}\)

“\textit{There is a special need in the military to make the killing of another as a result of simple negligence a criminal act. This is because of the extensive use, handling, and operation in the course of official duties of such dangerous instruments as weapons, explosives, aircraft, vehicles, and the like. The danger to others from careless acts is so great that society demands protection}”\(^\text{67}\)

However, it is my contention that it is unlikely that a civil lawsuit would be filed if a LAWS fails against a corporation or the developer(s) if the harm has been done in, let’s say, the Democratic Republic of the Congo. This would mean that they would be letting persons from outside their national legal system to file a lawsuit against an (for the purpose of the argument) an American company and its developer(s). In the case they would be willing to let this happen, does every state have the necessary mechanism for such civil lawsuits to be filed?

Nonetheless, the development of international criminal tribunals has come naturally for the prosecution of serious violations of IHL. The Rome Statute of the International Criminal Court prosecutes individuals (and not states), therefore it is an individual who has to be prosecuted for the crimes he/she is being charged with and the person may only be prosecuted if he/she is criminally culpable.\(^\text{68}\) Article 25 of the Rome Statute is explicit; the court only has jurisdiction over

\(^{66}\) Ibid. 333


‘natural persons’69, therefore it is inconceivable to have a machine facing trial, it would also be against habeas corpus and more generally what kind of punishment would a machine face? A definitive disconnect of its systems? Some sort of electroshock ‘therapy’? It is essential that there is a sense of justice for the victims and simply holding a machine accountable for the crimes it has committed because it is autonomous would be science fiction.

As mentioned earlier, LAWS are made up of complex lines of code made by various coders, and its complexity makes it impossible to be made by a sole coder or programmer. Therefore, there is not a single person who is ‘in charge’ of the LAWS. Moreover, for individual criminal liability to arise there needs to be ‘intent and knowledge’ of the crime, and because of the various actors present at different stages of the development of laws it would difficult to find a singular culprit who acted with the specific intent and knowledge to bring about the crime. In essence, the culprit must have acted with dolus directus. The intent necessary to be held criminally responsible is defined under art. 30 of the Rome Statute:

1. Unless otherwise provided, a person shall be criminally responsible and liable for punishment for a crime within the jurisdiction of the Court only if the material elements are committed with intent and knowledge.

2. For the purposes of this article, a person has intent where:

   (a) In relation to conduct, that person means to engage in the conduct;

   (b) In relation to a consequence, that person means to cause that consequence or is aware that it will occur in the ordinary course of events.

---

3. For the purposes of this article, “knowledge” means awareness that a circumstance exists or a consequence will occur in the ordinary course of events. “Know” and “knowingly” shall be construed accordingly.

Therefore, it is not sufficient that the outcome is possible for the developers to be held accountable. In this light, in the Pre-Trial Chamber of the Bemba Case, it was argued that the language of Article 30 “[did] not accommodate a lower standard than the one required by dolus directus in the second degree (oblique intention)”, and that “[t]his standard is undoubtedly higher than the principal standard commonly agreed upon for dolus eventualis – namely foreseeing the occurrence of the undesired consequences as a mere likelihood or possibility.”

Developers who satisfy the mens rea could be held to account under article 25(3)(c) of the Rome Statute for aiding or abetting or otherwise assisting in the commission or attempted commission of a crime. Nonetheless, considering that some of these weapon systems would be built prior to the start of the conflict and that the Rome Statute requires that the conduct took place in the context of an armed conflict, it might seem hard to find that nexus. However, these weapon systems need constant care and updates, and therefore any updates to LAWS after the beginning of the armed conflict would satisfy the nexus to an armed conflict requirement.

Another issue in ascribing responsibility to developers, is that unless the developers have specifically aimed at creating an inherently indiscriminate weapon and that can be seen in the line of code, they cannot be said to be aiders or abettors. Malfunctions, errors and mistakes of LAWS are, at the end of the day, part of the challenges of using machines.

---

70 Afonso Seixas-Nunes, Autonomous Weapons Systems and Deploying States. Making Designers and Programmers Accountable, No. 161, April 2022, 84-85
4.2 State Responsibility

The law of state responsibility is relevant to the discussion of Lethal Autonomous Weapon Systems (LAWS) because it addresses the legal accountability of states for violations of international law. If a state deploys a LAWS that causes harm or violates international law, the state may be held responsible under this legal framework. This means that states must ensure that they are in compliance with international law when developing, acquiring, deploying, and using LAWS. They must take appropriate measures to prevent and address any harm or violations of international law that may occur as a result of LAWS use. This includes establishing appropriate regulations and guidelines for LAWS use, as well as ensuring that LAWS are designed and programmed to operate in compliance with international law.

State responsibility is triggered by state organs, these are persons or entities who are under the control or authority of the state, and whose actions are endorsed by the state. States as such are not responsible, it is the human agents whose actions and/or omissions trigger the responsibility of the state.\(^71\) The Draft Articles on the Responsibility of States for Internationally Wrongful Acts (ARSIWA) are not legally binding as they are not treaty, they are nonetheless considered to be part of customary international law. IHL being the body of law governing armed conflict, violations of IHL by the actions or omissions of an agent of the state will naturally trigger state responsibility.

The armed forces of a state function as an official entity and operate in that capacity when involved in conflicts on behalf of the state.\(^72\) While there might be occasional uncertainty regarding

\(^{71}\) Marta Bo, Laura Bruun, Vincent Boulanin, Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On accountability for Violation of International Humanitarian Law Involving AWS, SIPRI, October 2022, 19

whether an individual member of the armed forces is acting in an official capacity while performing an action, it can be generally attributed to the state when weapons are used in combat.\footnote{See, Article 91 AP1: “A Party to the conflict which violates the provisions of the Conventions or of this Protocol shall, if the case demands, be liable to pay compensation. It shall be responsible for all acts committed by persons forming part of its armed forces.”} Now that it has been established that the armed forces and the weapons that they are employing fall under the terminology of agent of the state; we will be looking at how malfunction, accidents and system errors of LAWS may be attributable to the state, or not.

To begin with, article 2 of ARSIWA established when there is an internationally wrongful act of a state: “There is an internationally wrongful act of a State when conduct consisting of an action or omission: (a) is attributable to the State under international law; and (b) constitutes a breach of an international obligation of the State.”

A breach of an international obligation in the context of LAWS is a violation of IHL as a result of the use of the weapon. This might take the form of a failure to take the necessary precautions in attack which results in the death of civilians. In this context we are equating the use of lethal force of LAWS to the firing of regular, non-autonomous weapons by a soldier. However, most of the discourse of responsibility of LAWS lies on the fact that these are autonomous weapon systems which are taking decisions on their own.\footnote{Tim McFarland, “Accountability,” Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law (Cambridge University Press 2020), 134} Nonetheless, if we go back to the beginning, the LAWS is being deployed by a member of the armed forces of a state, and as we have developed before, acts of members of the armed forces of a states are attributable to the state. As such, even if there is no human individually targeting a civilian population, the state is still responsible for the deployment of the weapon resulting in the death of civilians.
Within the context of LAWS, the increasing reliance on novel technologies aims to enhance accuracy and minimize human errors, they are expected to be able to provide rapid and precise information. Nevertheless, while the integration of AI in the military domain offers significant benefits, it will also bring substantial risks.\(^75\) One significant area of concern is the susceptibility of computer systems to hacking, errors, malfunctions and mistakes. The crux of the problem lies in establishing accountability for such errors. Do errors of LAWS give rise to state responsibility? This issue has been the subject of debate, including at the Group of Governmental Experts (GGE) in the context of AWS. The United States takes the stance unintended harm arising from an accident or equipment malfunction is ‘not a violation of IHL as such’. In contrast, Switzerland argues that states remain legally responsible for unlawful acts and resulting harm caused by AWS they employ, including due to malfunction or other undesired or unexpected outcomes.\(^76\)

4.2.1 Legal Review of New Weapons Under IHL

IHL refers to the need to review new weapons technology in order to determine whether the new weapon is IHL compliant. In this view, article 36 of AP 1 provides:

“In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.”


\(^76\) Marta Bo, Laura Bruun, Vincent Boulanin, Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On accountability for Violation of International Humanitarian Law Involving AWS, SIPRI, October 2022, 14
The article does not propose any specific procedures which must be undertaken in order to conduct the legal review. However, generally a legal review is done by assessing the design, the characteristics of the weapons and the manner in which they are intended to be used. The weapon is to be assessed based on existing relevant rules of IHL on the means and methods of warfare as well as of the prohibition or limitation of use of certain means and methods of warfare. The review provides military commanders with assurance as to the legality of weapons and ammunition. Therefore, since there is no specific test which can be applied for the weapons review, the test requires states to take into account the overarching principles of International Humanitarian Law (IHL) and any other relevant international laws, including those that prohibit specific weapons and methods of warfare, or limit their permissible use. Unfortunately, weapons reviews are very seldomly done. While weapons review is enshrined in AP1 and the United States is not a party to AP1, they are considered among the most rigorous in conducting weapons reviews.

William Boothby proposed the following 5 criteria as basis to conduct weapon reviews:

1. Whether the weapon is of a nature to cause superfluous injury or unnecessary suffering in its normal or intended circumstances of use
2. Whether the weapon is intended or likely to cause widespread, long-term, and severe damage to the natural environment.
3. Whether the weapon is indiscriminate by nature

---

77 Tim McFarland, Zena Assaad, Legal reviews of in situ learning in autonomous weapons, Ethics and Information Technology (2023) 25:9, 10 February, 2
78 Ibid. 3
4. The existence of specific rules (treaty or customary) that prohibit or restrict the use of a weapon.\textsuperscript{80}

5. The existence of likely future developments in international humanitarian law that may be expected to affect the weapon being reviewed.

If the weapon is capable of meeting these standards, then it may be deployed in accordance with the law, thereby providing an additional level of protection for the commander. The primary purpose of the weapon's review is to ensure that it is capable of functioning and adhering to the laws of warfare, as well as to prevent commanders from using a weapon that cannot comply with the law, which would leave them vulnerable to criminal prosecution in the event of a breach of the LOAC. Furthermore, depending on the degree of adherence to these criteria, the weapon's use may be restricted to ensure maximum compliance with the LOAC.

In conducting the legal review of LAWS, it is important to use data and empirical evidence of the system in order for the review to take into account all of the characteristic, whether they be actual but also expected. SIPRI has correctly put forward the importance of taking into account the empirical evidence as it will allow to determine whether\textsuperscript{81}:

1. The weapon system has been adequately programmed to respect the requirements of international law.

2. The system performs as anticipated and intended in normal conditions— that is, it is capable, effective, reliable and suitable for the task assigned;

\textsuperscript{80} Art. 35(2), Art. 35(3), Art. 55, Art. 51(4)(b), Art. 35(2), Art. 51(4)(b), Art. 51(4)(c), Art. 51(5)(a), Art. 51(5)(b), Art. 35(2), Art. 35(3), Art. 55

\textsuperscript{81} Vincent Boulanin, implementing article 36 weapon reviews in the light of increasing autonomy in weapon systems, SIPRI Insights on Peace and Security, NO. 2015/1, November 2015, 14
3. The intended effect of the weapon does not cause indiscriminate damage, unnecessary suffering or superfluous injury, or have a long-term, widespread or severe impact on the environment;

4. The weapon includes safety or anti-tamper mechanisms that will minimize the probability or consequences of unintended loss of control due to system failure or cyber-attack.

5. The probability of system failure of critical and non-critical functions is at the required or acceptable level.

6. The system will remain under adequate/meaningful human control—that is, it is understandable to trained operators, it can provide traceable feedback on its status, and an operator can activate or deactivate system functions or override decisions; and

7. The potential consequences of accidental misuse, loss of control, systems failure or cyber-attack are foreseeable.

4.2.2 Challenges of LAWS weapons review

LAWS are equipped with complex machine learning algorithms which have the ability to learn from battlefield events to optimize their behavior. While LAWS will serve the same purpose as they were intended to be used for during the weapons review, they will be able to optimize how they achieve the task. For instance, McFarland and Assaad provide an example of an AWS designed to locate and shoot targets. Thanks to machine learning, the system can learn the most effective way to shoot a moving target while still performing its intended task to ‘target and shoot’. The AWS’s goal remains the same, but its output is adjusted to account for a moving targets distance, speed, direction and timing.82

82 Ibid. 3
However, designing such learning algorithms present a challenge, as the purpose and task of the weapon remains the same, while the methods for achieving that task may differ and optimize. The weapons review, as it stands now, does not account for such learning algorithms. Therefore, it is necessary to consider additional legal review considerations during the weapons review of LAWS, particularly in relation to the aforementioned aspects, to ensure that the weapons reflect their intended purpose and enable full compliance with IHL.
5. Conclusion

As mentioned throughout the thesis, Lethal Autonomous Weapon Systems are relatively new to the military domain. They are nonetheless developing fast and seeing them on the battlefield in the next couple of years does not seem to be part of science-fiction. The issue is that these machine raise many questions, whether it is about the lack of international consensus on what defines LAWS or who is responsible for them. Nonetheless, if these weapons are to be fielded it is important to understand their limitations and ensure that accountability mechanisms are put in place prior to the inevitable happening.

On the chapter on command responsibility, it was seen that owing to the lack of human agency of LAWS, the superior-subordinate relationship cannot be satisfied in order to hold the commander responsible. Moreover, since a commander is responsible for chargeable offences of its subordinates but that machines are not able to satisfy the mens rea of war crimes, there are technically no chargeable offence on the machine. Command responsibility is nonetheless an interesting avenue to look at because the commanders are effectively the ones giving the green light to these machines to act, and in doing so taking the risk of malfunctions or system errors. Nonetheless, ascribing command responsibility in the case of LAWS would mean that every commander must be extremely technologically savvy so as to understand the inner working of these machines. It would simply be unrealistic that all commanders would possess such knowledge.

If commanders are not responsible for violations of IHL by LAWS, there must be someone who is. We have noted that the issue with ascribing responsibility to programmers, coders or engineers is that it would be extremely difficult to first find a single person to hold accountable, as well as for that person to satisfy the mens rea requirement for a war crime. Unless the coder has specifically changed the code to target unlawful target, or other persons involved in the development phase has tampered with the machine in an attempt to commit said violations, ascribing responsibility would be a daunting task for the courts to take upon.
Luckily, responsibility of states for internationally wrongful acts are the basis which would place some liability on someone. Indeed, at the end of the line of these singular individual is the state, and it is the states which are responsible for their agents, especially its armed forces. There are obligations to which the state has to comply such as the weapons review which means that even though the mens rea for a war crime will not be satisfied by a singular individual, the state still bears the responsibility for wrongdoing by its agents.

Nonetheless, this poses an issue as while it might seem positive that the state bears responsibility at the end of the line. Part of the sense of justice which is thought by victims is the visual that a person or persons are being condemned for the crime. With state responsibility victims and their families might not have the visual sense of justice since nothing would happen to the state.

It is however food for thought, it is essential that international law evolves with the current advancements of technology to allow for accountability mechanisms to be put in place before LAWS result in death of civilians. This issue has been discussed at lengths and will be for the decades to come, our fast-paced technological revolution has to been followed with a fast-paced legal evolution.
BIBLIOGRAPHY

BOOKS:

JOURNAL ARTICLES:
- Gunawan Y, Haris Aulawi M, Anggriawan R & Anggoro Putro T, (2022) Command responsibility of autonomous weapons under international humanitarian law, Cogent Social Sciences, 8:1, DOI: 10.1080/23311886.2022.2139906


GOVERNMENT PUBLICATIONS:


- United Kingdom, Joint Doctrine Publication 0-30.2 Unmanned Aircraft Systems (2017)

WEBSITES:

- 'AEGIS Combat System'
- 'Goalkeeper close-in weapon system (CIWS)'
  https://www.seaforces.org/wpnsys/SURFACE/Goalkeeper-CIWS.htm
- John K. Hawley, 2017, 'Patriot Wars: Automation and the Patriot Air and Missile Defense System' Center for a New American Security,
  https://www.cnas.org/publications/reports/patriot-wars
- 'MK 15 – Phalanx Close-In Weapon System (CIWS)'
- Reuters, 'Explainer: What is the Patriot missile defense system?'
  https://www.reuters.com/world/europe/what-is-patriot-missile-defense-system-2022-12-21/
- 'SeaRAM Close-In Weapon System (CIWS) Anti-Ship Missile Defense System'

- 'The Ooda Loop' (The Decision Lab) https://thedecisionlab.com/reference-guide/computer-science/the-ooda-loop

REPORTS:
- Boulanin V, implementing article 36 weapon reviews in the light of increasing autonomy in weapon systems, SIPRI Insights on Peace and Security, NO. 2015/1, November 2015
- McFarland T, Assaad Z, Legal reviews of in situ learning in autonomous weapons, Ethics and Information Technology (2023) 25:9, 10 February
- UNIDIR, United Nations Institute for Disarmament Research, The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches, UNIDIR Resources

CASES:
- The Prosecutor v. Delalic et al (Appeals Chamber) 2001, IT-95-14-T
- The Prosecutor v. Jean-Pierre Bemba (Judgment), ICC-01/05-01/08, 21 March 2016
- The Prosecutor v. Jean-Pierre Bemba (Pre-Trial Chamber), ICC-01/05-01/08, 15 June 2009
- The Prosecutor v. Mucic et al (Celebeci)(Trial Chamber), 1998 IT-96-21-A
- The Prosecutor v. Radovan Karadzic (Judgement), IT-95-5/18-T, 24 March 2016, 586
- Supreme Court of the United States 327 U.S 1, Yamashita, 15