



Strathmore University

Law School

**CODE AND COMPLIANCE: A LEGAL CAUSATION
FRAMEWORK FOR FULLY AUTONOMOUS WEAPON
SYSTEMS DEPLOYED IN URBAN SETTINGS**

Submitted in partial fulfilment of the requirements of the Bachelor of Laws Degree,
Strathmore University Law School

By

Trevor Chege Kariuki

145873

Prepared under the supervision of

Cecil Abungu

February 2025

Word count: 14112

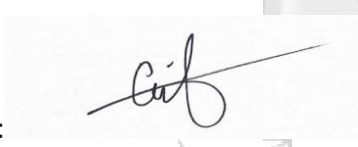
Declaration

I, TREVOR DAVID CHEGE KARIUKI, do hereby declare that this research is my original work and that to the best of my knowledge and belief, it has not been previously, in its entirety or in part, been submitted to any other university for a degree or diploma. Other works cited or referred to are accordingly acknowledged.

Signed: 

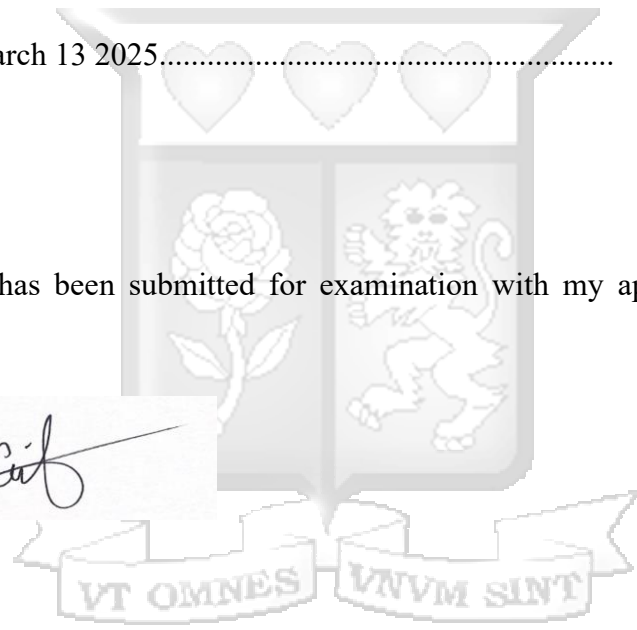
Date:March 13 2025.....

This dissertation has been submitted for examination with my approval as University Supervisor.

Signed: 

Cecil Abungu

Date:March 13 2025.....



Acknowledgements

First and foremost, I am profoundly grateful to God for granting me the strength, wisdom, and perseverance to complete this dissertation. Without His guidance and grace, this journey would not have been possible. Especially considering the general difficulty numerous students have in finding a research problem that is not only meaningful but also personally engaging and inspiring.

I would also like to express my deepest appreciation to my supervisor, Cecil Abungu, for his invaluable guidance, patience, and unwavering support. His insightful feedback and encouragement have been absolutely instrumental in shaping this research.

Thank you.



List of legal instruments

1. *Articles on state responsibility for internationally wrongful acts*, ILC 53rd Report, 2001, UN Doc A/RES/56/83 (2001).
2. *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (protocol I)*, 8 June 1977, 1125 UNTS 3.
3. *Treaty on the Prohibition of Nuclear Weapons*, 7 July 2017, Treaty XXVI.9.



List of cases

Foreign Case Law

1. *Barnett v Chelsea and Kensington Hospital Management Committee* (1969), Queen's Bench of The United Kingdom.
2. *Blyth v Birmingham Waterworks Company* (1856), England and Wales High Court Exchequer Division.
3. *Derdarian v Felix Contracting Corp* (1980), The New York Court of Appeals of the United States.
4. *Donoghue v Stevenson* (1932), The United Kingdom House of Lords.
5. *Glasgow Corporation v Muir* (1943) The United Kingdom House of Lords.
6. *Hines v Garrett* (1921), The Supreme Court of Virginia in the United States.
7. *Overseas Tankship (UK) Ltd v Morts Dock & Engineering Co Ltd (The Wagon Mound No 1)* (1961), The Supreme Court of New South Wales.
8. *State Rail Authority of New South Wales v Wiegold* (1991), High Court of Australia.

ICJ Case Law

9. *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, ICJ Reports 1996.
10. *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda)*, Judgement, ICJ Reports 2005.

Inter-American Court of Human Rights

11. *Santo Domingo Massacre v Colombia*, IACtHR Judgement of 30 November 2012, (Preliminary Objections, Merits and Reparations).

List of abbreviations

AWS – Autonomous Weapon Systems

AI – Artificial Intelligence

IACtHR – Inter-American Court of Human Rights

ICJ – International Court of Justice

ICRC – International Committee of the Red Cross

IHL – International Humanitarian Law

LAM – Loitering Attack Munitions

MHC – Meaningful Human Control

MSF – Médecins Sans Frontières

UPDF – Uganda People’s Defence Forces

WMD – Weapons of Mass Destruction



Abstract

The advancement of fully autonomous weapon systems presents unprecedented legal and ethical challenges, particularly concerning accountability under International Humanitarian Law. These weapons, characterised by deep-learning-led capabilities, can function without direct human intervention, raising concerns about compliance with established norms of warfare. The study explores the necessity of an effective liability framework under International Humanitarian Law for when these weapons inflict harm on protected persons, particularly in urban combat settings.

Current legal frameworks within International Humanitarian Law struggle to address accountability gaps introduced by these weapons. The study critically examines the existing liability mechanisms for conventional weapons and weapons of mass destruction, highlighting their inadequacy in dealing with fully autonomous systems. Through an analysis of legal precedents, the study argues that a causation framework exactly like the one in common law negligence—specifically, factual and proximate causation—provides a robust foundation for attributing liability in the context of this study’s problem.

Methodologically, this research employs a qualitative deductive approach, drawing from case law, international treaties, scholarly discussions, and more. The study finds that a causation framework exactly like the one in common law negligence should be integrated into International Humanitarian Law to hold users of these weapons (considered to be States) liable for harm they cause in urban areas. In instances where the weaponry is deployed directly to urban areas, liability should be established using the ‘but-for’ test. In cases where it is not deployed to urban areas but takes action within them, the foreseeability test should apply.

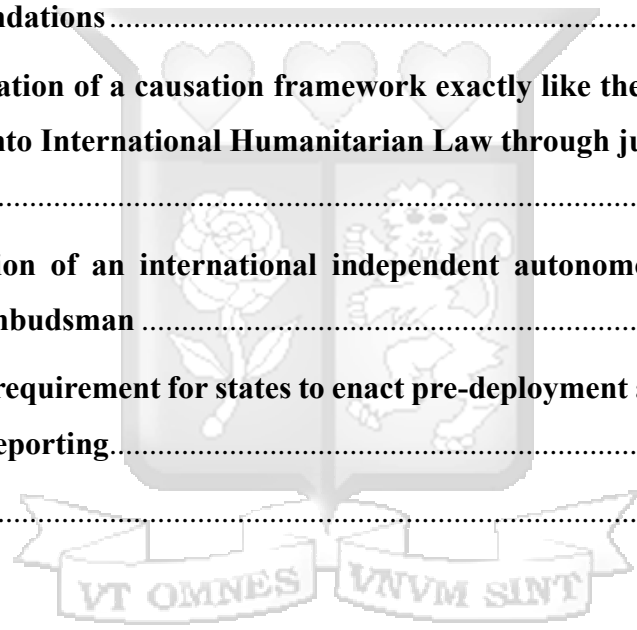
Ultimately, the study proposes the above integration to ensure the use of these weapons aligns with fundamental humanitarian principles, advocating for judicial interpretation, institutional oversight, and reporting mechanisms to mitigate risks posed by fully autonomous warfare technologies.

Table of contents

Declaration	i
Acknowledgements	ii
List of legal instruments	iii
List of cases	iv
List of abbreviations	v
Abstract	vi
1.0. Introduction	1
1.1. Problem Statement	5
1.2. Research Objectives	5
1.3. Research Questions	5
1.4. Hypothesis	6
1.5. Justification of the study	6
1.6. Conceptual Framework	6
1.7. Literature Review	8
1.7.1. On how autonomous weapon systems work	8
1.7.2. On the causation standard of common law negligence	9
1.7.3. On the necessity of a framework to hold the user of the fully autonomous weapon system liable for harm the weapon causes in urban areas	10
1.7.4. Contribution	11
1.8. Methodology	12
1.9. Chapter Breakdown	13
2.0. Understanding the functionality, appeal, and risks posed by autonomous weapon systems	14
2.1. Introduction	14
2.2. Technical functionality of autonomous weapon systems	14
2.2.1. What is artificial intelligence?	14

2.2.2. Artificial intelligence in autonomous weapon systems	15
2.3. The appeal of fully autonomous weapon systems	16
2.4. The risks posed by fully autonomous weapon systems	18
2.4.1. Explainability, transparency, and accountability.....	18
2.4.1.1. The ‘black-box’ problem	18
2.4.1.2. The multiple actors problem.....	19
2.4.2. Unpredictability	20
2.5. The importance of a liability framework.....	24
2.6. Conclusion	25
3.0. The liability framework governing the use of weapons in International Humanitarian Law.....	27
3.1. Introduction.....	27
3.2. The role of human control and the principles of International Humanitarian Law.....	27
3.2.1. What is control over weaponry understood to be?	27
3.2.2. How have courts attributed liability for harm by weapons based on what control is understood to be?	29
3.3. The challenge faced in proving causation under International Humanitarian Law in the context of fully autonomous weapon systems.....	32
3.4. Conclusion	34
4.0. The causation standard of common law negligence as the way forward.....	36
4.1. Introduction.....	36
4.2. Defining negligence and understanding the causation standard	37
4.2.1. Introduction to negligence.....	37
4.2.2. The causation standard	38
4.3. The causation standard of common law negligence as the ideal framework ..	39
4.3.1. The criteria for a desirable liability framework	39

4.3.2. Why a causation standard exactly like the one in common law negligence offers a desirable path forward	40
4.3.2.1. Application of the ‘but-for’ test	40
4.3.2.2. Application of the foreseeability test	41
4.3.2.3. A positive case for a causation standard exactly like the one in common law negligence.....	42
4.4. Conclusion	44
5.0. Conclusion and recommendations.....	45
5.1. Conclusion	45
5.2. Recommendations.....	46
5.2.1. Integration of a causation framework exactly like the one in common law negligence into International Humanitarian Law through judicial interpretation	46
5.2.2. Creation of an international independent autonomous weapon systems reporting ombudsman	46
5.2.3. Legal requirement for states to enact pre-deployment and post-deployment evaluation reporting.....	47
Bibliography.....	49



1.0. Introduction

“An autonomous weapon...has an independent capacity and it’s self-directed and whether it can kill depends on whether it is armed or not.”¹

Upon the commission of a crime, man seeks solace in knowing redress can be sought and that justice can prevail.² However, a lurking threat to this is a soulless imperfect weapon of international war that brings destruction of its own accord: the fully autonomous, human-out-of-the-loop, weapon system. It operates under no law—only an unwavering duty to fulfil an objective, with neither man nor State in control of its decisions and actions.

Autonomous Weapon Systems (AWS) lack a specific and universally accepted definition.³ However, one of the most critical concepts in each of these definitions, is the concept of human intervention. The presence of a human actor in the operation of AWS provides a subject to hold accountable. For instance, in weapon systems that are operated by remote operatives, by the push of a button that engages the weapon, simultaneous responsibility is incurred by the actor for whatever action that weapon partakes on the battlefield because they are the initiator and propeller of its act.⁴ However, AWS fully driven by artificial intelligence (AI), lack clarity with regards to accountability for their actions.⁵

¹ Quote by Stacy Pettyjohn, Director of the Defense Program at Center for a New American Security; ‘Unknown: Killer Robots’ directed by Jesse Sweet in 2023.

² UNGA, *Declaration of basic principles of justice for victims of crime and abuse of power*, UN A/Res/40/34 (29 November 1985).

³ This is evident in the non-exhaustive list of definitions by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System;

Convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects, 10 March 2023, CCW/GGE.1/2023/CRP.1, 1-8; Definition by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States: “...weapon systems that, once activated, can identify, select, and engage targets with lethal force without further intervention by an operator...”

Definition by China:

“Basic characteristics of Unacceptable Autonomous Weapons Systems should include but not limited to the following: Firstly, lethality, meaning sufficient lethal payload (charge) and means. Secondly, autonomy, meaning absence of human intervention and control during the entire process of executing a task. Thirdly, impossibility for termination, meaning that once started, there is no way to terminate the operation. Fourthly, indiscriminate killing, meaning that the device will execute the mission of killing and maiming regardless of conditions, scenarios and targets. Fifthly, evolution, meaning that through interaction with the environment, the device can learn autonomously, expand its functions and capabilities in a degree exceeding human expectations.”

⁴ Horowitz M, Scharre P, ‘Meaningful human control in weapon systems: A primer’ Center for a New American Security, Working Paper March 2015, 10-13 –< [Meaningful Human Control in Weapon Systems: A Primer | CNAS](#)> on 22 September 2024.

⁵ –< [Lethal Autonomous Weapon Systems \(LAWS\) – UNODA](#)> on 22 September 2024.

The fully autonomous weapon system, though not developed yet, is argued to be inevitable. The evolution of warfare has been marked by significant transformations driven by advancements in technology.⁶ From the primitive weaponry of the pre-modern era such as spears,⁷ to the Cold War's introduction of nuclear weapons as a result of the arms race,⁸ and the dawn of deadly high precision weaponry such as the switchblade 600 loitering munition,⁹ military technologies have regularly developed based on technological advancements and need for military advantage.¹⁰ A shared feature in these eras was that humans were the decision-makers, they had unequivocal control over the actions of the weapon.

Despite these advancements, there persists an unsatisfiable pressing need for increased efficiency, precision in military operations, and reduced risks to soldiers in hazardous environments. This necessity has led to a growing demand for autonomous weaponry. This has given rise to two distinct types: supervised autonomous weapons, where human operators retain control and intervene as necessary, and semi-autonomous weapons, which have partial independence but engage targets selected by human operators.¹¹

However, for the very same needs mentioned above, there remains a need to develop a "perfect soldier" capable of rapidly adapting to dynamic situations without human intervention or operation in international armed conflicts. Fully autonomous weapons, or "human-out-of-the-loop" systems, driven by deep learning,¹² offer the potential to increasingly satisfy this need. Despite this weaponry not being fully developed yet, and such development being opposed,¹³ precursors to such systems are already in development, indicating a trajectory toward their eventual realisation. For instance, the United States'

⁶ Buchanan R, 'History of technology' *Encyclopaedia Britannica*, 2023, 1 –<[History of technology | Evolution, Ages, & Facts | Britannica](#)> on 22 September 2024.

⁷ Buchanan R, 'History of technology' *Encyclopaedia Britannica*, 2023, 8 –<[History of technology | Evolution, Ages, & Facts | Britannica](#)> on 22 September 2024.

⁸ –<[Arms Race: Definition, Cold War & Nuclear Arms | HISTORY](#)> on 22 September 2024.

⁹ –<[Switchblade® 600 Loitering Munition Systems | Kamikaze Drones | Suicide Drone | Tactical Missile Systems | AeroVironment, Inc. \(avinc.com\)](#)> on 22 September 2024.

¹⁰ –<[What you need to know about autonomous weapons | ICRC](#)> on 22 September 2024.

¹¹ Kallenborn Z, 'Meet the future weapon of mass destruction, the drone swarm' *Bulletin of the Atomic Scientists*, 5 April 2021 –<[Meet the future weapon of mass destruction, the drone swarm - Bulletin of the Atomic Scientists \(thebulletin.org\)](#)> on 22 September 2024.

¹² –<[What is Deep Learning? - Deep Learning Explained - AWS \(amazon.com\)](#)> on 22 September 2024;

Deep learning allows the weapon system to process data in a way that is inspired by the human brain such that it can assess vast data, determine patterns and plausible actions, and act with timely precision.

¹³ *Convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects*, 8 August 2022, CCW/GGE.1/2022/WP.2, 4.

unmanned US X-47B combat aircraft has an autonomous flight control system designed to take off, autonomously aerial refuel, and land.¹⁴ An additional precursor is the AI pilot called Hivemind.¹⁵ Despite having about 10 months of learning through simulations against other AI, it was pinned against a highly experienced fighter pilot and absolutely dominated the human fighter pilot, taking no losses, even though the human pilot had about 25 years of experience and undertook 250 combat missions.¹⁶ It has a 99%-win rate and though it is for now being integrated into F-16s to read and react to the battlefield without requiring GPS waypoints or communications to make decisions, the goal is to rapidly scale it across next-generation aircrafts, F-22s, F-18s, and other advanced fighter jets.¹⁷ The inevitability of the realisation of fully autonomous weapon systems (fully AWS) is backed by the scaling hypothesis.¹⁸ The hypothesis suggests that current deep learning techniques, with increased computing power and data, will progress towards human-level AI and beyond,¹⁹ reducing the need for human intervention in operation of such systems.²⁰

No weapon is inherently perfect, and this equally applies to fully AWS. They are dependent on development by an inherently imperfect human being such as programmers, data scientists, manufacturers, etc.²¹ Intrinsic to human work is bias, lack of omniscience, limitation in resources, and much more. Despite the extraordinary efforts to fine-tune the AI model and weapon system to as optimal an output as possible, fully AWS cannot be perfect in all their engagements. This is more so the case when it comes to target identification, particularly in urban areas.²² Machine learning algorithms naturally do not

¹⁴ <[X-47B Unmanned Combat Air System \(UCAS\) \(naval-technology.com\)](#)> on 22 September 2024.

¹⁵ <[Hivemind - Shield AI](#)> on 22 September 2024.

¹⁶ ‘Unknown: Killer Robots’ directed by Jesse Sweet in 2023.

¹⁷ <[Hivemind - Shield AI](#)> on 22 September 2024.

¹⁸ <[The Scaling Hypothesis · Gwern.net](#)> on 22 September 2024;

“Once we find a scalable architecture like self-attention or convolutions, which like the brain can be applied fairly uniformly we can simply train ever larger neural networks and ever more sophisticated behaviour will emerge naturally as the easiest way to optimise for all the tasks & data.”

¹⁹ Rushing B, Lavin J, ‘Is the scaling hypothesis falsifiable?’ PhilSci-Archive, 2024, 1.

²⁰ Bashir D, Kurenkov A, ‘The ai scaling hypothesis’ Last Week in AI, 5 August 2022 <[The AI Scaling Hypothesis \(lastweekin.ai\)](#)> on 22 September 2024.

²¹ Lazarsfeld S, ‘The courage for imperfection’ 47(1), *Individual Psychology: Journal of Adlerian Theory, Research & Practice*, 1991, 93–96.

²² Target identification in this sense involves the complex process of an artificial intelligence system accurately recognising and classifying objects within a given environment as military objectives or not; Schwarz E, ‘Gaza war: Israel using AI to identify human targets raising fears that innocents are being caught in the net’ Queen Mary University of London, 17 April 2024 <[Gaza war: Israel using AI to identify human targets raising fears that innocents are being caught in the net - Queen Mary University of London \(qmul.ac.uk\)](#)> on 22 September 2024;

An urban area is a densely populated area that serves as a hub for economic, social, and/or cultural engagement, with extensive infrastructure including roads and buildings. It comprises of people of various

give output with 100% certainty and the reliability and success rate of AI machinery becomes more challenged as autonomy increases.²³ These issues are further exacerbated by dynamic and inherently unpredictable environments such as urban areas.²⁴ The resultant effect is a prominent hindrance of the ability of the fully autonomous weapon to successfully distinguish between protected persons and military targets.²⁵

Furthermore, the “learning” capabilities of these weapons make them increasingly adaptive which not only poses concerns regarding their predictability and trustworthiness,²⁶ but also makes unclear the subject to be held accountable in case of harm caused to protected persons by the weapon. The autonomy of fully AWS, characterised by deep learning capabilities and inscrutable decision-making processes, makes adequate supervision and understanding by human’s difficult, and intervention during operation impossible.²⁷ The subject of accountability is increasingly more difficult to establish due to the involvement of various actors in development of fully AWS at various interlinked stages such as the data acquisition and preparation stage, model development and training stage, etc.²⁸ As a result, identifying who is to be held accountable for harm caused by the weapon in urban areas during urban warfare is rather challenging because the weapon’s fault could very well be caused by the quality of data used in training, or a malfunction of hardware, or unsatisfactory programming, or various other possible causes from various other stages involving various different actors.

backgrounds, lifestyles, professions, genders, and ages. These environments are complex and uncertain due to high population density, complex infrastructure, and the dynamic nature of civilian and military movements, which complicate the ability to distinguish between combatants and non-combatants.

²³ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 37-39.

²⁴ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 37-39.

²⁵ –<[Protected persons | ICRC](#)> on 22 September 2024;

Protected persons are individuals who, due to their particular circumstances, are entitled to special protection during armed conflict. This includes civilians, combatants who are hors de combat (out of action), medical personnel, and cultural property. Civilians are generally protected from direct attacks, unless they take a direct part in hostilities. Combatants who are hors de combat, such as wounded, sick, or captured soldiers, are also protected and must be treated humanely. Medical personnel and cultural property, such as hospitals, schools, and places of worship, are also afforded special protection under IHL;

Kallenborn Z, ‘Meet the future weapon of mass destruction, the drone swarm’ *Bulletin of the Atomic Scientists*, 5 April 2021 – < [Meet the future weapon of mass destruction, the drone swarm - Bulletin of the Atomic Scientists \(thebulletin.org\)](#)> on 22 September 2024.

²⁶ Taddeo M, Blanchard A, ‘A comparative analysis of the definitions of autonomous weapons systems’ 28 (37) *Science and Engineering Ethics*, 2022, 12.

²⁷ Hua S, ‘Machine learning weapons and international humanitarian law: Rethinking meaningful human control’ 15(1) *Georgetown Journal of International Law*, 2019, 126-127.

²⁸ Saltz J, ‘What is the ai life cycle’ Data Science Process Alliance, 31 March 2024 –< [What is the AI Life Cycle? - Data Science Process Alliance \(datascience-pm.com\)](#)> on 22 September 2024.

International law and International Humanitarian Law (IHL) are ill-equipped to determine accountability for harm caused to protected persons by fully AWS due to key limitations. Firstly, it is crucial to note that there is no authoritative legal framework in place that specifically governs the development, production, and use of AWS.²⁹ Secondly, the existing regimes and international discourse regarding weaponry are hinged on human oversight, control, and intent, which fully AWS either lack or make inadequate, further complicating accountability. Thirdly, legal responsibility is traditionally assigned to individuals or states, but with numerous actors being involved in development and use of fully AWS, it is unclear if liability should fall on programmers, operators, commanders, the state itself, or any other actor(s).

1.1. Problem Statement

This study assesses what causation standard IHL ought to adopt to establish liability for harm caused by fully AWS in urban settings.

1.2. Research Objectives

- 1) To investigate the functioning of AWS, identify the factors that make them appealing for warfare, and assess the risks they pose.
- 2) To investigate the mechanisms by which IHL attributes liability for harm caused to protected persons by conventional weapons and weapons of mass destruction, and to explore the challenges IHL encounters in attributing liability in the context of fully AWS.
- 3) To evaluate the suitability of a causation framework exactly like the one in common law negligence as the appropriate framework for holding users of fully AWS liable for harm the weaponry causes in urban settings.

1.3. Research Questions

- 1) How do AWS function, what makes them appealing for warfare, and what risks do they pose?

²⁹ Lestari A, 'AWS: A threat to international humanitarian law or a necessary technological evolution?' *Modern Diplomacy*, 4 April 2024 < [AWS: A Threat to International Humanitarian Law or a Necessary Technological Evolution? - Modern Diplomacy](https://doi.org/10.18356/29a571ba-en) > on 10 November 2024; Davison N, 'A legal perspective: Autonomous weapon systems under international humanitarian law' *United Nations Office for Disarmament Affairs, UNODA Occasional Papers Number 30, 2018, 7* < <https://doi.org/10.18356/29a571ba-en> > on 10 November 2024.

- 2) How has IHL attributed liability for harm caused to protected persons by conventional weapons and weapons of mass destruction, and what challenges arise in IHL when it comes to attributing liability in the context of fully AWS?
- 3) Why might a causation framework exactly like the one in common law negligence be the appropriate framework for holding users of fully AWS liable for harm the weaponry causes in urban settings?

1.4. Hypothesis

A causation framework exactly like the one in common law negligence should be adopted under IHL to hold users of fully AWS liable for harm the weaponry causes in urban settings.

1.5. Justification of the study

This study primarily benefits six parties. Firstly, policy makers and international organisations stand to gain a deeper understanding of how existing legal frameworks can be adapted to address the emerging challenges posed by fully AWS to international armed conflicts, thereby informing the development of new regulations and international treaties. Secondly, legal scholars and researchers benefit by gaining a novel perspective on integrating the domestic tort principle of negligence into IHL which could not only influence academic discourse but also future legal interpretations. Lastly, military strategists and defence policymakers can use the insights to refine deployment protocols and accountability mechanisms for fully AWS, ensuring that the use of such technology aligns with international legal and ethical standards.

1.6. Conceptual Framework

Meaningful human control as the ability to foresee International Humanitarian Law violations and decide on deployment of the fully autonomous weapon

Meaningful human control (MHC) means that future weapons systems must have humans rather than computers and their algorithms³⁰ remain in control of, and thus morally responsible for, relevant decisions regarding lethal military engagement.³¹ This concept has struggled to reach a consensus among the United States and its international partners due to disagreements in terminology and operational scope of human involvement. While the

³⁰ –< [Algorithm | AI Glossary | OpenTrain AI](#)> on 22 September 2024;

Algorithm refers to a set of rules to be followed by a computer or machine in action, reasoning, computation, calculations, or other problem-solving operations. They are the backbone of AI and machine learning.

³¹ Article 36, *Killer Robots: UK Government Policy on Fully Autonomous Weapons*, 19 April 2013, 1-5.

United Nations Convention on Certain Conventional Weapons, specifically its Group of Governmental Experts, has employed the phrase, US Department of Defence Directive 3000.09 uses the term “appropriate human judgment” instead.³²

For the former, MHC ensures that use of AWS is lawful through: the AI system being predictable, reliable and transparent; the clarity and accuracy of information on the outcome sought and the context of the system’s use; the existence of the possibility of timely human intervention; and the ability to attribute legal responsibility for outcomes.³³ The human operators must have a significant degree of control and involvement throughout the decision-making process of AWS. This version leans towards maintaining human judgment in all critical aspects of AWS deployment and target engagement, reflecting a more cautious approach to autonomous systems.³⁴ For the latter, more emphasis is placed on a flexible approach that gives room for an increase in autonomy. Human oversight would only be involved in critical decisions about how, when, where, and why the weapon will be employed, and not require manual human control of the weapon.³⁵

However, these two views fail to consider fully AWS which make human intervention impossible due to their degree of autonomy and high complexity.³⁶ Therefore, the above views fall short in this study. However, partially drawing from Schuller’s movement from ‘control as human intervention’ to ‘control as predictability’,³⁷ and partially from understanding fully AWS, this study posits a new conceptualisation of MHC that best accommodates fully AWS.

Alan Schuller put forward the reasonable predictability test to assess whether compliance with IHL can be reasonably predicted/foreseen whereby if so, then control is maintained

³² Trabucco L, ‘What is meaningful human control, anyway? cracking the code on autonomous weapons and human judgment’ Modern War Institute, 2023 –< [What is Meaningful Human Control, Anyway? Cracking the Code on Autonomous Weapons and Human Judgment - Modern War Institute \(westpoint.edu\)](https://www.warinstitute.org/insights/what-is-meaningful-human-control-anyway-cracking-the-code-on-autonomous-weapons-and-human-judgment-modern-war-institute-westpoint.edu)> on 22 September 2024.

³³ Hua S, ‘Machine learning weapons and international humanitarian law’, 131.

³⁴ Hua S, ‘Machine learning weapons and international humanitarian law’, 131-132.

³⁵ Congressional Research Service, *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems*, 1 February 2024, 1.

³⁶ Matthias A, ‘The responsibility gap: Ascribing responsibility for the actions of learning automata’ 6(3) *Ethics and Information Technology*, 2004, 182-183;

MHC “is impossible when the machine has an informational advantage over the operator . . . [or] when the machine cannot be controlled by a human in real-time due to its processing speed and the multitude of operational variables. . .”

³⁷ Hua S, ‘Machine learning weapons and international humanitarian law’, 139.

no matter the level or type of interaction with the machine.³⁸ However, in stating that lethal decisions shouldn't be fully handed over to a machine, he fails to consider that in fully AWS, as is elaborated upon in this study, users³⁹ can only exercise precaution prior to deployment, deploy the weapon, and observe the engagements,⁴⁰ and that these are forms of meaningful human control in themselves.⁴¹

Therefore, this study posits the concept of MHC as the user's ability in the pre-deployment stage, to foresee that deploying the weapon would either result in its actions violating IHL or not, and the user's ability in the deployment stage, to decide on deployment or not. It highlights these two stages as the critical moments of control.

1.7. Literature Review

1.7.1. On how autonomous weapon systems work

AWS represent a significant evolution in warfare, utilising sophisticated sensor suites and computer algorithms to independently identify and engage targets without direct human intervention.⁴² The International Committee of the Red Cross (ICRC) recognised that these systems are highly complex in their autonomy because of machine learning algorithms.⁴³ Kelleher states that this is a system's ability to analyse and find patterns and relationships in data,⁴⁴ and the ICRC divides it into three: supervised learning which uses a set of examples with a label informing the algorithm of the expected output; reinforcement learning whereby algorithms learn how to choose between a set of actions to accomplish a task that will maximise some reward, this is done through trial and error; and unsupervised learning where algorithms learn by themselves without any external goal and which would be able to come up with their own goals.⁴⁵ With an increase in interconnected neural

³⁸ Schuller A, 'At the crossroads of control: The intersection of artificial intelligence in autonomous weapon systems with international humanitarian law' 8(8) *Harvard National Security Journal*, 2017, 409.

³⁹ This refers to States. This includes a military man/commander's actions because attribution to a State occurs when either: the entity is an organ of the state (structure); it is empowered to "exercise elements of the governmental authority" (function); or it is controlled by the state (control); *Articles on state responsibility for internationally wrongful acts*, ILC 53rd Report, 2001, UN Doc A/RES/56/83 (2001).

⁴⁰ Gaeta P, 'Who acts when autonomous weapons strike? the act requirement for individual criminal responsibility and state responsibility' 21(5) *Journal of International Criminal Justice*, 2023, 1053.

⁴¹ Taddeo M, Blanchard A, 'A comparative analysis of the definitions of autonomous weapons systems', 11.

⁴² Congressional Research Service, *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems*, 1 February 2024, 1.

⁴³ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 24.

⁴⁴ Kelleher J, *Deep learning*, The MIT Press, Cambridge, 2019, 6-7.

⁴⁵ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 37-39.

networks, various scholars argue that the weapon system becomes capable of analysing vast complex data amounts, processing them, and determining patterns and relationships essentially becoming capable of deep learning.⁴⁶ Hua acknowledges that deep learning allows this without requiring explicit programming for every potential scenario, making AWS increasingly adaptable to uncertainties in the environment and capable of highly complex decision making.⁴⁷ Furthermore, Taddeo and Blanchard acknowledge that deep learning is utilised in the weapon's intelligence gathering and target recognition and identification.⁴⁸ However, the ICRC outlines that it is crucial to note that these algorithms inherently do not give output with 100% certainty.⁴⁹

1.7.2. On the causation standard of common law negligence

Negligence is the omission to do something which a reasonable man, guided by those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do.⁵⁰ It is used to hold a person liable for carelessly causing harm and traditionally entailed four elements: duty, breach, cause, and damage. Owen, because of the misleading conflation of factual causation and proximate cause under the blanket term “cause”, states these elements as five, namely: duty of care, breach of such duty, factual causation, proximate cause, and damage.⁵¹ In expounding on the causation standard,⁵² he states that factual causation looks at whether the wrongful act actually caused the harm, determined by the ‘but-for’ test⁵³, while on the other hand, proximate cause looks at whether the act and the harm were sufficiently close rather than remote, determined by the foreseeability test.⁵⁴

Yannoulidis expounded on factual causation, stating that a defendant's act would be the factual cause of a plaintiff's loss if it was a necessary condition of such loss.⁵⁵ Furthermore,

⁴⁶ Kelleher, *Deep learning*, 21;

Piper K, ‘The case for taking ai seriously as a threat to humanity’ Vox, 15 October 2020 –< [What is artificial intelligence? Your AI questions, answered. | Vox](#)> on 13 November 2024.

⁴⁷ Hua S, ‘Machine learning weapons and international humanitarian law’, 119, 124-125.

⁴⁸ Taddeo M, Blanchard A, ‘A comparative analysis of the definitions of autonomous weapons systems’, 12.

⁴⁹ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 37-39.

⁵⁰ The Black's Law Dictionary, 2nd ed.

⁵¹ Owen D, ‘The Five Elements of Negligence’ 35 (4) *Hofstra Law Review*, 2007, 1672-1673.

⁵² This refers to factual causation and proximate cause.

⁵³ The test posits that the harm wouldn't have occurred were it not for the negligent act.

⁵⁴ The foreseeability test refers to contemplation of the plausibility of an act's consequence by the actor, at the time the choice was to be made;

Owen D, ‘The Five Elements of Negligence’, 1679-1685.

⁵⁵ Yannoulidis S, ‘Causation in the law of negligence’ 27(2) *Monash University Law Review*, 2001, 320;

that in every application of the ‘but-for’ test, an evaluation occurs of what probably would have happened if the defendant had not been negligent, and a comparison of that situation with what actually happened.⁵⁶ In proximate cause, he stated that one is liable if a reasonable person in his position would have foreseen the harm incurred, and it is irrelevant what the defendant did or could foresee, because what matters is that the type of harm be foreseeable to a reasonable person hence informing them to take the precautions the defendant neglected.⁵⁷

1.7.3. On the necessity of a framework to hold the user of the fully autonomous weapon system liable for harm the weapon causes in urban areas

Various scholars have acknowledged that technological progress in the field of AI continues to make possible the emergence of machines capable of taking human lives without human control thus representing a threat to humanity for which no legal framework currently exists.⁵⁸ Langley observes that the numerous ethical insights and debates regarding AWS, though useful and necessary, do not bring about the urgently required solutions for minimising the associated risks.⁵⁹ Kallenborn identifies one such risk as the incapacity of this weaponry, to adequately discriminate between protected persons such as civilians and military targets, which exacerbates the potential for unlawful harm and humanitarian crises.⁶⁰

In addition to the above risk, the ICRC recognised that these weapons pose challenges to the determination of responsibility because of the high degree of autonomy, the inherent uncertainty of output by highly complex algorithms (which increases as autonomy increases), and dynamic and unpredictable environments.⁶¹ The absence of clear liability structures could lead to a dangerous accountability gap, undermining fundamental

Barnett v Chelsea and Kensington Hospital Management Committee (1969), Queen’s Bench of The United Kingdom.

⁵⁶ Yannoulidis S, ‘Causation in the law of negligence’, 327.

⁵⁷ Yannoulidis S, ‘Causation in the law of negligence’, 326.

⁵⁸ Langley B, ‘The weaponization of artificial intelligence: What the public needs to be aware of’ *Frontiers in Artificial Intelligence*, 2023, 1-4;

Armitage R, ‘We must oppose lethal autonomous weapons systems’ 69(687) *British Journal of General Practice*, 2019, 510-511.

⁵⁹ Langley B, ‘The weaponization of artificial intelligence: What the public needs to be aware of’ *Frontiers in Artificial Intelligence*, 2023, 2.

⁶⁰ Kallenborn Z, ‘Meet the future weapon of mass destruction, the drone swarm’ *Bulletin of the Atomic Scientists*, 5 April 2021 –< [Meet the future weapon of mass destruction, the drone swarm - Bulletin of the Atomic Scientists \(thebulletin.org\)](https://thebulletin.org)> on 22 September 2024.

⁶¹ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 37-40.

principles of IHL. Furthermore, it identifies that these weapons reduce the cost of waging war for technologically advanced countries with high financial stability,⁶² and as per Taddeo and Blanchard, allow the waging of war without the need to overcome potential objections by military personnel or of a democratic populace.⁶³ The ICRC saw this as paving way for these countries to use force instead of peaceful means to settle international disputes resulting in catastrophe in humanitarian terms.⁶⁴

Hua identifies that fully AWS are systems that are capable of “learning how to learn”, which makes their decision-making process during intense time-critical combat scenarios, inscrutable to a human and difficult to subject to human supervision.⁶⁵ Despite all this, Gaeta outlines that these weapons are still considered “tools” of the user, and they are the ones that deploy the weapon system.⁶⁶ However, without a liability framework in place, the unchecked deployment of fully AWS could lead to severe and unintended consequences, undermining global security and the fundamental principles of warfare. Given these profound ethical, legal, and humanitarian concerns associated with fully AWS, the establishment of a comprehensive legal framework for holding users of fully AWS liable for harm the weaponry causes in urban settings is imperative.

1.7.4. Contribution

Scholars recognise that fully AWS present significant legal and ethical challenges due to their autonomous decision-making and the inherent uncertainty of algorithmic outputs, making it difficult to assign responsibility. Despite their independence in combat, fully AWS are still viewed as tools of the user. This study contributes to the discourse by proposing that a causation framework exactly like the one in common law negligence should be integrated into IHL to hold users of fully AWS liable for harm the weaponry causes in urban settings. The proposed legal framework attributes liability to the user based on responsibilities of a state and the proximity of their deployment decision to the harm caused. As a result, it addresses the current gap in accountability for violations of IHL by

⁶² International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 40.

⁶³ Taddeo M, Blanchard A, ‘A comparative analysis of the definitions of autonomous weapons systems’, 3.

⁶⁴ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 40.

⁶⁵ Hua S, ‘Machine learning weapons and international humanitarian law’, 126-127.

⁶⁶ Gaeta P, ‘Who acts when autonomous weapons strike? the act requirement for individual criminal responsibility and state responsibility’, 1053.

fully autonomous weapons while duly considering the complexity, autonomy, and other realities the weapon brings.

1.8. Methodology

This study employs a qualitative research approach. To answer how AWS function, their appeal, and the risks they pose, the study draws from secondary sources such as journal articles, research reports, and books on AI, machine learning, and autonomous weapon system technology. These sources provide a comprehensive understanding of the technical aspects of AWS, including how they and fully AWS operate. Additionally, these sources highlight both the advantages of AWS and their limitations. As a result, clarity would be gained on the high risk they pose to IHL-compliant target identification in urban warfare and make a case for why attribution of liability is needed for when fully AWS cause harm.

To investigate the mechanisms by which IHL attributes liability for harm caused to protected persons by conventional weapons and weapons of mass destruction (WMD), the study relies on case laws from the International Court of Justice (ICJ), among other bodies, to draw out what is considered in attribution of liability in the context of conventional weapons and WMD. To answer what challenges arise in IHL when it comes to attributing liability in the context of fully AWS, this study majorly uses scholarly work to show how the novelty of these weapons in warfare exposes the inadequacies of existing IHL regulations used in the context of conventional weapons and WMD.

Finally, to explore why a causation framework exactly like the one in common law negligence might be the appropriate framework for holding users of fully AWS liable for harm the weaponry causes in urban settings, this study uses case law and scholarly work to assess the common law causation standard in negligence, making a positive case for the standard and evaluating its suitability. Building on this, it uses scholarly works to outline a clear structure for addressing liability when fully AWS indiscriminately attack with respect to 2 instances: (1) when the weapon is deployed to an urban area and harms protected persons, and (2) when the weapon takes lethal action causing harm to protected persons in an urban area despite not being deployed to such a setting.

Therefore, to answer the main question, this study employs deductive reasoning, using both primary and secondary sources to draw connections between the various concepts and principles analysed, ultimately determining the possibility of, and making

recommendations for, integrating a causation framework exactly like the one in common law negligence under IHL.

1.9. Chapter Breakdown

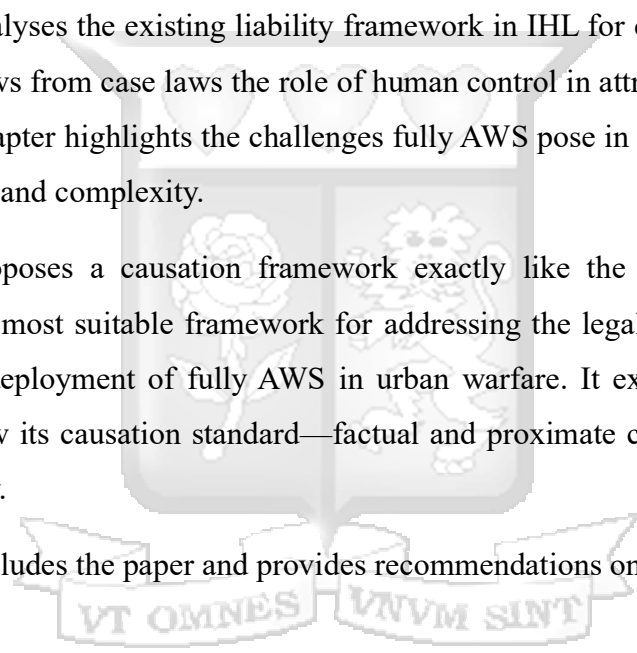
Chapter One lays out the background and context, the problem statement, the research objectives and questions, the hypothesis, among others.

Chapter Two examines the technical functionality, appeal, and risks of fully AWS. It explores AI's role in AWS, the operational advantages of these weapons, and the ethical and legal risks they pose. It also outlines the importance of a liability framework for holding users of fully AWS liable for harm the weaponry causes in urban settings.

Chapter Three analyses the existing liability framework in IHL for conventional weapons and WMD. It draws from case laws the role of human control in attributing responsibility. Thereafter, the chapter highlights the challenges fully AWS pose in proving causation due to their autonomy and complexity.

Chapter Four proposes a causation framework exactly like the one in common law negligence as the most suitable framework for addressing the legal and ethical concerns surrounding the deployment of fully AWS in urban warfare. It explains negligence, its elements, and how its causation standard—factual and proximate cause—can effectively determine liability.

Chapter Five concludes the paper and provides recommendations on the way forward.



2.0. Understanding the functionality, appeal, and risks posed by autonomous weapon systems

2.1. Introduction

Conflict has been a persistent aspect of human history, shaping societies and driving technological advancements in warfare. As the prevalence of conflicts continues in various forms, nations continuously seek to refine their military capabilities. This evolution of warfare has led to the growing desire to develop fully AWS, which promise to enhance operational efficiency, reduce human casualties, enable faster decision-making in complex combat environments, and much more.

This chapter seeks to expound on the ‘what’ and the ‘whys’ of fully AWS in 4 parts. The first part discusses the technical functionality of AWS by outlining what AI is, how it acts as the driving force of AWS, how fully AWS would arise, and the various capabilities the weaponry possesses. The second part analyses why these weapons are so attractive to users by expounding on their operational capabilities and advantages. The third part outlines the various risks posed by fully AWS. The fourth part makes a case for why a liability framework holding users of fully AWS liable for harm the weaponry causes in urban settings, is needed. The chapter then closes with a conclusion summarising the discourse.

2.2. Technical functionality of autonomous weapon systems

2.2.1. What is artificial intelligence?

AI is the ability of a machine to execute cognitive functions typically linked to the human mind, such as perception, reasoning, learning, environmental interaction, problem solving, decision making, and even exhibiting creativity.⁶⁷ It involves machine learning which according to Kelleher, focuses on the development and assessment of algorithms that enable the system to autonomously learn and discern intricate relationships, patterns, and structures within vast datasets.⁶⁸ However, from this broader field of machine learning came deep learning. Kelleher notes that deep learning allows the system to make data driven decisions by not only identifying but also extracting patterns that manifest relationships between complex inputs and outcomes.⁶⁹ It achieves this through the

⁶⁷ Rai A, Constantinides P, Sarker S, ‘Next-generation digital platforms: Toward human-ai hybrids’ 43(1) *MIS Quarterly*, 2019, 3.

⁶⁸ Kelleher, *Deep learning*, 6-7.

⁶⁹ Kelleher, *Deep learning*, 4.

utilisation of neural networks, complex structures composed of artificial neurons that process and transmit information in a manner strikingly analogous to the way the human brain functions.⁷⁰ Daei points out that the network is organised into layers which contain multiple interactive neurons that process the input data, learn patterns, and pass the transformed data to the next layer.⁷¹ The layers in a neural network may be multiple depending on the complexity of the problem at hand,⁷² thus the more layers present, the higher the depth of deep learning.⁷³

However, Kelleher warns that deep learning has a high affinity for overfitting which occurs when a model learns the training data too well, capturing noise and specific patterns that don't generalise to new data, resulting in high accuracy on the training data but poor performance on unseen data.⁷⁴ He notes that despite being a danger, this also reflects the suitability of neural networks to focus on large amounts of data.⁷⁵ Additionally, Piper asserts that deep learning has catalysed remarkable advancements by offering scalable and highly adaptable solutions to intricate, multifaceted problems, distinguishing itself from earlier AI techniques that, despite their theoretical potential, faced considerable challenges when applied to real-world complex and unpredictable scenarios.⁷⁶ The rise in computing power has been crucial in this progress, as it allows AI models to handle more extensive data and achieve greater sophistication.⁷⁷

2.2.2. Artificial intelligence in autonomous weapon systems

Hua asserts that deep learning is particularly suitable for AWS because it enables these weapons to possess the unique capacity to learn and adapt without the necessity for detailed, explicit programming for every conceivable scenario, thereby offering a level of autonomy that surpasses traditional weapon systems in terms of flexibility and decision-making

⁷⁰ Kelleher, *Deep learning*, 65.

⁷¹ Daei B, 'A layman's intro to machine learning — part 3: Neural networks and their applications' Medium, 7 May 2023 —< [A Layman's Intro to Machine Learning — Part 3: Neural Networks and Their Applications | by Bijan Daei | Medium](#)> on 13 November 2024.

⁷² Daei B, 'A layman's intro to machine learning — part 3: Neural networks and their applications' Medium, 7 May 2023 —< [A Layman's Intro to Machine Learning — Part 3: Neural Networks and Their Applications | by Bijan Daei | Medium](#)> on 13 November 2024.

⁷³ Kelleher, *Deep learning*, 68.

⁷⁴ Kelleher, *Deep learning*, 21.

⁷⁵ Kelleher, *Deep learning*, 21.

⁷⁶ Piper K, 'The case for taking ai seriously as a threat to humanity' Vox, 15 October 2020 —< [What is artificial intelligence? Your AI questions, answered. | Vox](#)> on 13 November 2024.

⁷⁷ Piper K, 'The case for taking ai seriously as a threat to humanity' Vox, 15 October 2020 —< [What is artificial intelligence? Your AI questions, answered. | Vox](#)> on 13 November 2024.

capacity.⁷⁸ This ability is made possible by neural networks in that they learn by backtracking predictions and adjusting to ensure minimal error between predicted output and the actual output hence increasing their accuracy in a manner and speed that occasionally outperforms humans.⁷⁹ Taddeo and Blanchard contend that the integration of deep learning algorithms, coupled with the enhancement of the weaponry through the incorporation of more advanced and sophisticated sensor hardware, significantly bolsters the weapon's ability to gather critical intelligence, while simultaneously improving its capacity for precise target recognition and identification, thereby allowing the system to discern and evaluate potential threats with remarkable efficiency and accuracy.⁸⁰

The development of autonomous weapons has currently given rise to two distinct types: supervised autonomous weapons, where human operators retain control and intervene as necessary, and semi-autonomous weapons, which have partial independence but engage targets selected by human operators.⁸¹ However, considering the scalable character of deep learning, the enhancement offered by rising computing power, and the scaling hypothesis, these weapons may inevitably evolve into fully AWS. These are weapons that, according to Sharp, leverage the powerful capabilities of deep learning algorithms and neuromorphic computing which mimics the functions of a human brain, allowing the weapon to expeditiously plan, anticipate, respond to complex and unstructured data from unstructured environments such as urban battlefields.⁸² Additionally, they can independently make critical decisions concerning their operational functions all without requiring human intervention.⁸³

2.3. The appeal of fully autonomous weapon systems

AWS possess a remarkable versatility that makes them highly attractive to militaries across the globe, because they offer a broad range of capabilities that can be adapted for numerous

⁷⁸ Hua S, 'Machine learning weapons and international humanitarian law', 124-127.

⁷⁹ Daei B, 'A layman's intro to machine learning — part 3: Neural networks and their applications' Medium, 7 May 2023 —< [A Layman's Intro to Machine Learning — Part 3: Neural Networks and Their Applications | by Bijan Daei | Medium](#)> on 13 November 2024.

⁸⁰ Taddeo M, Blanchard A, 'A comparative analysis of the definitions of autonomous weapons systems', 12.

⁸¹ Kallenborn Z, 'Meet the future weapon of mass destruction, the drone swarm' Bulletin of the Atomic Scientists, 5 April 2021 —< [Meet the future weapon of mass destruction, the drone swarm - Bulletin of the Atomic Scientists \(thebulletin.org\)](#)> on 22 September 2024.

⁸² Sharp C, 'Cognitive lethal autonomous weapons systems (claws)' Lieber Institute West Point, 5 November 2021 —< [Cognitive Lethal Autonomous Weapons Systems \(CLAWS\) - Lieber Institute West Point](#)> on 13 November 2024.

⁸³ Operational functions include but are not limited to the identification of potential targets, the continuous tracking of those targets, and the subsequent engagement with them; Taddeo M, Blanchard A, 'A comparative analysis of the definitions of autonomous weapons systems', 12.

applications in both civilian and military sectors. Various scholars refer to this as a case of dual-use capabilities of AI whereby tracking tools, object identification capabilities, and sophisticated visual perception algorithms initially designed for self-driving cars are now being utilised to enhance the identification, tracking, and engagement of military targets on the battlefield.⁸⁴

Greenberg highlights that militaries around the world are intrigued and drawn to these weapon systems because they reduce operational risk by removing human soldiers from hazardous situations and perform tasks faster with higher precision than humans.⁸⁵ Moreover, various scholars emphasise several other advantages including lower operational costs, greater geographical accessibility in communication-limited and/or contested territories, decreased human error, and accelerated reaction times, which all enable utility of these weapons in navigation, detection, intelligence, surveillance, reconnaissance, and so on.⁸⁶ Johnson adds that AI is increasingly being integrated into these weapon systems as a force multiplier of their remote and acoustic sensing, data processing, communications, and cyber capabilities.⁸⁷ Notably, this is all with respect to AWS in existence today: supervised autonomous weapons, and semi-autonomous weapons.

However, the growing need for more military advancements could see to the realisation of fully AWS. Sharp and Hua emphasise that, in addition to the previously mentioned appeal and capabilities AWS have, fully AWS introduce a further layer of sophistication by incorporating the critical abilities of highly efficient adaptability in unpredictable and constantly shifting environments, alongside their capacity to rapidly engage in highly complex, multi-faceted decision-making processes.⁸⁸

⁸⁴ Longpre S, Storm M, Shah R, 'Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies' MIT Science Policy Review, 2022, 48.

⁸⁵ Greenberg E, 'Wrestling with killer robots: The benefits and challenges of artificial intelligence for national security' MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021), 10 August 2021 -< [Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security · Summer 2021](#)> on 13 November 2024.

⁸⁶ Longpre S, Storm M, Shah R, 'Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies' MIT Science Policy Review, 2022, 47-49.

⁸⁷ Johnson J, 'Artificial intelligence: A threat to strategic stability' 14(1) *Strategic Studies Quarterly*, 2020, 18-19.

⁸⁸ Sharp C, 'Cognitive lethal autonomous weapons systems (claws)' Lieber Institute West Point, 5 November 2021 -< [Cognitive Lethal Autonomous Weapons Systems \(CLAWS\) - Lieber Institute West Point](#)> on 13 November 2024;

Hua S, 'Machine learning weapons and international humanitarian law', 119.

2.4. The risks posed by fully autonomous weapon systems

2.4.1. Explainability, transparency, and accountability

2.4.1.1. The 'black-box' problem

Various scholars argue that the continuous advancements in both hardware capabilities and machine learning techniques have enhanced the overall performance and efficiency of autonomous systems, but at the expense of a human's ability to interpret their decision-making processes and a corresponding diminishment in the human operator's ability to impose timely and effective intermediate interventions.⁸⁹ Various scholars argue that one of the most critical risks posed by fully AWS is their 'black-box' nature whereby the analytical process between complex inputs and the outputs is opaque and inhibits the explainability of the actions of these weapons.⁹⁰

Bathae states that the 'black-box' problem arises from the immense complexity and the high-dimensional nature of deep neural network computations.⁹¹ Various scholars note that these networks distribute their learning across countless parameters, with no single neuron or subset of neurons encoding a specific decision but instead all hundreds or thousands working together to do so, making it nearly impossible to trace how a specific decision is made.⁹² As information flows through layers, the model refines patterns in ways that are mathematically meaningful but unintelligible to humans, obscuring the reasoning behind its outputs.⁹³ This is because while humans can visualise patterns in two or three dimensions, neural networks operate in spaces with hundreds or even thousands of variables, where relationships between data points exist beyond human perception.⁹⁴

This inherent complexity makes deep learning models extraordinarily powerful but fundamentally unexplainable. This raises critical concerns in high-stake applications. This

⁸⁹ Longpre S, Storm M, Shah R, 'Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies' MIT Science Policy Review, 2022, 51.

⁹⁰ Greenberg E, 'Wrestling with killer robots: The benefits and challenges of artificial intelligence for national security' MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021), 10 August 2021 –< [Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security · Summer 2021](#)> on 13 November 2024;

Longpre S, Storm M, Shah R, 'Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies' MIT Science Policy Review, 2022, 49.

⁹¹ Bathae Y, 'The artificial intelligence black box and the failure of intent and causation' 31(2) *Harvard Journal of Law and Technology*, 2018, 901.

⁹² Bathae Y, 'The artificial intelligence black box and the failure of intent and causation', 902; Goodfellow I, Bengio Y and Courville A, *Deep learning*, The MIT Press, Cambridge, 2016, 17.

⁹³ Bathae Y, 'The artificial intelligence black box and the failure of intent and causation', 902-903.

⁹⁴ Bathae Y, 'The artificial intelligence black box and the failure of intent and causation', 903.

‘black-box’ nature exacerbates the accountability problem as decision-makers struggle to determine whether an action reflects an adversary’s intent, or whether the autonomous weapon carried out an attack that was inconsistent with the rival leader’s policies.⁹⁵ This profoundly impacts the targeted state's decision-making process regarding its response, as political and military leaders may be inclined to limit or restrain their retaliatory actions if they perceive the attack as unintentional or inadvertently carried out.⁹⁶ The challenge is further compounded by the fact that rival states or actors can intentionally misrepresent or obscure whether they were the ones who authorised the attacks, skilfully exploiting the ‘black-box’ nature of fully autonomous systems.⁹⁷

2.4.1.2. The multiple actors problem

These weapons exacerbate the accountability problem by making it increasingly difficult to hold a single actor accountable for violations of IHL due to the great multiplicity of actors involved in development and use of these weapon systems at various interlinked stages.⁹⁸ Various scholars argue that the operations of an agentic AI system are shaped by three key parties: the model developer, the system deployer, and the user.⁹⁹ The model developer creates the AI model that powers the agentic system, determining its overall capabilities and behaviours; the system deployer takes this model and integrates it into a larger system, which involves providing the model with instructions, connecting it to tools for action, creating interfaces for user interaction, and as a result of their domain-specific expertise compared to the model developer or the user, fine-tune the system for particular

⁹⁵ Greenberg E, ‘Wrestling with killer robots: The benefits and challenges of artificial intelligence for national security’ MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021), 10 August 2021 –< [Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security · Summer 2021](#)> on 13 November 2024.

⁹⁶ Greenberg E, ‘Wrestling with killer robots: The benefits and challenges of artificial intelligence for national security’ MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021), 10 August 2021 –< [Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security · Summer 2021](#)> on 13 November 2024.

⁹⁷ Greenberg E, ‘Wrestling with killer robots: The benefits and challenges of artificial intelligence for national security’ MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021), 10 August 2021 –< [Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security · Summer 2021](#)> on 13 November 2024.

⁹⁸ Christian R, ‘Mind the gap: The lack of accountability for killer robots’ Human Rights Watch, 9 April 2015 –< [Mind the Gap: The Lack of Accountability for Killer Robots | HRW](#)> on 4 December 2024.

⁹⁹ Agentic AI systems are systems that are capable of autonomously adapting their actions to achieve complex goals in dynamic environments over time, with minimal direct human supervision and without predefined behaviours.

use; and the user operates the AI system by setting its specific goals and directly overseeing its behaviour and interaction with for instance, humans.¹⁰⁰

Various scholars assert that, in some instances, a single entity may assume multiple roles within the lifecycle of an AI system, such as when a company both develops and deploys the model, thereby overseeing the entire process from creation to implementation.¹⁰¹ In other situations, roles may be distributed or shared between different entities, as exemplified by cases where one entity is responsible for training the model while another entity handles the fine-tuning, ensuring that the system performs optimally for its intended application.¹⁰² However, various scholars present a broader view where they argue that the list of actors provided above is far from exhaustive, as there are additional relevant actors that play crucial roles in the system's functioning, such as the compute providers, who supply and manage the physical infrastructure necessary for the AI system to run effectively, and various third parties that may interact with the system during its operation, contributing to its overall performance and potentially influencing its decisions through external inputs or interactions.¹⁰³

2.4.2. Unpredictability

The unpredictability of these weapons can cause fatal poor performance. This is argued by various scholars to stem from the high-level autonomy these weapon systems have. Deep learning models, by their very nature, operate autonomously by continuously reassessing their actions, modifying their approach based on new data, and iterating towards optimal solutions without human oversight.¹⁰⁴ As autonomy increases, these models transition from merely executing predefined tasks at specified times to independently determining what actions to take and how to implement them due to their capacity to write and execute their own code beyond the constraints imposed by developers, effectively overriding human control.¹⁰⁵ While developers may define high-level functions, the system itself dictates the

¹⁰⁰ Shavit Y, Agarwal S, Brundage M, 'Practices for governing agentic ai systems' OpenAI, 14 December 2023 –<[Practices for Governing Agentic AI Systems | OpenAI](#)> on 4 December 2024.

¹⁰¹ Shavit Y, Agarwal S, Brundage M, 'Practices for governing agentic ai systems' OpenAI, 14 December 2023 –<[Practices for Governing Agentic AI Systems | OpenAI](#)> on 4 December 2024.

¹⁰² Shavit Y, Agarwal S, Brundage M, 'Practices for governing agentic ai systems' OpenAI, 14 December 2023 –<[Practices for Governing Agentic AI Systems | OpenAI](#)> on 4 December 2024.

¹⁰³ Shavit Y, Agarwal S, Brundage M, 'Practices for governing agentic ai systems' OpenAI, 14 December 2023 –<[Practices for Governing Agentic AI Systems | OpenAI](#)> on 4 December 2024.

¹⁰⁴ Mitchell M, Ghosh A, Luccioni S and Pistilli G, 'Ai agents are here. What now?' Hugging Face, 13 January 2025 –<[AI Agents Are Here. What Now?](#)> on 7 March 2025.

¹⁰⁵ Mitchell M, Ghosh A, Luccioni S and Pistilli G, 'Ai agents are here. What now?' Hugging Face, 13 January 2025 –<[AI Agents Are Here. What Now?](#)> on 7 March 2025.

execution of all possible functions, making decisions dynamically in response to changing conditions.¹⁰⁶ The ability to self-modify and adapt allows deep learning models to act with increasing independence, reducing the need for human intervention while also introducing risks associated with unpredictable behaviour and potential loss of control.¹⁰⁷

Furthermore, unpredictability and fatal poor performance is consequential of the following factors. Firstly, the ICRC cautions that as these deep learning algorithms increase in complexity with the corresponding increase in autonomy, they inherently do not guarantee outputs with absolute certainty.¹⁰⁸ In 2.2.1., this study pointed out that according to Kelleher, deep learning has a high affinity for overfitting which occurs when a model has too many parameters that learn the training data too well, capturing noise and specific patterns that don't generalise to new data, resulting in high accuracy on the training data but poor performance on unseen data.¹⁰⁹ This causes the model to become overly confident in its predictions even when faced with unfamiliar data, or reinforce misleading or noisy patterns in the training set, making the model less reliable in real-world applications. This is highly common in models that have high-level depth of deep learning such as is the case for these weapon systems.¹¹⁰

Secondly, these systems rely on hardware—such as cameras or sensors—to collect input data and interact with dynamic real environments which, as noted by various scholars, can introduce ambiguities that lead to data uncertainty.¹¹¹ For instance, a sudden change of weather that obstructs the sensors hence limiting target identification. These uncertainties stem from inherent variability in the environment and limitations of the hardware and scholars highlight that they cannot be eliminated.¹¹²

Thirdly, Johnson further notes that although algorithms outpace humans with their speed, access to diverse data, and processing power, complex AI-augmented systems still rely on human-programmed assumptions and training, making them susceptible to biases and

¹⁰⁶ Mitchell M, Ghosh A, Luccioni S and Pistilli G, 'Ai agents are here. What now?' Hugging Face, 13 January 2025 –<[AI Agents Are Here. What Now?](#)> on 7 March 2025.

¹⁰⁷ Mitchell M, Ghosh A, Luccioni S and Pistilli G, 'Ai agents are here. What now?' Hugging Face, 13 January 2025 –<[AI Agents Are Here. What Now?](#)> on 7 March 2025.

¹⁰⁸ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 37-39.

¹⁰⁹ Kelleher, *Deep learning*, 21.

¹¹⁰ –< [Challenges in Deep Learning - GeeksforGeeks](#)> on 6 March 2025.

¹¹¹ Gawlikowski J, Tassi C, Ali M, Lee J, Humt M, Feng J, Kruspe A, Triebel R, Jung P, Roscher R, Shahzad M, Yang W, Bamler R and Zhu X, 'A survey of uncertainty in deep neural networks' 56(1) *Artificial Intelligence Review*, 2023, 1520-1521.

¹¹² Gawlikowski J *et al.*, 'A survey of uncertainty in deep neural networks', 1514.

errors.¹¹³ Stop Killer Robots warns that as a result, these systems may reinforce inequality through pre-programmed labels, corrupting system reliability and model integrity, amplifying discrimination, embedding systemic unfairness, and more.¹¹⁴

Lastly, there is the questionable capacity of these systems to effectively make qualitative assessments considering their natural mode of operation through numerical data.¹¹⁵ Qualitative assessments are integral to adequately discriminating between protected persons and military targets. However, Kallenborn argues that these systems are incapable of this, and the suggestion that this principle of distinction should be converted to digital format is contested by Wagner who outlines that the conversion is doubted precisely because of the aforementioned questionable capacity.¹¹⁶ Furthermore, such conversion compounds the concern by Stop Killer Robots that fully AWS reduce complex human interactions into data points causing digital dehumanisation. They argue that such systems would rather than recognise human beings in their fullness as complex rational beings, process them as objects identified by stereotypical labels or other limited programmed criteria.¹¹⁷

As a result of the above factors, coupled with understanding the geographical reach of these weapons and the absence of human intervention in their operations, there is exacerbation of the concern elaborated in a report by CNAS¹¹⁸ that a failure of the weapon could lead to multiple unintended engagements across a wide geographic area until the platform exhausts its ammunition.¹¹⁹ Their damage potential is further amplified by factors such as the speed and scale of engagements, and the volume of ordnance they can deploy.

The gravity of these concerns is exacerbated by the following considerations. Firstly, it has been noted that momentum is growing among technologically advanced nations to develop and deploy these weapons, driven by the weapon's ability to reduce operational costs and

¹¹³ Johnson J, 'Artificial intelligence: A threat to strategic stability', 29.

¹¹⁴ —<[Problems with autonomous weapons - Stop Killer Robots](#)> on 3 December 2024.

¹¹⁵ Qualitative assessments such as interpreting subtle, context-dependent clues, such as tone of voice, facial expressions, or body language;

Wagner M, 'The dehumanization of international humanitarian law: Legal, ethical, and political implications of autonomous weapon systems' 47(5) *Vanderbilt Journal of Transnational Law*, 2014, 1388.

¹¹⁶ Kallenborn Z, 'Meet the future weapon of mass destruction, the drone swarm' *Bulletin of the Atomic Scientists*, 5 April 2021 —< [Meet the future weapon of mass destruction, the drone swarm - Bulletin of the Atomic Scientists \(thebulletin.org\)](#)> on 13 November 2024;

Wagner M, 'The dehumanization of international humanitarian law', 1388.

¹¹⁷ —< [Problems with autonomous weapons - Stop Killer Robots](#)> on 4 December 2024.

¹¹⁸ This refers to the Center for a New American Security.

¹¹⁹ Center for a New American Security, *Autonomous weapons and operational risk: Ethical autonomy project*, 29 February 2016, 18-23.

bypass objections from military personnel or the public.¹²⁰ Secondly, various scholars also note that nations involved in active development of AWS are progressively focusing on the offensive applications of these advanced technologies, specifically tailoring them for use in urban environments, as opposed to more traditional or open battlefields.¹²¹

Thirdly, Klare argues that military organisations' risk-averse nature drives the trend towards greater autonomy in lethal decision-making, potentially eliminating human oversight.¹²² This is exemplified by Robert Work's assertion that the US will not delegate lethal authority to machines in military force but this stance could be challenged if a strategic competitor adopts a more permissive approach, requiring the US to reassess its position.¹²³ Greenberg and the ICRC echo Klare's concern, warning that fully AWS could lead to increased easy reliance on force by nations and a heightened risk of humanitarian catastrophe.¹²⁴

Notably, efforts to delegate lethal authority to fully AWS with no human involvement are already underway. This is evident from precursors such as Loitering Attack Munitions (LAM) such as Israel's Harop,¹²⁵ the Kargu-2,¹²⁶ United States' unmanned US X-47B

¹²⁰ International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 40;

Taddeo M, Blanchard A, 'A comparative analysis of the definitions of autonomous weapons systems', 3.

¹²¹ Longpre S, Storm M, Shah R, 'Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies' MIT Science Policy Review, 2022, 47.

¹²² Klare M, 'Autonomous weapons systems and the laws of war' 49(2) *Arms Control Today*, 2019, 9-10.

¹²³ Robert Work is the former United States Deputy Secretary of Defense;

—< [WATCH: David Ignatius and Pentagon's Robert Work on the latest tools in defense - The Washington Post](#)> on 14 November 2024.

¹²⁴ Greenberg E, 'Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security' MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021), 10 August 2021 —< [Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security · Summer 2021](#)> on 13 November 2024;

International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016, 40.

¹²⁵ Johnson J, 'Artificial intelligence: A threat to strategic stability', 20;

Systems that can partially operate in a fully autonomous, human-out-of-the-loop, mode. They autonomously pursue targets based on pre-programmed criteria, detecting enemy air defence radars and solely utilise AI to shoot down incoming projectiles at a speed far exceeding the reaction times of human operators.

¹²⁶ —< [Engineering the Next](#)> on 25 December 2024;

A weapon armed with both armor-piercing warheads and anti-personnel munitions on the same platform with an endurance of close to an hour. It was claimed to be a system that operates according to the 'man in the loop' principle but was shown to be capable of hunting down and attacking targets independent of any kind of pilot or control scheme, all without requiring data connectivity between the operator and the munition;

—< [STM Kargu loitering munition | Automated Decision Research](#)> on 1 December 2024;

—< [Doctors without borders | The Practical Guide to Humanitarian Law](#)> on 22 December 2024; *Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011)*, 8 March 2021, 17.

combat aircraft,¹²⁷ and the AI pilot, Hivemind,¹²⁸ that are not only operational but undergoing further development.

2.5. The importance of a liability framework

The need for human accountability in the use of fully AWS arises from the severe and unpredictable harms these systems can cause, the foundational principles of legal and justice theory, and the impracticality of attributing liability to the machine itself.

Firstly, as AI continues to develop at an accelerating pace, as was noted by Piper’s reference to Stuart Russell’s concern that AI’s trajectory is akin to the unforeseen discovery of nuclear chain reactions—an advancement that profoundly altered global security landscapes,¹²⁹ the risks outlined in 2.4 of this study and the potential of fully AWS to cause unintended yet catastrophic consequences remain ever-present.¹³⁰ However, various scholars note that not only does international law lag behind these advancements but IHL’s applicability to AWS also remains rather ambiguous.¹³¹ Without accountability, these harms would go unaddressed.

Secondly, legal and justice theory generally recognises that accountability must exist for harm caused to individuals who bear no fault. As Matthias highlights, the very nature of fully AWS—high complexity and full autonomy—precludes meaningful human intervention.¹³² This absence creates an accountability gap but does not negate the necessity of liability. The foundational principles of law dictate that when harm occurs, there must be a responsible party, particularly in cases where victims cannot have prevented or consented to the harm they suffer.

Finally, it is legally and conceptually nonsensical to hold fully AWS itself liable. A machine cannot possess intent, responsibility, or the ability to engage in reparative justice.¹³³ Responsibility must, therefore, rest with either those who design, deploy, or authorise the

¹²⁷ —< [X-47B Unmanned Combat Air System \(UCAS\) \(naval-technology.com\)](#)> on 13 November 2024.

¹²⁸ —< [Hivemind - Shield AI](#)> on 13 November 2024.

¹²⁹ Piper K, ‘The case for taking ai seriously as a threat to humanity’ Vox, 15 October 2020 —< [What is artificial intelligence? Your AI questions, answered. | Vox](#)> on 13 November 2024.

¹³⁰ Longpre S, Storm M, Shah R, ‘Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies’ MIT Science Policy Review, 2022, 53.

¹³¹ Longpre S, Storm M, Shah R, ‘Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies’ MIT Science Policy Review, 2022, 52.

¹³² Matthias A, ‘The responsibility gap’, 182-183.

¹³³ Christian R, ‘Mind the gap: The lack of accountability for killer robots’ Human Rights Watch, 9 April 2015 —< [Mind the Gap: The Lack of Accountability for Killer Robots | HRW](#)> on 4 December 2024.

use of fully AWS.¹³⁴ Without a robust liability framework, there would be no mechanism for redress, effectively rendering the notion that “justice will prevail” an empty ideal rather than an enforceable reality. This is especially troubling given the increasing prevalence of urban warfare where, as Bolaños notes, conflicts have been rising globally due to the tactical advantages urban settings offer over enemies, resulting in fast-paced engagements where combatants and civilians are increasingly entangled, and outcomes are less predictable.¹³⁵

Therefore, holding humans accountable for the actions of fully AWS is not only a moral and legal imperative but also an essential safeguard against the unchecked consequences of autonomous warfare. Without such accountability, victims of harm caused by fully AWS will have no avenue for justice, and the deployment of such weaponry will proceed without meaningful constraint.

2.6. Conclusion

This chapter has provided an in-depth analysis of the technical functionality, appeal, risks, and legal considerations surrounding fully AWS. As discussed, the integration of AI, particularly deep learning, has enabled these systems to process vast amounts of data, identify targets, and make operational decisions without human intervention. Their efficiency, rapid response, and reduced risk to military personnel make them attractive for modern warfare. However, these advantages come with significant ethical, legal, and strategic concerns that require rigorous scrutiny.

A key issue is the opacity of deep-learning-led decision-making. The inability to trace and explain the rationale behind actions of AWS undermines transparency and accountability. This is particularly problematic in armed conflict, where IHL mandates clear responsibility for actions taken on the battlefield. The unpredictability and uncertainty of these systems further exacerbate these concerns, particularly as military strategies increasingly favour engagements in complex urban environments where the distinction between combatants and civilians is blurred. The deployment of fully AWS in such settings heightens the risk

¹³⁴ Christian R, ‘Mind the gap: The lack of accountability for killer robots’ Human Rights Watch, 9 April 2015 –< [Mind the Gap: The Lack of Accountability for Killer Robots | HRW](#)> on 4 December 2024.

¹³⁵ Bolaños P, ‘Lawyers, guns and AI: Gaza’s new urban warfare?’ Security Distillery, 26 July 2024 –< [Lawyers, Guns and AI: Gaza’s New Urban Warfare? — The Security Distillery](#)> on 26 July 2024.

of unintended engagements, indiscriminate harm, and violations of fundamental IHL principles such as distinction, proportionality, and humanity.

Moreover, the challenges associated with these weapons extend beyond legal considerations to the broader implications of delegating lethal decision-making to machines. The inability of fully AWS to interpret nuanced human behaviours, the potential for algorithmic bias, and the risk of digital dehumanisation all raise profound questions about the moral legitimacy of their use. The diffusion of responsibility across multiple actors compounds this issue, making it difficult to determine liability when operations of fully AWS result in unlawful harm.

Given the accelerating development of these weapons as exemplified by the precursors and inclination of various nations, this chapter underscores the urgent need for a robust liability framework for holding users of fully AWS liable for harm the weaponry causes in urban settings. Without clear regulatory mechanisms, these systems risk undermining legal and ethical norms governing warfare. Given the growing complexity and prevalence of urban combat, regulatory interventions are essential to safeguard humanitarian protections. While fully AWS may offer strategic advantages, their potential to disrupt existing legal and ethical frameworks demands immediate scholarly, legal, and policy engagement to ensure their responsible use.



3.0. The liability framework governing the use of weapons in International Humanitarian Law

3.1. Introduction

In the preceding chapters, it has been hinted at that the existing liability framework in IHL is inadequate in the face of fully AWS. Through 2 parts, this chapter probes into this inadequacy by assessing how IHL attributes liability for harm caused to protected persons by conventional weapons and WMD, and what challenges arise in IHL when it comes to attributing liability in the context of fully AWS. The first part explores the role of human control and the principles of IHL in the attribution of liability for harm caused by conventional weapons and WMD. It begins by discussing what control over weaponry is understood to be, followed by an analysis of how international courts, particularly the ICJ and the Inter-American Court of Human Rights (IACtHR), have attributed liability for harm caused by conventional weapons and WMD based on the aforementioned understanding of control over weapons. This part also illustrates how such control relates to the IHL principles in such attribution of liability.

The second part critically examines the challenge of proving causation in the context of fully AWS. Building on the findings from the first part, it highlights the key differences between fully AWS, and conventional weapons and WMD in terms of human control and use in accordance with IHL principles. The part then closes by emphasising the need for IHL to address these challenges and develop a coherent legal approach to determining liability in the context of fully AWS. The chapter then closes with a summary of the findings discussed.

3.2. The role of human control and the principles of International Humanitarian Law

3.2.1. What is control over weaponry understood to be?

Frei notes that control of weaponry is in selecting the type of weapons, directing how they are used in war, and managing their impact during conflicts to prevent legal violations.¹³⁶ This is an understanding of control as involving 3 stages. The significance of control over weaponry lies in its role in upholding the fundamental principles of IHL. It ensures that

¹³⁶ Frei D, 'International humanitarian law and arms control' 28(267) *International Review of the Red Cross*, 1988, 493-494.

military operations remain targeted at combatants and military objectives, thereby minimising harm to civilians—aligning with the principle of distinction.¹³⁷ Additionally, control over weaponry reinforces the principle of humanity, which requires all humans to care for and respect one another, whether they are enemies or not.¹³⁸ As a fundamental pillar of IHL, the principle of humanity ensures that armed conflict remains guided by considerations that seek to prevent excessive destruction and harm, a crucial function of control.¹³⁹ Thus, control over weaponry is not merely a technical or strategic measure but a crucial mechanism for ensuring that warfare remains constrained and responsible.

Considering Frei's understanding of control, Médecins Sans Frontières (MSF) notes that regulation of weaponry in IHL is generally known to be in 2 parts. Firstly, it regulates the weapons themselves, for some it goes as far as to strictly prohibit production, transfer, and stockpiling.¹⁴⁰ This particularly affects the 1st stage of control because any weapon can be selected for warfare, but regulation adds the consideration of whether such selection would be permitted by law in the first place. Secondly, it regulates the use of weapons in a bid to avoid indiscriminate, inhumane, or disproportionate use.¹⁴¹ Notably, the 2nd and 3rd stages are linked such that use must always be in accordance with lawful impact. MSF relies on Article 35 of Additional Protocol I, noting that IHL limits the means and methods of warfare by prohibiting the use of weaponry to cause superfluous injury or unnecessary suffering.¹⁴²

MSF notes that conventional weaponry¹⁴³ and WMD¹⁴⁴ are governed by rules set forth in specific international conventions, including the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons and its Protocols, which only apply to states that are parties to them, with the exceptions of the Geneva Conventions and Additional Protocols which amount to customary international law that bind all states.¹⁴⁵ These

¹³⁷ Article 48, *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (protocol 1)*, 8 June 1977, 1125 UNTS 3;

Frei D, 'International humanitarian law and arms control', 494.

¹³⁸ —< [Basic principles of IHL - Diakonia International Humanitarian Law Centre](#) > on 23 December 2024.

¹³⁹ —< [Basic principles of IHL - Diakonia International Humanitarian Law Centre](#) > on 23 December 2024.

¹⁴⁰ —< [Doctors without borders | The Practical Guide to Humanitarian Law](#) > on 22 December 2024.

¹⁴¹ —< [Doctors without borders | The Practical Guide to Humanitarian Law](#) > on 22 December 2024.

¹⁴² Article 35, *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (protocol 1)*;

—< [Doctors without borders | The Practical Guide to Humanitarian Law](#) > on 22 December 2024.

¹⁴³ Weapons whose ability to damage comes from kinetic, incendiary, or explosive energy such as small arms, defensive shields, light weapons, sea and land mines, as well as bombs, shells, rockets, missiles, and cluster munitions.

¹⁴⁴ Weapons such as nuclear, biological, and chemical weapons.

¹⁴⁵ —< [Doctors without borders | The Practical Guide to Humanitarian Law](#) > on 22 December 2024.

regulations are further supplemented by various case laws. These case laws, as shown in 3.2.2, manifest how this understanding of control is instrumental in attribution of liability for harm caused by these types of weaponry and by extension, how justice is attained.

3.2.2. How have courts attributed liability for harm by weapons based on what control is understood to be?

In its advisory opinion on the Legality of the Threat or Use of Nuclear Weapons, the ICJ recognised that while no specific prohibition against nuclear weapons exists in international treaties, their use is subject to fundamental customary principles of IHL, particularly distinction, proportionality, necessity, and humanity.¹⁴⁶ The Court underscored that the impacts of nuclear weapons—explosive force, heat, and extreme radiation—are inherently indiscriminate and uncontrollable, making it impossible to distinguish between civilians and combatants or between civilian objects and military targets.¹⁴⁷ Judge Weeramantry reinforced this point by arguing that nuclear weapons, by their very nature to cause genetic damage and obliteration, defy the principles of IHL due to their indiscriminate, excessive, and devastating consequences.¹⁴⁸ This underscores that weapons whose effects cannot be controlled are inherently incompatible with IHL.

Drawing from Frei's understanding, control operates at three critical stages: selection, use, and impact. The ICJ's reasoning regarding the uncontrollable nature of the impact nuclear weapons have directly implicates the concept of control as it relates to the regulation of weaponry. The Court implicitly recognised that nuclear weapons lack the necessary degree of control at the stage of impact—such that their imprecision, ability to cause indiscriminate harm, and ability to cause large-scale unnecessary and excessive suffering cannot be limited—rendering them incapable of compliance with IHL. Their effects cannot be contained spatially or temporally, resulting not only in superfluous injury and unnecessary suffering but also environmental devastation.

While the opinion did not directly address individual liability, it suggested that states bear primary responsibility for ensuring compliance with IHL and for preventing the use of weapons that defy the principles of control.¹⁴⁹ Ultimately, the ICJ's advisory opinion highlights control of impact as a fundamental criterion in determining the legality of

¹⁴⁶ *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, ICJ Reports 1996, 52-73, 86.

¹⁴⁷ *Legality of the Threat or Use of Nuclear Weapons*, ICJ, 92-95.

¹⁴⁸ *Legality of the Threat or Use of Nuclear Weapons*, ICJ, 105.

¹⁴⁹ *Legality of the Threat or Use of Nuclear Weapons*, ICJ, 99-105.

weapons. By demonstrating that nuclear weapons fail the test of control at the level of impact, the opinion affirms that any weapon whose effects inherently cannot be meaningfully controlled, falls outside the bounds of lawful warfare, and users incur liability for their launch.¹⁵⁰ Notably, control's role and importance played a significant role in shaping the legal and political discourse that eventually led to the Treaty on the Prohibition of Nuclear Weapons that prohibits the development, testing, production, acquisition, possession, stockpiling, use, and threat of use of nuclear weapons.¹⁵¹

In *Democratic Republic of the Congo v Uganda*, the violations committed by the Uganda People's Defence Forces (UPDF) underscore the central role of control particularly at the stage of impact, as outlined by Frei, when conventional weaponry is used. In this case, every aspect of the UPDF's use of military force was determined by human decision-makers who exercised full control over their weapons. Each act of shelling,¹⁵² firing, and destruction was directed, reinforcing the idea that the harm inflicted was a direct result of deliberate human agency rather than the autonomous action of the weapons themselves.

Control over weaponry is essential because it ensures compliance with fundamental IHL principles, particularly distinction, proportionality, and necessity. Despite having full control over the use of their arms, the UPDF failed to uphold these obligations, leading to indiscriminate attacks. The shelling of Kisangani in 2000, which resulted in over 760 civilian deaths, widespread destruction of homes and public infrastructure, and the displacement of thousands, exemplifies this failure.¹⁵³ The ICJ held Uganda responsible under the principle that a State is accountable for the conduct of its military forces, irrespective of whether individual actors exceeded their authority.¹⁵⁴ This underscores the legal obligation that accompanies control: the ability to regulate the use of force imposes a duty to ensure that such force remains within the confines of lawful military engagement.

The importance of control in this context aligns with the broader regulatory framework of IHL, which prohibits the use of weapons in a manner that causes superfluous injury and

¹⁵⁰ *Legality of the Threat or Use of Nuclear Weapons*, ICJ, 99-105.

¹⁵¹ Article 1, *Treaty on the Prohibition of Nuclear Weapons*, 7 July 2017, Treaty XXVI.9.

¹⁵² The act of firing artillery or weapons at an enemy, or the heavy fire of artillery to saturate an area instead of hitting a specific target.

¹⁵³ *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda)*, Judgement, ICJ Reports 2005, 206-208.

¹⁵⁴ *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda)*, ICJ, 213-214.

unnecessary suffering.¹⁵⁵ This further demonstrates that the 2nd and 3rd stages of control are linked such that use must always be in accordance with lawful impact. The case illustrates that control is not merely a technical capability but a mechanism that should ensure warfare remains constrained and responsible. By improperly exercising their control over conventional weaponry, Ugandan forces not only violated legal obligations but also undermined the foundational IHL principles that seek to minimise harm in armed conflict, resulting in Uganda's liability.¹⁵⁶

The IACtHR, alongside the ICJ, has affirmed the salience of control as demonstrated in *Santo Domingo Massacre v Colombia*. This case underscores how control, as understood by Frei, determines compliance with IHL obligations, particularly the principles of distinction and humanity. On December 13, 1998, the Colombian Air Force deployed an AN-M1A2 cluster bomb on the village of Santo Domingo. This weapon, by design, disperses multiple submunitions over a wide area, rendering its impact inherently indiscriminate, imprecise, and beyond human control.¹⁵⁷ Despite operators exercising control in its selection and release—via a manual mechanism pulled by the pilot—the inability to direct its effects in a manner that spares civilians and civilian objects contravenes the principle of distinction, which mandates that military operations be directed solely at combatants and military objectives.¹⁵⁸ The Santo Domingo case thus exemplifies how weapon systems that largely lack precision preclude effective human control at the stage of impact, making them inherently unlawful under IHL as was the case in the advisory opinion on the *Legality of the Threat or Use of Nuclear Weapons*.

Further exacerbating the violation, survivors fleeing the attack or aiding the wounded were reportedly subjected to machine-gun fire from a helicopter.¹⁵⁹ Unlike the cluster bomb, machine guns require continuous human operation, necessitating direct and immediate control over their discharge.¹⁶⁰ Colombian Air Force regulations explicitly restricted such

¹⁵⁵ Article 35(2), *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (protocol 1)*.

¹⁵⁶ *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda)*, ICJ, 213-214.

¹⁵⁷ *Santo Domingo Massacre v Colombia*, IACtHR Judgement of 30 November 2012, (Preliminary Objections, Merits and Reparations), 215-217, 228.

¹⁵⁸ *Santo Domingo Massacre v Colombia*, IACtHR, 217.

¹⁵⁹ *Santo Domingo Massacre v Colombia*, IACtHR, 232.

¹⁶⁰ —< [Machine gun | History, Description, & Facts | Britannica](#) > on 24 December 2024.

attacks in populated areas, barring their use unless it could be ensured that civilians would not be affected.¹⁶¹ The disregard for these restrictions underscores a failure of control.

Frei's conceptualisation of control extends across three dimensions: selection, use, and impact. The Santo Domingo case illustrates that while Colombian forces retained control over the selection of the cluster bombs, they could not exercise control over their impact, resulting in imprecise and indiscriminate harm. Furthermore, despite their control over the machine gun's selection and use, they improperly exercised control over such use, resulting in indiscriminate effects. This disjunction renders the attack unlawful, as IHL does not merely demand human oversight at the point of selecting the weapon to operate but requires that control further translates into proper use and controllable effects that align with the principles of IHL. Consequently, Colombia was ordered by the Court to provide reparations to the victims, take measures to restore the community, and ensure non-repetition.¹⁶² The case thereby highlights that liability is incurred for harm by weapons unless control over weaponry is substantive, ensuring that military operations remain restrained and proportionate.

3.3. The challenge faced in proving causation under International Humanitarian Law in the context of fully autonomous weapon systems

As noted by Ighobor, autonomous weapons are new, very potent weapons majorly different from the other weapons referenced in 3.2 because of the place and role of humans in the engagements and operations of AWS.¹⁶³ As was covered in chapter 2, fully AWS present major challenges. Firstly, they inhibit explainability and transparency because of their 'black-box' nature. This is where the analytical process between complex inputs and outputs, though mathematically meaningful, is unintelligible to humans, obscuring the reasoning behind its outputs. Secondly, there is a great multiplicity of actors involved in development and use of these weapon systems at various interlinked stages hence each substantially influences the weapon. Thirdly, they have a high-level autonomy that diminishes human involvement because they can continuously reassess their actions, self-modify their approach based on new data, write and execute their own code independently, and iterate towards optimal solutions. This also makes their actions unpredictable.

¹⁶¹ *Santo Domingo Massacre v Colombia*, IACtHR, 236.

¹⁶² *Santo Domingo Massacre v Colombia*, IACtHR, 292.

¹⁶³ Ighobor K, 'Unregulated autonomous weapons systems pose risk to Africa' Africa Renewal, 17 May 2024 –<[Unregulated Autonomous Weapons Systems pose risk to Africa | Africa Renewal](#)> on 3 December 2024.

Considering the above, this study argues that the first challenge is that the understanding of control outlined in 3.2 is traditional and cannot be said to entirely apply in the context of fully AWS. One cannot directly control the use and impact of something that determines its own actions and determines such actions in a manner they cannot fully understand. This is unfortunate considering that amongst the international community, there is a great discomfort with the idea of this weaponry placing the use of force and the ability to make life or death decisions beyond the understanding of human control.¹⁶⁴

Furthermore, in 3.2.2, it was outlined that a weapon whose impact cannot be controlled is inherently unlawful and should not be selected for warfare. However, this study contests this position when it comes to fully AWS. This is because fully AWS lack legal frameworks built on decades of established practice and precedent as is the case for other weapons. As various scholars have noted, the rapid pace of technological advancement in AWS complicates the development of legal mechanisms, making it challenging to establish standards that align with their evolving capabilities.¹⁶⁵ Moreover, existing treaties and conventions regulate other types of weapons ranging from edged weapons and firearms to WMD and cluster bombs, but none specifically address fully AWS, creating a regulatory and enforcement gap.¹⁶⁶ Instead, their lawfulness is assessed based on compliance with established IHL principles—more so because the principles are considered customary international law.

However, this exposes the second challenge. Kallenborn argues that fully AWS are incapable of, for instance, adequately discriminating between protected persons such as civilians and military targets.¹⁶⁷ To adequately do so, as noted by Wagner, the principle, as well as others, would need to be converted into a digital format in order to apply it in its engagements.¹⁶⁸ He however argues that the encoding of this principle is uncertain, especially considering the questionable capacity of these weapons to effectively make qualitative assessments.¹⁶⁹ Despite all this, however, development of these weapons is

¹⁶⁴ Laufer H, 'War, weapons and watchdogs: An assessment of the legality of new weapons under international human rights law' 6(1) *Cambridge International Law Journal*, 2017, 68.

¹⁶⁵ Longpre S, Storm M, Shah R, 'Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies' MIT Science Policy Review, 2022, 52.

¹⁶⁶ —<[Doctors without borders | The Practical Guide to Humanitarian Law](#)> on 22 December 2024.

¹⁶⁷ Kallenborn Z, 'Meet the future weapon of mass destruction, the drone swarm' Bulletin of the Atomic Scientists, 5 April 2021 — <[Meet the future weapon of mass destruction, the drone swarm - Bulletin of the Atomic Scientists \(thebulletin.org\)](#)> on 13 November 2024.

¹⁶⁸ Wagner M, 'The dehumanization of international humanitarian law', 1388.

¹⁶⁹ Wagner M, 'The dehumanization of international humanitarian law', 1388.

progressively gaining momentum for the benefits they give. Given these challenges, it is imperative for IHL to develop a solution to the causation problem posed by these weapons, ensuring accountability and compliance with fundamental legal and ethical principles.

3.4. Conclusion

This chapter has demonstrated that the existing liability framework under IHL is deeply rooted in the concept of traditional control, which operates at three critical stages: the selection of a weapon, its use, and the impact of use. These stages carry specific obligations—certain weapons are inherently unlawful due to their inherently indiscriminate nature, while others, though permissible, must be used in strict compliance with IHL principles. As seen in the cases regarding conventional weapons and WMD, control over these aspects ensures that armed forces operate within the boundaries of distinction, proportionality, and humanity, thereby minimising harm to protected persons. Through an examination of international jurisprudence, particularly from the ICJ and the IACtHR, it is evident that violations stemming from failures in stage of control results in state liability.

However, fully AWS challenge this liability framework in fundamental ways. As outlined in this chapter, they introduce three major complications: (i) a lack of transparency and explainability due to their ‘black-box’ nature, (ii) a multiplicity of actors involved at various interlinked stages of development and deployment, and (iii) an unprecedented degree of autonomy that diminishes human involvement in decision-making, rendering their actions unpredictable. Unlike conventional weapons, where humans exercise direct control over targeting, firing, and cessation of attacks, fully AWS remove these critical decision-making elements from human hands. This makes it significantly more difficult to attribute liability in the same manner as other weapons, especially if liability is to be assigned based on the traditional understanding of control.

Furthermore, while the ICJ holds the position that weapons whose impact cannot be controlled are inherently unlawful, applying this reasoning to fully AWS is not straightforward. Unlike WMD, which are prohibited due to their inherently indiscriminate impact, fully AWS are not outrightly banned and lack a legal framework built on decades of established practice and precedent. Their compliance with IHL is instead assessed based on customary IHL principles. However, these weapons would then need to have these principles converted to digital format which is arguably doubted considering their questionable capacity for qualitative assessments.

Thus, fully AWS introduce additional considerations that make the current status quo on attribution of responsibility unfair and legally insufficient due to oversimplifying the issue and being dismissive of the complexities, advancements in, and benefits of AWS technology. The absence of a clear legal framework tailored to fully AWS necessitates urgent international efforts to establish precise legal standards that balance accountability with the realities of fully autonomous warfare. Without such measures, IHL risks being inadequate in addressing the profound implications of AWS, ultimately undermining its core objective of protecting civilians and ensuring the humane conduct of war.



4.0. The causation standard of common law negligence as the way forward

4.1. Introduction

As argued in the preceding chapters and as is noted by Malik, an increase in the autonomy of machines is directly proportional to a critical decline in the attribution of human responsibility, leading to a steady erosion of accountability for the outcomes produced by such systems.¹⁷⁰ Such erosion manifests itself as a grave risk considering the risks discussed in 2.4.1. and 2.4.2. on the nature of fully AWS to inhibit explainability and accountability, and be unpredictable, uncertain, and risk an escalation of conflict coupled with a high risk of failure to comply with IHL. Boulanin further notes that there are still many uncertainties about what IHL demands, allows, or forbids, particularly when it comes to interactions between humans and machines such as AWS.¹⁷¹ This ambiguity makes it challenging to determine definitively when IHL has been breached.¹⁷²

Considering this, this chapter examines how a causation framework exactly like the one in common law negligence provides the most viable approach to addressing accountability gaps when fully AWS violate IHL in international armed conflicts, particularly in urban warfare. It does so in 3 parts. The first begins by defining negligence and expounding on its core elements while demonstrating that negligence is a shared legal concept among common law jurisdictions. It then further delves into the causation standard, particularly focusing on factual causation and proximate causation, and highlights their applicability in assessing liability when deployment of fully AWS results in engagement in urban settings that causes harm to protected persons under IHL.

The second part then builds a case for a causation framework exactly like the one in common law negligence as the appropriate framework for holding users of fully AWS liable for harm the weaponry causes in urban settings. It begins by outlining the criteria that

¹⁷⁰ Malik S, 'Autonomous weapon systems: The possibility and probability of accountability' 35(3) *Wisconsin International Law Journal*, 2018, 626.

¹⁷¹ Boulanin V, Bo M, 'Three lessons on the regulation of autonomous weapons systems to ensure accountability for violations of IHL' *Humanitarian Law & Policy*, 2 March 2023 –<[Three lessons on the regulation of autonomous weapons systems to ensure accountability for violations of IHL - Humanitarian Law & Policy Blog](#)> on 25 November 2024.

¹⁷² Boulanin V, Bo M, 'Three lessons on the regulation of autonomous weapons systems to ensure accountability for violations of IHL' *Humanitarian Law & Policy*, 2 March 2023 –<[Three lessons on the regulation of autonomous weapons systems to ensure accountability for violations of IHL - Humanitarian Law & Policy Blog](#)> on 25 November 2024.

constitute a desirable liability framework before discussing how the causation standard can be applied in instances of deployment of fully AWS. The part then analyses the merits of this standard essentially making a positive case for it. Finally, the chapter concludes with a summary of the analysis and discourse of the chapter.

4.2. Defining negligence and understanding the causation standard

4.2.1. Introduction to negligence

Drawing from Alderson B in *Blyth v Birmingham Waterworks Company*, negligence is the failure to act in a manner that a reasonable person would undertake, or the act of doing something that such a person would avoid.¹⁷³ As was noted by Lord Macmillan in *Donoghue v Stevenson*, it is used to hold a person liable for carelessly causing harm and traditionally entailed four elements: duty, breach, cause, and damage.¹⁷⁴ In common law jurisdictions these elements are fundamentally similar, as courts and legal commentators have consistently recognised them as such.¹⁷⁵ This four-part framework is consistently applied across common law systems demonstrating a fundamentally similar legal foundation.¹⁷⁶ The widespread acceptance of this structure underscores the doctrinal unity of negligence across common law jurisdictions.

However, Owen, because of the misleading conflation of factual causation and proximate cause under the blanket term “cause”, states these elements as five, namely: duty of care, breach of such duty, factual causation, proximate cause, and damage.¹⁷⁷ He notes that duty refers to an obligation owed to another that binds human beings in a community, ensuring socially responsible behaviour before the fact and a basis for judging propriety of behaviour after the fact.¹⁷⁸ He posits that people have a duty imposed on them to act with reasonable care for the safety of others and themselves.¹⁷⁹

¹⁷³ A reasonable person is a person who is presumed to have the knowledge, skill, and awareness of an ordinary, prudent individual in similar circumstances and is guided by the considerations that typically govern human conduct. In situations encompassing expertise or technical knowledge and skill, the reasonable person is understood to have such level of expertise as is expected of an ordinary individual in that particular field in similar circumstances;

Blyth v Birmingham Waterworks Company (1856), England and Wales High Court Exchequer Division.

¹⁷⁴ *Donoghue v Stevenson* (1932), The United Kingdom House of Lords.

¹⁷⁵ Owen D, ‘The Five Elements of Negligence’, 1671-1672.

¹⁷⁶ This is evident in States such as the United Kingdom, United States, Canada, Australia, New Zealand, India, and more.

¹⁷⁷ Owen D, ‘The Five Elements of Negligence’, 1672-1673.

¹⁷⁸ Owen D, ‘The Five Elements of Negligence’, 1674-1675.

¹⁷⁹ Owen D, ‘The Five Elements of Negligence’, 1677.

The Blyth case outlines that to say a duty has been breached is to say that an action or conduct was unreasonable and done without due care.¹⁸⁰ This is assessed not only by demonstrating that the conduct caused harm through an analysis of how a reasonable person would have acted in the circumstances concerning the imposition of risks on others, but also by establishing a clear causal link between the damage and the defendant's negligent conduct.¹⁸¹ The latter is where factual and proximate cause come in. *Barnett v Chelsea and Kensington Hospital Management Committee* states that factual causation looks at whether the wrongful act actually caused the harm, determined by the 'but-for' test,¹⁸² while on the other hand, *Overseas Tankship (UK) Ltd v Morts Dock & Engineering Co Ltd* and *Hines v Garrett* outline that proximate cause looks at whether the act and the harm were sufficiently close rather than remote, determined by the foreseeability test.¹⁸³ The final component of a negligence claim, as noted by Owen, is harm. It refers to the injury or loss the plaintiff incurs because of the defendant's breach of duty with the fundamental purpose being to ensure the defendant compensates the plaintiff for harm unjustly caused.¹⁸⁴

4.2.2. The causation standard

Yannoulidis, relying on the Barnett case, expounded on factual causation, stating that a defendant's act would be the factual cause of a plaintiff's loss if it was a necessary condition of such loss.¹⁸⁵ Furthermore, with reliance on *State Rail Authority of New South Wales v Wiegold*, that in every application of the 'but-for' test, an evaluation occurs of what probably would have happened if the defendant had not been negligent, and a comparison of that situation with what actually happened.¹⁸⁶ Scordato further adds that this test yields largely determinate results in that they are predictable and consistent in largely similar

¹⁸⁰ *Blyth v Birmingham Waterworks Company* (1856), England and Wales High Court Exchequer Division of the United Kingdom.

¹⁸¹ *Blyth v Birmingham Waterworks Company* (1856), England and Wales High Court Exchequer Division of the United Kingdom.

¹⁸² The but-for test posits that the harm wouldn't have occurred were it not for the negligent act; *Barnett v Chelsea and Kensington Hospital Management Committee* (1969), Queen's Bench of The United Kingdom.

¹⁸³ The foreseeability test refers to contemplation of the plausibility of an act's consequence by the actor, at the time the choice was to be made;

Overseas Tankship (UK) Ltd v Morts Dock & Engineering Co Ltd (The Wagon Mound No 1) (1961), The Supreme Court of New South Wales;

Hines v Garrett (1921), The Supreme Court of Virginia in the United States.

¹⁸⁴ Owen D, 'The Five Elements of Negligence', 1685.

¹⁸⁵ Yannoulidis S, 'Causation in the law of negligence', 320;

Barnett v Chelsea and Kensington Hospital Management Committee (1969), Queen's Bench of The United Kingdom.

¹⁸⁶ Yannoulidis S, 'Causation in the law of negligence', 327;

State Rail Authority of New South Wales v Wiegold (1991), High Court of Australia.

conditions.¹⁸⁷ However, the downside of this test is that, as is put by Yannoulidis, any event has numerous antecedent conditions.¹⁸⁸ Conditions that may not be legally significant hence necessitating courts to distinguish between legally significant conditions and mere background circumstances.

In proximate cause, Lord Macmillan in *Glasgow Corporation v Muir* stated that one is liable if a reasonable person in his position would have foreseen the harm incurred, and it is irrelevant what the defendant did or could foresee, because what matters is that the type of harm be foreseeable to a reasonable person hence informing them to take the precautions the defendant neglected.¹⁸⁹ Notably, proximate cause is considered by various scholars and cases such as *Derdiarian v Felix Contracting Corporation* as being elusive and uncertain in that it cannot be reduced to absolute rules because it is determined on the facts of each case.¹⁹⁰ The focus is on whether the defendant's actions can be considered blameworthy within the broader context of justice, societal norms, and other considerations rather than relying solely on a rigid or overly technical assessment of what was foreseeable in a literal sense.¹⁹¹ In essence, it is about persuading the court that holding the defendant accountable aligns with principles of fairness and responsibility.

4.3. The causation standard of common law negligence as the ideal framework

4.3.1. The criteria for a desirable liability framework

This study holds the express position that a causation framework exactly like the one in common law negligence is the most appropriate liability framework to be used to hold users of fully AWS liable in the context of harm caused by the weaponry to protected persons due to engagement in urban settings. Therefore, it is important to understand the metric this study uses to hold such a position. A desirable causation framework should clearly establish the causal link between the actions of a subject and the resulting harm while ensuring fairness by duly considering relevant circumstances, interdisciplinary insights, and fault. It should also be flexible enough to adapt to future changes, such as technological

¹⁸⁷ Scordato M, 'Three kinds of fault: Understanding the purpose and function of causation in tort law' 77(1) *University of Miami Law Review*, 2022, 202.

¹⁸⁸ Yannoulidis S, 'Causation in the law of negligence', 320.

¹⁸⁹ *Glasgow Corporation v Muir* (1943) The United Kingdom House of Lords.

¹⁹⁰ Owen D, 'The Five Elements of Negligence', 1681-1682;

Scordato M, 'Three kinds of fault', 202;

Lior A, 'The "accident network": A network theory analysis of proximate causation' 106(2) *Marquette Law Review*, 2023, 416;

Derdiarian v Felix Contracting Corporation (1980), The New York Court of Appeals of the United States.

¹⁹¹ Scordato M, 'Three kinds of fault', 208.

advancements and evolving common practices, in a timely manner. Additionally, the framework must provide certainty and predictability to enable individuals and entities to anticipate legal repercussions. Finally, it must not deprive society of substantial benefits.

4.3.2. Why a causation standard exactly like the one in common law negligence offers a desirable path forward

4.3.2.1. Application of the ‘but-for’ test

This study posits the following application of factual causation. When a user deploys fully AWS directly to an urban area, resulting in the indiscriminate attack of civilians, the ‘but-for’ test would be satisfactory in attributing liability because the harm they cause would not have occurred were it not for their deployment. However, this test is not to be applied in isolation. This study argues that the test’s over-inclusivity as outlined in 4.2.2 works in this paper’s favour because it allows the inclusion of insights that would complement the application of the test in various scenarios. Insights from logic, common sense, circumstance, various disciplines including international law and technology pertaining to fully AWS, and more.

One such complementary insight is knowledge of the functionality of fully AWS. Article 36 of Additional Protocol 1 obligates states to have sufficient technical and operational knowledge of how a weapon works in order to assess its legality under IHL prior to determining its usage.¹⁹² Therefore, it is legally expected that the user knows the functionality of fully AWS, their capabilities, risks such as high degree of opacity, unpredictability and potential for grave harm, and the role multiple actors play in developing these weapons.

Another insight is understanding the reality fully AWS bring with respect to how control of the weapon is to be understood. It mandates a new understanding of control being through the pre-deployment stage. Limitations, such as biases introduced by programmers and limitations of hardware, ought to be known by the user and potential consequential risks be foreseeable to them due to the pre-deployment stage necessitating an evaluation and assessment of the weapon. The stage also requires an assessment of the environment the weapon is to be deployed to in a bid to ensure adherence to IHL.

¹⁹² Article 36, *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (protocol 1)*.

The above nullify an attempt by the user of fully AWS to absolve themselves of liability or deem the ‘but-for’ test unfair by stating that civilians acted beyond the manner initially predicted resulting in the harm inflicted by the weapon in an urban setting. This is because the supplementary insights necessitate a user to know that the weapon is limited, and through an assessment of an urban area, to know that it is an area inherently ridden with unpredictable dynamic behaviour of people. Furthermore, considering the user of fully AWS is expected to know the weapon’s functionality, limitations and actors who may influence those limitations, and exercise control over these weapons in the pre-deployment and deployment stages, they are placed in a position where the ‘but-for’ test most effectively attributes direct causal responsibility to them rather than another actor. This direct causal link between the act of deployment and the harm inflicted by the fully autonomous weapon strengthens the case for their liability, rather than that of other actors involved in the weapon’s development.

The application of the test alongside complementary insights and factors, such as the one above, strengthen the application of the test against possible counterarguments and allows the test to fairly and justly isolate the user’s decision to deploy the weapon in an urban area as the sole causal link to the resulting harm. However, this paper concedes that the test’s over-inclusive nature could also capture numerous insignificant factors hence the determination of which factors are legally relevant should be made by the court on a case-by-case basis.

4.3.2.2. Application of the foreseeability test

This study suggests that should a target be in an environment where, though not within an urban area, they can be reasonably foreseen to relatively easily find their way to such area upon their attempt to survive, yet fully AWS—given their capabilities—are still deployed and cause harm to civilians in the urban area, then proximate cause would suffice in attribution of liability for such harm.

This foreseeability test allows the determination of cause to be supplemented by other disciplines and knowledge because of its nature to encompass broader considerations as stated in 4.2.2. The test can therefore be supplemented by knowledge on fully AWS and the realities it brings such as the understanding that one of the crucial ways fully AWS are controlled is through the pre-deployment stage where evaluations of the weapon and location the weapon is to be deployed to are conducted. The test is also enhanced by

disciplines such as law such as the IHL obligation that all parties to a conflict must take all feasible precautions to avoid or minimise civilian harm.¹⁹³ These insights are further buttressed by the reasonable person standard. In situations encompassing expertise or technical knowledge and skill, a reasonable person in this standard is understood to have such level of expertise as is expected of an ordinary individual in that field in similar circumstances.

The above manifest an expectation that the user should understand the functionality, capabilities, limitations, and risks of fully AWS. Additionally, that users are expected to understand that as a result of deep learning, these weapons, by design, possess advanced capabilities to independently search, track, and attack targets with a degree of autonomy that eliminates human intervention. Furthermore, that they can adapt to dynamic environments, including densely populated urban areas, without direct human input. Therefore, the potential for harm in urban areas should not only be anticipated but also deemed an inherent risk of deploying such systems, regardless of their initial deployment location.

4.3.2.3. A positive case for a causation standard exactly like the one in common law negligence

The above applications would allow a nuanced, fair, satisfactorily substantiated and reasonable approach when it comes to attribution of liability when fully AWS cause harm in urban areas. It achieves this through its utility of other fields of knowledge when assessing actions or omissions that were unreasonable as outlined in 4.3.2. This causation standard ensures that liability is not imposed in ignorance because of its interdisciplinary approach such as knowledge regarding fully AWS and the interconnected roles played by various actors in the system's deployment and operation. This prevents courts from imprecisely attributing fault.

For instance, it is suggested that programmers should be singled out and held accountable for harm caused by these systems despite the self-learning capacity of the weapon that causes the programmer to be incapable of seeing beyond the original program design because it no longer remains the same, rendering the succeeding actions entirely

¹⁹³ Article 57, *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (protocol I)*.

unforeseeable to them.¹⁹⁴ This causation standard, through its utility of complementary insights, is enlightened. It duly considers that states (the users as per this study) are the most appropriate subjects of liability because they are obligated by Article 35 of Additional Protocol 1 to know the functionality of fully AWS, their capabilities, limitations and risks, and the role multiple actors play in developing these weapons thus making them accountable for using the weapon in a manner that breaches a duty under IHL. Moreover, it attributes liability to the user not only based on responsibilities of a state but also the proximity of their deployment decision to the harm caused.

Additionally, fairness is upheld through a contextual assessment of each case, considering factors such as the foreseeability of harm, the reasonableness of precautions taken, and the broader context of justice and societal norms. Such an approach prevents unjust outcomes that might arise from rigid or overly technical liability standards in turn preventing the hindrance of innovation and progress as is the case in strict liability.¹⁹⁵ This allows the causation standard to be better placed to gain traction considering Abungu and Iradukunda's note that the US and the UK have embraced a pro-innovation strategy focused on managing AI risks without hindering its development and use, emphasising that regulations, including those addressing liability, must be crafted to support, rather than constrain, AI innovation.¹⁹⁶

This standard is inherently well-suited to keep up with the rapid advancement of technology and warfare. Its reliance on principles such as foreseeability and reasonableness allow it to evolve alongside advancements in AWS. As these systems become more sophisticated, the legal reasonable person standard can be adjusted on a case-by-case basis without necessitating a complete overhaul of the legal framework. This ensures that liability determinations remain relevant as technology progresses.

The interdisciplinary approach harnessed by the causation standard not only ensures that emerging practices in AI weapon development and deployment are reflected in legal

¹⁹⁴ Mann J, 'Autonomous weapons systems and the liability gap, part two: Civil liability and state responsibility' Rethinking SLIC, 3 October 2019 –< [Autonomous Weapons Systems and the Liability Gap, Part Two: Civil Liability and State Responsibility - Rethinking SLIC](#)> on 25 November 2024.

¹⁹⁵ Asaro P, 'The liability problem for autonomous artificial agents', Association for the Advancement of Artificial Intelligence, 2016 AAAI Spring Symposium Paper Number 4, 2016, 193–<[The Liability Problem for Autonomous Artificial Agents - AAAI](#)> on 29 November 2024.

¹⁹⁶ Abungu C, Iradukunda M, 'Some underrated reasons why the ai safety community should reconsider its embrace of strict liability' EA Forum, 8 April 2024 –<[Some underrated reasons why the AI safety community should reconsider its embrace of strict liability — EA Forum](#)> on 8 January 2024.

assessments but also tackles the reality fully AWS bring out such as a new understanding of control. Moreover, the interdisciplinary approach prevents oversimplified or one-dimensional liability determinations ensuring a just and equitable application of the law. This approach prevents the presumption of the user being overridden by the imperfection of the weapon, fault be damned, as is the case in strict liability, because it would be unreasonable and excessively onerous.¹⁹⁷

4.4. Conclusion

This chapter has demonstrated that a causation framework mirroring the common law negligence causation standard presents the most effective approach to addressing accountability gaps in the use of fully AWS in international armed conflicts, particularly in urban warfare. By establishing liability through the core principles of negligence—duty, breach, factual causation, proximate cause, and harm—this framework ensures a fair, balanced, and legally sound method for attributing responsibility to users of fully AWS.

The analysis has illustrated that the factual causation standard, through the ‘but-for’ test, effectively isolates the deployment decision as the critical causal link to the harm inflicted by fully AWS. This is further reinforced by interdisciplinary insights regarding the nature and functionality of fully AWS, and the expectation that users must be fully aware of the weapon's capabilities, risks, and operational constraints. Likewise, the proximate causation standard, guided by foreseeability, ensures that liability is appropriately assigned when harm to civilians is a reasonably predictable consequence of deployment decisions.

The chapter has also established that causation standard offers a principled and flexible approach that evolves with technological advancements while ensuring accountability aligns with fundamental legal and ethical considerations. Ultimately, a causation standard exactly like the one in common law negligence provides a robust and pragmatic liability framework. It acknowledges the complexities of AWS deployment, upholds the principles of fairness and justice, and ensures that users remain accountable for their decisions in compliance with IHL. By adopting such a framework, legal systems can better address the challenges posed by fully AWS while maintaining a coherent and enforceable structure of responsibility.

¹⁹⁷ Abungu C, Iradukunda M, ‘Some underrated reasons why the ai safety community should reconsider its embrace of strict liability’ EA Forum, 8 April 2024 —<[Some underrated reasons why the AI safety community should reconsider its embrace of strict liability — EA Forum](#)> on 8 January 2024.

5.0. Conclusion and recommendations

5.1. Conclusion

This study has systematically explored the legal, ethical, and operational challenges posed by the deployment of fully AWS in urban warfare when engaging in international armed conflict, alongside the inadequacies of the existing IHL liability framework.

The findings of this study reveal that the integration of AI, particularly deep learning, into AWS fundamentally alters the dynamics of warfare by creating a pathway to fully AWS, weapons entirely driven by deep learning. While these systems offer significant operational advantages, such as enhanced efficiency, rapid decision-making, and reduced risk to military personnel, they also introduce profound ethical, legal, and strategic challenges. The opacity of deep-learning-led decision-making, often referred to as the "black-box" problem, complicates efforts to ensure transparency and accountability. This unpredictability heightens the risk of unintended engagements, indiscriminate harm, and violations of core IHL principles such as distinction, proportionality, and humanity, particularly in dynamic and unpredictable environments.

The existing liability framework under IHL is deeply rooted in a traditional understanding of the concept of human control where a human exercises their will and decision-making capability in every aspect of selection, use, and impact of the weaponry. However, this traditional understanding is ill-equipped to address the unique challenges posed by fully AWS. It has been argued that unlike conventional weapons or WMD, fully AWS operate with a high degree of autonomy, removing critical decision-making elements from human hands. This shift necessitates a re-evaluation of how liability is attributed in the context of fully AWS. When it comes to fully AWS, control can only be understood to be at the stage of pre-deployment and the stage of deployment itself. Without a clear and coherent legal approach that has such an understanding of control, the deployment of fully AWS risks undermining the fundamental principles of IHL and eroding accountability for unlawful harm.

To bridge this gap, this discourse has argued for the adoption of a causation framework exactly like the one in common law negligence as the most viable framework for holding users of fully AWS liable for harm the weaponry causes in urban settings. This approach offers a structured and principled method for attributing liability when fully AWS cause harm to protected persons in urban settings. The causation standard provides a balanced,

fair, contextually sensitive and adaptable framework that ensures accountability without stifling technological innovation. This framework is equipped with the reasonable persons test and harnesses an interdisciplinary and logical approach that allows for a case-by-case assessment that is responsive to the evolving nature of AWS technology and the new understanding of control that fully AWS demand. It ensures that accountability is not imposed arbitrarily or ignorantly.

In conclusion, this dissertation has made a compelling case for the adoption of a causation framework exactly like the one in common law negligence under IHL to address the accountability gaps arising from the deployment of fully AWS in urban warfare. As autonomous warfare advances and continuously demands acceptance of certain realities, this adaptable approach will help maintain legal relevance, protect civilians, and uphold the fundamental principles of responsible warfare.

5.2. Recommendations

5.2.1. Integration of a causation framework exactly like the one in common law negligence into International Humanitarian Law through judicial interpretation

Given the challenges of securing multilateral consensus for a dedicated treaty on AWS accountability, this study proposes that a causation framework exactly like the one in common law negligence be incorporated into IHL primarily through judicial interpretation in international legal judgments and advisory opinions issued by bodies such as the ICJ, IACtHR, and more. Such legal integration would allow for progressive adaptation of liability principles to technological developments without the extensive political and procedural hurdles associated with treaty negotiations.

5.2.2. Creation of an international independent autonomous weapon systems reporting ombudsman

To ensure adherence to IHL and prevent accountability gaps arising from the use of fully AWS, an international independent AWS reporting ombudsman ought to be established. This body will function as an impartial oversight mechanism, free from state, military, or private sector influence, ensuring transparency, accountability, and legal compliance in the development, deployment, and use of AWS.

States and entities deploying fully AWS in international armed conflict must submit comprehensive pre-deployment and post-deployment evaluation reports to the ombudsman. These reports will be reviewed for completeness, accuracy, and adherence to legal standards. The ombudsman will have the authority to initiate independent investigations into alleged fully AWS-related IHL violations, based on formal complaints or credible information. Investigations will include analysing operational data, such as engagement logs, video footage, and sensor recordings. Where evidence of violations is found, the ombudsman may refer cases to international courts, providing documented findings and recommendations for legal action or policy reform.

To ensure ongoing compliance, the ombudsman will conduct periodic reviews of AWS operations, requiring states and entities to maintain documentation on system functionality and corrective measures taken in response to past violations. Proof of adherence to legal standards governing AWS procurement, testing, and battlefield use must also be provided. In cases of non-compliance, the ombudsman may issue formal warnings, publish public reports, and, where necessary, recommend sanctions or legal action.

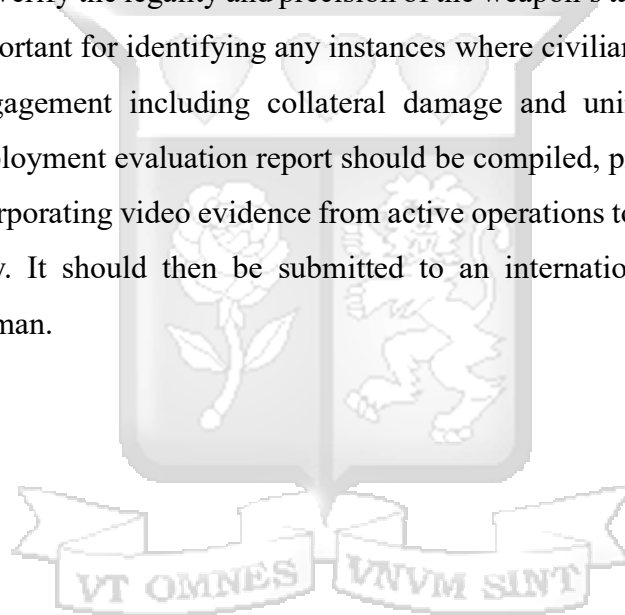
Operating under a legally binding framework, the ombudsman will be governed by a panel of legal, military, human rights, and technology experts, ensuring a well-rounded and informed oversight approach. It will possess full investigative authority, including the power to request classified documents, and conduct field inquiries, thereby strengthening accountability and safeguarding IHL principles in autonomous warfare.

5.2.3. Legal requirement for states to enact pre-deployment and post-deployment evaluation reporting

Prior to deployment, a rigorous assessment must be conducted to verify the readiness of fully AWS and evaluate the operational environment. The pre-deployment evaluation should encompass a thorough weapon system readiness assessment, ensuring that the weapon's targeting capabilities comply with IHL principles. This includes testing the accuracy of its sensors to prevent unlawful engagements. Additionally, an environmental and operational risk assessment must be conducted. This assessment should include an analysis of proximity to urban settings, civilian population density, proximity to protected infrastructure such as hospitals and schools, and the potential risk of unintended engagements due to dynamic battlefield conditions. Finally, a pre-deployment evaluation report must be submitted to an international independent AWS reporting ombudsman,

detailing the results of the technical readiness tests, findings from the environmental risk assessment, precautionary measures employed, and any and all previous instances of poor performance of the weapon in prior engagements in war and the corrective measure taken.

Following deployment of the weapon, a post-deployment evaluation is crucial to assess the system's battlefield performance, compliance with IHL, and any instances of unintended harm. This process should include a battlefield performance analysis, reviewing whether the weapon's operations adhered to the principles of proportionality, humanity, and distinction. The accuracy and effectiveness of its engagements must be scrutinised to determine if they were consistent with IHL guidelines and established rules of engagement. A video and sensor data analysis must also be conducted, using recorded battlefield footage and sensor logs to verify the legality and precision of the weapon's targeting decisions. This is particularly important for identifying any instances where civilian harm occurred due to the weapon's engagement including collateral damage and unintended engagements. Finally, a post-deployment evaluation report should be compiled, providing a summary of all the above, incorporating video evidence from active operations to enhance transparency and accountability. It should then be submitted to an international independent AWS reporting ombudsman.



Bibliography

Books

1. Goodfellow I, Bengio Y and Courville A, *Deep learning*, The MIT Press, Cambridge, 2016.
2. Kelleher J, *Deep learning*, The MIT Press, Cambridge, 2019.

Dictionaries

3. The Black's Law Dictionary, 2nd ed.

Foreign Case Law

4. *Barnett v Chelsea and Kensington Hospital Management Committee* (1969), Queen's Bench of The United Kingdom.
5. *Blyth v Birmingham Waterworks Company* (1856), England and Wales High Court Exchequer Division.
6. *Derdiarian v Felix Contracting Corp* (1980), The New York Court of Appeals of the United States.
7. *Donoghue v Stevenson* (1932), The United Kingdom House of Lords.
8. *Glasgow Corporation v Muir* (1943) The United Kingdom House of Lords.
9. *Hines v Garrett* (1921), The Supreme Court of Virginia in the United States.
10. *Overseas Tankship (UK) Ltd v Morts Dock & Engineering Co Ltd (The Wagon Mound No 1)* (1961), The Supreme Court of New South Wales.
11. *State Rail Authority of New South Wales v Wiegold* (1991), High Court of Australia.

Hard Copy Journals

12. Armitage R, 'We must oppose lethal autonomous weapons systems' 69(687) *British Journal of General Practice*, 2019.
13. Bathaee Y, 'The artificial intelligence black box and the failure of intent and causation' 31(2) *Harvard Journal of Law and Technology*, 2018.
14. Frei D, 'International humanitarian law and arms control' 28(267) *International Review of the Red Cross*, 1988.
15. Fuzaylova E, 'War torts, autonomous weapon systems, and liability: Why a limited strict liability tort regime should be implemented' 40(3) *Cardozo Law Review*, 2019.
16. Gaeta P, 'Who acts when autonomous weapons strike? the act requirement for individual criminal responsibility and state responsibility' 21(5) *Journal of International Criminal Justice*, 2023.

17. Gawlikowski J, Tassi C, Ali M, Lee J, Humt M, Feng J, Kruspe A, Triebel R, Jung P, Roscher R, Shahzad M, Yang W, Bamler R and Zhu X, 'A survey of uncertainty in deep neural networks' 56(1) *Artificial Intelligence Review*, 2023.
18. Hua S, 'Machine learning weapons and international humanitarian law: Rethinking meaningful human control' 15(1) *Georgetown Journal of International Law*, 2019.
19. Johnson J, 'Artificial intelligence: A threat to strategic stability' 14(1) *Strategic Studies Quarterly*, 2020.
20. Klare M, 'Autonomous weapons systems and the laws of war' 49(2) *Arms Control Today*, 2019.
21. Lazarsfeld S, 'The courage for imperfection' 47(1) *Individual Psychology: Journal of Adlerian Theory, Research & Practice*, 1991.
22. Laufer H, 'War, weapons and watchdogs: An assessment of the legality of new weapons under international human rights law' 6(1) *Cambridge International Law Journal*, 2017.
23. Lior A, 'The "accident network": A network theory analysis of proximate causation' 106(2) *Marquette Law Review*, 2023.
24. Malik S, 'Autonomous weapon systems: The possibility and probability of accountability' 35(3) *Wisconsin International Law Journal*, 2018.
25. Matthias A, 'The responsibility gap: Ascribing responsibility for the actions of learning automata' 6(3) *Ethics and Information Technology*, 2004.
26. Owen D, 'The Five Elements of Negligence' 35(4) *Hofstra Law Review*, 2007.
27. Rai A, Constantinides P, Sarker S, 'Next-generation digital platforms: Toward human-ai hybrids' 43(1) *MIS Quarterly*, 2019.
28. Schuller A, 'At the crossroads of control: The intersection of artificial intelligence in autonomous weapon systems with international humanitarian law' 8(8) *Harvard National Security Journal*, 2017.
29. Scordato M, 'Three kinds of fault: Understanding the purpose and function of causation in tort law' 77(1) *University of Miami Law Review*, 2022.
30. Taddeo M, Blanchard A, 'A comparative analysis of the definitions of autonomous weapons systems' 28(37) *Science and Engineering Ethics*, 2022.
31. Wagner M, 'The dehumanization of international humanitarian law: Legal, ethical, and political implications of autonomous weapon systems' 47(5) *Vanderbilt Journal of Transnational Law*, 2014.

32. Yannoulidis S, 'Causation in the law of negligence' 27(2) *Monash University Law Review*, 2001.

ICJ Case Law

33. *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda)*, Judgement, ICJ Reports 2005.
34. *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, ICJ Reports 1996.

Inter-American Court of Human Rights

35. *Santo Domingo Massacre v Colombia*, IACtHR Judgement of 30 November 2012, (Preliminary Objections, Merits and Reparations).

International Instruments

36. *Articles on state responsibility for internationally wrongful acts*, ILC 53rd Report, 2001, UN Doc A/RES/56/83 (2001).
37. *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (protocol 1)*, 8 June 1977, 1125 UNTS 3.
38. *Treaty on the Prohibition of Nuclear Weapons*, 7 July 2017, Treaty XXVI.9.

Institutional Authors

39. Article 36, *Killer Robots: UK Government Policy on Fully Autonomous Weapons*, 19 April 2013.
40. Center for a New American Security, *Autonomous weapons and operational risk: Ethical autonomy project*, 29 February 2016.
41. Congressional Research Service, *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems*, 1 February 2024.
42. International Committee of the Red Cross, *Autonomous weapon systems implications of increasing autonomy in the critical functions of weapons*, 16 March 2016.

Online Journals

43. Buchanan R, 'History of technology' *Encyclopaedia Britannica*, 2023 —<[History of technology | Evolution, Ages, & Facts | Britannica](#)>

Other Internet Resources

44. Abungu C, Iradukunda M, 'Some underrated reasons why the ai safety community should reconsider its embrace of strict liability' EA Forum, 8 April 2024 —<[Some](#)

- [underrated reasons why the AI safety community should reconsider its embrace of strict liability — EA Forum](#)>
45. —< [Algorithm | AI Glossary | OpenTrain AI](#)>
 46. —<[Arms Race: Definition, Cold War & Nuclear Arms | HISTORY](#)>
 47. Bashir D, Kurenkov A, ‘The ai scaling hypothesis’ Last Week in AI, 5 August 2022 —< [The AI Scaling Hypothesis \(lastweekin.ai\)](#)>
 48. —<[Basic principles of IHL - Diakonia International Humanitarian Law Centre](#)>
 49. Bolaños P, ‘Lawyers, guns and AI: Gaza’s new urban warfare?’ Security Distillery, 26 July 2024 —<[Lawyers, Guns and AI: Gaza’s New Urban Warfare? — The Security Distillery](#)>
 50. Boulanin V, Bo M, ‘Three lessons on the regulation of autonomous weapons systems to ensure accountability for violations of IHL’ Humanitarian Law & Policy, 2 March 2023 —<[Three lessons on the regulation of autonomous weapons systems to ensure accountability for violations of IHL - Humanitarian Law & Policy Blog](#)>
 51. Christian R, ‘Mind the gap: The lack of accountability for killer robots’ Human Rights Watch, 9 April 2015 —< [Mind the Gap: The Lack of Accountability for Killer Robots | HRW](#)>
 52. —< [Challenges in Deep Learning - GeeksforGeeks](#)>
 53. Daei B, ‘A layman’s intro to machine learning — part 3: Neural networks and their applications’ Medium, 7 May 2023 —< [A Layman’s Intro to Machine Learning — Part 3: Neural Networks and Their Applications | by Bijan Daei | Medium](#)>
 54. —<[Doctors without borders | The Practical Guide to Humanitarian Law](#)>
 55. —< [Engineering the Next](#)>
 56. Greenberg E, ‘Wrestling with killer robots: The benefits and challenges of artificial intelligence for national security’ MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021), 10 August 2021 —< [Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security · Summer 2021](#)>
 57. —<[Hivemind - Shield AI](#)>
 58. Ighobor K, ‘Unregulated autonomous weapons systems pose risk to Africa’ Africa Renewal, 17 May 2024 —<[Unregulated Autonomous Weapons Systems pose risk to Africa | Africa Renewal](#)>

59. Kallenborn Z, 'Meet the future weapon of mass destruction, the drone swarm' Bulletin of the Atomic Scientists, 5 April 2021 –< [Meet the future weapon of mass destruction, the drone swarm - Bulletin of the Atomic Scientists \(thebulletin.org\)](#)>
60. Lestari A, 'AWS: A threat to international humanitarian law or a necessary technological evolution?' Modern Diplomacy, 4 April 2024 –< [AWS: A Threat to International Humanitarian Law or a Necessary Technological Evolution? - Modern Diplomacy](#)>
61. –< [Lethal Autonomous Weapon Systems \(LAWS\) – UNODA](#)>
62. –< [Machine gun | History, Description, & Facts | Britannica](#)>
63. Mann J, 'Autonomous weapons systems and the liability gap, part two: Civil liability and state responsibility' Rethinking SLIC, 3 October 2019 –< [Autonomous Weapons Systems and the Liability Gap, Part Two: Civil Liability and State Responsibility - Rethinking SLIC](#)>
64. Mitchell M, Ghosh A, Luccioni S and Pistilli G, 'Ai agents are here. What now?' Hugging Face, 13 January 2025 –< [AI Agents Are Here. What Now?](#)>
65. Piper K, 'The case for taking ai seriously as a threat to humanity' Vox, 15 October 2020 –< [What is artificial intelligence? Your AI questions, answered. | Vox](#)>
66. –< [Problems with autonomous weapons - Stop Killer Robots](#)>
67. –< [Protected persons | ICRC](#)>
68. Quote by Stacy Pettyjohn, Director of the Defense Program at Center for a New American Security in the documentary 'Unknown: Killer Robots' directed by Jesse Sweet in 2023.
69. Saltz J, 'What is the ai life cycle' Data Science Process Alliance, 31 March 2024 –< [What is the AI Life Cycle? - Data Science Process Alliance \(datascience-pm.com\)](#)>
70. Schwarz E, 'Gaza war: Israel using AI to identify human targets raising fears that innocents are being caught in the net' Queen Mary University of London, 17 April 2024 –< [Gaza war: Israel using AI to identify human targets raising fears that innocents are being caught in the net - Queen Mary University of London \(qmul.ac.uk\)](#)>
71. Sharp C, 'Cognitive lethal autonomous weapons systems (claws)' Lieber Institute West Point, 5 November 2021 –< [Cognitive Lethal Autonomous Weapons Systems \(CLAWS\) - Lieber Institute West Point](#)>

72. Shavit Y, Agarwal S, Brundage M, ‘Practices for governing agentic ai systems’ OpenAI, 14 December 2023 –<[Practices for Governing Agentic AI Systems | OpenAI](#)>
73. –<[STM Kargu loitering munition | Automated Decision Research](#)>
74. –<[Switchblade® 600 Loitering Munition Systems | Kamikaze Drones | Suicide Drone | Tactical Missile Systems | AeroVironment, Inc. \(avinc.com\)](#)>
75. –< [The Scaling Hypothesis · Gwern.net](#)>
76. Trabucco L, ‘What is meaningful human control, anyway? cracking the code on autonomous weapons and human judgment’ Modern War Institute, 2023 –< [What is Meaningful Human Control, Anyway? Cracking the Code on Autonomous Weapons and Human Judgment - Modern War Institute \(westpoint.edu\)](#)>
77. ‘Unknown: Killer Robots’ directed by Jesse Sweet in 2023.
78. –< [WATCH: David Ignatius and Pentagon’s Robert Work on the latest tools in defense - The Washington Post](#)>
79. –< [What you need to know about autonomous weapons | ICRC](#)>
80. –< [What is Deep Learning? - Deep Learning Explained - AWS \(amazon.com\)](#)>
81. –< [X-47B Unmanned Combat Air System \(UCAS\) \(naval-technology.com\)](#)>

Other UN Documents

82. *Convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects*, 8 August 2022, CCW/GGE.1/2022/WP.2.
83. *Convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects*, 10 March 2023, CCW/GGE.1/2023/CRP.1.
84. *Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011)*, 8 March 2021.

Self-Published Articles

85. Langley B, ‘The weaponization of artificial intelligence: What the public needs to be aware of’ *Frontiers in Artificial Intelligence*, 2023.
86. Longpre S, Storm M, Shah R, ‘Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies’ *MIT Science Policy Review*, 2022.
87. Rushing B, Lavin J, ‘Is the scaling hypothesis falsifiable?’ *PhilSci-Archive*, 2024.

UN General Assembly

88. UNGA, *Declaration of basic principles of justice for victims of crime and abuse of power*, UN A/Res/40/34 (29 November 1985).

Working Papers, Discussion Papers and Research Papers

89. Asaro P, 'The liability problem for autonomous artificial agents' Association for the Advancement of Artificial Intelligence, 2016 AAAI Spring Symposium Paper Number 4, 2016—<[The Liability Problem for Autonomous Artificial Agents - AAAI](#)> on 29 November 2024.
90. Davison N, 'A legal perspective: Autonomous weapon systems under international humanitarian law' United Nations Office for Disarmament Affairs, UNODA Occasional Papers Number 30, 2018—< <https://doi.org/10.18356/29a571ba-en>> on 10 November 2024.
91. Horowitz M, Scharre P, 'Meaningful human control in weapon systems: A primer' Center for a New American Security, Working Paper March 2015—< [Meaningful Human Control in Weapon Systems: A Primer | CNAS](#)> on 22 September 2024.

