

Northumbria University
Law School



Research Question:

“The laws surrounding responsibility and accountability of autonomous weapons systems are insufficient: An analysis of legal and ethical implications of autonomous weapons systems.”

Word count: 16,463

This Project is submitted for the qualification of MLaw
Date of Submission: Monday 11th May 2020

Name: Sophie Quince

Supervisor: Professor Christopher Newman

Research Declaration

I confirm that I have already submitted my Project Synopsis and Ethical Approval Form, which has been signed by my supervisor. I further confirm that this project is entirely my own work and that the research undertaken for the completion of this project was based entirely on secondary material or data already in the public domain (case law, journal articles, published surveys etc). It did not involve people in data collection through empirical research (eg, interviews, questionnaires or observation).

Signed: SOPHIE QUINCE

Dated: 11/05/2020

Table of contents:

Introduction

Chapter 1: Self-awareness of Autonomous Weapons (development of AWS)

1.1 Defining Autonomy in Weapons

1.2 Self-awareness in AWS

Chapter 2: Responsibility and Accountability

2.1 Criminal Responsibility

2.2 Developer/Manufacturers' Liability

2.3 State liability

2.4 Command responsibility

2.5 The accountability gap

Chapter 3: Legal and Ethical Implications

3.1 International Humanitarian Laws

3.2 Distinction

3.3 Proportionality

3.4 Human Rights Agreements

3.5 Ethical and Moral Implications

Conclusion

Bibliography

Introduction:

This dissertation is going to look at the law surrounding responsibility and accountability for Autonomous Weapons Systems (AWS). The basic hypothesis for this work is that the laws are insufficient and the legal frameworks on a national and international level are not equipped to deal with the challenges posed by AWS. This dissertation will look at the legal and ethical challenges of these weapons systems, with the aim of posing some solutions to these challenges.

Throughout history there has been a race to make technological advancements, some of which are merely trivial, such as a light that is activated by someone clapping or a device that will recite your daily itinerary by simply saying “*good morning google*”; however more serious advancements can be seen in computer technology and life-saving medical equipment. At some point during the development of all this technology, it is likely that developers had to contemplate the legal and ethical implications of mass-producing these products for consumer use. However, what if we were to consider developments in technology that have lethal application, limited human control and the potential to be fully autonomous in the future?

The development of such technology is currently a reality and is already being used in combat. Therefore, this dissertation aims to answer the question of whether or not the laws regarding responsibility and accountability of autonomous weapons systems are sufficient and what the legal and ethical implications of this technology are.

The majority of the materials available surrounding Autonomous Weapons indicate that there is yet to be full autonomy attributed to robots, however contrary to this evidence one of the most distinctive fears surrounding this topic are the implications of such autonomy occurring and how this would affect humanity. Fear of the unknown is not an unusual human trait and the preservation of life is something that is a global value, because of this the concept of human dignity is seen to be one of the most important human rights, with the rules of war being implemented in order to reduce civilian casualties as well as unlawful action against soldiers in time of war. The new generation of autonomous weapons threatens such concepts and opens up a gateway into unknown territory where AWS technology is not solely

dependent on human operation. An interesting article was published in September 2019 where an ex-google engineer disclosed how she resigned from Google last year in protest of being sent to work on a military drone project. Nolan stated that killer robots had the potential to do “calamitous things that they were not originally programmed for”.¹ This article caught my attention for several reasons. First of all, it highlighted how fear for the unknown is shaping the way people view AWS, and secondly that so called ‘killer robots’ merely have the potential to do calamitous things, not that they *are* doing calamitous things. The latter point is crucial to unravelling the questions that I am posing for my dissertation as it indicates there are gaps in the understandings of both the technology itself and the legislations and definitions surrounding them. The subject of Autonomous Weapons is fascinating to myself and academics alike, particularly because whilst there is a willingness to develop such technology, there is a level of ignorance when it comes to the legality and ethics of such progresses. Whilst this is the case, it is encouraging to note that the issues identified have previously been discussed by academics such as Rebecca Crootof² and Robert Sparrow³ as well as legal figures such as Christopher Heyns⁴. I found that these figures provided me with some stimulating perspectives on the subject that I hope to incorporate throughout my work.

Throughout this dissertation the intention is for there to be an in-depth discussion regarding the law of AWS, including a conversation regarding the definitional development and the ambiguities surrounding this. Moreover, great consideration will be given to the responsibility and accountability of such technology as well as what ethical and legal implications have emerged and implications that could develop in the future. It is hoped that problems will be identified and discussed for each element of this dissertation and a discussion will be had regarding what the implications of such problems are.

Methodology:

¹ Henry McDonald, 'Ex-Google Workers Fear 'Killer Robots' Could Cause Mass Atrocities' *The Guardian* (London 15 September 2019).

² Rebecca Crootof, 'The Killer Robots are Here: Legal and Policy Implications', [2015] *Cardozo Law review*

³ Robert Sparrow, 'Killer Robots' (2007) Vol.24 *Journal of applied science*.

⁴ Christof Heyns, 'Autonomous weapons systems and human rights law' presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons' (13 – 16 May 2014), Geneva, Switzerland. UN Doc A/HRC/26/36

The chosen methodology of my dissertation provides a framework that allows for the question at hand to be appropriately broken down into subsections, which combined allow the opportunity to answer the research question.

The information that I will collect throughout my research will be desk-based research and be a combination of peer review journals, news sources and legislation. My initial starting point has been the website ‘Campaign to stop killer robots’; this provided me with an important insight to the main issues as well as providing me with an exhaustive list of reputable sources. This is particularly important due to the rising tensions surrounding AWS which inevitably means there will be contradictory opinions and articles that are perhaps not fully grounded in facts; because of this, the validity and the provenance of the sources will be imperative.

This dissertation will by-in-large not incorporate the traditional doctrinal methods of research, instead we will look at the societal and cultural impact this technology has and the various writings around this. One of the main reasons for this is the fact that there isn’t that much ‘black letter’ law surrounding this area, which is perhaps, in some ways one of the problems. By integrating this research method into my dissertation, I hope it will allow for a discussion to be had regarding how the law of AWS or indeed the lack of laws for AWS has implications on society, therefore linking to the ethical consideration embedded within my research.

In terms of the structure of my dissertation I have broken it down into 3 chapters, each of which provide a different outlook into why the laws surrounding AWS are failing to address the issues of responsibility and liability, and how such failings have created ethical and legal problems in the process. The first chapter is ‘Self-awareness of Autonomous weapons (the development of AWS) and is designed to provide some background on the topic and discuss some potential definitions of AWS. By having this as the first chapter I am able to immediately address the underlying issue that runs throughout my research, that being the definitional ambiguity involved in AWS. Whilst definitions seem to be a somewhat trivial area, my research led to me to believe that this ambiguity is at the core of this issue.

Chapter Two is ‘Responsibility and Accountability of Autonomous Weapons Systems’; this will provide a large portion of the dissertation and will examine and compare the arguments for and against possible applications of responsibility and accountability. I hope to discuss a number of parties who are exposed to potential liability and the practicalities of doing so. The comparative nature of this chapter allows me to expand my research to international law in

order to identify possible ways in which legislation could be adapted in order to sufficiently address the accountability gap that is currently present.

The final chapter will expand on the previous 2 chapters, discussing the legal and ethical implications of using AWS. There are some arguments that this technology creates a more precise and humane method of warfare, however, others suggest that actually it creates a human rights dilemma and that it should not be an acceptable method of combat. This chapter will analyse these arguments and will expand on the concept of responsibility and accountability using a comparative approach.

The structure of my research will allow for each chapter to complement each other and for the main research question to run throughout.

Ethical Considerations:

I have considered the ethical dimension of this project and due to the fact that it is desk-based research, myself and my supervisor feel there is no need for any additional ethical safeguards. However, this will be monitored throughout the life cycle of this project and if any potential ethical issues arise, they will be brought to the attention of the appropriate committee.

Outcomes:

The intended outcomes of this piece of research is to provide a discussion regarding the practicality of allocating responsibility and accountability for the actions of AWS and whether or not current law provides adequate provisions for this. Following a review of the literature I hope that I will be able to identify an area that provides the potential to be adapted in order to amend any inconsistencies or failures in the law as it stands. As previously mentioned, chapter one discusses the definitional ambiguity of AWS, this is an element that I expect will run throughout my research due to these failings being embedded in every element. I hope to prove that by creating a universal definition of AWS it is possible to create a set of comprehensive laws that can govern this technology. Furthermore, I hope to identify how the inability to allocate responsibility or accountability for the actions of AWS will create ethical issues as well as legal issues relating to international humanitarian law and human rights.

Chapter One: Self-Awareness of autonomous weapons (the development of AWS)

Having outlined the initial areas of this enquiry, this chapter is going to address the first part of the research question and discuss the development of these weapons systems as well as considering the the definitional discussion surrounding this technology.

Recent years have been host to increasingly rapid technological advancements that incorporate varying levels of autonomy into machinery and robots, such advancements are shaping the way in which the modern world functions⁵. Technology is being pushed out of the realms of the impossible and into one of the most crucial debates of this century. The human ability to create biotechnology and to manipulate DNA all provide examples that show the human capability of engineering technology and genetics in in order to enhance the human body, with the human brain being the epicentre of such human intelligence. However, there have been extreme advancements in departments that are developing technology that can manifest artificial intelligence (AI) and is able to be programmed to act in a certain way that is independent to human controllers; such technology challenges the idea that the human brain is the unsurpassed form of cognitive intelligence.

The concept of the singularity hypothesis⁶ suggests that eventually there will be no reason for the ordinary human to be ‘in the loop’; the intelligence of the AI technology will have exceeded that of the normal functioning human brain and humans will no longer be able to keep up with the rate at which such technology makes decisions and reacts. When you apply this concept to the advancement of AWS, the possibility of humans no longer being in the loop becomes a much more frightening and potentially catastrophic prospect.

A crucial discussion in relation to the research question is that of the differentiation between legal responsibility and moral responsibility; whilst these are individual concepts one could argue that as we often view some legal actions to be immoral as well as some laws to be unjust, the two concepts conflate and compliment each other. And whilst it is true that one

⁵ The advancements referred to here can not only be seen in the rapid development in Automated vehicles and weapons, but Artificial intelligence had a vital impact on society, from medical innovations using AI to simple household gimmicks such as ‘Alexa’.

⁶ Murray Shanahan, *The Technological Singularity* (The MIT Press 2015).

can have law without morality and morality without law⁷, the combination of the two can provide for a more well-rounded set of legislations. If one is to consider the above in the context of the responsibility and accountability for AWS, it is clear that whilst laws and morals should be considered in conjunction with each other, due to the discrepancies in how others perceive morals, the legal arguments regarding AWS provide a better opportunity for a global agreement as to the responsibility and accountability of AWS.

This research will provide an insight into the possibility of allocating legal responsibility and accountability for the actions of AWS, whilst delving into the legal and moral responsibilities of developing this technology. In the process of doing this, consideration must be given to what exactly is meant by a weapons system acting ‘autonomously’ and the extent of self-awareness of AWS. The following chapter considers how defining Autonomy in weapons and what is considered by ‘self-awareness’ has an impact in the ability for responsibility and accountability to be allocated.

1.1 Defining Autonomy in Weapons

Accurate and comprehensive definitions are key when it comes to any product regardless of its function, as a definition provides certainty and the ability to accurately legislate such products if this is necessary. Before we consider the legal or ethical arguments surrounding the responsibility and accountability of AWS, it is important that we establish the definition parameters of this technology and identify any potential gaps in these definitions.

The notion of armed conflict is one that is familiar worldwide, yet in past years the nature of armed conflict has adapted, taking on a more technological form. Modern day warfare is no longer solely dependent on human soldiers, in fact it now has a heavy reliance on Artificial Intelligence (AI) and the development of autonomous weapons systems (AWS). Yet whilst we are advancing this technology at rapid rates, integrating it into armed conflict with seemingly little thought to the legal and ethical consequences of doing so, the lexicography surrounding this technology is failing to adapt as quickly as the technology itself.

‘Autonomous’ and ‘automation’ have been used with some degree of interchangeability, yet in reality they bear different meanings. The issue regarding definition is one that is extremely prevalent in the field of autonomous weapons, with scholars and policy makers alike failing

⁷ ‘The Difference Between The Moral And The Legal’ (Reason and Meaning 2020)

<<https://reasonandmeaning.com/2016/03/31/the-difference-between-the-moral-and-the-legal/>> Accessed 17 April 2020

to agree on a definition of autonomy in weapons. One could argue that this is a trivial matter and the need for a precise definition is redundant, yet this is actually an integral factor.

Without a working definition that all countries can implement, creating a worldwide consolidation as to which military weapons are deemed to be ‘autonomous’ is near impossible. The lack of legislation and definitional clarity regarding the use of AWS leaves the issue of responsibility and accountability unaddressed.

Currently technologies are being used that possess a certain level of autonomy including tracking, identifying, deciding when to fire a weapon or to detonate a device, with approximately 30 countries deploying or developing defensive systems⁸ that can be placed somewhere on the spectrum of ‘autonomous’. However, this autonomy is only actively in use in circumstances where engagement time is too narrow for human response. Such technology has been in use for the best part of 70 years⁹ and has naturally developed from being a fairly basic use of autonomy in weapons to much more advanced and precise equipment, but this is not the technology that is currently in the spotlight. Indeed, the debates regarding AWS and the concerns surrounding its integration into modern warfare at this present time are merely hypothetical and are based on the predicted advancement of AWS technology, and what these systems are predicted to be capable of in the future. Crotoft considered that “there is a nearly universal consensus, among both ban advocates and sceptics that autonomous weapons do not yet exist¹⁰”. Moreover, there is a further concern that AWS, whilst they are yet to exist on a fully autonomous scale, will fail to comply with the principles of the Law of Armed conflict and will lack the ability of proportionality, distinction and military necessity, alongside the application of the Martens clause. Such fears were set out in ‘*Loosing Humanity: The case Against Killer Robots*’ published by Harvard Law Schools International Human Rights Clinic in 2012¹¹. So why is it the case that countless NGOs including Campaign Against Killer Robots are calling for a pre-emptive ban on this technology? This could be explained in the point made by Paul Scharre and Michael Horowitz: they consider that the rapid advances in information technology makes the development of more advanced autonomous weapons something that may come to fruition sooner rather than later¹², and

⁸ Wg Cdr (Dr) U C Jha (Rhett), ‘*Killer Robots, Lethal Autonomous Weapon Systems: Legal, Ethical and Moral Challenges*’ (VIJ Books (India) Pty Ltd, 2016)

⁹ *ibid*

¹⁰ Rebecca Crotoft, ‘The Killer Robots are Here: Legal and Policy Implications’, [2015] Cardozo Law review

¹¹ Human Rights Watch & INT’L Human rights clinic, Harvard Law School., ‘*Loosing Humanity: The Case against Killer Robots*’ 1-2, 5 (2012)

¹² Paul Scharre and Michael C. Horowitz, ‘*An Introduction to Autonomy in Weapon Systems*’, (2015) Centre for a New American Security

whilst there is extreme importance in the awareness of this issue, there is a trend of leaping to the conclusion that AWS should be legally banned without having a fully comprehensive understanding of what is meant by ‘autonomous weapons systems’ and what their development would mean.¹³ One could argue that the apparent failure to develop a globally recognised definition of AWS allows the notion of “autonomous robot” to be subject to imagery of science fiction characters and anthropogenic robots. This knowledge gap is contributing to what could be described as a fear-based understanding of the consequences of introducing a more advanced form of autonomous technology, especially if that comes in the form of weaponry. With there already being a varying level of autonomy in existing technology, the question would be whether or not one definition would be able to encapsulate the complexity of autonomy. As previously discussed, simple forms of autonomous weapons and technology have been in use for approximately 70 years, but even seemingly innocuous examples generate these kinds of concerns; for example, existent ‘self-driving’ vehicles have the capability to manoeuvre themselves around obstacles and produce countermeasures to assist in the completion of its tasks, but do so with little concern for human safety. This demonstrates that although development of lethal autonomous weapons systems (LAWS) has produced the legal and ethical minefield we now find ourselves in, the legal grey area is by no means restricted to LAWS. There is no doubt that clear terminology is vital when it comes to providing clarification and understanding, without this the word “autonomy” will remain an umbrella term for a host of complex and varying AWS. Moreover, resolving the current debate on definition should be seen as a priority - once this has been established, issues regarding ethics and legality will be able to be considered.

The aforementioned definitional ambiguity has led to different ways of thinking about AWS. Whilst these definitions vary, a popular theory can be seen in the form of the ‘loop theory’, which is based on the amount of human input/supervision. The Human Rights Watch applies this theory within their definition regarding degree of control, and defines it as follows: *Human-in-the-loop weapons*: robots that can select targets and deliver force only under human command; *Human-on-the-loop weapons*: robots that can select targets and deliver force under the oversight of a human operator who can override the robots actions and

¹³ Whilst many academics such as Kenneth Anderson, Daniel Reisner, and Matthew Waxman (“adapting the Law of Armed Conflict to Autonomous Weapons Systems,” *international law studies*, Vol.90 (2014), p.406)., have argued that limiting the specific situation where the weapons could be used would reduce the risk to civilians, the opposition argue that a narrowly constructed hypothetical case cannot legitimise the use of these weapons, prompting for a pre-emptive ban.

Human-out-of-the-loop weapons: robots that are capable of selecting targets and delivering force without any human input or interaction¹⁴.

One could argue that this bears resemblance to many theories including that of John Boyd's observe-orient-decide-act paradigm¹⁵. In this paradigm, Boyd recognises that a military pilot is required to make decisions at a quicker rate than their opponent, and that this decision-making process can be condensed into his model of an observe-orient-decide-act paradigm (OODA). This concept is not to be considered as an entirely new concept; however, it does organise this way of thinking into a strategic system that essentially allows one to adapt to any given situation, thus coping with the uncertainties of war. When considering this paradigm in the context of AWS, the level of autonomy attributed to the AWS systems would in theory mean that they would be able to adapt to different strategic situations, however the autonomy of a robotic is vastly different to the autonomy of humans and so the approach to this paradigm would be significantly different. A human is able to apply the OODA paradigm to situations where ambiguity is clouding judgment and there is the need deliver a fast reaction. Whilst this is the case Boyd comments that our inability to properly make sense of our changing surroundings is a bigger hindrance and can potentially mean that rather than shifting their perspective of a situation, humans will simply try and address their situational ambiguity by creating solutions according to their personal experiences. However, AWS are yet to be given the self-awareness that the human mind possesses and so do not have the same perspective on a situation that a human may have. Applying this paradigm to AWS could therefore introduce a technology that has a higher effectiveness and a drastically quicker response time to any human soldier or pilot.

I am of the opinion that whilst the OODA paradigm doesn't necessarily provide a definition for AWS, it provides us with a mode of thinking that could allow us to understand the implication of AWS being developed and introduced into modern day warfare. One might argue that without this deeper understanding and the lack of comparison between the ways human soldiers and AWS would make decisions in an ambiguous situation, it is not possible to produce a definition that would sufficiently engage with the capability of such technology.

¹⁴ Human Rights Watch, '*Losing Humanity: The Case Against Killer Robots*' (19 November 2012) New York: Human Rights Watch and the International Human Rights Clinic,

¹⁵ William Marra and Sonia McNeil, "*Understanding 'The Loop': Regulating the Next Generation of War Machines*," (2013) 36 Harvard Journal of Law and Public Policy 36. Pages 1139–85,

One of the most widely recognised definitions of AWS is from the US Department of Defence (DOD) directive 3000.09 on AWS (2012), which defines an AWS as: *“a weapon system that, once activated can select and engage targets without further intervention by a human operator¹⁶”*. This definition distinguishes autonomous weapons from semi-autonomous weapons by providing a further definition of the latter as being *“a weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator¹⁷”*. The same directive defines AWS as being able to *“select and engage targets without further intervention¹⁸”*. Be that as it may, this definition is not without its ambiguities, the main concern being that the directive’s attempt at distinguishing semi-autonomous weapons from autonomous weapons somewhat tarnishes the clear definition above. The main distinguishing element of autonomous and semi-autonomous is the level of human responsibility involved in the selection of targets, and whilst the directive makes an effort to distinguish between the two, one could be of the same opinion of Crotoft in thinking that this element remains vague and unclarified. These definitions provided by the DOD show a similar way of thinking to that of the Human Rights Watch ‘loop’ theory. Moreover, the DODs definition is not dissimilar to the directive set out by the British Ministry of Defence¹⁹ in which it is suggested that an autonomous system is capable of understanding a higher level intent and direction which will allow it to take the desired action in order to bring about the intended state. The MoD directive continues to state that *‘autonomous systems will, in effect, be self-aware and their response to inputs indistinguishable from, or even superior to, that of a manned aircraft. As such, they must be capable of achieving the same level of situational understanding as a human...²⁰’*. The concept of AWS being ‘self-aware’ is widely regarded as one of the main concerns in the development of this technology, as removing humans from the loop creates a whole host of legal and moral conundrums that will be discussed further in the following chapters. The 2014 International Committee of the Red Cross (ICRC) expert meeting on “Autonomous Weapons Systems” maintained that there was yet to be an agreed upon definition of AWS.

¹⁶ U.S Dep’t of Defense Directive No. 3000.09, Autonomy in Weapon Systems, (Nov.21, 2012) , <<http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf>>

¹⁷ *Ibid*

¹⁸ *ibid*

¹⁹ UK Ministry of Defence, Joint Directive Note 3/10, ‘Unmanned Aircraft systems: Terminology, definitions and classification’, (May 2010)

²⁰ *ibid*

Following the 2014 ICRC meeting, Rebecca Crootoof made an attempt to further address the ambiguity facing defining AWS²¹: her new definition took into account the law in relation to armed conflict rather than simply the practicality of having a working definition. Her definition stated the following: “an ‘Autonomous weapons system’ is a weapons system that, based on conclusions derived from gathered information and pre-programmed constraints, is capable of independently selecting and engaging targets.” The way in which this definition has been phrased appears to address the individual concerns regarding defining AWS and takes into account the vital distinction between automated and autonomous, as well as accounting for the machine’s ability to, in some sense, ‘improvise’ within the constraints of its programming.

Whilst there is still an element of ambiguity, it is clear that the numerous definitions that have been created all share the common theme of AWS having the ability to independently select and attack a target without human intervention; suggesting the aforementioned ‘loop’ principal could be seen as one of the most agreed upon definitions at present. When developing a globally recognised definition there should be a balance between the recognition of the different domestic laws that may affect its implementation and the need for clarification regarding what an autonomous weapons system is classified as. A pre-emptive ban on the use of AWS in armed conflict is, as considered above, based on hypothetical developments of this technology. Yet it is seemingly impossible and somewhat impractical to ban a system that is yet to be provided with a recognisable legal definition.

1.2 Self Awareness in AWS

In view of the above it is clear that defining AWS is at present, one of the predominant discussions amongst academics and authorities, yet one could argue that the real concern lies within the diminishing human control that AWS possess and the technology’s increasing self-awareness. The extent to which humans control the decisions of AWS is diminishing so rapidly that it is not unreasonable to assume that at some point in the future AWS will possess what could be considered as complete self-awareness. This is a quality that Crootoof explains provides a threat to one of the laws of armed conflicts most fundamental assumptions; that ultimately, a human being decides whether another human being lives or

²¹ Rebecca Crootoof, ‘The Killer Robots are Here: Legal and Policy Implications’, [2015] *Cardozo Law Review*, Vol 36, p.1854

dies²². This is a point that Asaro considers when stating that, “*it is the delegation of the human decision-making responsibilities to the autonomous system designed to take human lives that is the central moral and legal issue*”²³. Whilst this is a point that potentially has closer links to the liability and accountability aspects in respect of decision making, the element of machine liability is what ties this to the attribution of self-awareness to AWS. One of the oldest questions in relation to the use of AWS is its capability to function like a human, make ‘human like’ decisions and to think and reflect on its actions in the way in which a human is capable of doing. By giving technology the ability to essentially make its own decisions, developers risk removing things such as the degree of empathy, common sense, and creativity exhibited by a human soldier on the battlefield. Moreover, if AWS achieve a level of self-awareness that a human may possess, we leave ourselves with a number of questions that touch on a somewhat philosophical approach yet will also provide a unique perspective for the legal responsibility that could be given to AWS. For example, if the level of self-awareness increases, would this therefore mean that it is capable of making moral judgements, and if so, could it be held responsible for its actions? Murray Shanahan discusses the idea of Brain-inspired AI, and what the consequences of this would mean for humans and AI alike. However, more relevant to AWS would be the AI that has been manufactured from scratch and has not been modelled with the intention of replicating the way in which the human brain functions and is more focused on ‘artificial general intelligence’. AWS are not designed to mimic the appearance and behaviour of humans, but to resemble and act as weapons, as the name suggests. Whilst this means AWS do not emulate a traditional form of consciousness, they would be capable of showing basic cognitive qualities such as awareness, purpose and integration - all qualities that are inevitably going to be a product of artificial general intelligence.

As with most technologies, there are varying levels of capability and advancement, and this is certainly the case for autonomy and self-awareness of AWS. If we take the basic example of a landmine, it is an entirely independent device that is capable of performing the task it has been assigned without the need for further human intervention. It will be activated when someone walks over it or if a vehicle drives over it, yet this is the extent of its autonomy. It cannot discriminate between those who it has been placed there to kill or those who are innocent in this scenario. It simply knows two states, either exploding according to the

²² Ibid, p. 1845

²³ Peter Asaro, ‘On Banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making’, (International Review of the Red Cross, Summer 2012) Volume 94 Number 886

pressure applied, or not exploding. This device would not be placed in the same category as a targeting system, for example. A targeting system would be functioning on a much higher level of autonomy, as it is designed to distinguish between many states and possibilities in order to achieve its desired outcome²⁴; it therefore demonstrates a level of autonomy that links to cognitive skills. From these simple examples a clear trend emerges: the greater the responsibility given to an AWS, the greater the degree of cognitive skill required, and thus the higher the level of autonomy. By assigning a device a higher level of autonomy we are essentially reducing how much it depends on humans when performing its tasks.

The theory behind autonomous vehicles can also be closely linked to that of autonomous weapons due to the similarity behind their operational functions. The prospect of autonomous vehicles on our roads does not seem as deadly as an autonomous drone designed for warfare, yet as alluded to earlier, the legal and moral questions remain the same. Moreover, the in the loop system can also be applied here: with vehicles becoming increasingly autonomous, the purpose of the driver changes from being in charge of operating the vehicle to essentially monitoring how the vehicles system is operating. Therefore, like the relationship between operators and autonomous weapons the ‘driver’ of an autonomous car is essentially removed of its purpose to drive; the driver has to be assumed as completely out of the loop²⁵. Both autonomous weapons and vehicles must be able to act autonomously with a level of self-awareness in order to deal with critical situations and assess what the appropriate actions and responses are for any situations it may come across.

As stated by the authors of “self-awareness in Autonomous Automotive systems”, a general challenge for self-aware autonomous systems is the fact that they are operated in an environment that allows only limited predictability. In most cases not all the effects that impact the system can be fully anticipated²⁶. Armed forces around the world already have some autonomous functions, such as navigation, communication and detection, and the level of autonomy these systems show ranges from remotely piloted to fully autonomous. Whilst these systems are highly advanced, there will still be a certain level of unpredictability that will only be noticed when the system faces a certain degree of uncertainty where it is challenged to adapt and apply the relevant countermeasure to combat any issues that arise. If we take the example of an autonomous missile device that is programmed to get to a certain

²⁴ Giulio Tononi, ‘Consciousness as integrated information: a provisional manifesto’, (2008) *The Biological Bulletin*, Vol. 215 No. 3

²⁵ Johannes Schlatow et al., "Self-awareness in Autonomous Automotive Systems." *Design, Automation & Test In Europe Conference & Exhibitin*. (2017).

²⁶ *ibid*

location, it is impossible to predict all scenarios which could occur on its journey and so a certain level of self-awareness is vital to allow such systems to operate effectively.

A final point to make in relation to the self-awareness of AWS is the notion of the bar being set too high in regard to what is deemed as autonomous. The legal view of fully autonomous weapons being merely hypothetical is not misplaced and is grounded in the confusion surrounding differentiating between ‘automated’, ‘autonomous’ and ‘semi-autonomous’ and the conflicting definitions of what classifies as an AWS. An example of this can be seen in the U.K MoDs definition of “autonomous systems” in which it states that “...as long as it can be shown that the system logically follows a set of rules or instructions and is not capable of human levels of situational understanding, then they should only be considered to be automated²⁷”. One could argue that this definition fails to take into account that as it stands, AI is unlikely to reach the same level as a human in terms of understanding of its situation and its surroundings, this remains a skill unique to the cognitive ability of humans.

Having considered the definitional ambiguity surrounding AWS, it seems as though this is one of the main contributing factors that has led to the laws in this area being insufficient. A clear lack of structure and understanding is preventing a universally recognised definition from being developed, in turn one could argue that this has the potential to contribute to an accountability gap where responsibility and accountability is concerned. The following chapter is going to concentrate on the concepts of responsibility and accountability and how they are applicable to the discussion at hand.

²⁷ UK Ministry of Defence, Joint Doctrine note, 2/11: “The UK Approach to unmanned aircraft systems” 2-3 (2011)

Chapter 2: Responsibility and Accountability

The discussion in the previous chapter regarding self-awareness of AWS posed interesting questions both legal and ethical. However, once we emerge from the conversation regarding definitional ambiguity, we are left with the development of weapons systems that can act without direct human intervention. Thus, providing the prospect of technology advancing to the level at which AWS could be fully autonomous and thus rendering human input surplus to requirement. We are then faced with yet another legal ambiguity which opens up a unique conversation regarding the challenges of attributing individual responsibility. Put simply, who is responsible for the actions of AWS?

By design, AWS and LAWS have a whole host of human agents who contribute to its functioning; this considered we are then presented with various individuals that could be candidates for legal responsibility of the actions of AWS²⁸. It has been identified by UN Special Rapporteur Christof Heyns that those who could face individual responsibility include ‘the software programmers, those who build or sell hardware, military commanders, subordinates who deploy these systems and political leaders who authorize them’²⁹. Whilst this is true, one must consider the practicalities as well as the legal boundaries involved in holding these parties responsible.

The previous chapter touched on the issue of AWS becoming fully self-aware and capable of making its own decisions, however, this is not reality at present. Because of this, the attribution of responsibility to AWS itself will be difficult as it is incapable of acting in a manner that could be subject to criminal liability. It is inevitable that the software running this technology will continue to be developed, increasing its complexity, therefore adding to the fact that not one person is involved in its creation or its functioning. It is thought that with the increase in people working on this software, it will reduce the likelihood that one individual will have complete understanding on how the software functions as a whole, it therefore follows that its functioning could be unpredictable as well³⁰. We must also then consider the prospect that in the future, near or distant, fully autonomous weapons systems

²⁸ Nehal Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Clause Kreß, *Autonomous Weapons Systems, Law, Ethics, Policy*, (Cambridge University Press 2016), pages, 303-324.

²⁹ United Nations General Assembly, ‘Report of The Special Rapporteur on Extrajudicial Summary or Arbitrary Executions’ (United Nations General Assembly, Human Rights Council 2013) A/HRC/23/47

³⁰ James G Foy, ‘Autonomous Weapons Systems: Taking the Human out of International Humanitarian Law’, (2014) Vol.23, Dalhousie Journal of Legal Studies p.58

may be reality, removing the human input element and creating a further concern regarding attributing responsibility to technology.

Considering the above, in the situation where the conduct of AWS violates the laws of war or if a malfunctioning of the technology, faulty programming or misguided deployment results in international crimes being committed, who are we holding responsible for such violations? More importantly who should be held responsible?

2.1 Criminal Responsibility

When considering the practical and legal standing of applying responsibility then a logical starting point would be to consider the elements of criminal responsibility both domestic and international.

The basic principal of criminal responsibility is that when a person commits an offence that is deemed to be criminal in nature, they will be held criminally responsible for the commission of this crime. The perpetrator will then receive a penalty in the form of a fine, a community order or imprisonment. The presumption that someone can be held responsible for their actions can be rebutted in certain circumstances, for example if the perpetrator is a group, corporation or state, when the subject of the blame is an animal or non-human object, or where exemptions apply which mean that the individual responsible for committing the crime cannot be subject to blame; this can be in situations where the perpetrator is a minor or has diminished mental capacity.

Current laws regarding responsibility and accountability have long been integrated into criminal and civil law both domestically and internationally. However, such laws vary depending on which country they belong to, with Germany having a more complex and versatile set of laws that recognise the concept of direct and indirect perpetrators. German law recognises that a person who commits and act through another is an indirect perpetrator³¹, signifying that the indirect perpetrator (*Hintermann*) has control over the direct perpetrator (*Vordermann*). The Hintermann often exploits a certain deficit that the Vordermann possess, this can be something as simple as lacking the intent for the offence. The German concept of someone being held criminal responsible by acting through another person is also acknowledged on an international level, with Article 25(3) (a) of the Rome Statute of the International Criminal Court providing ‘whether as an individual, jointly with

³¹ German Criminal Code, in the version promulgated on 13 November 1998, *federal Law Gazette [bundersgesetzblatt] I*, p. 3322, last amended by article 1 of the law of the 24 September 2013, *federal Law gazette I*, p. 3671 and with the text of Article 6(18) of the Law of 10 October 2013, *federal gazette I*, p. 3799

another or through another person, regardless of whether that other person is criminally responsible³². In view of this, an analogous comparison can be made between AWS and the Vordermann; whilst AWS are programmed to adapt and to make decisions based on the algorithms it has been assigned, at present it lacks the human ability to intend to commit an offence. Moreover, with intent comes the element of *Mens Rea*, an element which is designed to establish whether there was intent surrounding an action. Article 30 of the Rome Statute states that in order to act with intent, the perpetrator must mean to engage in the conduct and mean to cause or be aware that the consequences will occur in the ordinary course of events³³. AWS have been developed to operate in hostile environments and the functioning of the AWS itself is inherently unpredictable, this considered it is unlikely that one would be able to argue that its actions fit the requirements of the *Mens Rea*. Whilst there are inherent difficulties in applying laws that have been designed for human application only to AWS, the general principals embedded within them provide potential foundations to re-design the laws in the context of AWS.

We are all aware that the development of AWS was facilitated in order to provide strategical advantages in armed conflict; with this in mind it seems only natural that they could be used to commit crimes capable of being recognised as war crimes. With this comes further questions regarding not only the legal implication of responsibility and accountability but also the ethical whirlwind surrounding AWS being in charge of kill decisions. This is an area that will be discussed in detail in a later chapter. Given that this is a very real prospect, attention should be given to Criminal responsibility being applied on the international stage in relation to international crimes such as war crimes and crimes against humanity.

Individual criminal responsibility for war crimes committed in International Armed Conflicts has been the basis for prosecutions under the Charters of the International Military Tribunal of Nuremberg and at Tokyo, under the statute of the International Criminal Tribunal for the former Yugoslavia and the statute of the International Criminal Court³⁴. There have been countless examples of war criminals having been tried on the basis of this principle. Of some significance is the case of Tadić in 1995 which was seen in the Appeals Chamber of the International Criminal Tribunal for the former Yugoslavia; the conclusion in this case was that there was in fact individual criminal responsibility for war crimes committed in non-

³² Rome Statute of the International Criminal Court, UN Doc A/CONF. 183/9, 1 July 2002

³³ Article 30 of the Rome Statute

³⁴ IMT Charter (Nuremberg), Article 6; IMT Charter (Tokyo), Article 5; ICTY Statute, Articles 2–3; ICC Statute, Articles 5 and 25.

international armed conflicts³⁵. One could argue that the significant developments in the 1990s of individual criminal responsibility in non-international armed conflict was somewhat of a turning point in IHL (international humanitarian law), allowing those who had been at the forefront of such internal atrocities to be held responsible for their actions. Amongst these developments, the case of Tadić now provides us with a precedent for individual criminal responsibility for war crimes being committed in non-international armed conflicts. However, as with the majority of the laws governing armed conflict (international or non-international), the issues arise when it comes to determining responsibility for war crimes committed by the likes of AWS and LAWS. Based on the fact that individuals who commit war crimes are individually responsible for them under IHL and ICL, one could argue that those who deploy an autonomous weapon system which has been programmed to carry out acts that amount to crimes under either domestic or international law should therefore be criminally liable. This is an issue that will be discussed later on as whilst the solution seems somewhat simple, prosecuting such individuals would prove extremely difficult due to the necessary levels of understanding and proving that the intention for the crime in question to be committed was present.

Elements of the ICTY decision can be seen to have been adopted from The Elements of Crimes document which considered the substantive crime within the Rome Statute. Within this, it stated that for every war crime there was a requirement that the alleged conduct “took place in the context of and was associated with an [international or non-international depending upon the precise provision of the statute] armed conflict³⁶”.

Moreover, it is also the case that criminal responsibility in the context of a war crime has also been applied to individuals who have attempted to commit war crimes, as well as for those individuals who have assisted in, facilitated or aided or abetted the commission of a war crime. One could argue that this bares close resemblance to the approach seen in German law.

One could consider this by approaching AWS as being akin to a Vordermann, who possesses some defect or deficit, such as the lack of capacity to act intentionally. The fact that AWS lacks the human qualities it takes to carry out an intentional act separate to its programming makes it an innocent agent as its actions can be controlled or caused by a human agent. It is

³⁵ Prosecutor v Dusko Tadic a/k/a “DULE”, ‘Decision on the defence motion for interlocutory appeal on jurisdiction’ (International Criminal Tribunal for the former Yugoslavia)

³⁶ International Criminal Court, Elements of Crimes art. 8 into, U.N. Doc. PCNICC/2000/1/Add.2 (June 30, 2000)

no mystery that AWS and LAWS have been developed in order to be integrated into warfare to reduce human involvement and in theory soldier casualties and fatalities. Naturally this comes with the benefit that such technology has a precision that cannot be obtained by a human and so a more accurate form of warfare can be engaged. Considering the above it is possible to argue that these elements could be adapted and applied in order to allocate responsibility and accountability, be it criminal or civil, to those involved in the commission and actions of AWS.

2.2 Developer/Manufacturers' Liability

The concept of criminal responsibility is evidently well established within both domestic and international policy; however, the dynamic is bound to shift when this concept is required to be applied to AWS. As stated previously, Christof Heyns identified that there are a number of parties who are exposed to potential liability, this includes the developer/manufacturer of this technology.

AWS have been developed to be able to operate with a capacity to manage their own operation, with various components working together to amplify its effectiveness. AWS technology was developed to provide specific advances, by creating algorithms that allow AWS to manage the operation of itself without explicit human operator input, moreover the speed and accuracy of modern computation means that AWS are able to function more efficiently in situations where a human operator lacks the capability of making such rapid fire decisions, especially in the heat of battle.³⁷ The developers are programming AWS to be able to process information at a greater rate than a human would ever be capable and in turn creating what could be considered as more efficient warfare. This demonstrates the operational advantages driving the development in this area.

The process is undoubtedly complex, and it would not be possible for those involved to accurately estimate the possible consequences of their deployment,³⁸ therefore making it extremely difficult to attribute responsibility to the manufacturer and developer. One could argue that if you were to hold the developer accountable for each death or war crime committed using one of the AWS weapons they helped to develop, then you could also hold

³⁷ Tim McFarland and Tim McCormack, 'Mind the Gap: Can developers of autonomous weapons systems be liable for war crimes?', (2014) International Law studies, U.S Naval War college. volume 90

³⁸ Geiss Robin, 'The international-law Dimension of Autonomous Weapons Systems', (October 2015), International Policy Analysis, Germany, p.209.

the developer of every gun, grenade, explosive and generic weapon accountable for the deaths, injuries and crimes their development has contributed to. However, it is worth noting that the Protection of Lawful Commerce in Arms Act (PLCAA)³⁹ is implemented in US laws and was created in order to protect firearms manufacturers and also the dealers of these products from being held liable if crimes are committed using the products they have produced or sold.⁴⁰ Whilst this is the case, the above parties will not escape liability if damage is caused by defective products, breach of contract or criminal misconduct. Similarly, they may also have liability for negligent entrustment,⁴¹ which involves them having reason to believe the gun is intended to be used for criminal purposes. From the implementation of the PLCAA we can see that it is in fact a legal possibility to exempt manufacturers from being held accountable for the use of their technology, could it therefore be said that a similar set of laws should be created that have the same effect to manufactures of AWS?

Furthermore, if this became a reality and we used the PLCAA as a precedent, and manufactures sold AWS to various Armed Forces or leaders of countries involved in non-international armed conflict and were fully aware that they could be used to engage in War Crimes or strikes on civilians, would they be liable for negligent entrustment?

The difficulty and impracticability in holding manufacturers responsible for the actions of AWS is discussed by Sparrow, who considers that if you were to hold programmers or manufacturer's responsible for the actions of their creation once it becomes autonomous, it "would be analogous to holding parents responsible for the actions of their children once they have left their care".⁴²

A Human Rights Watch report stated that it would not be possible to hold the manufacturer liable for any harm caused, if: (i) the specification for LAWS was approved by the government, (ii) the weapons conformed to those specifications, and (iii) the manufacturer did not deliberately fail to inform the government of any expected or known danger from the weapon system.⁴³

³⁹ "S.397 – 109th Congress: The Protection of Lawful Commerce in Arms Act", 2005

⁴⁰ Vivian S. Chu, 'The Protection of Lawful Commerce in Arms Act: An Overview of Limiting Tort Liability of Gun manufacturer', (2012) Congressional Research Service

⁴¹ PLCAA defines "negligent entrustment" as "the supplying of a qualified product by a seller for use by another person when the seller knows, or reasonably should know, the person to whom the product is supplied is likely to, and does, use the product in a manner involving unreasonable risk of physical injury to the person or others", a plaintiffs claim of negligent entrustment will be asserted under state law.

⁴² Robert Sparrow, 'Killer Robots', (2007) Journal of Applied Philosophy, Vol. 24, No. 1.

⁴³ Human rights watch, 'Making the case: The Dangers of Killer Robots and the Need for a Pre-emptive Ban' (2016)

If we are to search for individuals to direct responsibility and accountability to, then one should consider that there would be overwhelming challenges in applying this to the developer. There is such uncertainty surrounding this field that to predict future developments and the way that the technology will be used would be ill-advised and challenging. Further to this, without knowing the outcome of future developments the task of saying how activities of the developers may constitute acts proscribed by the law of armed conflict is also extremely difficult.⁴⁴

2.3 State Liability

The development and deployment of AWS undoubtedly provides a new level of precision when it comes to armed conflict, the technology that is being developed allows AWS to react at greater speeds than any human soldier, providing strategic advantages and potentially reducing the mortality rate of human soldiers. However, this all comes at a cost; despite numerous states having commissioned the development and use of AWS technology there has been seemingly little consideration to the legal and ethical issues that come with them. As with any weapon commissioned by the state, there should be an obligation to ensure that the weapons are not being used in a way that violates International Humanitarian Laws or the human rights of those that come into contact with it.

However, the conversation regarding state responsibility and risk management for when things go wrong is highly controversial. Robert Geiss considered that whilst the deployment of AWS is not unlawful when used responsibly, it is a high-risk activity that is not fully understood and so there is a level of predictable unpredictability.⁴⁵ Considering this, Geiss states that a State that benefits from the strategic gains associated with AWS should be held responsible whenever such unpredictable risks are realised.⁴⁶ It is clear from this that state liability is something that is being taken seriously and developments are being made in this area. One could argue that the State should hold a high level of responsibility for issues brought about by AWS as in theory, they are the party that is making the decision to integrate them into their own militaries and deploy them into combat.

⁴⁴ Tim McFarland and Tim McCormack, 'Mind the Gap: Can developers of autonomous weapons systems be liable for war crimes?', (2014) International Law studies, U.S Naval War college. volume 90

⁴⁵ Robert Geiss, 'Autonomous Weapons Systems: Risk Management and State Responsibility' (2016) Third CCW meeting of experts on lethal autonomous weapons systems (LAWS) Geneva, 11-15

⁴⁶ *Ibid*

If States are capable of being held responsible for the actions of AWS in the event an unpredictable risk occurs, then it would make sense to implement preventative measures that would reduce the harm that such occurrences could cause. An essential component of this is for States to acknowledge their due diligence obligations that are aimed at risk prevention and harm reduction – state responsibility arises in the event that such obligations are violated. The Law of Armed Conflict clearly provides for the way in which states should conduct themselves in conflict and the repercussions of breaching such conduct, building on the element of ‘due diligence’. Furthermore, Article 1 of the Geneva Convention I-IV requires States to ‘ensure respect’ for the laws of armed in conflict in *all circumstances*.⁴⁷ There is therefore no reason that it would be misplaced in interpreting this to include armed conflict that involves the use of AWS. This being said there are still elements of ambiguity that call for clarification in order to diminish the opportunity of any State claiming that their actions do not amount to a violation of the Law of Armed Conflict. Geiss considers that the problem is not that there is a lack of legal basis per se, but that there is a lack of clarity surrounding the meaning of ‘*due diligence to ensure respect*’ in the context of autonomous weapons systems.⁴⁸ If we examine this, it is apparent that the issue here stems from the fact that this legal basis has been created in order to address human error not that of AWS, and that due diligence requires there to be a given circumstance, such as combat, and what a reasonable party would do if placed in these circumstances. The problem therefore arises when this is applied to the conduct of AWS, it is difficult to know what could and should be considered as reasonable when the technology in use is new and there are no precedents set to use as a point of reference when making this decision.

An interesting point that Geiss raises is that whilst human beings are in charge of the deployment of AWS, accountability can be determined using the pre-established rules of attribution.⁴⁹ This essentially means that if a member of the military of any given state decides to deploy AWS in the course of combat then the activities the AWS carries out will be attributable to that State and not the AWS itself – a weapons system possessing some autonomous capabilities does not alter this. It can be said that this accurately deals with the fact that whilst AWS exist, they are not fully autonomous (as previously discussed) and therefore humans still remain in the loop to a certain extent.

⁴⁷ Article 1 of the Geneva Convention I-IV

⁴⁸ Robert Geiss, ‘Autonomous Weapons Systems: Risk Management and State Responsibility’ (2016) Third CCW meeting of experts on lethal autonomous weapons systems (LAWS)

⁴⁹ *ibid*

A further obstacle to state liability is that of jurisdiction; only states are able to submit contentious cases for ICJ adjudication, because of this, the Courts lack jurisdiction to deal with any applications from individuals, non-governmental organisations, corporations or private parties.⁵⁰ Simply being a member of the U.N does not mean that automatically gives the Court jurisdiction, it would in fact require both states to have given consent to the Courts Authority.⁵¹ If we consider the consequences of a state submitting itself to the court's jurisdiction, it is evident that there is little to no incentive for them to do so as simply admitting liability would increase the risk of significant adjudication costs, costs that would increase substantially if the State in question was found guilty.

2.4 Command responsibility

Command responsibility, also known as indirect responsibility, is another branch of International Criminal Law that presents the possibility of holding a military commander or a civilian superior liable (criminally). It could potentially occur if a superior is found to have failed to take reasonable measures when preventing or punishing a subordinate if they have committed a criminal act.⁵² This is due to the fact that the superior effectively has control over the actions of the subordination, and therefore as soon as they are aware that one of their subordinates has committed a criminal act, they have a duty to act on this. This was discussed in the case of *Prosecutor v Halilović*,⁵³ which stated that command responsibility will hold the superior accountable for dereliction of duty. Superior or command responsibility is a concept that has been historically been applied to a number of cases following the Second World War.⁵⁴ Article 28 of the Rome Statute provides a provision for the 'Responsibility of Commander and Other Superiors'. The acts states that: *'a military commander or person effectively acting as a military commander shall be criminally responsible for crimes within the jurisdiction of the court committed by forces under his or her effective command and control, or effective authority and control as the case may be, as a result of his or her failure*

⁵⁰ Daniel N, Hammond, 'Autonomous Weapons and the Problem of State Accountability' (2015), Chicago Journal of International Law: Vol 15, No.2

⁵¹ Statute of the International Court of Justice, 59 Stat, 1055 (1945), arts 34-39; Practical information: frequently asked questions, The international court of justice

⁵² *Prosecutor v Delalić and others*, The International Criminal Tribunal for the Former Yugoslavia 1997-1999, page 462

⁵³ *Prosecutor v Halilović*, ICTY, Case No. IT-01-48-T, Judgment (Trial Chamber), Nov, 26, 2005, para. 54

⁵⁴ Wg Cdr (Dr) U C Jha (Retd), *'Killer Robots: Lethal Autonomous Weapons Systems Legal, Ethical and Moral Challenges'* (VIJ Books (India) Pty Ltd, 2016), Page 80

*to exercise control properly over such forces.....*⁵⁵ The accountability discussed here is activated if the superior should have known that the forces were committing or about to commit such crimes or that he had failed to take all necessary and reasonable measures to prevent or repress their commission. An important element to note if you were to pursue this level of responsibility would be the fact that it must be proven that there was a commander/superior and subordinate relationship between the parties. If this is unable to be shown, then the necessary elements would not have been satisfied and criminal responsibility of the superior would be unsuccessful.

On one hand, this could be seen as analogous of AWS due to the nature of them being weapons and so functioning in a similar context to soldiers (subordinates) – this considered it could be said that there are potential grounds to trigger the doctrine or superior responsibility. However, AWS, if fully autonomous, have a certain ability to make their own independent decisions and judgments relating to target selection and engaging with a threat, with advanced technology embedded in them specifically designed in order for them to operate independently from a superior. If they are not under the direct command of a superior it would be unreasonable for one to expect superiors to have anywhere near the appropriate level of knowledge to trigger the doctrine of command/superior responsibility. It could also be argued that the rapid rate at which AWS technology makes its decisions would make it near impossible for superiors to foresee that the forces were about to commit such a crime that could then make them subject to command liability.

If we were to consider that the robot was able to communicate its decision to the commander prior to acting, then this would in theory mean the commander would have sufficient knowledge of the impending criminal act and it could therefore could be held responsible. Be that as it may, this would only be in the case of a ‘human-in-the-loop’ system, if however, the system was fully autonomous, there would be a dramatic reduction in communication with a superior and the decisions would be made without the need of approval. Moreover, it has been established that a commander must have sufficiently alarming information in order to investigate,⁵⁶ without receiving such information they cannot be held liable for negligently failing to find out information. Furthermore, in *Prosecutor v Strugar*, it was noted that knowledge of past offenses by a particular subordinate may constitute as ‘sufficiently

⁵⁵ The Rome Statute of The International Criminal Court, Article 28 (Responsibility of Commanders and other superiors)

⁵⁶ *Prosecutor v Strugar*, ICTY, Case No. IT—1-42- A, Judgement (appeals chamber), July 17, 2008, paras. 297-98

alarming' information, therefore the superior would have sufficient knowledge that the subordinate in question could commit future criminal acts,⁵⁷ thus the mens rea of command responsibility would be satisfied.

Taking this into account, we are then left with the confusion surrounding what the threshold is for 'past unlawful acts' that have been committed by AWS. AWS technology has been developed primarily as an aid to combat and due to its adaptability, there are different scenarios that this technology can be placed in and different tasks that it can be assigned. If we therefore consider a situation where one AWS robot engages in conduct that violates international laws are we to assume that this conduct is setting a precedent for a past unlawful act that can be applied to all robots of that variety and with that same programming or is this only going to be applicable to that individual robot? We could then contemplate a scenario where one specific algorithm has been programmed into a model of AWS robots that means they are to target combatants - however one of them mistakes a civilian as a combatant, strikes and kills the civilians. This scenario has the capability of becoming reality, so in the event that this does in fact happen, are we to consider its inaccurate judgment of proportionality to be unique to that individual robot, or should the entire model have this on their record as a 'past unlawful act', given the unpredictability of it has already been demonstrated by one robot. Moreover, as previously discussed, the AWS technology is rapidly advancing and is complex in nature, in theory, in order for a commander to be able recognise that an AWS robot is going to commit a criminal act imminently, then they would have to have a some-what in depth understanding on the way the technology works, the nature of its programming and the level of autonomy it presents. It could be argued that this is highly unrealistic as well as unreasonable, there for it seems likely that commander responsibly could be avoided.

The issue of fast development and decision making in AWS also links to the concept that there must be effective control over a subordinate in order for commander responsibility to be applicable. According to the judgement in *Prosecutor v Delalicet al*,⁵⁸ effective control comes specifically from the "material ability to prevent or punish criminal conduct"; punishment of AWS is not only pointless but also fairly impossible and the preventative aspect of this has already been identified as difficult due to all the varying qualities of AWS. Further to this, the

⁵⁷ *ibid*

⁵⁸ *Prosecutor V Delalicet. Et al, ICTY Judgement of 16 November 1998*

unpredictability of AWS means that there are any number of circumstances that cannot be foreseen, or alternatively, can be foreseen but nothing can be done about it, this all contributes to the application of effective control being questionable. Examples of the unpredictability mentioned could be signal interference, errors in its programming and the fast processing speed of AWS. If one of these situations occurred then the ability for a commander to intervene and call off the attack is diminished, demonstrating that effective control is not in place.

One could therefore reach the conclusion that the application of command responsibility would be confronted with numerous issues revolving around insufficient knowledge and lack of effective control and consequently command responsibility would be an unrealistic form of responsibility to pursue.

2.5 The accountability gap

In the discussion surrounding responsibility and accountability, whether we are looking at state, commander or even criminal responsibility, a common feature is the gap in accountability. This gap has emerged due to the fact that the laws surrounding legal responsibility are yet to be adapted to accommodate the advanced nature of AWS and the fact that they are not human entities and so do not possess the human qualities referred to in countless laws regarding responsibility and accountability. Whilst one could analogously apply laws to the functioning of AWS, the accuracy of the conclusions would be heavily scrutinised, and it is unlikely that it would stand in a court of law.

If we are to consider that AWS are likely to become more advanced and will be deployed more regularly, the likelihood of an international law violation is somewhat inevitable – when this violation occurs those involved will search for someone to be held accountable. Accountability is one of the most essential elements in international law, it aims to deter and prevent violations, and so protecting potential victims of human rights abuses and war crimes.⁵⁹ The rules of International Humanitarian Law and Human Rights law are in place to protect people against violations to their rights, including the rights to life; yet the unpredictability of AWS technology means that they do not have the capability to comply

⁵⁹ Robert Sparrow, 'Killer Robots' 24, *Journal of Applied Philosophy* (Sparrow can be seen to argue that being able to attribute legal or moral responsibility to someone forms one of the fundamental elements of a just war)

with these rules, leaving people vulnerable to immense violations of their human rights. If civilians are wrongly targeted by AWS, the accountability gap means that there would be disputes regarding who is legally responsible for these violations. An accountability gap is dangerous for several reasons; however, one could argue that the most important reason is that if there are no consequences for human operators or commanders in the circumstances where such violations occur, then there is no deterrent for other states regarding future criminal acts. Bonnie Docherty, senior arms division researcher at Human Rights Watch suggested that “no accountability means no deterrence of future crimes, no retribution for victims, no social condemnation of the responsible party”.⁶⁰ If this is the case and there is a failure to produce a deterrent, then the same violations will be repeated by the AWS as there has been no intervention with its programming that would direct it away from the decisions it was making. Moreover, the violations that are being referred to are those that are inflicted on civilians, the lack of actual meaningful human control possessed by the weapons would therefore make it near impossible to hold anyone criminally or civilly liable for such violations and unlawful acts.

Academics and experts alike argue that the presence of an accountability gap gives weight to the argument that there should be a pre-emptive ban on AWS production and deployment until the law can accommodate all eventualities of AWS. Human Rights Watch has said that whilst military commanders could potentially be found liable if they intentionally deployed a fully autonomous weapon to commit a crime, such justice would likely be eluded as the most likely situation is where there is a violation due to unforeseen actions of AWS that human intervention wasn’t able to stop.⁶¹ Whilst one could predict the potentially dramatic outcomes the accountability gap will create, the unknown still remains one of the biggest enemies, until predictions become reality and the violations have been committed, it is unlikely that a proactive effort to fill the void will be attempted. In the meantime, any incidents caused by AWS will continue to evade justice and be marked down as an accident or even a glitch,⁶² in doing so the accountability gap is being allowed to act as a scapegoat for potential violations and by continuing to acknowledge this we merely risk trivializing the serious harm that could be done.

⁶⁰ Bonnie Docherty, Senior Arms Division researcher at Human Rights Watch and the reports lead author to ‘The ‘Killer Robots’ Accountability Gap: Obstacles to Legal Responsibility Show Need for Ban’ (2015)

⁶¹ Human Rights Watch, “‘The ‘Killer Robots’ Accountability Gap: Obstacles to Legal Responsibility Show Need for Ban’”. (2015).

⁶² Bonnie Docherty, Senior Arms Division researcher at Human Rights Watch and the reports lead author to ‘The ‘Killer Robots’ Accountability Gap: Obstacles to Legal Responsibility Show Need for Ban’ (2015)

Chapter 3: Legal and Ethical Implications

So far, this dissertation has provided an insight into how the ambiguity and lack of clarity surrounding the definition of AWS has further contributed to the development of problems in this area. The previous chapter drew on the points made in chapter 1 and explored how the lack of a clear universal definition has contributed to insufficient laws surrounding responsibility and accountability as well as how elements of international law could be used to somewhat fill the void in this area. This chapter aims to consider the legal and ethical implications of AWS, contemplating the impacts of this technology on a global scale rather than limiting it to domestic implications.

The integration of AWS into armed conflict inevitably provides a cross-over between the legal and ethical implications of AWS; however, the legal aspects involved in the deployment and development of this technology seem to be overshadowing the ethical implications that have and are likely to continue to increase as this technology advances. To say that legality is overshadowing ethics is not to say that the legality is insignificant, rather, that legality and ethics should be considered simultaneously in order for AWS to be assigned with appropriate regulations. As discussed in the previous chapter, the accountability gap provides the perfect opportunity for ethical implications to go somewhat under the radar. The unpredictability of AWS not only means that they lack the capability of adhering to the legality of International Humanitarian Laws (IHL), but their inability to simulate human qualities such as empathy and morality means that they are also incompatible with the ethical consideration embedded in IHL and Human Rights Laws.

One could argue that ‘Human dignity’ is one of the most important formulas embedded in international politics – once introduced in Article 1 of the United Nations Universal Declaration of Human rights, the concept of human dignity became an umbrella term that was used to bridge the gaps in ideological gulfs.⁶³ In turn, this provided a unified terminology allowing the concept of human dignity open to a certain level of interpretation that allows each member state to speak with one voice. Human dignity as a concept has become increasingly incorporated in international documents and treaties, underpinning the majority of IHL elements such as distinction and proportionality; because of this I believe that it is a principle that forms the foundations for the legal, political and ethical implications of AWS.

⁶³ Universal Declaration of Human Rights, UN GA Res. 217 (III) A, 10 December 1948

3.1 International Humanitarian Laws

International Humanitarian Law (IHL), stems from international law that has been put in place in order to manage and monitor the use of violence in armed conflict; the two-fold aim of this is to save civilians from the consequences of armed conflict and to protect soldiers from unnecessary suffering. The rate at which this technology is now advancing therefore brings new challenges to the basic principles of IHL. Whilst there are internationally recognised agreements to ban or regulate a number of problematic weapons such as expanding bullets, poisonous gases, antipersonnel landmines, biological and chemical weapons, blinding lasers, incendiaries and cluster munitions,⁶⁴ these principles were created at a time where the technology involved in LAWS and AWS had not been developed and so making provisions for them in IHL would have been a moot point. This has now led to a reality where such technology is advancing at such a rapid rate that it is now exceeding the parameters of IHL.

3.2 Distinction

A core element of IHL that is threatened by the advancements of AWS technology is the principle of Distinction. This principle is made up from two components: combatants must be able to distinguish (i) between civilians and enemy combatants,⁶⁵ and (ii) between civilian and military objects.⁶⁶ Moreover, this principle has been codified in Article 48 of the Additional Protocol I to the Geneva Convention: *‘in order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.’*⁶⁷

When we consider distinction in relation to AWS, the main problem is that in an area of conflict the people present are not restricted to soldiers, there are a whole host of people involved including civilian workers, medics and injured combatants. The concern lies with

⁶⁴ Wg Cdr (DR) U C Jha (Retd) ‘Killer Robots: Legal Autonomous Weapons Systems, Legal, Ethical and Moral Challenges’ 2016

⁶⁵ International Humanitarian Law, Volume II, Chapter 1, Section A, Rule 1. ‘The Principle of Distinction between Civilians and Combatants’

⁶⁶ International Humanitarian Law, Volume II, Chapter 2, Section A, Rule 7, ‘The Principle of Distinction between Civilian Objects and Military Objects’

⁶⁷ *Protocol Additional to the Geneva Convention of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 8 June 1977, 1125 UNTS 3, art 48, (entered into force 7 December 1979) [additional Protocol I]

the fact that AWS⁶⁸ and LAWS technology is unable to discriminate between combatants and non-combatants that may be present during conflict. Furthermore, when we consider that the ability to exercise the principle of distinction requires an individual to have their own set of moral and ethical guidelines in order to make decisions, we are once again confronted with the fact that AWS technology does not have the ability to simulate the relevant human qualities which could allow it to construct its own set of moral and ethical principles. Human soldiers are able to take positive steps in coming to their decision as to whether the subject they are engaging with is a combatant or a non-combatant, something that also requires a certain amount of common sense. Further to this, another important feature that relies on distinction is the ability to recognise when someone is surrendering; Sparrow successfully identifies this as an important shortcoming of AWS, discussing that because this technology does not have the capacity to recognize someone surrendering, there are ethical implications in relation to the deployment of such weapons systems.⁶⁹ Being able to differentiate between someone surrendering legitimately and someone using their body language to deceive the opposition is an extremely subjective decision and involves having the capacity to interpret the actions and intentions of the other person – this is not a quality that AWS technology currently possesses.

A further issue is presented in the form of definition ambiguity surrounding the legal definition contained in the Geneva Conventions of 1949 and AP I, such ambiguities mean that elements are left open to a certain level of subjectivity. If we are unable to implement a clear definition that it follows that it would be virtually impossible to integrate the concept of discrimination into the programming of AWS. Moreover, in order for a robot to be able to discriminate, they must first have the ability of recognition. It wouldn't necessarily be beyond the capability of designers to allow AWS to have the ability of recognising basic signals (such as surrendering), however, the difficulty of this is the fact that war is volatile and unpredictable and takes place in unpredictable terrain. If you consider all the environmental factors involved, then programming that allows AWS to have basic object recognition and classification skills are rendered useless.⁷⁰ Whilst programming basic recognition/distinction software may not be entirely unrealistic, creating software that can be implanted in AWS programming that gives them the same sophisticated capacity of interpretation of human

⁶⁸ Robert Sparrow, 'Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender', *international law studies*, Vol. 9, 205, p.299 - 728

⁶⁹ *ibid*

⁷⁰ Krishnan Armin, 'Killer robots: legality and ethicality of autonomous weapons, USA: Ashgate' (2009), p.98

reactions is something that is extremely complex and unlikely to be reality in the near future.⁷¹

One could argue that because of this factor, the ability for AWS to comply with IHL is unrealistic and its lack of ability to distinguish between combatant and non-combatants is the reason that could be deemed unsuitable for integration into armed conflict.

3.3 Proportionality

The prospect of civilian casualties is inevitable in armed conflict and there is no way in which this can be eliminated from war altogether despite the best efforts of IHL implementing concepts such as proportionality in an effort to protect the civilian population in war zones. The rule of proportionality is defined in Article 51 (5) (b) of Additional Protocol I. It states that a violation of proportionality will be “an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian object, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated”.⁷² It is clear from this definition that the principal of proportionality seeks to mandate that where there is collateral damage to civilian population, it must be proportional to military advantage.

The concept of proportionality is applied in a military context when considering if the damage to civilian objects, civilian death or injury is anticipated prior to targeting a military object, an assessment is given in which the anticipated military advantage is weighed against the anticipated ‘collateral’ damage to protected civilians or civilian objects.⁷³ In comparison to distinction, proportionality provides us with a more realistic opportunity to create programming that would enable AWS to make basic decisions that would present a certain level of compliance to the rules of proportionality; however, this would be very limited. Even if we were to assume that such programming could successfully embed basic proportionality understanding in AWS technology, we are once again faced with the same issue that can be seen to underlie many legal and technological hurdles facing AWS. Noel Sharkey also identifies this as a crucial barrier in AWS proportionality, the technology simply does not

⁷¹ Robert Sparrow, ‘Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender’, *international law studies*, Vol. 9, 205, p.299 - 728

⁷² *Protocol Additional to the Geneva Convention of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 8 June 1977, 1125 UNTS 3, art 48, (entered into force 7 December 1979) [additional Protocol I]

⁷³ Wg Cdr (DR) U C Jha (Retd) ‘Killer Robots: Legal Autonomous Weapons Systems, Legal, Ethical and Moral Challenges Killer robots’, (VIJ Books (India) Pty Ltd, 2016) p.73

possess the relevant human qualities to decide on when the damage to civilians would exceed the aforementioned anticipated military advantage provided by the attack.⁷⁴ Considering this, one could say that Kastan is correct in saying that despite potential technological advancements the future may hold, the necessary analysis and assessment involved in the principle of proportionality would have to be left to humans.⁷⁵ Additionally, one must also take into account the infinite number of scenarios that AWS might be confronted with, this alone makes it almost impossible to program this technology to replicate the decision process of a combatant.

If there is no way in which this issue can be resolved, then it appears as though AWS violates IHL in such a serious way that it could be considered as a war crime under the 1998 Rome Statue of the International Criminal Court.⁷⁶

3.4 Human Rights Agreements

With the rapid growth of AWS development over the past few years the debate surrounding their use has intensified, with the primary concern being the legality of them. Despite the fact that legality has seemingly taken centre, the implication arising from potential human rights violation is an issue that deserves equal attention. The lack of importance allocated to human rights in this conversation seems completely misguided; civilians are the unwilling victims of armed conflict, by deploying AWS into combat there is an automatic increase risk and concern regarding human rights violations. The use of force human soldiers apply will eventually be applied by AWS, and with the rapid development of technology this could occur sooner rather than later. One could argue that the consideration of Human Rights should not be considered as a separate issue and should in fact be considered alongside IHL due to the fact that they are complementary of each other when applied in the situation of armed conflict. As discussed by Christof Heyns, the question we should be concerned with is: ‘is the use of AWS to apply force permissible under human rights law, and is so under what

⁷⁴ Noel E. Sharkey, ‘The Evitability of Autonomous Robot Warfare’, *international review of the red cross* (2012), Vol. 94, No. 886, p.789

⁷⁵ Kastan Benjamin, ‘Autonomous Weapons Systems: A coming Legal “singularity”?’ , 2013 Journal of Law, Technology and Policy, 45 (2013) 45-82

⁷⁶ Article 85, para 3 (b) of the additional protocol I and article 8 (b) (iv) of the Rome statue of the international criminal court. See also; Margalit Alon, ‘The Duty to Investigate Civilian Casualties During Armed Conflict’, Yearbook of the international humanitarian law, Vol. 15, 2012, Springer: Cambridge University Press, P. 155 - 186

circumstances’.⁷⁷ In his report Heyns breaks this question down into 7 areas that all hold equal importance, however it could be argued that the main point of focus should be which human rights are at risk of being infringed if AWS are to dispense force either in a lethal or non-lethal way. The right to life and the right to human dignity present themselves as the most noticeable rights that are exposed to potential infringements, these two rights are widely recognised and are not only included in the main human rights treaty but can also be found in customary international law.

If we first consider the right to life, according to article 6 (1) of the ICCPR, ‘every human being has the inherent right to life. This shall be protected by law. No one shall be arbitrarily deprived of his life’.⁷⁸ The potential violation here occurs due to the fact that one of the assumptions under international human rights laws is that the ‘kill decision’ must be reasonable and carried out by a human. The concept of ‘reasonableness’ is inherently human and stems from a combination of strategy and emotion that can only be displayed by a human. Despite machines having a certain level of self-development in the form of machine learning, the extent of this does not reach to emotional development, and so it is not possible for machines to ‘reason’ in the same way a human does, and it therefore follows that if a machine cannot ‘reason’ it also cannot take a ‘reasonable’ decision. Like Heyns, if we consider Article 1 of the Universal Declaration of Human Rights, it provides that ‘all human beings are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.’⁷⁹ Whilst the terminology is somewhat archaic, the meaning remains relevant and maintains a strong emphasis on the human being the party that exercises reasoning and interaction, not AWS or any other technology that may be developed to act autonomously in combat. This considered, if we allow AWS to make their own decision when it comes to force then there will certainly be an infringement to the right to life. With the inability to feel emotion and make reasonable decisions, it is extremely doubtful that that AWS will be able to distinguish between a person surrendering and a person whom has the intention to attack. Without this element of distinction, we are allowing the machines programming to determine whether to act with lethal effect, therefore presenting a grave risk to the right to life.

⁷⁷ Christof Heyns, ‘Autonomous Weapons Systems and Human Rights Law, presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons’ 13 – 16 May 2014, Geneva, Switzerland

⁷⁸ International Covenant on Civil and Political Rights, Adopted and opened for signature, ratification and accession by general assembly resolution 2200A (XXI) of 16 December 1966 entry into force 23 March 1976, in accordance with Article 49

⁷⁹ Article 1 of the Universal Declaration of Human Rights

The second human rights issue that should be considered is the potential violations to the right to human dignity; this right is seen to be at the heart of all human rights and should be considered in conjunction with other human rights, such as the right to life. Article 1 of the Universal Declaration of Human Rights provides the following: ‘all human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.’⁸⁰ The concept of human dignity has distinct moral grounds and in saying that there has been a violation of human dignity you are implying that the action committed is morally problematic. When considering the violation of human dignity at the hands of AWS, a common theme amongst academics is that it is the removal of human agents from the decision to kill is the ultimate indignity; Heyns uses the example of the ‘Riobot’⁸¹ to demonstrate this. Whilst it is not currently autonomous in relation to the use of force, it is not farfetched to assume that this could become a reality. The idea that the miners are being herded like cattle by an autonomous robot strips the miners of their dignity and de-humanises them. The so called ‘death by algorithm’⁸² essentially means that AWS would be treating people as interchangeable entities and so will have total disregard for their dignity. In his report to the United Nations General Assembly Heyns, voices his concern regarding this and writes: ‘*delegating this process [of deciding on targets] dehumanizes armed conflict even further and precludes a moment of deliberation in those cases where it may be feasible. Machines lack morality and mortality and should as a result not have life and death powers over humans*’.⁸³ One could say that ‘machines lack morality and mortality’ is the most significant aspect of this discussion. We are aware that AWS do not have the emotional capacity of a human, nor do they possess any quality that resembles empathy or reasonableness; this is simply something that cannot be programmed and remains a distinctive human feature. Because of this, AWS cannot possibly comprehend the implications of killing or injuring a human being.

There are, of course, various candidates that could be affected by the use of AWS in combat, if we split these into two categories: active players (combatants) and passive players (civilians), we can begin to see whose dignity would be at stake. One could argue that the civilians are at greater risk of being stripped of their dignity due to them being passive

⁸⁰ Article 1 of the Universal Declaration of Human Rights, UN GA Res. 217 (III) A, 10 December 1948

⁸¹ The ‘Riobot’ is being developed in order to control unrest on the mines in Africa.

⁸² Christof Heyns, ‘Autonomous weapons systems and human rights law, presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons’ 13 – 16 May 2014, Geneva, Switzerland

⁸³ Christof Heyns, ‘Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions’, UN Doc. A/HRC/23/47, 9 April 2013, 17.

players in combat, they have little regard to the rule of war due to the fact that they are not actively engaging in the war, unlike the combatants, they are unable anticipate what could happen and taking refuge is often not an option. Furthermore, if we consider that both passive and active players in combat will be inhabiting the same geographical area, the issues of distinction (as discussed above) becomes an issue. Without the capacity to discriminate between combatants and non-combatants and to combine this distinction with the rules of proportionality, how are we to expect that AWS will not violate the human right to dignity by indiscriminately killing a civilian trying to surrender?

The issue of AWS violating human dignity is complex, whilst there are unknown factors and limitations to AWS, it does not take away the fact that they are programmed to have a high level of accuracy and in some senses provide more benefits than human combatants. There is a great deal of focus on the fact that human dignity is violated because a human commander is able to display mercy and compassion and robots do not possess such abilities, whilst this is true it does not stand to mean that human commanders *will* show mercy and compassion. A question raised by Dieter Birnbacher⁸⁴ is whether or not an attack by a terror bomber is less cruel because the commander of the aeroplane might in principle be merciful whereas an autonomous system would not.⁸⁵ In answer to this, it could be argued that there is a higher level of cruelty involved if an individual possesses the ability to be merciful yet actively chooses not to. In the discussion regarding AWS violating human dignity, it is imperative that there is a realisation that the mere capability to display mercy and compassion does not equate to a 'safer' warzone and a reduced risk to violations of human rights. By dismissing AWS because of their lack of human emotion is in some ways idealising human warfare by making the assumption that humans will always exercise correct and merciful judgment in the heat of a battle. It could be argued that humans are capable of violating the dignity of other humans in more ways than AWS are; robots lack the emotional capacity of humans and are yet to have full autonomy and so are limited to their programming. Humans, however, are capable of making their own decisions, and both historic and current war crimes demonstrate horrific violations of dignity and other human rights violations that were inflicted by humans without the aid of AWS.

⁸⁴ Dieter Birnbacher, (2016). Are autonomous weapons systems a threat to human dignity? In N. Bhuta, S. Beck, R. Geiß, H. Liu, & C. Kreß (Eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (pp. 105-121). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316597873.005

⁸⁵ *Ibid*

There is no doubt that AWS present us with dangers and uncertainties that should be dealt with and many of these include ethical considerations, however we should be cautious in thinking that AWS introduces a new quality of warfare.⁸⁶

3.5 Ethical and Moral implications

The concept of ethics and morals hold great significance in the discussion regarding AWS; as previously discussed, the human rights implications and potential violations create a worrying narrative when it comes to introducing this technology into modern warfare and integrating it into society. Some may argue that these concepts have been pushed to one side in a bid to advance this technology as quickly as possible and with limited hurdles. Whilst the terms are often used with a great level of interchangeability, ‘ethics’ and ‘morals’ are in fact two distinct concepts. When we consider what is meant by morals, we are referring to notions that can be seen as right and wrong, such notions influence us individually in our daily life and should be seen as subjective;⁸⁷ ethics however are norms that are shared by a group and are formed around mutual reciprocal recognition.⁸⁸ Clearly the differences here are somewhat subtle yet they are important non-the-less; this chapter will focus mainly on the ethical implications of AWS and will consider how such ethics exist in the broader international community.

The legality of implementing AWS has been at the forefront of the discussion regarding their development and deployment into armed conflict, however, the technological developments involved in this weaponry raise some serious concerns in relation to IHL, human rights and disarmament agreements. As with any new weapons development, each country strives to be in possession of the most advanced technology in order to put themselves at a strategic advancement. With the US government making the prediction of the automation of armed conflict for the year 2032,⁸⁹ it would be logical to assume that the 40 countries currently developing AWS for their militaries will be anxious to be in possession of the most up to date technologies. This considered, the continued development of AWS and LAWS has the potential to trigger global arms race, a concern raised by several experts, including Robert

⁸⁶ Nehal Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu and Clause Kreß, ‘Autonomous Weapons Systems, Law, Ethics, Policy’, (2016) Page 121

⁸⁷ Geoffrey C Jr Hazard, ‘Law, Morals and Ethics’ (1995) 19 S III U L J 447, Page 451

⁸⁸ Geoffrey C Jr Hazard, ‘Law, Morals and Ethics’ (1995) 19 S III ULJ 447, Page 453

⁸⁹ Wg Cdr (DR) U C Jha (Retd) ‘Killer Robots: Legal Autonomous Weapons Systems, Legal, Ethical and Moral Challenges Killer robots

Sparrow.⁹⁰ The Human Rights Watch made the following statement regarding this point: *“But the temptation will grow to acquire fully autonomous weapons, also known as ‘lethal autonomous robotics’ or ‘killer robots’. If one nation acquires these weapons, others may feel they have to follow suit to avoid falling behind in a robotic arms race”*,⁹¹ if one considers what is meant by this, we are brought back to the reason why AWS were developed in the first place and that is to play a valuable role in armed conflict and reduce the mortality rate of soldiers. Autonomous systems as a whole provide countless advantages to the human race, the introduction of Autonomous systems into biology, neuroscience and cybernetics have made the impossible a possibility and the advancements in these areas are not showing any signs in slowing down. However, this is not the area of robotics that has sparked calls for a pre-emptive ban; the robotic revolution in military weaponry is where the core fear stems from and the developments in this area raise numerous ethical and moral questions.

The kind of autonomous weapons that feature in science fiction films are of course not the technology that we are dealing with as it is yet to be developed, however one could argue that the calls for a pre-emptive ban are in fact based on these sci-fi orientated depictions of AWS becoming a reality and the ethical minefield that will be attached to this.

Attempts have been made to create some kind of laws or set of rules that would be able to govern AI in order for them to make ethical decisions for their own actions. One attempt comes in the form of The Three Laws of Robotics put forwards by Isaac Asimov. These rules provide us with an elegant set of ethical principals in relation to robotics, the three laws are as follows: (i) “a robot may not injure a human being, or through inaction, allow a human being to come to harm”, (ii) “A robot must obey the orders given it by human beings, except where such orders would conflict with the first law”, (iii) “a robot must protect its own existence as long as such protection does not conflict with the first or second law.”⁹² Whilst these laws are successful in the realms of science fiction, they have very little practical implications for the contemporary autonomous military reality that we are being faced with, and academics such as Ronald Arkin generally regard the rules as an inadequate basis for machine ethics.⁹³

⁹⁰ Robert Sparrow, Killer Robots, *Journal of Applied Philosophy*, Vol. 24, No. 1, 2007, p.62-77

⁹¹ A quote given by the Human Rights Watch on the discussion regarding the advancement of AI in the development of AWS and how these advancements are likely to revolutionize the way wars are fought, October 21, 2003.

⁹² Isaac Asimov, The Three Laws of Robotics

⁹³ Ronald Arkin, ‘Governing Lethal Behaviour in Autonomous Robots’ 48 (2009); see also Aaron Saenz, ‘Myth of Three Laws’ (May 10, 2011) < <https://singularityhub.com/2011/05/10/the-myth-of-the-three-laws-of-robotics-why-we-cant-control-intelligence/> >

Whilst Asimov's Laws can essentially be disregarded in this discussion, the concept of applying some degree of ethical thinking and moral responsibility to AWS is something that cannot be ignored, and one would argue that the ethics surrounding AWS hold a higher level of importance to the actual legality of them. This seems like a somewhat drastic view to take however, the implementation of AWS, and indeed the rapid development of AWS technology, provides a legal anomaly and an accountability gap. The race to keep up with the development in this area has seemingly blinded the parties involved to their moral and legal responsibilities and whether or not this technology would be able to comply with international humanitarian laws, human rights law or the laws of armed conflict. Having said this, there has not been a complete failure in this department; The Martens Clause,⁹⁴ which appears in the Geneva Convention, creates a legal obligation for states to consider moral implications when assessing new technology. Its application becomes necessary in the event there is no specific existing law on the topic, therefore would be applicable in relation to AWS and LAWS due to the limited laws in place to deal with them. In particular, this clause highlights the requirement for new technology to comply with the principles of humanity and dictates a public conscience. Many of the AWS that are in use are Human-on-the-loop weapons,⁹⁵ yet with the technological advancements that in progress, the machines have a quicker reaction time than humans; this therefore removes the ability for a human operator to override the robots actions before impact. If we consider this fact in conjunction with the Martens clause, we notice that the technology indicates there is a gap in international law that threatens the human right to dignity and provides an ethical conundrum.

Fundamental ethical implications of AWS have been highlighted by Lin in 2008, this came in the form of three core ethical implications:⁹⁶ (i) whether LAWS would be able to follow established guidelines of IHL and the rules of engagement, as specified in the Geneva Conventions; (ii) whether they would know the difference between military and civilian personnel; (iii) whether they would recognize a wounded soldier and refrain from shooting⁹⁷.

⁹⁴ The Martens Clause is part of the laws of armed conflicts, appearing in the preamble to the 1899 Hague Convention (II) with respect to the Laws and Customs of War on Land.

⁹⁵ robots that can select targets and deliver force under the oversight of a human operator who can override the robots actions and

⁹⁶ Whilst the three core ethical principles make reference to LAWS, due to the inherent similarities between LAWS and AWS, for the purpose of ethical consideration Lin's three fundamental ethical implications can be used in relation to the conversation on AWS.

⁹⁷ Lin Patrick, George Bekey and Keith Abney, 'Autonomous Military Robotics: Risk, Ethics and Design, report prepared for the US Department of Navy, Office of Naval research', 2008, p.21

When considering these implications, it becomes apparent that these all relate back to issue of responsibility and accountability for AWS; all of the questions posed by Lin's ethical implication can be answered if asked in relation to humans due to the fact that they are able to process information in both a logical and emotional way, something that AWS lacks. If we are to discuss these matters in the context of AWS, we start asking the following questions: if a robot acts for itself or in an unpredictable way that the human controllers don't understand, who will be held responsible for it? This is a question that has been considered in chapter 2 in relation to the legality of assigning responsibility and accountability to AWS, however if we look at it from an ethical standpoint, we are dealing with the lives of innocent civilians whom the technology could not identify as such. If we take the example of an AWS in the form of a drone that has been programmed to protect a boarder, this drone then erroneously identifies a mother and child seeking asylum as a hostile threat and shoots them both dead. If this duty was given to a human soldier then they would be able to distinguish between a threat and a civilian, yet at present AWS do not have this capability.

Separate, yet not dissimilar to the ethical concerns raised by AWS technology is the morality of AWS, with one of the main concerns being whether or not their algorithms can be considered to be discriminatory enough. This takes us back to the issue regarding distinction discussed above and provides us with one main question that is integral to this discussion: should we relinquish the decision to kill a human to a non-human machine?⁹⁸ This is a question that Aaron M. Johnson and Sidney Axinn discuss in their paper on the morality of autonomous robots, a question that one may argue is at the centre of all discussion be that legal, ethical, political or moral. The engineering involved in the development of AWS allows this technology to advance in a way that in theory makes them more compliant with IHL, calculating and performing tasks with levels of accuracy and efficiently than any human could possibly achieve. Yet despite this, it still fails to consider the moral judgments involved in these decisions; regardless of the progress made in machine development, the concept of integrating empathy and real human emotions such as love, guilt and mercy into a machine, is nothing short of science fiction. For example, if we consider a situation in armed combat where an enemy soldier finds themselves in the firing line of the oppositions AWS and makes an attempt to surrender, or pleads to be taken as a prisoner, the robot simply does not have an

⁹⁸ Aaron M. Johnson & Sidney Axinn (2013) 'THE MORALITY OF AUTONOMOUS ROBOTS', *Journal of Military Ethics*, 129-141, DOI: 10.1080/15027570.2013.818399

emotional capacity to make a judgment based on morals. The enemy soldier will consequently end up being shot and killed as this is the programming of the AWS.

It is not unreasonable to assume that this scenario has the capability of becoming a reality, nor is it unreasonable to assume that at some point in the future, robotics will develop to the stage at which they have an artificial sense of morality and are able to make decisions based on basic human morals that have been embedded into their algorithms. Hellstrom takes the view that futuristic robots will be programmed with highly advanced cognitive abilities to perceive, plan and learn as well as a multitude of complex behaviour. He argues that this increase in autonomy could open up the possibility of Military commanders treating them as though they were human soldiers, directing orders at them and expecting tasks to be carried out in the same way in which a human would.

CONCLUSION:

The questions this dissertation was seeking to address regarded the way in which AWS sit within legal frameworks and ethical frameworks and if these were suitable. The work will look at and evaluate these findings, then postulate some future suggestions that may be appropriate.

Chapter one provided some definitional parameters that allowed the discussion regarding AWS to develop, and in doing so provided us with foundations of the issues at hand. Autonomous weapons are not an entirely new form of technology, with basic forms such as antipersonnel land mines first used on a wide scale in World War II. Whilst these land mines work on the basis that there need not be any direct human intervention in setting it off, the Autonomous weapons that are being developed currently are embedded with extremely sophisticated software that allow for a more effective and advanced weapon. At the forefront of this discussion was the fact that AWS technology is being developed at an increasingly rapid rate with little consideration regarding the legal or ethical implications of such developments. Furthermore, the lexicography surrounding AWS is failing to adapt at an efficient rate, consequently paving the way to a dramatic definition ambiguity. An example of such ambiguity can be seen by the way in which ‘autonomous’ and ‘automation’ are being used interchangeably yet have different meanings and applications.

Whilst the misuse of certain terminology could be seen as trivial, it actually provides us with a serious issue that is deeply embedded in the legislation governing the use of AWS. With approximately 30 countries either deploying or developing defensive systems that can be placed somewhere on the spectrum of ‘autonomous’, this is not simply a domestic issue, but an issue that has international implications. One of the most central points in relation to definitional parameters is the fact that at present, fully autonomous weapons are essentially a thing of science fiction and do not currently exist in the way people believe they do. The prospect of a ‘killer robot’ and the potential dangers they pose has sparked fear and unease into civilians and academics alike, with a widespread call for a pre-emptive ban. However, the research that was conducted for the purpose of this dissertation has led me to believe that whilst there should be caution and perhaps a certain level of apprehension regarding the development of AWS, a pre-emptive ban could be seen as a fear-based decision based on the imagery of anthropogenic robots depicted in science fiction. A global consensus on what is deemed to be an Autonomous Weapon would provide much needed clarity, allowing

legislation to be effectively implemented and thus governing the use of such weaponry. A further concern discovered in chapter one was regarding self-awareness of AWS and the subsequent diminished human control that would follow. The delegation of the human decision-making responsibilities to AWS is a concern that many academics share, including Crootof and Asaro. Whilst it may be a trivial solution, introducing clarity surrounding what is classified as an Autonomous weapon provides us with a potential solution to reducing the uncertainty surrounding the development of self-aware AWS and whether or not they would be able to comply with the laws of war. Fully autonomous weapons are, again, merely hypothetical and the confusion here is once again embedded in conflicting definitions of what classifies as an AWS. Without definitional clarity it seems unlikely that any form of responsibility or accountability could be enforced, therefore leaving the floodgates open to a whole host of detrimental ethical implications.

The issues regarding responsibility and accountability of AWS formed one of the key discussions at the centre of this dissertation. The wide-spread definitional ambiguity involved in AWS has consequently led to holding any human agent responsible or accountable becoming extremely difficult. Chapter two was able to identify the possible parties that could be exposed to potential liability, as well as a consideration to potential criminal responsibility and the accountability gap that has developed. Any violations committed by AWS technology are likely to contain criminal elements; however, the application of criminal responsibility to AWS faces various difficulties due to the fact that this legislation was designed for human application and elements of criminal responsibility such as *mens rea* are not compatible with the capabilities of AWS. It was interesting to note however, that the general principles of criminal responsibility, specifically those relating to German law, and the individual criminal responsibility for war crimes provided an opportunity to re-design these existing laws in the context of AWS. This was a somewhat unexpected discovery due the fact that the majority of current laws have no apparent leeway to integrate provisions to deal with AWS being implemented and developed. I found however, that if one was to approach AWS as being akin to a Vordermann,⁹⁹ it removes the issues of lack of capacity to act with intent. The lack of human qualities AWS possesses have been the route to an extensive range of problems

⁹⁹ In this situation the consideration of AWS being akin to a Vordermann (the direct perpetrator of the criminal act) is based on the concept of a person committing a crime through another being indirect perpetration. The Hintermann controls the Vordermann in such a way that he is manipulated and used as a tool. Analogous to the way in which AWS are used as tools.

covered within this research, but if we were to consider AWS as an innocent agent, we then allow ourselves the possibility of attributing the responsibility and accountability to the human agent.

Further to the responsibility and accountability concerns addressed, I hoped to discuss the legal and ethical implication that AWS posed, chapter 3 provided an outlook on these issues. Considering the global impact that the deployment of AWS has, a discussion surrounding the International Humanitarian Laws and Human Rights implication provided an appropriate angle to fully consider the legal effects that AWS threatens. The right to human dignity emerged as being one of the most prevalent issues in relation to both the legal and ethical implications of AWS, this could be due to the fact that it is a right that each nation has in common and is something that humanity holds dear to them.

It seems to me that the development of this technology will continue regardless of whether the legislation is willing to keep up with these advancements. If we are to ask whether or not laws regarding responsibility and accountability for the actions of AWS are sufficient or not, the simple answer would be no. International and domestic laws provide the potential for appropriate adaptation yet a large proportion of the relevant parties are pre-occupied with either calling for a pre-emptive ban or engaging in preparation for a potential arms race.

Bibliography:

Cases:

- Prosecutor v Delalić and others, The International Criminal Tribunal for the Former Yugoslavia 1997-1999, page 462
- Prosecutor v Dusko Tadic a/k/a “DULE”, ‘Decision on the defence motion for interlocutory appeal on jurisdiction’ (International Criminal Tribunal for the former Yugoslavia)
- Prosecutor v Strugar, ICTY, Case No. IT—1-42- A, Judgement (appeals chamber), July 17, 2008, paras. 297-98
- Prosecutor v Halilović, ICTY, Case No. IT-01-48-T, Judgment (Trial Chamber), Nov, 26, 2005, para. 54
- Prosecutor V Delalicet. Et al, ICTY Judgement of 16 November 1998

Legislation:

- “S.397 – 109th Congress: The Protection of Lawful Commerce in Arms Act”, 2005
- International Covenant on Civil and Political Rights, Adopted and opened for signature, ratification and accession by general assembly resolution 2200A (XXI) of 16 December 1966 entry into force 23 March 1976, in accordance with Article 49
- international Criminal Court, Elements of Crimes art. 8 into, U.N. Doc. PCNICC/2000/1/Add.2 (June 30, 2000)
- UK Ministry of Defence, Joint Doctrine note, 2/11: The UK Approach to unmanned aircraft systems 2-3 (2011)
- *Unmanned Aircraft systems: Terminology, definitions and classification*, the UK Ministry of Defence, Joint directive Note 3/10, May 2010

International Sources:

- Alon M, ‘The Duty to Investigate Civilian Casualties During Armed Conflict’, Yearbook of the international humanitarian law, (Springer: Cambridge University Press 2012)
- Asaro P, ‘On Banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making’, International Review of the Red Cross, Volume 94 Number 886 (Summer 2012)

- Heyns C, ‘Autonomous weapons systems and human rights law, presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons’ (13 – 16 May 2014, Geneva, Switzerland)
- Heyns C, ‘Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions’, UN Doc. A/HRC/23/47, (9 April 2013, 17.)
- Human Rights Watch, “The ‘Killer Robots’ Accountability Gap: Obstacles to Legal Responsibility Show Need for Ban’. (2015).
- Human Rights Watch, Losing Humanity: The Case against Killer Robots (New York: Human Rights Watch and the International Human Rights Clinic, 19 November 2012)
- Human Rights Watch. Mind the Gap, The lack of accountability for killer robots, (April 9, 2015)
- ICRC, Customary IHL Database, ihl-databases-icrc.org/customary-ihl/eng/docs/citation, (30/01/20)
- Statute of the International Court of Justice, 59 Stat, 1055 (1945), arts 34-39; Practical information: frequently asked questions, The international court of justice

Books:

- N. Bhuta, S. Beck, R. Geiß, H. Liu, & C. Kreß (Eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (pp. 105-121). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316597873.005
- Shanahan M, The Technological Singularity (2015)
- Wg Cdr (Dr) U C Jha (Retd), ‘Killer Robots: Lethal Autonomous Weapons Systems Legal, Ethical and Moral Challenges’ (2016)

Journal Articles:

- Arkin R, Governing Lethal Behaviour in Autonomous Robots 48 (2009); see also Saenz, *Myth of Three Laws*
- Armin K, ‘Killer robots: legality and ethicality of autonomous weapons, USA: Ashgate’ (2009)
- Leys N, ‘Autonomous Weapons Systems and International Crises’, Strategic Studies Quarterly, Vol. 12, NO. 1 (SPRING 2018),
- Kastan B, ‘Autonomous Weapons Systems: A coming Legal “singularity”?’ ‘2013 Journal of Law, Technology and Policy, 45 (2013)

- Chu V S, 'The protection of lawful commerce in arms act: an overview of limiting tort liability of gun manufactures', Congressional Research Service (2012)
- Crootoof R, 'The Killer Robots are Here: Legal and Policy Implications', [2015] *Cardozo Law Review*, Vol 36
- Docherty B, 'The 'Killer Robots' Accountability Gap: Obstacles to Legal Responsibility Show Need for Ban' (2015)
- Foy J, 'Autonomous Weapons Systems: Taking the Human out of International Humanitarian Law'. *Dalhousie Journal of Legal Studies*, 23 (2014)
- Hammond D N, 'Autonomous Weapons and the Problem of State Accountability', *Chicago Journal of International Law*: Vol 15, No.2 (2015)
- Hazard G C Jr, 'Law, Morals and Ethics' (1995)
- Horowitz M C, 'Why words matter: the real-world consequences of defining autonomous weapons systems' 30 *Temp Int'l & Comp LJ* 85 (2106)
- Jain N, 'Perpetrators and Accessories in International Criminal Law, Individual Mode of Responsibility for Collective Crimes' (2014)
- Johnson A M. & Axinn S, 'THE MORALITY OF AUTONOMOUS ROBOTS', *Journal of Military Ethics*, 129-141, (2013)
- Marra W and McNeil S, "Understanding 'The Loop': Regulating the Next Generation of War Machines," 36 *Harvard Journal of Law and Public Policy* 36. (2013)
- McFarland T and McCormack T, 'Mind the Gap: Can developers of autonomous weapons systems be liable for war crimes?', *International Law studies*, U.S Naval War college. volume 90 (2014)
- Patrick L, Bekey G, and Abney K, 'Autonomous Military Robotics: Risk, Ethics and Design, report prepared for the US Department of Navy', Office of Naval research, (2008)
- Scharre P and Horowitz M C, 'An introduction to Autonomy in Weapon Systems', Centre for a New American Security, (2015)
- Schlatow J et al., "Self-awareness in Autonomous Automotive Systems." *Design, Automation & Test In Europe Conference & Exhibitin*. (2017).
- Tononi G, 'Consciousness as integrated information: a provisional manifesto', *The Biological Bulletin*, Vol. 215 No. 3 (2008)

Newspaper Article:

- McDonald H, 'Ex-Google Workers Fear 'Killer Robots' Could Cause Mass Atrocities' *the Guardian* (2019)

Online sources:

- Seumas Miller, 'Autonomous Weapons: An Introduction',
<<http://counterterrorismethics.com/autonomous-weapons-an-introduction/>>
- 'The Difference Between The Moral And The Legal' (Reason and Meaning 2020)
<<https://reasonandmeaning.com/2016/03/31/the-difference-between-the-moral-and-the-legal/>> Accessed 17 April 2020