COMMAND ACCOUNTABILITY FOR AI WEAPON SYSTEMS IN THE LAW OF ARMED CONFLICT*

Silahlı Çatışma Hukukunda Yapay Zeka Silah Sistemleri İçin Komuta Sorumluluğu

Berkant AKKUŞ[1]

ABSTRACT

Autonomous weapon systems can select and engage targets without meaningful human intervention. In light of developments in artificial intelligence and machine learning technology, autonomous weapon systems are likely to become a central feature of today's armed conflicts. Who is responsible when autonomous weapons systems commit a war crime is still being debated. In this article, to whom responsibility should be attributed will be examined in the context of command responsibility. Autonomous weapon systems destroy the chain of command in line with their ability to make decisions on their own. While the commander is expected to discipline his troops, to what extent will it be possible for autonomous systems consisting of algorithms. Military discipline systems prescribe punitive measures to ensure that the rules are followed. But artificial intelligence cannot be punished therefore it will not be able to respond to disciplinary actions. There are two main research questions that will be examined in this article. Because of systems that can make their own decisions through autonomy and artificial intelligence, when can a reasonable commander knew or should have known that an autonomous weapon system is about to commit a crime? What necessary and reasonable measures should commanders take to carry out their preventive duties? The concept of command responsibility regulates the relationship between human commanders and their subordinates. In the Rome statute, the commander is the person who exercises authority over military activities, but the commanders and subordinates are always considered human beings therefore the concept of command responsibility should be interpreted in light of the new developments in military technology.

ÖZET

Uluslararası insancıl hukuk ve uluslararası ceza hukuku uyarınca, bireyler işledikleri tüm savaş suçlarından cezai olarak sorumludur. Otonom silah sistemleri tarafından işlenen savaş suçlarından kimin sorumlu tutulabileceği ise belirsizdir. Silah sistemlerinde yapay zekanın kullanılması; operasyonel kuvvetlerin savaş ortamını anlamak, hedefleri belirlemek, izlemek ve seçmek, onları en uygun etkilerle yok etmek için sensörleri birleştirme yeteneğini geliştirir. Yapay zekanın ölüm zincirini kapatmaya yardımcı olma potansiyeli, insanların kararları ile makinelerin eylemleri arasında hesap verebilirlikte bir boşluk yarattığı ve insanların artık silahlı çatışmalar sırasında alınan kararlardan sorumlu olmadığı endişesini artırdı. Bu çalışma, askeri komutanın tüm savaş yöntem ve araçlarının kullanılmasından her zaman doğrudan ve bireysel olarak sorumlu olduğu için bir boşluk olmadığını göstermektedir. Komutanın askeri sorumluluğu savaş alanını kaplar. Bu sorumluluk, yapay zeka silah sistemlerinin kullanımı da dahil olmak üzere savaşta kullanılan kuvvet yapısına, silah sistemlerine ve taktiklere bağlıdır. Askeri hesap verebilirlik, askeri görevin temelidir ve silahlı çatışma hukukuna veya uluslararası insancıl hukuka uymaya ilişkin yasal yükümlülüğü içerir. Komutan, üstü askeri, sivil liderlere karşı sorumludur ve uluslararası ceza hukuku da dahil olmak üzere askeri düzen ve disiplin yoluyla uygulanan siyasi, kurumsal ve yasal yaptırımlara tabidir. Komutanın doğrudan ve bireysel sorumluluğu doktrini, üst düzey askeri liderlerin, otonom silah sistemlerinin gözetimi, seçimi ve kullanımı dahil olmak üzere hukuk ve liderlik ihlallerinden sorumlu olmasını sağlar. Bu makale, yapay zekanın kullanıldığı otonom silah sistemlerinin, komuta sorumluluğunda yarattığı problemleri inceleyecektir.

Keywords: Autonomous weapon systems, artificial intelligence, command responsibility.

Anahtar Kelimeler: Otonom silah sistemleri, yapay zekâ, komuta sorumluluğu.

1. ORCID: 0000-0001-6652-2512

1. Dr. Öğr. Üyesi, İstanbul Okan Üniversitesi, berkantakkus91@gmail.com

GENİŞLETİLMİŞ ÖZET

Dolaylı sorumluluk olarak da bilinen komuta sorumluluğu, Uluslararası ceza hukukunun bir askeri komutanı veya bir sivil üstü cezai olarak sorumlu tutma olasılığını sunan bir dalıdır. Suçlu bir eylemde bulunan astı, engellerken veya cezalandırırken, üstünün makul önlemleri almadığı tespit edilirse potansiyel olarak ortaya çıkabilir. Bunun nedeni üstün, bağlı olanların eylemleri üzerinde etkin bir kontrole sahip olması ve bu nedenle, astlarından birinin suç işlediğini öğrenir öğrenmez, bu konuda harekete geçme görevidir. Komuta sorumluluğu, görevi ihmalden üstün sorumlu tutulacağını belirten Prosecutor v Halilović davasında tartışılmıştır. Roma Statüsü'nün 28. Maddesi, komutan ve diğer üstlerin sorumluluğuna ilişkin bir hüküm getirmektedir. Bir askeri komutan veya askeri komutan olarak fiilen hareket eden bir kişi, mahkemenin yargı yetkisine giren suçlardan, kendi etkin komuta ve kontrolü veya duruma göre etkin yetki ve kontrolü altındaki kuvvetler tarafından işlenen suçlardan, bu tür güçler üzerinde düzgün bir şekilde kontrol sağlayamamasının bir sonucu olarak, cezai olarak sorumlu olacaktır. Burada tartışılan hesap verme sorumluluğu; üstün, kuvvetlerin bu tür suçları işlediğini veya işlemek üzere olduğunu bilmesi veya bunların işlenmesini önlemek, bastırmak için gerekli ve makul tüm önlemleri almamış olması durumunda devreye girer. Bu düzeyde bir sorumluluk için önemli bir husus, taraflar arasında bir üst ve ast ilişkisi olduğunun kanıtlanması gerektiği gerçeği olacaktır. Bu kanıtlanmazsa, gerekli unsurlar yerine getirilemez ve üstün cezai sorumluluğu ortadan kalkar.

Silah olmaları ve dolayısıyla askerlere astlara benzer bir bağlamda işlev görmeleri nedeniyle, otonom silah sistemlerinin üstlerin sorumluluğunu tetiklemek için potansiyel nedenleri olduğu tartışılabilir. Silah sistemi tamamen otonomsa, hedef seçimi ve bir tehditle etkileşime girmeyle ilgili olarak, kendi bağımsız kararlarını ve yargılarını verme konusunda belirli bir yeteneğe sahiptir. Üstün emirlerinden bağımsız olarak çalışabilmeleri için özel olarak tasarlanmış gelişmiş teknolojiye sahiptir. Üstün doğrudan komutası altında değillerse, üstlerin komuta sorumluluk doktrinini tetikleyecek uygun bilgi düzeyine yakın herhangi bir yerde olmasını beklemek mantıksız olacaktır. Otonom silah sistemleri teknolojisinin karar verme hızının, üstlerin, astların böyle bir suç işlemek üzere olduğunu ve daha sonra onları komuta sorumluluğuna tabi kılabileceklerini öngörmelerini neredeyse imkânsız kılacağı da iddia edilebilir.

Robotun harekete geçmeden önce kararını komutana iletebildiğini düşünürsek, bu teorik olarak komutanın yaklaşmakta olan suç eylemi hakkında yeterli bilgiye sahip olacağı ve dolayısıyla sorumlu tutulabileceği anlamına gelir. Bu durum yalnızca bir döngüdeki insan sistemi durumunda mümkündür, ancak sistem tamamen otonom olsaydı, bir üstle iletişimde çarpıcı bir azalma olurdu. Onaya gerek kalmadan hareketler yapılabilir ve kararlar olumsuz olurdu. Ayrıca, bir komutanın soruşturma yapabilmesi için bilgilere sahip olması gerektiği, bu bilgileri almadan, ihmal nedeniyle bilgi edinememesinden sorumlu tutulamayacağı tespit edilmiştir. Prosecutor v Strugar davasında, belirli bir astın geçmişteki suçlarına ilişkin bilgisinin yeterli bilgi oluşturabileceği, bu nedenle üstü, söz konusu astın gelecekte cezai fiiller işleyebileceği konusunda yeterli bilgiye sahip olacağından, komuta sorumluluğunun manevi unsuru yerine getirilecektir.

Otonom silah sistemleri tarafından işlenen geçmiş yasa dışı eylemler için eşiğin ne olduğu konusu belirsizdir. Otonom silah sistemler teknolojisi öncelikle yardımı amacıyla geliştirilmiştir ve uyarlanabilirliği nedeniyle bu teknolojinin yerleştirilebileceği farklı senaryolar ve atanabileceği farklı görevler vardır. Otonom robotların uluslararası hukuku ihlal eden bir davranışta bulunduğu bir durumu ele alırsak, bu davranışın, o türden tüm robotlara ve aynı programlamaya uygulanabilen geçmişteki yasa dışı bir eylem için emsal teşkil ettiğini veya aynı programlamaya sahip olduğunu varsayabiliriz. Belirli bir algoritmanın, savaşanları hedef alacakları anlamına gelen bir otonom robot modeline programlandığı bir senaryo tasarlayabiliriz ancak bunlardan biri sivilleri savaşan olarak karıştırır, sivilleri öldürebilir. Bu senaryo gerçeğe dönüşme yeteneğine sahiptir.

Otonom silah sistemleri teknolojisi hızla ilerleyen doğası gereği karmaşıktır, teoride, komutanın otonom robotun yakında bir suç eylemi gerçekleştireceğini fark edebilmesi için, programlamasının doğası ve sunduğu otonomi düzeyi nedeniyle derinlemesine bir teknolojik anlayışa sahip olması gerekir. Bunun mantıksız olduğu kadar gerçekçi de olmadığı tartışılabilir, çünkü komutanın sorumluluktan bir şekilde kaçınması muhtemel görünüyor. Otonom silah sistemlerinde hızlı karar verme konusu, komuta sorumluluğunun uygulanabilir olması için bir ast üzerinde etkili bir kontrolün olması gerektiği kavramıyla da bağlantılıdır. Prosecutor v Delalicet al davasındaki karara göre, etkili kontrol özellikle suç davranışı önleme veya cezalandırmaya yönelik maddi yetenekten gelir. Otonom silah sistemlerinin cezalandırılması sadece anlamsız değil, aynı zamanda oldukça imkansızdır. Otonom sistemlerin öngörülemezliği, etkili kontrolün uygulanmasının sorgulanabilir olmasına katkıda bulunur. Öngörülemezlik

örnekleri, siber müdaheleler, programlama hataları ve otonom silah sistemelernin makine öğrenmesini kullanması olabilir. Bu durumlardan biri meydana gelirse, komutanın müdahale etme ve saldırıyı durdurma yeteneği azalır ve bu da etkin kontrolün yerinde olmadığını gösterir.

Komuta sorumluluğunun uygulanmasının, yetersiz bilgi ve etkin kontrol eksikliği etrafında dönen sayısız sorunla karşı karşıya kalacağı ve sonuç olarak komuta sorumluluğunun izlenmesi, gerçekçi olmayan bir sorumluluk biçimi olacağı sonucuna varılabilir.

INTRODUCTION

Over time, the nature of war has significantly changed. Whilst warfighting once marked a clash between states and was typically conducted away from densely populated areas, non-state conflict is ever more dominant, and non-combatants are becoming ever more involved in clashes, whether willingly or otherwise. As a result, or sometimes as a direct cause of these shifts, the ways and means of war have also taken on a new form (Anderson, 2009:344). The laws of armed conflict (LOAC) have also evolved, with the prominence of international criminal law (ICL) marking one of the most notable outcomes:

"Since the laws of war are for the most part still premised on the concept of classic international armed conflict, it proved difficult to fit this law into modern war crimes trials dealing with crimes committed during non-international armed conflicts. The criminal law process has therefore updated the laws of war. The international criminal judge has brought the realities of modern warfare into line with the purpose of the laws of war the prevention of unnecessary suffering and the enforcement of fair play. (van Sliedregt, 2012:3)"

This article will examine the impact of autonomous systems as one of the latest wartime developments, and the impact of their technological advances. On this subject, there are two key considerations to be made. The first is whether or not an autonomous weapon system would jeopardise the ICC's powers of prosecution when considering offences defined by the Rome Statute. Most importantly, the necessary mens rea requirements will be difficult to prove when a highly autonomous weapon system is considered at fault. Secondly, it is important to examine how to address such an issue. This article will consider the option of adjusting the mental element (mens rea) requirements and attributing accountability to the autonomous weapon systems' (AWS) commander. There are plenty of possible actors to whom responsibility can be attributed for violations committed by a fully autonomous weapon. Individuals involved in the development and deployment of fully autonomous weapon systems, such as software programmers, hardware manufacturers and vendors, political leaders, military commanders, and their subordinates, may be held responsible for violations of IHL or international human rights law. However, in the context of international law, the main addressees of IHL rules are the planners or executors of military activities (Güneysu, 2022:12). For this reason, this article solely focuses on command responsibility under international law.

1. Liability and Autonomous Weapon Systems

As technology advances, so too does the debate around the use of AWS. In this section, this work will address the concept of criminal liability and its application to the world of AWS.

Academics have approached AWS and criminal liability from a plethora of angles. At one end of the spectrum, studies have suggested that humans should be accountable for every single action of an AWS, whether legal or illegal. This approach considers AWS to be deterministic, owing to the fact that they are a human creation. As such, it considers that AWS are unable to exceed the expectations or intelligence of their human creators (Schmitt, 2013:4). At the other end of the scale, research has suggested that it would be dangerous to attribute complete responsibility for an AWS's actions to its human creators or operators, as their autonomy creates a level of unpredictability and lack of control (Human Rights Watch, 2014). Others consider it too difficult to attribute criminal liability to an AWS, suggesting instead that a utilitarian approach would be more appropriate, assigning collective responsibility to the state (Hammond, 2014).

This last view, as supported by Waxman and Anderson, appears to be the most dangerous. It supports a repeat of past performances where wartime actions have relied entirely on a regime of state liability. Further, it represents a total regression in terms of the movement towards individual liability, initiated during the Nuremberg Trials after WWII. This movement remains important today and has been instrumental in recent tribunals and courts, including the ICC,

ICTR, and ICTY (Anderson & Waxman, 2013:6). Before considering this position in further detail, it is first prudent to describe the technology behind AWS and its relationship with ICL. For the most part, an autonomous system is able to conduct its specified tasks without any human interaction. It can identify circumstances, make relevant decisions, and act upon them without the need for human intervention.

According to Scharre and Horowitz, autonomy has a different meaning in the context of each of the specific actions listed above. The pair believe autonomy to mean the level of control exercised by humans over an AWS's decision-making. That is to say, autonomy measures whether or not humans are actively involved in a system's decision-making process (known as the man in the loop), merely supervise it (known as the man on the loop), or have no involvement at all (known as the man out of the loop) (Horowitz, 2016:85). In the case of an AWS's actions, the pair measure autonomy in terms of a system's ability to execute an action that its programmers could not directly predict. Otherwise put, this considers the foreseeability of an AWS's behaviour. When considering the most advanced systems, AI and machine learning negate the predictability of, and human control over, such system's actions. Whilst machine-learning may generate a better result than human operators for instance IBM's Watson won Jeopardy! and correctly diagnosed cancer, it offers no form of a black box, allowing humans to examine the causal logic for certain results or to foresee the outcome (Scharre & Horowitz, 2015:9).

Considering international criminal law and individual criminal liability, the Prosecutor v. Furundzija International Criminal Tribunal for the Former Yugoslavia (ICTY) specified that the accused needs a criminal intent (mens rea) by way of "knowledge that the accomplice's action will assist the perpetrator in the commission of the crime" (Prosecutor v Furundžija, 1998:245). If there were no malice in the programming of an AWS, it would not be possible to demonstrate sufficient criminal intent where a programmer was unable to predict an AWS's actions. Further, when the concept of command responsibility is considered in greater detail, the ICTY in Prosecutor v. Delalic (Celebici) specified that the requisite actus reus would be effective control "…material ability to prevent and punish the commission of these offences" (Prosecutor v Mucić , 1998:377). In Prosecutor v. Halilovic, the court explained the necessary mens rea as being the understanding of, or belief of knowing about, a crime (Prosecutor v. Halilovic, 2005:54). It is improbable that a commander could be considered criminally responsible for failing to ex ante prevent or ex post facto punish an AWS if they had no material ability to do so.

These issues of individual liability have already been examined in the context of AWS, with Waxman and Anderson recognising but downplaying their relevance. According to the pair, state liability is a more appropriate model, with it being more traditionally compliant with ICL. Further, they suggest that society should not necessarily sacrifice any potential gains offered by AWS, simply on account of a lack of human accountability (Anderson, Reisner, & Waxman, 2014: 399). Whilst, much like any utilitarian argument, their claims may appear encouraging, they are impractical in both a legal and political sense. Furthermore, they disregard the important post-WWII shift towards individual criminal liability.

In accordance with Art. 43 of Additional Protocol I, states are only accountable for attributable illegal actions performed by their military personnel. As such, no offence committed by an AWS would result in direct state liability. Further, since commanders would be unable to foresee or fully control an AWS's actions in the battlespace, states would not be in a position to assume liability by proxy.

Furthermore, it would be impractical for states to be held responsible for the actions of their AWS on the international stage. This would be particularly relevant for the five nations which hold veto powers in the U.N. Security Council. The lack of effective international establishments and political considerations already negate such possibilities, and would continue to do so (Trahan, 2020:51). This can be seen clearly in the example of Russia's annexation of Crimea. Furthermore, in terms of the use of AWS in the near future, these five nations are most likely to be at the helm.

Post-war years have seen a notable shift towards individual criminal liability. Such an approach was born at the Nuremberg and Tokyo Trials, during which Nazi and Japanese officials were tried for WWII war crimes. This movement later saw the birth of a number of other one-off trials, including the ICTY and ICTR's conviction of individuals for crimes against humanity, war crimes, and genocide during the 1990s Balkan Crisis and in Rwanda in 1994.

In more recent times, this movement has seen the birth of the ICC, the (first permanent) International Criminal Court. The ICC's purpose is to hold perpetrators to account and to prevent the reoccurrence of criminal acts. The institutional

power, case selection, and effectiveness of the ICC are, however, widely criticised (Badar & Porro, 2015:665). Nonetheless, its creation at the turn of the century, and prompt support from key state actors including the UK, USA, and Israel of which a number went on to withdraw their support, highlighted that the leaning towards individual liability was still very present.

Complete reliance on state liability for the guidance and mitigation of an AWS's forecast shortcomings would be both impractical and counterproductive. Notwithstanding the solutions available for AWS, such an approach would go against the principles of ICL and hamper its current progression. According to Sparrow, "it is a fundamental condition of fighting a just war that someone may be held responsible for the deaths of enemies killed in the course of it." (Sparrow, 2007:67) Thus, when considering any crime committed by AWS, all future solutions, be they technological, political or legal, should endeavour to support the attribution of criminal liability at both a state and individual level.

2. Understanding Autonomous Technology

First and foremost, it is important to understand autonomous technologies and the responsibility gap. A typical image conjured up by proponents of this gap sees a military commander authorise the use of an AWS in theatre. That AWS then goes on to use lethal force in such a way that its actions, if performed by a human soldier, would be considered illegal, or a violation of LOAC.

This scenario presents two liability gaps. The first exists between the system and the outcome of its actions. This results from the fact that the system cannot be held legally responsible for its actions. One day, it could be possible that such a system would be considered to have sufficient moral agency to be culpable. However, this is a possibility in the distant future and could prove altogether impossible. The second gap exists between the system's actions and its commander. This results from a normative separation between the commander's authorisations and the resulting actions of the AWS. Arguably, in this instance, the AWS's actions were not predictable by the commander, or the AWS acted outside of the commander's orders or control. As such, the commander cannot be justly held accountable for the system's behaviour. The key question that arises from these gaps is: what does autonomy mean in the context of these technologies? There are two alternative views. The first presents system autonomy as strong autonomy, equivalent to that of humans.

Strong system autonomy: A technological system is considered to be strongly autonomous when it can act based on its own internal reasoning and previous experience. If the system possesses this level of autonomy, it is, in theory, a moral agent. It could, however, have a number of deficiencies that render it an irresponsible moral agent. The responsibility gap between a commander and such a robot is akin to the responsibility gap between two humans. An alternative argument disagrees with this approach, claiming that the term autonomous, when used to describe technologies, has a unique and non-analogous meaning (Hoven & Santoni de Sio, 2018).

Emergent system autonomy: A technological system is considered to be emergently autonomous when it acts according to its sensory data and according to probability-based reasoning that it uses for machine learning and fault correction. This sensory data can be unpredictable. Such a mode of autonomy is more relevant to its dynamic abilities than to its moral-based decision-making and free will. A system can be considered autonomous if it is deployed in a range of areas and is able to adapt and respond to the specifics of each environment (Marra & McNeil, 2013:61). Because the behaviour of such systems is unpredictable, emergent autonomy produces a responsibility gap. This work examines legal doctrines that would allow for this gap to be bridged, regardless of the level of autonomy given to the system.

2.1. Strong Autonomy: Bridging the Gap

Consider a strongly autonomous system. This strong autonomy does not mean that the deploying commander is not accountable for the system's action. This issue has long been addressed by the courts. As a result, two defined doctrines have emerged. The first is called innocent agency or perpetration. The second is called command responsibility.

Innocent agency or perpetration is, perhaps, a less familiar concept. It considers a situation in which one individual, the principal, employs another individual, or as will be shown, a human-led organisation, to perform an illegal act

for them (Williams, 1992:291). Consider a situation in which a woman has grown tired of her husband and decides to poison him. She does not, however, want to give him the poison herself. As such, she elects to combine the poison with sugar and asks her young son to add sugar to his father's tea. The child adds the sugar and, unbeknown to him, the poison, to the tea.

In this scenario, the wife has employed another individual to perform an illegal act for her. Evidently, the second individual is innocent of the crime. As the son was unaware of the criminal act, it would be unjust to hold him accountable for the crime. If the wife has hired a hitman to commit the crime, it would be appropriate to hold the second individual accountable. Under common law, the wife can be held liable for the crime, according to the innocent agency. The innocent agency exists when an individual uses another individual, who has some form of incapacity for responsibility, to commit a crime for them. Classic examples include those who exploit ignorant, juvenile, or mentally unwell individuals.

Further, there is the legal notion of perpetration which means the act of committing a crime. Perpetration applies in cases where an individual, or indirect perpetrator, uses another individual, or direct perpetrator, to perform an illegal act for them. In such cases, the direct perpetrator is used as an instrument of crime. As such, the direct perpetrator must, in some way, be at a deficit to the indirect perpetrator. Section 25 of the German Criminal Code details the scope of perpetration and contains some interesting elements. The indirect perpetrator has acted hegemony and dominates the direct perpetrator's will (Bohlander, 2009).

This can happen in three distinct ways. The first is coercion. The second is the use of a mistake by the direct perpetrator. The third is control over an organisation (Jain, 2013:831). This final method is of particular interest as it proposes a scenario in which a perpetrator employs a bureaucratic organisation to fulfil their aims. Further, Article 25 of the Rome Statute acknowledges perpetration, and the ICC covers perpetration via an organisation.

How, then, does this doctrine relate to lethal autonomous weapon systems and the responsibility gap? The answer appears obvious. If such systems have strong autonomy but also are deficient in such a way that they cannot act morally, they are effectively akin to innocent agents or direct perpetrators. According to the doctrine of perpetration, human commanders should be held accountable for their actions, provided they were able to predict the systems' actions, or they met the commanders' intent. The issue here, however, is that innocent agency and perpetration do not apply to situations in which these actions extend beyond the commander's permissions and orders. To bridge this gap, the doctrine of command responsibility would offer the most appropriate solution. This doctrine is more widely understood, albeit it has been met with some controversy. Command responsibility is broken down into three key features, all allowing for the allocation of responsibility. The first applies when a commander has effective control over their subordinates (Cassese, 2003:205). The second is when a commander knew or should have known, that their subordinates were acting illegally. The third is when a commander failed to deter or stop illegal behaviour, or discipline subordinates for it (Garraway, 1999:787).

Command responsibility applies equally to civilian and military commanders, although its application is typically stricter for military commanders. In order for command responsibility to be applied to civilian commanders, they must have been aware of the actions of their subordinates. Military commanders, however, can be held to account for failing to know about their subordinates' actions when it is deemed that they should have known. This measurement of their failure is known as the negligence standard. International law recognises command responsibility, and it is documented in Article 28 of the ICC's Rome Statute. To be applicable, the court must prove a causal link between the commander's actions or lack of action and those of their subordinates. There also has to be a temporal link between the commander's period of control and their subordinate's illegal conduct.

This could also be easily applied to the context of a strongly autonomous system. Any commander that deploys such a system would be accountable for its actions, provided they have effective control over it, they know of, or should know of, its illegal conduct, and they do not prevent or curtail its criminal behaviour. The issue here is that the system is presumed to act in a predictable and reasonable way. It is presumed that the system's actions are ones that the commander could have been, or indeed should have been, conscious of. If a system has strong autonomy, this could be possible. However, for systems that have emergent autonomy, this might not be the case.

2.2. Emergent Autonomy: Bridging the Gap

In this case, the challenge arises when a system's behaviour is dynamic and adaptable and when it responds to its changing environment in an intricate, unforeseeable way. The method by which such a system will adapt its behaviour could be difficult for a human commander to understand. Indeed, it could even be difficult for its programmers to understand, if the system has machine-learning capabilities. As such, commanders will be hesitant and less able to predict AWS's actions. In such circumstances, it is particularly difficult to bridge the responsibility gap. Even if the doctrines of command responsibility and perpetration seem relevant and applicable, it is debatable whether the application of these doctrines would be appropriate or fair. An important discussion will centre on the mental component inherent in each doctrine.

In terms of criminal law, an individual can only be held accountable for an offence if they meet specific mental conditions of mens rea. Otherwise put, a responsible individual must have had intended for, or knowledge of, a crime, or have been reckless or negligent in some way (Ohlin, 2016:22). In murder cases, individuals are considered responsible if they intended to kill someone or to seriously injure them. In terms of manslaughter, they need to have been negligent or reckless to such an extent that their behaviour was likely to cause the death of another. If command responsibility is to be applied to emergently autonomous technologies, it will need to be applied through recklessness or negligence standards. Traditionally, the application of perpetration does not cater for recklessness or negligence. A direct perpetrator must have been cognisant of the indirect perpetrator's crime, or elements thereof. Command responsibility does offer this provision for the recklessness and negligence standards. Concerning civilian commanders, recklessness would need to be proven. Concerning military commanders, however, negligence will suffice. As such, if perpetration were to be applied to emergently autonomous systems, the mens rea standard would need to be reduced accordingly.

Nevertheless, this adjustment may fail to bridge the gap. If we first consider recklessness. There is no unilateral consensus on the requirements of this mental component. What is agreed, is that for someone to be considered reckless, they must have known about and ignored a significant risk of criminality (Bo, 2021). What is not agreed upon, is the framework within which the level of risk is assessed. Perhaps the individual should have acknowledged that the risk was significant. Alternatively, perhaps they simply needed to have noted that there was a level of risk, with courts responsible for assessing the significance of this risk. This is an important distinction. Individuals may disagree on what constitutes a significant risk and what does not. Military commanders could have different standards from the general population and their civilian counterparts in terms of risk. Something commonly considered to be a significant risk could form part of standard practice for a military commander.

Further, there is disagreement about whether or not an individual needs to have understood the specific level of harm that would be caused, or whether it would suffice for them to have known the general category within which the harm would fall. Therefore, in terms of a military operation that has derailed, should a commander have accepted the risk of collateral damage, or should they have accepted a specific risk of a specific group of people being harmed? Again, this is an important distinction. If a commander simply needs to recognise and ignore the risk to be reckless, their responsibility, according to this standard, will be easier to prove. The negligence standard gives rise to similar considerations. Negligence applies to circumstances in which risks were not noted and ignored but should have been recognised. Negligence concerns duties of care and individuals' failure to meet them. The consideration, here, is what a reasonable individual, or professional, would have been able to foresee. In the case of AWS, it would examine whether a military commander could have predicted the deviant behaviour of the system. This behaviour could even be entirely unprecedented.

Naturally, a reasonable military commander should foresee some risk when deploying an AWS. Military missions always have associated risks and AWS are, after all, lethal weapons. Does this alone fulfil the negligence standard? If it becomes too simple for a military commander to be held responsible for the actions of an AWS, their development, and employment could be severely obstructed. Although the Campaign against Killer Robots may welcome such an outcome, there are many others who would not. Proponents of the technology will argue that it brings potential advantages, much like those who champion self-driving cars. Further, they will argue that by setting the mens rea standard too low, we will not be able to reap these rewards.

3. Superior Responsibility and Autonomous Weapon Systems

If AWS technologies are not subject to a blanket ban, their use will become the guiding issue. Defined as the linchpin of international law, individual responsibility would if absent, collapse the rule of law (Prosecutor v. Tadić, 1997:665).

The leading thought on this, although incomplete, is that a lethal autonomous system would disconnect actus reus from mens rea (Sharkey, 2012:791). In the absence of either, no crimes would be committed. When a crime is committed by an unaccountable object, atrocities occur (Roff, 2013:354). Owing to this fact, victims would never receive any admissions of guilt, recourse, or reparation (Margulies, 2017). Ultimately, lethal machines would sever the chain of command in a way that limits compliance with the Law of Armed Conflict (LOAC) or the practice of just war. As such, the systems could be considered unlawful (Sparrow, 2007:62). This opinion can be summarised as such: if a machine makes decisions that its commanders cannot predict or understand, the mechanism of command responsibility fails. This, however, implies two things. The first is that such an attenuation prohibits command responsibility. The second is that AWS are both unreliable and unpredictable. Neither of these statements are true.

3.1. Attenuation Does Not Prohibit Command Responsibility

Aside from discussions on the coding of attribution or rules of engagement (ROE) into an AWS's software, no accountability gap will exist while either strand of command responsibility exists. An incorrect assumption would suggest that by disconnecting actus reus from mens rea, liability is removed. This assumption does not cater for circumstances in which illegal orders are issued (Arkin, 2009:211).

3.2. Issuing Orders: Direct Command Responsibility

According to both the ad-hoc tribunals and the ICC's Rome Statute, commanders are considered individually liable for any illegal order they issue to an autonomous system. By issuing such an order, the commander has directly contributed to a crime. To illustrate such a scenario, consider a caravan that is transiting through Taliban territory.

The commander is aware that the caravan is full of civilians. Rather than a Predator drone, controlled by a remote pilot in Nevada, consider that the commander, themselves, has control over an autonomous drone that is loitering in the vicinity. If the commander called forth the autonomous drone and ordered a strike upon the caravan, they would be guilty of a crime. Through their actions in this scenario, the commander would have proven the requisite actus reus by issuing an illegal order with de jure or de facto command. This would be proven by their ability to issue orders to the drone. Secondly, they would have proven the requisite mens rea as they knew the caravan to be full of civilians. According to the ICTY, this commander would have acted "with the awareness of a substantial likelihood" that the drone strike would constitute a crime against humanity or a war crime (Prosecutor v. Kordić, 2004).

Conversely, it could be argued that the commander's strike order would not have determined the actions subsequently taken by the machine. Albeit the commander would not necessarily have known precisely how the machine would execute the strike, the fundamental facts remain the same. First and foremost, the commander issued the order to strike. Had the commander not done so, the drone would not have engaged the target. Secondly, the commander knew, and had sufficient mens rea, of the fact that this order was illegal, as they knew that the caravan was full of non-combatants and that an attack upon them would likely cause their death. The fact that the commander was not able to accurately predict the specific manoeuvres of the machine does not negate their responsibility for the strike. The commander was only required to know that the order was illegal and to then issue the order regardless.

If each of the requisite elements were proven, a defence council would struggle to apportion blame to the drone. As such, autonomous weapon systems offer no obstacle to the proof of command responsibility. The second strand of command responsibility, however, is more complex.

3.3. Dereliction of Duty: Indirect Command Responsibility

The second strand of command responsibility focuses on the notion that a machine's decisions isolate responsibility, and that no other parties can be held accountable for its actions. As shown by academic concern about responsibility for omission (Meloni, 2007:620), this viewpoint considers it unjust to hold a commander accountable for any action that takes place beyond the boundaries of their personal awareness. Nevertheless, in such circumstances, a

commander is not responsible because of their knowledge of actions, but because of their dereliction of duty (Prosecutor v. Krnojelac, 2003:171).

Consider the previous example and assume that the commander is unaware that the caravan is full of non-combatants. The commander does, however, understand that there may be civilians in the area and that by activating the drone, they risk ordering a strike upon human activity near targeted agents (Cockburn, 2016:10). The commander would not be culpable through direct command responsibility as their activation of the machine in itself would not be illegal. They did not, after all, know that the civilians were indeed in the area, or that the system would target them (Prosecutor v. Blaškić, 2004:41).

Moreover, consider that an autonomous drone fires upon a caravan full of civilians. The fact that the drone was autonomous, rather than a remotely piloted system, does not affect the matter of liability as the crime does not rest upon the commander knowing the illegal outcome of their order. Rather, the cornerstone of the crime is the commander's dereliction of duty, deploying the drone in the knowledge that there are civilians in the vicinity and that the system may fire upon them. What is most important is whether this understanding and the cost-benefit analysis made when considering humanity and military necessity, as defined by LOAC, constitutes liability. It is not the case that these machines provide an escape route for commanders simply because deaths are caused by the weapon systems.

Otherwise put, a commander who instigates a remotely piloted system in the same circumstances, with the same contextual understanding and result, would hold the same level of liability for the outcome. The commander would have committed a crime by means of dereliction of duty. They would have deliberately ignored the risk of there being civilians in the area and chosen to deploy the drone irrespective of this risk. Such systems do not initiate their own actions. This is done by commanders. The chain of command, therefore, is not severed.

If it is impossible for a software programmer to understand precisely how a weapon system's coding will perform in the battlespace, it could have arguably been impossible for the commander to know that the system would fire upon the caravan (Roff, 2013:357). Although there is always a level of unpredictability with such machines, this argument is misinformed and excessive. It is excessive because the law has never required military personnel to understand the precise inner-workings of their weaponry. It is misinformed because autonomous weaponry would never be employed if its behaviour were entirely unreliable or unpredictable. This would be a question of effective control and not of the commander's mens rea.

3.4. AWS Are Neither Unreliable nor Unpredictable

The argument that command responsibility fails when machines make decisions that their superiors cannot foresee also promotes the belief that autonomous weapon systems cannot be effectively controlled (Human Rights Watch, 2014:20). This belief is unfounded, not because it creates polarity, suggesting that AWS are either useless or have fourth-loop AI, but also because it does not acknowledge the initial decision to engage as the moment of responsibility.

At the bottom end of the scale, there are in loop systems. Such systems do not make any autonomous decisions and pose no risk to effective control. At the top end of the scale are fourth-loop systems which do not exist yet. AWS with fourth-loop AI would be the only systems that would validate Sparrow's argument that: "If we hold anyone else responsible for the actions of an agent i.e., an autonomous weapon, we must hold that, in relation to those acts at least, they were not autonomous." (Sparrow, 2007:65) As a commander's accountability either relies on an illegal order or their dereliction of duty, out of the loop, or third-loop, systems are no different from any other (Anthony & Holland, 2014:423). As seen in the decisions made in the case of General Yamashita and the subsequent ICTY and ICTR, even without full understanding of their subordinates' actions, commanders are criminally liable for them (Parks, 1973:22).

Further, although AWS may operate in a way that humans cannot understand, and while it may be true that autonomous machines operate at levels incomprehensible to humans, they do not prevent effective control (Anderson & Waxman, 2013:2). The system's commander does not delegate their decision to either strike or refrain from striking a target. Until human intelligence can be fully replicated artificially, culpability will remain with the human who decides to deploy the system—a commander. Whilst the concept of robotic choice may seem relevant when

considering the blurred lines between attribution and responsibility (Schulzke, 2013), examination and correct application of law silence this argument. According to the doctrine surrounding superior responsibility, a commander remains liable even if they did not order, actin, or know the outcome of an event before it took place (Bassiouni, 2013:336).

What is important here, is that commanders perform a duty. That duty, in itself, serves a role in subsequent actions (Henckaerts & Beck, 2005:556). This role is recognised by customary international law and a high premium is placed upon the attached liabilities (D'Amato, 1986:607). A system's operator is responsible for assessing the capability of the machine faced with the threat, and deciding whether or not the system's autonomous use is appropriate to the scenario (U.S. Department of Defense, 2012:2).

CONCLUSION

One of the key arguments against lethal autonomous systems is the fact that they will lead to responsibility gaps in military operations. This is problematic on two counts. Firstly, the systems themselves will not hold any responsibility for their actions. Secondly, as their autonomy grows, so too does the distance between their actions and those of their commanders and developers. As such, it becomes ever more difficult to hold those individuals to account for the systems' actions. The responsibility gap then grows.

Robert Sparrow presented a classic view of this responsibility gap argument:

"…the more autonomous these systems become, the less it will be possible to properly hold those who designed them or ordered their use responsible for their actions. Yet the impossibility of punishing the machine means that we cannot hold the machine responsible. We can insist that the officer who orders their use be held responsible for their actions, but only at the cost of allowing that they should sometimes be held entirely responsible for actions over which they had no control. For the foreseeable future then, the deployment of weapon systems controlled by artificial intelligences in warfare is therefore unfair either to potential casualties in the theatre of war, or to the officer who will be held responsible for their use." (Sparrow, 2007:74)

These claims have been subject to much discussion since Sparrow's initial musings. Often, these debates fail to consider the legal application of the responsibility doctrine. Similar situations, in which individuals direct others to commit legal acts, have long been the subject of legal debate. As such, a number of doctrines have arisen, bridging the resulting responsibility gap. Further, legal philosophers and theorists have given much consideration to the moral relevance of these doctrines, illustrating their shortcomings and proposing reforms to bring them in line with our notion of justice. Further interaction with this legal discourse could progress the debate on lethal autonomous systems and the responsibility gap.

Consequently, military command responsibility does not apply to the relationship or interaction between man and machine/system/robot because there is no subordinate-superior relationship between man and machine/system/robot. Autonomous weapon systems cannot be considered subordinate and it is difficult for military commanders to have effective control and supervision over such systems. Indeed, it is not possible to punish an autonomous weapon system and it is difficult to prevent unforeseen actions. The individual behaviour of autonomous weapon systems disrupts the chain of command between the military commander and the system, which means that the military commander does not have effective control over his subordinates. Autonomous weapon systems can obtain, process, and decide on a large amount of information within nanoseconds. Therefore, the speed of the human controller prevents effective control over the system. As a result, the doctrine of command responsibility is not an appropriate mechanism for imposing criminal liability on commanders of such complex weapon systems.

REFERENCES

ANDERSON, K., and WAXMAN, M. (2013). *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can.* California: Hoover Institution.

ANDERSON, K. (2009). The Rise of International Criminal Law: Intended and Unintended Consequences. *European Journal of International Law*, 331-358.

ANDERSON, K., REISNER, D., and WAXMAN, M. (2014). Adapting the Law of Armed Conflict to Autonomous Weapon Systems. *International Law Studies*, 386-406.

ANTHONY, I., and HOLLAND, C. (2014). *The Governance of Autonomous Weapons.* Stockholm: Stockholm International Peace Research Institute.

ARKIN, R. (2009). *Governing Lethal Behavior in Autonomous Robots.* London: Routledge.

BADAR, M., and PORRO, S. (2015). Rethinking the Mental Elements in the Jurisprudence of the ICC. C. Stahn içinde, *The Law and Practice of the International Criminal Court* (s. 665). Oxford: Oxford University Press.

BASSIOUNI, M. C. (2013). *Introduction to International Criminal Law.* Leiden: Brill.

BO, M. (2021). Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute. *Journal of International Criminal Justice*, 289.

BOHLANDER, M. (2009). *Principles of German Criminal Law.* Oxford: Hart Publishing.

CASSESE, A. (2003). *International Criminal Law.* Oxford: Oxford University Press.

COCKBURN, A. (2016). *Kill Chain: The Rise of the High-Tech Assassins.* London: Picador.

D'AMATO, A. (1986). Superior Orders vs. Command Responsibility. *American Journal of International Law*, 607.

GARRAWAY, C. (1999). Superior Orders and the International Criminal Court: Justice Delivered or Justice Denied. *International Review of the Red Cross*, 787.

GÜNEYSU, G. (2022). *Otonom Silah Sistemleri: Bir Uluslararası Hukuk İncelemesi.* Eskişehir: Nisan Yayınları.

HAMMOND, D. (2014). Autonomous Weapons and the Problem of State Accountability. *Chicago Journal of International Law*, 652-687.

HENCKAERTS, J. M., and BECK, L. D. (2005). *Customary International Humanitarian Law.* Geneva: International Committee of the Red Cross.

HOROWITZ, M. (2016). Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems. *Temple International and Comparative Law Journal*, 85.

HOVEN, J., and SANTONI DE SIO, F. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 4.

HUMAN RIGHTS WATCH. (2014). *Shaking the Foundations: The Human Rights Implications of Killer Robots.* Geneva: HRW.

JAIN, N. (2013). Individual Responsibility for Mass Atrocity: In Search of a Concept of Perpetration. *American Journal of Comparative Law*, 831.

MARGULIES, P. (2017). Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts. J. D. Ohlin in, *Research Handbook on Remote Warfare* (s. 405). Cheltenham: Edward Elgar Publishing.

MARRA, W., and MCNEIL, S. (2013). Understanding The Loop: Regulating the Next Generation of War Machines. *Harvard Journal of Law and Public Policy*, 61.

MELONI, C. (2007). Command Responsibility: Mode of Liability for the Crimes of Subordinates or Separate Offence of the Superior? *Journal of International Criminal Justice*, 620.

OHLIN, J. D. (2016). The Combatant's Stance: Autonomous Weapons on the Battlefield. *International Law Studies*, 22.

PARKS, W. (1973). Command Responsibility for War Crimes. *Military Law Review*, 22.

PROSECUTOR V FURUNDŽIJA, IT-95-17/1-T (International Criminal Tribunal for the Former Yugoslavia 12 10, 1998).

PROSECUTOR V MUCIĆ , IT-96-21-T (International Criminal Tribunal for the Former Yugoslavia 11 16, 1998).

PROSECUTOR V. BLAŠKIĆ, IT-95-14-A (International Criminal Tribunal for the Former Yugoslavia 07 29, 2004).

PROSECUTOR V. HALILOVIC, IT-01-48-T (International Criminal Tribunal for the former Yugoslavia 11 16, 2005).

PROSECUTOR V. KORDIĆ, IT-95-14/2-A (International Criminal Tribunal For the Former Yugoslavia 12 17, 2004).

PROSECUTOR V. KRNOJELAC, IT-97-25-A (International Criminal Tribunal for the Former Yugoslavia 09 17, 2003).

PROSECUTOR V. TADIĆ, IT-94-1-T (International Criminal Tribunal for the Former Yugoslavia 05 07, 1997).

ROFF, H. (2013). Killing in War: Responsibility, Liability, and Lethal Autonomous Robots. F. Allhoff, N. Evans, and A. Henschke içinde, *Routledge Handbook of Ethics and War: Just War Theory in the 21st Century* (s. 354). London: Routledge.