The Dilemma of Deploying Lethal Autonomous Weapons: Balancing Moral Responsibility and Technological Advancement

Hoek, Sophia van den

The Dilemma of Deploying Lethal Autonomous Weapons: Balancing Moral Responsibility and Technological Advancement

Universiteit Leiden

Sophia H.E. van den Hoek
Student number: 1558803
Leiden University
Master Thesis Legitimacy and Justice
Political Theory: Legitimacy and Justice
Supervisor: Dr. J. Belic and Dr. T. Theuns
Word count: 10808

Abstract

The (hypothetical) deployment of Lethal Autonomous Weapons (LAWs) challenges the way in which we conceptualize moral responsibility. The emergence of LAWs have added an autonomously acting non-human entity to a moral responsibility framework which is inextricably linked to human nature and moral capacity, which LAWs neither have nor possess. This leaves open a responsibility gap in which it becomes unclear who exactly is responsible for the outcome of the decisions made by LAWs. Although several solutions have been proposed to solve the gap, such as the concept of meaningful control or role-specific responsibility, I find that they cannot sufficiently address the responsibility gap. The concept of meaningful human control is inadequate for the complex and chaotic environment of warfare, particularly when introducing powerful weapons that push the boundaries of human capability. While role-responsibility considers the collective nature of the military and the entire chain of command, it faces challenges in accounting for the problem of many hands and the emergent behavior of autonomous weapons that cannot be directly attributed to a specific part of the system or individual. Especially in a value-loaded and ethically charged environment such as war, where choices regarding life or death are a routine matter, there is no room for obscured responsibility. Without proper responsibility, one cannot justify the introduction of LAWs onto the battlefield.

Table of contents

Introduction

The development of lethal autonomous weapons (LAWs) has been one of the most significant technological developments in modern warfare. Driven by advancements in artificial intelligence (AI) and robotic technologies, LAWs can make decisions on their own, including the selection of targets and the use of lethal force. This can potentially shift the locus of control from human operators to machines. Some speak of ‘super soldiers’ that can replace human soldiers on the battlefield; others speak of ‘killer robots’ that ‘are a war crime waiting to happen’ (Vaccari 2015, 10-21; Crootof 2015; Parkin 2015; Sparrow 200).

Although this might sound as a mere dystopian nightmare to some, some countries are already developing autonomous weapon systems that can operate independently in a variety of combat situations (Galliott and Lotz 2015, 2-3). In October 2018, Zeng Yi, a senior executive at the Chinese defense firm Norinco, gave a speech in which he said that "on future battlegrounds, there will be no people fighting", and that the use of LAWs in warfare is "inevitable" (Allen 2019). In 2020, the US Defense Advanced Research Projects Agency (DARPA) announced the completion of project OFFensive Swarm-Enabled Tactics (OFFSET), the development of 250 aerial and land-based AI-drones, with the aim of finding “ways for small unit teams to benefit from advances in swarm autonomy and human-swarm teaming” (DARPA 2021).

In fact, some of these developments are already being implemented: according to a UN report, a drone airstrike in Libya in 2020 was conducted without ‘humans in the loop’ (United Nations Security Council 2021). The Super aEgis II, South-Korea’s best-selling automated turret that has been installed in many places around the world, was initially designed to track, identify, and decide whether to fire at intruders, without any intervention by humans (Parkin 2015). Interestingly enough, human control was added later on: “our costumers asked for a safeguard to be implemented, because they were concerned the gun might make a mistake, even though the need for human intervention is close to be eliminated”, senior research engineer Jungsuk Park explained.

The example of the Super aEgis II as mentioned above is exemplary for the concerns surrounding the (potential) development and deployment of LAWs, which essentially coalesces around the tension between LAWs, autonomy, and moral responsibility. LAWs rely solely on their programming and algorithms to make decisions, but are unable to consider ethical or moral

implications, a trait that is indispensable in a morally charged arena such as during warfare. The use of these kinds of weapons therefore leads to worries of an increasing number of moral wrongs, such as disproportional use of force or an increased risk of civilian casualties (Crootof 2016, 1347). Likewise, the deployment of LAWs makes questions of who bears responsibility for a certain outcome on the battlefield more complex. Many scholars have argued that the deployment of LAWS results in a so-called responsibility gap, in which it becomes unclear who is responsible for the outcome (see Matthias 2004, Sparrow 2007 and Smith 2022). If a LAW malfunctions and mistakenly identifies a group of civilians as enemy combatants, would it be the operator, the developer, or the system itself who's responsible, as the decision was made without direct human intervention?

This complexity raises questions about the applicability of the concepts of autonomy and responsibility to LAWs (Vaccari 2015, 9) and challenges the philosophical basis of ethical conduct (Ulgen 2017, 60). Autonomy and responsibility are almost always specified within an ethical framework that places human nature at the center, which is not applicable to LAWs (Guyer and Matthews 2000, 56). Being able to use LAWs in a justified matter therefore seems to require a reassessment and revision of our traditional understanding of responsibility. This research is therefore centered around the following main research question:

How does the (hypothetical) deployment of Lethal Autonomous Weapons (LAWs) challenge our traditional understanding of moral responsibility, and how can we address tensions that follow from this, such as the responsibility gap?

The research is built up as follows: in chapter one, I will provide a brief overview of relevant literature, specifically focused on the definition and taxonomy of LAWs and the different standpoints within the debate regarding possible deployment. In the second chapter, I will analyze the relationship between LAWs and the concept of autonomy, and show how this relationship leads to a so-called responsibility gap. In the third chapter, I will introduce some proposed solutions to the responsibility gap, and critically analyze whether they offer an adequate answer.

Chapter 1 | Literature review

In this chapter, I will give a brief overview of the technical definitions and different categories of LAWs and look into the debate on the permissibility of the deployment of LAWs by focusing on three views, respectively the deontologists' stance, the consequentialist stance, and the legalist stance.

1.1 Technical definitions and the taxonomy of LAWs

When it comes to LAWs, the central issue at hand can be put into a rather straightforward 'yes or no' question: is it morally acceptable to deploy LAWs or not? This question is preceded by another question: what does it mean to be an *autonomous* weapon? An "autonomous" agent or to "have autonomy" can take on many forms in a variety of contexts, especially when it comes to non-human robotic agency and autonomy. The existing body of literature lacks a consensus, with numerous definitions being put forward.

LAWs are most commonly defined in a technical sense, with a focus on different characteristics that enable LAWs to act autonomous, such as a good level of task-execution, goal-achievement, and information processing skills for technological devices (Tamburrini 2016, 122-127). Following Asaro's (2012) definition, any weapon that is capable of "releasing lethal force without the operation, decision or confirmation of a human supervisor can be deemed autonomous" (Asaro 2012, 687-94). The US Department of Defense proposed a definition which states that autonomous weapon systems must be capable, once activated, "to select and engage with targets without further intervention by a human operator" (Schmitt 2013, 1-4). Gubrud (2014) even broadens the notion of robotic autonomy a little more, as he states that "[a weapon] should be treated as autonomous as long as it is autonomously involved in one or multiple parts of the preparation process" (Gubrud 2014). On these accounts, LAWs are deemed autonomous agents in the sense that they are able to perceive their environment, make decisions based on their perception and take actions to achieve their aims, without (direct) human intervention. I will consider this as the foundations when discussing LAWs, since it appears to be a fundamental aspect present in all proposed definitions.

Following this, Beard (2014) formulated a taxonomy that categorizes LAWs according to different degrees of autonomy and severity of impact. He distinguishes between 'humans in-the-

loop' (HITL) weapons, 'humans on-the-loop' (HOTL) weapons and 'humans out-of-the-loop' (HOOTL) weapons (Beard 2014, 627). HITL-weapons require direct human control and decision-making. Here, the autonomous weapon is still able to operate independently, but a human operator is always present and has complete control over the weapon's actions. There is interactivity involved, meaning that the agent and its environment (can) act upon each other, such as with unmanned aerial vehicles or drones (Floridi and Sanders 2004, 353-7). HOTL-weapons require human control and oversight, but the human operator is not directly controlling the weapon. The weapon can act independently, without direct response to interaction, but the human operator *can* still intervene, such as with missile defense systems or so-called fire and forget loitering munitions, weapons that do not *require* external intervention after the launch (Beard 2014, 627-630; Floridi and Sanders 2004, 357; Amaroso and Tamburrini 2018, 3).

Both HITL- and HOTL-weapons are theoretically considered morally permissible to the extent that there is still human autonomy and control involved, leaving no room for eschewed responsibility. However, HOOTL-weapons, weapons that can make decisions about target selection, firing and other critical aspects of an operation without human control, are devoid of any human intervention (Beard 2014, 625-30). The human operator is involved in setting parameters or rules of engagement for the weapon, but does not have any control over its actions. Humans no longer exercise meaningful control once the weapon is *in use* (Heyns 2016, 14). An important concept here is adaptability, meaning that the LAWs's interactions (can) change the transition rules by which it changes its state. This property ensures that LAWs might be viewed as learning its own mode of operation in a way which depends critically on its experience (Floridi and Sanders 2004, 357). Examples of HOOTL-weapons include stationary robotic sentinels or automated turrets (Amoroso and Tamburrini 2018, 3).

Especially in the case of HOOTL-weapons, it can be difficult to attribute responsibility for the weapon's actions, because a human operator is not present in the decision-making process. This can be compounded by issues such as the complexity of the technology, the involvement of multiple parties in the development and deployment of the weapon, and the potential lack of transparency in decision-making processes (Matthias 2004, 175-82). Therefore, whilst acknowledging the wide variety of definitions, this research uses HOOTL-weapons as the basis for distinguishing between autonomous weapons and other weapon systems. These weapons represent the most significant departure from the traditional mode of human control and raise unique ethical concerns, particularly related to moral responsibility and autonomy, as these

concepts consistently emphasize the central role of *human* nature. Researching this specific type of autonomous weapon can help us understand the implications these emerging technologies have on longstanding moral concepts.

Although the definitions and taxonomy as stated above are useful for understanding the basic elements and categories of LAWs, they do not yet clearly specify the moral standing of LAWs. They only tell us what qualifies as autonomous *behavior*, not what qualifies as being an autonomous *agent*, and therefore avoid the question on the relationship of LAWs to the moral framework. This becomes more apparent when turning to the philosophical definition of autonomy. I will address this problem in the next chapter.

1.2 Overview of the debate: deploying LAWs

Debates on the moral permissibility of using LAWs on the battlefield are generally speaking threefold in nature and can be characterized as deontological, consequentialist and legalist (Sharkey 2018, 78). In the following section, I will give a brief description of the three views.

Deontological view

Following a so-called 'duty-based' ethics, where what one morally ought to do is led by certain moral norms, deontologists argue that LAWs, "to the extent that they take humans 'out-of-the-loop'", are unable to fulfill the obligations that are attached to agents operating in warfare or conflict (Amaroso and Tamburrini 2018, 4). Individuals, deontologists argue, have moral permissions and obligations relative to their situation, and therefore have to be able to recognize a moral context. LAWs, on the contrary, do not develop any moral character, and are therefore unable to judge a context according to their moral duties and moral reasoning. Here, "the nature of the mechanized decision maker" will ipso facto violate the agent-relative duty attached to the role as a commander or soldier in warfare, for they cannot access moral intuitions that guide one's conscience for or against certain actions (Kant 1981, 141-44). Amaroso and Tamburini (2018) provide a strong example to illustrate this objection:

> Consider a simplified military scenario where the decision-maker faces the following two options: Option A will yield substantial military advantage and is likely to provoke a significant amount of civilian casualties; Option B will cause less damage to the enemy and is unlikely to harm civilians. A commander who has decided that the proportionality

> principle is respected in option A might still refrain from following this course of action on account of the nefarious impact which, knowing well her or his soldiers and sharing with them daily warfare hardships, she or he deems that the killing of civilians will have on the troops' morale. This negative effect will in fact, according to her or his best judgment, in the long run jeopardize the possibility of defeating the enemy in the given operational theatre. It is presently difficult to imagine technologically foreseeable machines which are capable of gauging the psychological state of the troops (Amaroso and Tamburini 2018, 6).

Therefore, given their robotic nature, LAWs can never develop a moral character that is needed to make moral decisions, the making of which is imperative in warfare[1]. LAWs are unable to cultivate 'human virtues' that are necessary for moral decision-making and ethical behavior. Therefore, their deployment is deemed unethical on this account.

Consequentialist view

Where deontologists focus on the requirement of moral character in order to be able to make a right decision, traditional consequentialist theories hold that that what is right is determined solely by its consequences. Hence, the consequentialist view on the deployment of LAWs merely focuses on that: what are the *consequences* of using LAWs on the battlefield?[2]

Two views can be identified within the consequentialist discourse. The so-called 'narrow' consequentialists focus on the consequences of deploying LAWs in the battlefield and argue that using these kinds of weapons brings advantages, the most important one being that it reduces casualties because these weapons are more precise and are not led by worries about self-preservation and 'human passions' such as anger or fear (Schmitt 2013, 30). Wide consequentialists, on the other hand, consider the effect of LAWs on a broader scale, for instance for global security and peace (Amaroso and Tamburrini 2018, 9). Several objections to the use of LAWs due to the impact of the consequences have been raised, such as the (mass) proliferation of these weapons by for instance terrorist groups or oppressive regimes, the incentives to start wars, the

[1] Although it is out of the scope of this research, opponents of LAWs tend to forget the moral wrongs that can only be committed by human soldiers, such as rape or looting. See Smith (2018).

[2] Given that as of yet autonomous weapon systems (AWS) where human control is still (in)directly present, the consequentialist scope seems to be limited to mostly HITL- and HOTL-weapons.

acceleration in the pace of battle and the use of disproportionate force (Sassòli 2014, 308-317; Walzer 2000). Although LAWs might have some advantages on a smaller, *jus in bello* scale, most consequentialists argue that the negative consequences in the geopolitical realm outweigh the small-scale advantages, making a case for the prohibition of LAWs (Amaroso and Tamburrrini 2018, 8-10).

Legalist view

Along the same lines as the consequentialists but focusing on the legal perspective, legalists argue that the use of LAWs will undermine international humanitarian law (IHL) as well as the rules governing the use of force, for instance the *jus in bello* principles underlying the UN Charter's prohibition of the use of force as laid down in Article 2(4) (Asaro 2008, 50-58). Firstly, most international humanitarian treaties create obligations for human combatants to carry out attacks during warfare. Whether use of force is justified depends on the 'Who, Against Whom and the Why', the circumstances in which the force is unleashed (Sparrow 2007, 67-9). These combatants have to follow core principles, respectively (i) precaution – taking measures if the principle of proportionality or distinction is violated –; (ii) proportionality, whether the attack would be excessive in relation to military advantage anticipated, and (iii) distinction between military objects and civilian objects (Davison 2016, 7-11). The combatant is responsible for following these rules and will be held accountable if he or she violates them. It is unclear whether this same responsibility can be transferred to LAWs as of yet (Davison 2016, 7).[3]

Secondly, because an autonomous weapon can operate on their own and without supervision, it can be hard for commanders to judge whether the rules of proportionality, distinction and necessity are applicable, for instance if one does not know when the weapon is going to attack, or where (Davison 2016, 8).

1.3 Concluding remarks

In this chapter, I highlighted the context of the debate on the permissibility of deploying LAWs in warfare. Although the three different views emphasize different concerns, they all – with exception of the narrow consequentialists – seem to argue against the deployment of LAWs. All three strands seem concerned about the shift from human to machine autonomy and the

[3] It therefore might be justified on this account to use HITL-and HOTL-weapons.

effect this has on traditional normative concepts and legislation that are used to evaluate outcomes in warfare. However, the variety of definitions of autonomy and the new developments of robotic autonomy make the debate rather diffuse, for what do we mean when we talk about machine autonomy?

Although it is quite evident that LAWs have the capacity to *act* autonomous, it is the question of their autonomous nature that makes their deployment such a complex matter, especially for the concept of (moral) responsibility. An autonomous agent is an agent that acts on his own motives, that must have a specific relation to its own actions and consequences: a soldier who decides to take a risky action in combat, such as charging an enemy position, must be aware of the potential consequences of their actions and take responsibility for the outcome (Taddeo and Blanchard 2022, 77-80). Although most authors hold that it is hardly defendable to say that LAWs are able to take responsibility for their actions, fact of the matter is that LAWs are capable of acting autonomously (see Sparrow 2007). In the following section, I will analyze how LAWs relate to the concept of autonomy, and show how this leads to a so-called responsibility gap.

Chapter 2 | LAWs and the moral responsibility gap

In this section, I will examine the relationship between LAWs and the concept of autonomy. Following this, I will demonstrate how this relationship gives rise to a responsibility gap.

2.1 On the question of moral autonomy

As mentioned in the previous chapter, the most common definitions of LAWs focus solely on the different characteristics of LAWs that enable them to *act* autonomous. Although this is useful for understanding the technical abilities of LAWs, the definitions leave out any moral considerations or questions regarding the position of LAWs within the framework of autonomy. Although we speak of *autonomous* lethal weapons, most people would find it absurd to talk about autonomous weapons with a moral duty or an understanding of moral concepts. Yet, especially in the morally loaded context of warfare, autonomy cannot be detached from moral concepts such as responsibility, accountability, and dignity. Therefore, there seems to be a tension between LAWs being able to act autonomous and LAWs as autonomous agents.

This tension becomes more apparent when we look at autonomy from a moral perspective, where the concept is specified within an ethical framework placing human existence and moral capacity at the center (Guyer and Matthews 2000, 56). Indeed, looking at the concept of autonomy as understood in Immanuel Kant's theory of moral philosophy[4], the first continental philosopher in whose work the concept was prominent, autonomy is always intrinsically linked with moral agency and responsiveness to moral reasons (Taylor 2005, 614-16). On the traditional Kantian account, autonomy is defined as self-legislation: to govern oneself according to moral, undisputable duties and norms (Bacin and Sensen 2018, 3-4; Kant 1781, 20; 25). According to Kant, these norms and duties are known to human beings through their ability for rational thinking, practical reason, and judgment (Taylor 2005, 614-16l Kant 1781, 141-44; Werkmeister 1973, 26-29; Allison 2011, 75-92).[5]

Especially the concept of judgement, which Kant defines as "the intermediary between understanding and reason", is important here, because it gives people the capacity for moral judgement, to judge something according to the concepts of right or wrong (Kant 1981, 140-141; Kant 1785, 177-179). This ability follows from two notions of judgements, respectively analytical judgement, *a priori* judgements that make individuals able to assess concepts where the relationship between concepts is already contained within the subject, such as "all wounded soldiers are harmed" (Kant 1981, 141-44; Werkmeister 1973, 26-29; Allison 2011, 75-92; Paton 1979, 20-21), and synthetic judgements, a posteriori judgements in which the predicate is external to the subject, and which can only be assessed through empirical knowledge and sensory experience, such as "the enemy combatant is a child" or "the enemy combatant is unarmed (Kant 1981, 141-144; Werkmeister 1973, 26-29; Allison 2011, 75-92; Paton 1979, 20-21).

Now for Kant, moral judgements are a combination of *a priori* and synthetic judgements, because they require application of reason and recognition of moral duties that go beyond empirical observations (Paton 1979, 21). Through this moral or practical judgement, an individual can recognize the "supreme principle of morality", also known as Kant's famous categorical imperatives, "moral demands or requirements that are unconditional for all rational agents in all circumstances" (Kant 1785, 5).

[4] I find that Kant's foundational works on ethics provide a solid foundation for assessing ethical conduct of individuals. His emphasis on the link between autonomy and morality highlights the problem of ascribing autonomy to a weapon characterized as autonomous. Additionally, his emphasis on achieving a practical philosophy might be helpful for coming up with rules for practical implementation.
[5] For a more elaborate work on Kantian ethics, see Allison (2011).

Yet, to what extent can LAWs mimic this ability of synthetic a priori judgement and moral reasoning accordingly? In the following section, I will show that LAWs are – perhaps unsurprisingly - unable to attain to the moral definition of autonomy, regardless of their autonomous capabilities.

2.2 On the question of robotic autonomy

Given the human-centered nature of Kant's moral framework, it is quite easy to see how LAWs are unable to fulfill the requirements of being autonomous agents. Firstly, LAWs do not possess the capacity for rational thinking and practical reason. While they can use logical deduction and analytical reasoning to inform their decisions, this data is always evaluated in relation to the pre-existing rules and criteria. A LAW cannot ask itself whether or why to act in a certain way, whereas human beings can formulate reasons for acting in a certain way. By using reason, human beings can set up systems of rules governing moral conduct, which LAWs are unable to do (Ulgen 2017, 77).

Secondly, the lack of a priori synthetic judgement or moral judgement denotes a clear difference between the rationality of a calculating machine and "wisdom of human judgment", led by intuitive feelings, notions of instinct, contextuality and 'unwritten' rules (Weizenbaum 1976; Ulgen 2017, 78). Although one could argue that LAWs can be programmed for analytic and synthetic judgement–let's say, they are able to link empirical knowledge to learned analytic judgements, for instance if an enemy combatant is unarmed or whether someone is a civilian– they lack the deliberative and sentient aspects of human reasoning that are necessary for moral reasoning. LAWs do not have an innate sense of right and wrong in the same way as human being do.[6] They are able to perform repetitive routine and control tasks, and to make reasoned decisions based on logical deduction, data sets and learned algorithms, but are unable to understand the moral component of these decisions. This involves the mere characteristics of human thinking, reasoning inductively and meta-cognition, which LAWs simply do not possess (Sharkey 2016, 27).

[6] One could argue that human beings are prone to error, irrational thinking and led by personal desires. But, Kant argues, given that every human being at least has to *capacity* for rational thinking – considering the possibility of human error, irrationality, and wrongdoing – every human is able to *recognize* moral universal rules. See Allison (2011).

To exemplify, let us consider a LAW that stumbles upon unarmed enemy combatant. The LAW has been programmed to identify enemy combatants and has learned that it cannot fire arms if the enemy combatant is unarmed. Here, the situation and outcome is probably similar to that of a human soldier, apart from perhaps the occurrence of human emotions such as fear or angriness.

In a second example, the LAW stumbles upon an armed child who is clearly an enemy combatant. The armed enemy combatant attacks, and the LAW decides to fire arms. A human soldier will at least feel uneasy about their action, even though what he or she did was considered right in combat. Their feelings of uneasiness follow from an understanding that killing another human is wrong, especially if it is a child, regardless of whether the action according to war law might be right. It seems evident that a LAW is not able to do so: it cannot feel any remorse about the *moral* outcome of its action, for it is unable to normatively reflect. It requires additional thinking as to how 'child' relates to the concept of weapon, enemy combatant, and the information the LAW uses to decide to attack. The LAW is not able to understand how abstract, perhaps unwritten moral rules such as 'killing', 'right and 'wrong', 'child soldier', 'international humanitarian law' and 'human dignity' shape the outcome as well. It only possesses the capacity to evaluate, and therefore only has agential inclinations to the outcome.

The moral definition of autonomy, to govern oneself according to moral, undisputable norms (Bacin and Sensen 2018, 3-4; Kant 1781, 20; 25) is therefore not applicable to LAWs, at most *only* insofar as they are programmed by a human being to learn to make certain decisions. LAWs can *act* autonomously, but they are unable to *be* autonomous. This distinction has tremendous implications for the relationship between humans and LAWs as it can become unclear who should be held responsible for the autonomous decision of the LAW. In the following section, I will explain this responsibility gap in further detail.

2.3 The moral responsibility gap

As explained in the previous section, LAWs cannot be characterized as moral autonomous agents. They lack key qualities to engage in moral reflection and reasoning, such as moral ra-

tionality and the capacity for a priori synthetic judgements. But even though they are not autonomous agents, they can *act* autonomously. Now what does that mean for the concept of responsibility?

Traditionally, (moral) autonomy is considered essential for the attribution of moral responsibility. When one says an individual is morally responsible for his or her action, one is essentially saying she or he had the ability to reason, reflect and decide on his or her action (Kant 1981, 141-44). In other words: for a person to be held morally responsible, "she must have control over the behavior and the resulting consequences in a suitable sense" (Fischer and Ravizza 1998, 13).

Now as established earlier, LAWs can make decisions, but not on their own volition: their decisions are limited to fixed parameters, programming, and algorithms. LAWs therefore seem to only be *causally* responsible in the sense that decision A caused outcome X. Given that LAWs do not have the capacity for moral judgment or reflection, we cannot hold them *morally* responsibility for the effects of said decision. The question that then remains is who we *can* hold responsible for the effects of decisions made by LAWs.

A frequently suggested solution to this problem is the attribution of responsibility to those who worked on or with LAWs. But this list can become quite extensive and cluttered: from programmers and quality assurance to testing specialists and military personnel (Amoroso and Tamburrini 2018, 7). And these are only the direct links: one could argue we can also add manufacturers, sellers, project managers, contractors, or even (high-ranking) officers or politicians who decide to deploy LAWs to this the chain of responsibility.

Furthermore, transferring this responsibility from LAWS to individuals is not that straightforward. Because LAWs can act autonomously and operate on their own, it can be hard for commanders to judge whether the *jus in bello* rules of proportionality, distinction and necessity are applicable, for instance if one does not know when the weapon is going to attack, or where (Davison 2016, 8). Especially in the light of so-called "black box" algorithms, where the inner workings of the model are opaque to the user, this seems rather odd, for it is difficult to understand how a LAW arrives at its decision. As a consequence, the deployment of LAWs can lead to "responsibility gaps", where it is either unsure if someone can be held responsible, where

there is a risk of over-ascription of individual responsibility or where responsibility is distributed over 'many hands' – responsibility as the sum of its parts –, in which a group of people can be held responsible, but none of them can be blamed individually (Amaroso and Tamburrini 2018, 7; Liu 2016, 326; Sparrow 2007, 69-73).

Especially the risk of over-ascription of individual responsibility is a focus point in this research. As mentioned earlier, self-learning and black box algorithms give LAWs the capacity to develop and learn based on experience. Their behavior cannot be predicted in advance, for the design may change in response to external factors (Noorman 2014, 51-54). Consequentially, some behavior might not be traced back to individual programming, as external factors also influence behavioral outcome (Matthias 2004, 175-82; Liu 2016, 330). This has implications for assessing liability in legislation, as Amoroso and Tamburrini (2018) rightfully point out:

> Indeed, since LAWs might be capable of taking courses of action whose reason "may be opaque even to the system's designers", there will be cases where it is impossible to ascertain the existence of the mental element (intent, knowledge, or recklessness), which is required under ICL [international criminal law, sec.] to ascribe criminal responsibilities. As a consequence, no one could be held criminally liable, notwithstanding the conduct at stake clearly amounts to a war crime or a crime against humanity (Amoroso and Tamburrini 2018, 7).

If not resolved whilst LAWs are being deployed, the responsibility gap can lead to serious threats to the moral framework that is used to assess moral conduct on the battlefield as well as to people involved with LAWs (Matthias 2004,176; Liu 2016, 327). As Schieman (2022) states: "Far scarier than a world in which no one could be held accountable is one in which many of us could come to shoulder the blame for no less than the wrongful killing done by machines acting on our behalf" (Schieman 2022, 229). If countries want to use LAWs in an accountable and justified manner, the responsibility gap should be solved first. Therefore, in the following chapter, I will assess two different solutions that have been proposed to solve the responsibility gap, respectively the concept of meaningful human control and the concept of role-responsibility.

Chapter 3 Solutions to the responsibility gap

In this chapter, I will introduce two proposed solutions to the responsibility gap that occurs when LAWs are (hypothetically) deployed during war, respectively the concept of meaningful human control and role-responsibility. I will analyze to what extent they can be applied to LAWs, and whether they provide an adequate answer to the responsibility gap.

3.1 Meaningful Human Control

Following from a shared concern about a non-human entity making a decision that could violate 'just war' requirements without any possibility of holding the violator responsible, the concept of meaningful human control was introduced by several states to highlight the importance of human control in regard to the deployment of LAWs (Sharkey 2016, 23; Smith 2022, 2; Ekelhof 2019, 344). Rather than starting from the question of what it means when we talk about an autonomous weapon, the focus is reframed towards the question of what it means to have *control* over these weapons (Ekelhof 2019, 344). This meaningful human control (MHC) view focuses on the ability of human operators to exercise appropriate judgement when it comes to the deployment of LAWs, and highlights the need for human involvement in critical aspects of warfare, such as target selection, the use of force and the eventual decision to engage in lethal actions (Amoroso and Tamburrini 201, 13).

But what does *meaningful* human control entail? Indeed, as Sharkey (2016) points out, MHC is quite abstract, and does not specify what type of human control will be employed. To say that there is a human 'in the loop' for instance, does not clarify the exact degree of human involvement: "It could mean a human programming a weapon system for a mission or pressing a button to activate it, or it could mean exercising full human judgement about the legitimacy of a target before initiating attack" (Sharkey 2016, 26). Therefore, Sharkey (2016) developed a classification system of levels of human supervisory control:

1. Human engages with and selects target and initiates any attack;
2. Program suggests alternative targets and human chooses which to attack;
3. Program selects target and human must approve before attack;
4. Program selects target and human has restricted time to veto;
5. Program selects target and initiates attack without human involvement (Sharkey 2016, 27).

With this classification, Sharkey shifts the focus from weapon – 'how much can an autonomous weapon do by itself' – towards human – 'how much control can an individual exercise over such a weapon'. The question that is at stake here is then the following: At what level of human control is the attribution of responsibility guaranteed, avoiding a responsibility gap?

Given the many different categories of LAWs, the MHC must be determined per case. Especially the goal and whether the effects of deploying the weapon can have a potential destabilizing outcome are important indicators: a drone primarily used for scanning purposes may have a relatively low potential for causing disruption compared to a supervisory automated turret designed to deter intruders from accessing a specific area. As a result, the drone might require less stringent human control or oversight in comparison to the automated turret, even though they might be autonomous in the same sense. I will exemplify the practical application of this with the following examples.

Aerial and land-based AI drones

As seen in the introduction, the US Defense Advanced Research Projects Agency (DARPA) announced the development of autonomous collaborative air and ground vehicles that are used for so-called swarm-enabling tactics. Produced in swarms, these machines can support ground forces operating in complex environments where communications and mobility is constrained (DARPA 2020). The goal of these weapons is to *assist* and provide critical insights to military units operating in these diffuse areas. The weapons are not attack-oriented by themselves, but can help troops locate an enemy combatant or develop an offensive counter attack. Given their support-function, these weapons can justifiably be deployed at level five, for the ultimate responsibility rests with the troops who utilize the information provided by the LAWs. LAWs are considered as tools or instruments that aid in gathering data, analyzing situations, and suggesting courses of action. However, the final decision and responsibility for the actions taken still lies with the human operators

The Super Aegis II

Initially designed as a HOOTL-weapon for supervisory purposes, the 'Super Aegis II' turret can autonomously detect and identify targets from miles away, regardless of the time of day and the weather conditions (Parkin 2015). Given its potential destabilizing effects –for instance wrongfully targeting civilians or friendly fire– and purpose of preventing intruders to enter a

certain area, with, in the most extreme case, the use of lethal force, there is a big risk of causing a responsibility gap when weapons of this sort are deployed. Therefore, this weapon can only be deployed up until level four, where at least some form of human control is guaranteed.[7] Here, human operators have the ability to intervene, review, and potentially override the decisions made by autonomous systems. This helps mitigate the risk of a complete delegation of authority to machines and helps ensure that human judgment is still a significant factor in determining the use of (lethal) force. If these weapons are deployed at a level where human control is diminished or absent, questions arise about who is responsible for the actions and consequences of the LAW, which will inevitably lead to a responsibility gap.

Now hypothetically, this shift towards MHC makes one able to identify the risks of potentially creating a responsibility gap and install the appropriate level of human control to avoid it. Yet, it does raise some dependence and epistemic objections, which I will address in the following section.

Dependence and epistemic objections

Although the introduction of MHC does highlight the importance of a proper distribution of responsibility, it does not pay attention to the complexities and potential limitations in achieving effective oversight as technology advances. The concept takes little notice of the division of labour between commanders and LAWs (Ekelhof 2019, 347). At least from level two to level four, much of the work that is done in preparation of an attack is not part of the operator's task. The weapon finds, tracks, targets, engages, and assesses, only for the operator to validate the information and make the ultimate decision. The operator does not gather any intelligence by volition of his or her own senses, and is therefore very dependent on the abilities of the weapon (Ekelhof 2019, 347-48). This dependence could eventually lead to automatic reasoning by operators, fast responses that are always cued automatically, rather than engaging in active deliberation. This is especially problematic in the context of life-or-death decisions (Sharkey 2016, 30).[8]

[7] The need for meaningful, human control was already acknowledged and implemented in the Super Aegis II. See Parkin (2015).

[8] Automatic reasoning neglects ambiguity, is biased to believe and confirm, focuses on existing evidence rather than looking for absent evidence and infers and invents causes and intentions by attributing intentions to actors that fit the causal story. See Carroll and Rosson (1987).

Looking back at the definition of autonomy in the normative sense, could we still say that an operator uses its practical reason and practical judgement if the information he or she gets does not follow from one's own accord? To what extent is 'pressing a button' based on information completely provided by the weapon MHC? As Ekelhof (2016) rightfully asks: "Does the human operator have sufficient information to make informed and conscious decisions about the use of weapons? Is trusting the process or system the same as exercising meaningful control?" (p. 346).

To counter this objection, Sharkey (2016) introduced an additional framework which rests on the notion of human or "deliberative reasoning", where he argues that meaningful human control over LAWs only occurs from slower, deliberative processes that follow from active participation in the whole preparation process leading towards the launch of an attack (Sharkey 2016, 29-30; 34-8). Here, the focus is again reframed to what Sharkey calls "an equal partnership between human and machine", in which, if "configuration of a human and machine partnership is done correctly", a commander would theoretically be able to perform in the following ways:

1. Have full contextual and situational awareness of the target area at the time of initiating a specific attack;
2. Be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack, such as changes in the legitimacy of the targets;
3. Have active cognitive participation in the attack;
4. Have sufficient time for deliberation on the nature of targets, their significance in terms of the necessity and appropriateness of an attack and the likely incidental and possible accidental effects of the attack;
5. Have a means for the rapid suspension or abortion of the attack (Sharkey 2016, 26-8).

Yet, questions arise as to how realistic this framework is, especially in regard to the reasons for deploying LAWs. LAWs are being developed to advance military technology *beyond* human capacity (Galliot and Lotz 2015, 1-3). These advancements can blur the lines of control and create difficulties in maintaining meaningful human oversight. As weapons become more sophisticated and capable, they may enable faster decision-making, precision targeting, and heightened situational awareness, potentially exceeding the capabilities of individual human operators. The "faster-than-human reaction times' of LAWs for instance would make human

control predominantly speculative (Crootof 2016, 1380). These time frame operational constraints do not allow for real-time judgements by human operators (Amoroso and Tamburrini 2018, 14). Furthermore, some of the unpredictable, self-learning AI-driven technology makes it difficult for individual commanders to make reliable, and effective judgements to prevent violations or just war non-compliance within a short period of time (Smith 2022, 3). Even if the algorithm was explainable, the process would go by too quickly for the commander to remain "in the loop".

An example might be at place here. Consider the Super Aegis II again. This turret has the capacity to maintain exceptional visibility even in challenging weather conditions like heavy fog or blizzards. During a heavy fog, the Super Aegis enhanced sight capabilities allow it to still identify targets with high accuracy, track their movements, and engage with them. However, a human operator relying solely on their natural vision would face significant difficulties in identifying targets, estimating distances, and ensuring precise engagements in such adverse weather conditions. Taking in consideration the second framework developed by Sharkey, it might be hard, perhaps impossible, for a commander to reach levels one up to level four, for he or she simply does not have the ability to do so.

The 'larger than life' capacities of LAWs that go beyond human sensory experience therefore make it likely to reduce the situational awareness of the operator to the extent that the operator is fully dependent on LAWs, and therefore stripped of their duty to assess risks and to set out orders accordingly (Tamburrini 2016, 131-34). This begs the question on how meaningful human control can actually be meaningful if human capacities are 'too slow' to control a weapon that is designed to advance human limits. Operators are unable to stop the decision-making process of LAWs without paying in terms of (combat) effectiveness (Smith 2018, 282). If the human capacities that are needed for meaningful, human control 'hinder' the technological advancements that are being developed to go beyond human limits, what is the use of deploying LAWs in the first place?

One could object to this simply by means that it is unreasonable and unfair to think individuals can make reliable, effective judgements in such a short time frame in the context of warfare. Warfare is by nature chaotic, unpredictable, and a complex area to operate in. Even without LAWs, individual human soldiers are often unable to make reliable, effective judgements in these settings in a short period of time. The objections regarding human limits and dependency

are intrinsic to human nature, and not necessarily to the relationship between LAWs and humans[9]. Therefore, one can perhaps not find a solution for the responsibility gap by shifting focus from weapon to human or from human to weapon, but by shifting from a narrow, command responsibility to a wider shared responsibility. In the following section, I will analyse whether the notion of role responsibility can address the responsibility gap.

3.2. Role Responsibility

As seen in the previous section, the notion of MHC seems too narrow to pose a solution for the responsibility gap. It focuses solely on the relationship between human commander(s) and LAWs, which makes responsibility an individualist notion, belonging solely to individual operators and their ability to exert MHC. However, given the epistemic and dependency objections raised in the previous sections, it is not that obvious that one person can be assigned the relevant moral status. It even seems rather unfair to grant individuals with such responsibility, given the complex and chaotic nature of war.

In order to propose a solution to the responsibility gap, we need to understand responsibility from precisely that complex and chaotic context of warfare and its leading institution, the military. The action of warfare and the institution that deals with it, the army, are not individual, but *collective* by nature (Smith 2022, 3-4; Schieman 2022, 234). The army is an institution with a collective identity, with collective objectives and a collective reflection on the ends it serves (Lucas 2015, 94). It consists of a structured, hierarchical framework that is composed of many parts to guide its functioning. Soldiers do not kill in their personal capacity, but as representatives of the army, a community that shapes the use of force. Substantial accountability mechanisms have been installed to mitigate this collective, military power and to avoid moral wrongs (Smith 2022, 4; Schieman 2022, 233-37). These 'checks and balances' come in the form of 'command responsibility', where the commanding officer(s) bear(s) the responsibility of maintaining discipline and regulating the conduct of the forces under their command. Any violations of ethics or misconduct by the troops reflect on the commanding officer, who is legally held responsible for such breaches (Schieman 2022, 236-237; Lucas 2015, 49-50; Thompson 1980, 906). Here, moral responsibility follows from role-specific duties to which commanders, in

[9] Although it is outside the scope of this research, some argue that the narrow responsibility on the command responsibility account would also cause a responsibility gap between human soldier and commander. See Schieman (2022) and Smith (2018).

their capacity as commander, ought to attend (Schieman 2022, 234; Eschleman 2016; Lucas 2015, 49).

Given that LAWs are also deployed in name of the collective, why not treat them as any other soldier –albeit in a different capacity– that becomes part of the military? We should move away from an interpersonal understanding of responsibility between LAWs and their operators through the mechanism of MHC, and focus on a collective and distributive conception of responsibility in the context of warfare. However, I argue that this also entails a broadening of command responsibility. Command responsibility is still limited to responsibility for (ordering) the deployment of LAWs. Yet, LAWs are embedded in a network of relationship formed prior to the military context. Especially the phases leading up to deployment can attribute significantly to the outcome of the use of LAWs, and these actors should be taken in consideration as well.

Therefore, rather than pinpointing people up in the chain of commands, we need an account that takes in consideration the entire chain that deals with LAWs, from development to deployment (Smith 2022, 4). This notion of role responsibility focuses on responsibility attached to an individual by virtue of his or her role and the function he or she is expected to fulfill (Liu 2016, 336-39). On this account, an individual is responsible for the way in which she carries out the obligations that are attached to her role within the collective. In the case of LAWs, this means that a programmer fulfills her role responsibly by exercising due diligence to ensure that algorithms operate in accordance with the prescribed requirements, a manufacturer fulfills her responsibility when LAWs are constructed in accordance with the specified design specifications, and a commander who takes appropriate measures in deploying the system fulfills the obligations associated with her position (Liu 2016, 337).

To exemplify the practical application of role-responsibility, let's consider the Super Aegis II once more. This automatic turret is programmed to track, identify, and decide whether to fire at intruders, initially without intervention by humans. However, due to a malfunction in its target recognition algorithms, accidently missed by the developer because she spilled her coffee during the trial run, the turret mistakenly identifies a group of innocent civilians as hostile targets and launches an attack, causing unintended casualties. Operator O, who is responsible for overseeing the actions of said LAW, notices the mistake of the LAW too late, and is unable to

counter the attack. Operator O has been dealing with insomnia, of which the army's psychologist is aware, but who did not yet mention it to operator O's overseeing commander C due to a sudden family emergency. However, overseeing commander C did have her doubts about the well-being of operator O, and asked military planner M to take operator O off the schedule. Military planner M writes it down but forgets it later on. A couple of days later, this tragic event occurs.

Who is responsible for the unintended causalities? In this scenario, responsibility for the mistake can be distributed among several roles. Firstly, the developers are, by nature of their role, responsible for ensuring accuracy and reliability in the design. Errors in programming or inadequate testing led to a significant portion of the responsibility for the mistake made. The operator also carries responsibility, by nature of his duty to monitor the system's performance, assess risks and overall supervision. The army's psychologist can also be held responsible for the outcome, for she did not fulfil the duty of psychological assessing soldiers and intervening when necessary. Interestingly enough, it seems as if the commander –whom would have carried most to full responsibility on the command responsibility framework– bears little responsibility due to lack of knowledge. The commander did fulfil her duty to judge the competence of his or her soldiers (Schieman 2022, 237).

Despite being a 'worst case scenario', this example shows that including a broad notion of responsibility distributed according to roles and duties can help avoid over-ascription of individual responsibility. If we only distribute responsibility according to the chain of command, all responsibility would be placed upon the operator or the overseeing commander. Yet, a big part of the outcome follows from malfunctioning, which can only be attributed to role-specific duties that do not follow from the chain of command, but rather from the chain of development and deployment.

However, the framework does give rise to certain objections, particularly concerning the issue known as the 'problem of many hands'. In the following section, I will briefly discuss the problem and apply it to the case of LAWs.

The problem of many hands

First coined by philosopher Dennis Thompson, the problem of many hands occurs when multiple individuals contribute to an action or decision that turned out to be morally unjustifiable,

but which cannot be traced back to a single individual (Thompson 1980, 905-7). Using the concept to explain how responsibility of public officials work, Thompson argues that political outcomes usually follow from actions of many different people whose individual contributions may not be identifiable or distinguished from other people's contributions. No individuals can therefore be morally blamed for these outcome, only the collective (Thompson 1980, 905-7).

Although originally designed to address the responsibility of public officials, I find the problem of many hands very applicable to when discussing LAWs and responsibility. Firstly, as already pointed out, in complex engineering projects, it is quite difficult to trace responsibility back to individuals. LAWs often consist of multiple, interconnected components that interact in complex ways, making it difficult to isolate responsibility for specific outcomes. Furthermore, complex projects involve multiple actors, in which each person's contribution may seem relatively small or insignificant compared to the overall outcome. This diffusion of responsibility can make it difficult to determine who should be held accountable for the consequences. As Nussenbaum puts it: "It's the collective effort of the team that gives rise to a product. When decisions translating into actions and consequences, the connection between the outcome and who are all responsible is difficult to make" (Nussenbaum 1994, 75).

Secondly, even with a diffuse division of responsibility among multiple actors, role-responsibility is unable to address concerns about the emergent behavior of LAWs that is not directly traceable to a particular part of the system, or which is caused by a combination of retraceable and external, untraceable factors. Looking back at the above-mentioned example, what if the malfunctioning of the LAW was caused by an unknown desert bug that had a taste for metal and munched away at the cables that were necessary for proper target selection as well as suspension of the attack by an operator? Although unique, these types of cases do raise questions whether role-responsibility really solves the potential problem of over-ascription. Additionally, if an outcome follows from a combination of actions that can individually be traced back but would not have caused the outcome on their own, would it still be fair to hold individuals responsible for those individual actions responsible?

3.3 Concluding remarks

In this section, I evaluated two proposed solutions, respectively meaningful human control, and role responsibility, to the responsibility gap that occurs when LAWs are being deployed. Although the concept of MHC can in theory solve the responsibility, it is hardly applicable in practice. Firstly, the framework raises questions about the dependency: it is hardly human control when an operator is almost fully dependent on the information a LAW provides in order to judge a situation accordingly. Secondly, LAWs are military advancements developed to go beyond human limits. In that regard, the idea of MHC seems undoable and rather odd, as it goes against the idea of an advancement that is not limited by human impairments. If the framework is applied correctly, the advantages that come from using LAWs, such as speed and response time, are being hindered. Why then use LAWs in the first place?

Moreover, the framework does not take in consideration the distinctiveness of the context of war. Warfare is unpredictable, chaotic, and diffuse. Even when LAWs are not present, individual human soldiers frequently encounter challenges in regard to reliable and effective judgments within a limited timeframe. Individual responsibility is hardly at play in a context such as war. The army is a collective, and responsibility should be considered in a matter that suits this context. Therefore, the notion of role-responsibility was introduced. Here, responsibility is disaggregated among different actors that are part of the chain of development up until the chain of deployment. Yet, from this framework arose the so-called problem of many hands, where the outcome follows from the sum of the actions of individuals, which cannot be traced back to a single individual. Secondly, the framework was unable to address previously mentioned concerns about behavior or outcomes of the deployment of LAWs that follow from interconnected internal and external components.

Conclusion

In this research, I analyzed the influence of the (hypothetical) deployment of Lethal Autonomous Weapons (LAWs) on our traditional understanding of moral responsibility on the basis of the following research question:

How does the (hypothetical) deployment of Lethal Autonomous Weapons (LAWs) challenge our traditional understanding of moral responsibility, and how can we address tensions that follow from this, such as the responsibility gap?

I started this research with a brief overview of relevant literature, from which followed that the main concern of the hypothetical deployment of LAWs centers around the changing relationship between human and machine autonomy, and the effect a "mechanized, non-human decision maker" has on our traditional understanding of normative concepts such as autonomy and responsibility. Although most authors hold that it is hardly defendable to say that LAWs are able to take responsibility for their actions, fact of the matter is that LAWs are capable of acting autonomously, which, on the traditional account, would be a relevant factor in attributing responsibility to an individual, as it reflects their capacity for independent decision-making and self-governance. This causes a so-called responsibility gap, in which it is unsure who can be held responsible for the outcomes of decisions made by LAWs, with a risk of under- or over ascription of responsibility to individuals.

Several scholars have proposed solutions to counter this gap. I highlighted two, respectively the concept of meaningful, human control and role-responsibility, and found that both cannot adequately address the responsibility gap: the notion of meaningful human control is not suited for the chaotic, complex, and collective context of warfare and the battlefield, and certainly not when introducing 'larger than life' weapons that are used to advance human limits. Although role-responsibility does take into consideration the collective nature of the army as well as the entire chain of command – from development to deployment – it encounters the problem of many hands and cannot accord for emergent behavior of LAWs that is not directly traceable to a particular part of the system or individual. It seems that the concept of responsibility, especially in a highly-morally loaded context of war, is not (yet) suited to address the emergence of LAWs on the battlefield. Therefore, I would argue that the deployment of LAWs as of this point is not justified.

In conclusion, we can answer the research question in the following way:

The (hypothetical) deployment of Lethal Autonomous Weapons (LAWs) challenges the way in which we conceptualize moral responsibility. The emergence of LAWs have added an autonomously acting non-human entity to a moral responsibility framework which is inextricably linked to human nature and moral capacity, which LAWs neither have nor possess. This leaves open a responsibility gap in which it becomes unclear who exactly is responsible for the outcome of the decisions made by LAWs. Although several solutions have been proposed to solve the gap, such as the concept of meaningful control or role-specific responsibility, I find that they cannot sufficiently address the responsibility gap. The concept of meaningful human control is inadequate for the complex and chaotic environment of warfare, particularly when introducing powerful weapons that push the boundaries of human capability. While role-responsibility considers the collective nature of the military and the entire chain of command, it faces challenges in accounting for the problem of many hands and the emergent behavior of autonomous weapons that cannot be directly attributed to a specific part of the system or individual. Especially in a value-loaded and ethically charged environment such as war, where choices regarding life or death are a routine matter, there is no room for obscured responsibility. Without proper responsibility, one cannot justify the introduction of LAWs onto the battlefield.

Bibliography

Allen, G.C. (2019). “Understanding China's AI Strategy Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security”. *Centre for a New American Security*. Retrieved from https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy

Asaro, P. (2012). “On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making”. *International Review of the Red Cross*, pp. 1-45.

Allison, H.E. (2011). *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press

Amoroso, D. and Tamburrini, G. (2018). “The Ethical and Legal Case Against Autonomy in Weapons System”. *Global Jurist*, Vol. 18 No. 1, pp. 1-20.

Bacin, S., & Sensen, O. (2018). *The Emergence of Autonomy in Kant's Moral Philosophy*. Cambridge: Cambridge University Press

Beard, J.M. (2014). "Autonomous Weapons and Human Responsibilities" *College of Law*, Vol 196, pp. 624-659.

Carroll, J.M. and Rosson, M.B. (1987). “The Paradox of the Active User.” In Carrol, J.M. (1987). *Interfacing Thought: Cognitive Aspects of Human–Computer Interaction*. Cambridge: MIT Press.

Crootof, R. (2015). “The Killer Robots Are Here: Legal and Policy Implications”. *Cardozo Law Review*, Vol. 33, pp. 1837, 1847.

Crootof, R. (2016). “War Torts: Accountability for Autonomous Weapons”. *University of Pennsylvania Law Review*, Vol. 164, No. 6, pp. 1347-1402.

Davison, N. (2016). “A legal perspective: Autonomous weapon systems under international humanitarian law”. *International Committee of the Red Cross*, No. 30, pp. 5-30.

DARPA (2021). “OFFSET Swarms Take Flight in Final Field Experiment: Program showcased four years of advancements in swarm capabilities”. *Defense Advanced Research Projects Agency*.

Ekelhof, M. (2019). “Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation”. *Global Policy*, Vol. 10, No. 3, pp. 343-48.

Fischer, J. M., and M. Ravizza. (1998). *Responsibility and Control*. Cambridge: Cambridge University Press.

Floridi, L. and Sanders, J.W. (2004). “On the Morality of Artificial Agents”. *Minds and Machines*, Vol. 14, pp. 349-379.

Gubrud, M. (2014). “Stopping killer robots”. *Bulletin of the Atomic Scientists*, Vol. 70, No. 1, pp. 32-41.

Guyer, P. and Matthews, E. (2000). *Kant’s Critique of the Power of Judgement*. Cambridge: Cambridge University Press.

Heyns, C. (2016). “Human Rights and the use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement”. *Human Rights Quarterly*, Vol. 38, No. 2, pp. 350-378.

Lucas, G. R. (2015). *Military Ethics. What Everyone Needs to Know*. Oxford: Oxford University Press.

Kant, I. (1785). *The Groundwork of the Metaphysics of Morals*. Cambridge: Cambridge University Press.

Kant, I. (1781). *Critique of Pure Reason*. Cambridge: Cambridge University Press.

Liu, H-Y. (2016). “Refining responsibility: differentiating two types of responsibility issues raised by autonomous weapons systems”. In Bhuta, N. et al. (2016). *Autonomous Weapon Systems: Law, Ethics, Policy*. Cambridge: Cambridge University Press.

Noorman, M. and Johnson, D. (2014). “Negotiating autonomy and responsibility in military robots”. Ethics and Information Technology, Vol. 16, No. 1, pp. 51-62.

Nussenbaum, H. (1994). “Computing and accountability”. *Communications of the ACM*, Vol. 37, No.1, pp. 72-80.

Matthias, A. (2004). “The responsibility gap: ascribing responsibility for the actions of learning automata”. *Ethics and Information Technology*, Vol. 6, pp. 175-183.

Majumdar Roy Choudhury, L. *et* al. (2021). “Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011)”. *United Nations Security Council*, pp. 1-548.

Paton, H.J. (1967). *The Categorical Imperative: a study in Kant’s moral philosophy*. London: Hutchinson.

Parkin, S. (2015). “Killer robots: The soldiers that never sleep”. BBC.
Retrieved from: https://www.bbc.com/future/article/20150715-killer-robots-the-soldiers-that-never-sleep

Santoni De Sio, F., & Van den Hoven, M. (2018). Meaningful Human Control Over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, Vol. 5, 15–45.

Sassòli, M. (2014). “Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified”. *International Law Studies*, Vol. 90, pp. 308-320.

Schiemann, K. (2022). “The Soldier’s Share: Considering Narrow Responsibility for Lethal Autonomous Weapons”. *Journal of Military Ethics*, Vol. 21, No. 3-4, pp. 228-245.

Schmitt, M. N. (2013). “Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics” *Harvard National Security Journal Features*, Vol. 1, pp.

Sharkey, N. (2016). “Staying in the loop: human supervisory control of weapons”. In Bhuta, E. (2016). *Autonomous Weapons Systems: Law, Ethics, Policy*. Cambridge: Cambridge University Press, pp. 23-38.

Sparrow, R. (2007). “Killer Robots”. *Journal of Applied Philosophy*, Vol. 24, No. 1, pp. 62–77.

Smith, P.T. (2018). “Just Research into Killer Robots’. *Ethics and Information Technology*, Vol. 21, pp. 281–29

Smith, P.T. (2022). “Resolving responsibility gaps for lethal autonomous weapon systems”. *Frontiers in Big Data*, Vol. 5, pp. 1-4.

Tamburrini, G. (2016). “On banning autonomous weapon systems: from deontological to wide consequentialist reasons”. In Bhuta, E. (2016). *Autonomous Weapons Systems: Law, Ethics, Policy*. Cambridge: Cambridge University Press, pp. 124–127.

Taddeo, M. & Blanchard, A. (2022). “Accepting Moral Responsibility for the Actions of Autonomous Weapons Systems—a Moral Gambit”. *Philosophy & Technology*, Vol. 35, No. 78, pp. 1-24.

Taylor, R. (2005). “Kantian Personal Autonomy,” *Political Theory*, Vol. 33, No. 5, pp. 602–628.

Thompson, D. F. (2014). *Responsibility for Failures of Government. American Review of Public Administration*, 44(3), 259–273

Ulgen, O. (2017). “Kantian Ethics in the Age of Artificial Intelligence and Robotics”. *Questions of International Law*, Vol. 2, pp. 59-83.

Vaccari, A. (2015) “Abjection Humanity: Dehumanising and Post-humanising the Military”. In Galliot, J. & Lotz, M. (2015). *Super Soldiers: The Ethical, Legal and Social Implications*. London: Routledge.

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgement to Calculation*. Freeman & Company.

Werkmeister, W.H. (1973). “Analytic and Synthetic Concepts According to Kant's LOGIK”. *The Southwestern Journal of Philosophy*, Vol. 4, No. 2, pp. 25-28.

Walzer, M. (2000). Just and Unjust Wars. New York: Basic Books.