Creating a future with Autonomous Weapon Systems

Kim-Morgaine Lohse

IT and Ethics

Dr. Pak-Hang Wong

The University of Hamburg

23.07.2018

Abstract: Is it legal and responsible to use Autonomous Weapon Systems (AWS) in war acts and who can take responsibility for the killings from machines? A majority of scientists argue against the use of AWS and are even engaged in anti-development movements. Others argue for the use but only based on specific terms of usability. In this paper, I will present and discuss the two positions on who can hold the responsibility of AWS: Human versus non-human. Thereby, I argue that when talking about killer machines, people who are involved in the development should take on the responsibility and voucher for them in advance. However, I will also debate that the crucial question in AWS might not be the responsibility, but rather a matter of how to design and implement the system morally and ethically. I will argue for the use of AWS, but only if the necessary changes in engineering are being made, resulting in becoming in line with the Just War Theory. AWS could revolutionize war if we start integrating moral development.

Introduction

Over the past years, research and development in Autonomous Weapon Systems (AWS) have increased exponentially and thereby changed warfare dramatically. AWS will be able to target and kill people without human help. China, United States, South Korea, Russia and the United Kingdom are the most prominent countries in developing these systems. (Un.org, 2018). This development has pushed a lot of debates about the ethical, moral and security challenges that follow with this development. Who will be in charge if something goes wrong? Who will be responsible? Machines or humans? Allowing machines to make decisions over life and death results in new and challenging moral questions. The use of AWS systems could result in a responsibility gap. Meaning that we just do not know, who to hold responsible.

This paper proceeds as following. I will start by explaining what is meant by the word "autonomous" within weapon systems and give examples in the current and future use of them. I then introduce the concept of "responsibility" in context with the Just War Theory and what ethical issues and dilemma we could face with AWS. Afterwards, I will present and discuss two positions on who could hold the responsibility for the behavior of AWS systems. Prominent scientists and philosophers argue for either human- versus non-human responsibility. After that, I present my own position and elucidate why I believe that the responsibility is not the most crucial question. Instead, I think that AWS should satisfy the requirement of the Just War Theory resulting in fairness and redistribution of risk.

How do we define AWS?

The idea of AWS often reminds us of action- or sci-fiction movies of the future. But the fact is, that this is no longer a future scenario. There are already AWS that are used in present war acts. To better understand and critically reflect on the use of AWS, we must be clear on what exactly is meant by these systems. I will dedicate the next paragraph to define the word "autonomous" in the context of weapon systems and give examples of recently developed AWS. The most common definition for AWS is currently from the U.S. Department of Defense's (DoD) and was defined in 2013. This definition states that "autonomous weapon systems once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon systems but can select and engage targets without further human input after activation" (Crootoft, 2015). However, I find this definition outdated. I argue for that because currently there is a lot of developing in both fully autonomous weapon systems and human-supervised autonomous weapon systems. I think these two cases are separate problems which overlap at some points. I personally think that it is "easier" to blame the responsibility at someone when discussing human-supervised AWS. Therefore, I believe these two cases should be discussed separately. In this paper, I will be concentrating on fully AWS. The PAX cofounder defines autonomous systems as "weapons which, once activated, using sensors and/or artificial intelligence, will be able to operate without meaningful human control over the critical function" (Wareham, 2018). I believe this sets a clear idea about autonomy in weapon system and I will from now on refer to this definition when I use the term 'autonomy'. Since we now have an idea about what the word stands for, I will give some examples of the use and power of this weapon systems.

Existing AWS include unmanned combat aircraft, ground vehicles and marine vehicles, border control, counter-terrorism and law enforcement, and anti-animal killer machines. The research is widespread and sometimes in areas that we cannot even imagine. Unmanned combat aircraft are armed drones that are equipped with missiles. These weapons are already used in warfare and researchers are trying to make the existing drones even faster with more advanced sensor systems and internal weapon bays. The nEUROn, for instance, was developed by a cooperation between the Franco-British government-supported Unmanned Combat Air System (UCAS). It is currently ranked as the most advanced AWS under development. It is a 10 m long aircraft with a 12.5 m wingspan and can fly for over 3

hours completely autonomous. It has target adjustment, which allows it to choose the target once it has reached the warfare area. Thereby it has the capabilities of a fully automated attack (Slijper, 2017). The new robotic weapons must handle a lot of different tasks in different environments. Consequently, there is an increasing development in AWS for air, land, and water. Unmanned ground vehicles, for example, can operate entirely without a human controller and they can be used to provide physical security on patrol, carry weapon systems and serve as a forward observation platform. An example is ROBATTLE. It is a 6 wheeled UGV which can be equipped with machine guns and can give protection, decoy, and attack. It is made by Israel forces, currently using it at their borders (Slijper, 2017). However, AWS are not only produced for traditional warfare, but there are also several inventions in security robots and anti-animal warfare. The ANBOT is a security robot for deployment for airports and can act around-the-clock. It can scan people with sound and light to detect a potential terrorist. This robot is only 1.5 m tall and weighs around 75 kg (Slijper, 2017). The computing warfare has even spread towards animals. It is a different but very related development where robots are developed to eliminate animals which are considered having diseases that can be spread. An example of an anti-animal robot is the Mosquito killer robot. It is a Chinese killer robot which recognizes mosquitos and immediately kills them with laser light. It is estimated to be able to shoot 30 to 40 mosquitoes in one second (Slijper, 2017). I think that this idea is a great way to combat diseases such as malaria. However, it also makes visibly how quickly the border between warfare and civil use can be fluent and thereby clearly crossed. It shows how important it is to act immediately, during the development of new AWS. We need new policies before the line between warfare and civil will be even more challenging to draw.

Responsibility and Just War Theory

Who will be held responsible if war crimes are being made by AWS? This is a question that a lot of researchers, philosophers, and scientist have been examined in recent years. I will be discussing two of the main outstanding sides of who should be held responsible if a robot is going to make a war crime and conclude with my own ideas to the two parties. Before I start discussing these sides, we must understand, what responsibility in the context of warfare even means. In this paragraph, I want to present the idea of the Just War Theory. The purpose of the Just War Theory is to ensure that war is morally justified. The theory states several criteria that are split into two groups. The “right to go into war” (jus ad bellum) and the “right to conduct war” (jus in bello). These criteria all have to be med to

have a war that is justified (Crawford, 2003). For a war to be just, it is evident, that someone must take responsibility for the deaths of the enemies who are being killed. Furthermore, in most warfare's, some civilians are being killed, and someone must take responsibility for these deaths. Therefore, responsibility is an essential part of the Just War Theory.

We have talked about the definition of autonomy within autonomous systems. However, Noorman & Johnson (2014) state, that if we look at the autonomy in the context of moral philosophy, autonomy implies acting on one's own, controlling one's self and being responsible for one's actions. I think this description is very interesting. It clearly shows the link between autonomy and responsibility. So, if robots are increasingly becoming autonomous, they will have more control, and therefore they should be able to increasingly take responsibility for their actions. Hence, the human beings should not be held responsible. But do the people who lost their closest friends and families feel justified with punishing the robots? Are we sure that it is not the fault of the programmers? Or the government, who allowed the use of the machines or maybe the military headquarters? It definitely still does not exist a consistent view of this responsibility distribution. In the following, I will present the arguments for human versus non-human responsibility.

The responsibility of Autonomous Weapon Systems

Robots will be responsible.

Let us start to discuss the idea of giving the responsibility to the robots. As mentioned earlier, if the robots act autonomously, they must also be able to take the responsibility. The author Asaro, (2017) for example, suggests that the concept of autonomies will end in fully autonomies, where robots will be able to think, sense and take their own decisions. Additionally, they will have human-like personalities. He believes that they will be able to imitate moral capacities and even be able to formulate own ethical rules, choices and reason those. Thereby, they will be able to acquire moral autonomy (Asaro, 2017). The Naval Research made an analysis of how autonomous robots could look like in the future. They postulate, that when robots become more autonomous, we could start to treat robots as "culpable legal agents." (Lin et. al, 2018). They believe, that we thereby give them legal responsibility, but maybe not yet moral responsibility. This idea of legal responsibility has been suggested by some authors. They argue that robots are 'artificial agents' and therefore having the ability of acting. I personally think it is challenging even to consider this idea or even take it seriously at this time. I think this statement could be interpreted as ignorant of me. Therefore, I devote the next paragraphs to describe my thoughts behind this

straightforward statement. Currently, robots only complement and make human activity more convenient. However, they do not replace human beings in thinking on their own. They learn algorithms and are able to recognize patterns. When they have learned the pattern, they are not able to develop new original ideas that are not related to the content of the algorithm. As I mentioned, I personally cannot take the idea seriously at the present time. Nevertheless, I think, that most specialized people in the different robotics fields, do not have an exact idea of, how the development of robots in the next 30 years will look like. In the past years, we have seen an exponential development in IT, and I think 10 years ago a lot of people would not have imagined having had machines such as "Alexa" at home with who we can 'interact' with. Even if we believe that robots will continue to develop in a continuum and end up being fully autonomous machines which can think on their own and show creativity, it is not essential for the discussion at the moment. It is inevitable that we already have AWS and therefore need new policies right away. At the present time, I think we are not ready to give the responsibility to robots. One reason is also that it is hard to think of how we could hold a machine responsible. Would we put a machine into a prison? I do not think that we are ready for that. I also do not believe, that victims who suffered from wrongdoings from killer robots, will be satisfied with a solution where we give the blame to the robots. If wrongdoing has been done, victims and their affiliated, need some sort of punishment for the loss of their loved ones. The affiliated strive after fairness for them. They need justice to work through emotional and psychological sadness which such a loss results in. The current system in the western society is built on punishing those, who have done wrong. Another discussion is whether this at all is necessary, but I will not take on this discussion in this paper, as it will be too broad a topic on its own.

Therefore, I see significant problems in the discussion whether we can blame the robots. Firstly, because of what psychological effects it will have on the people who lost their loved ones. Secondly, I don't think that AWS are so developed, that we make them responsible for their actions. Therefore, I gladly take the discussion up, in the future, if robots will develop in the scenario described above. However, right now, I do not believe that the responsibility can be given to the AWS.

Humans will be responsible.

Another position taken in the course of the responsibility question is that we humans should take responsibility for the products that we make. Champagne et al. (2015) discuss the possibility of the term "blank check" responsibility. They believe that owners of AWS should

accept responsibility for the actions before the machines are being used. In this way, the actions of the robots can be "justified, reasonable and fair." They point out that they do not believe that this idea should be thought of as pointing fingers at someone. The idea is, that if we are going to use these AWS, then we need someone "vouching for" the actions that can follow with the use of them. In case of war, they suggest that a person of high standing, such as a military officer, or even people in the government, could take on the responsibility of the actions of AWS. Thereby, social prestige would go in hand with signing away some of their own freedom to an unpredictable future. They believe, likewise, other authors, that blaming someone randomly, if something goes wrong when using AWS, is morally unacceptable. Thereby, only the "blank check" can make it possible for us to use the AWS in a morally acceptable way.

I find this consideration provides a convincing argument, but I also see problems with this allocation of responsibility. I agree that we only can implement AWS into warfare if someone takes responsibility for the actions, or worst-case war crimes, that could result from the use of these weapons. I think that the idea that someone vouchers for the risk, is very interesting. However, I do see some flaws that can result from this agreement. First, imagine being powerful, wealthy and well educated. I think these three criteria have a significant influence on whether you would or won't or whether you get forced to sign this contract. It might not be that way, but I could imagine two scenarios where this could influence the decision of blank check responsibility. Firstly, what if you are an IT-Manager who decides to produce AWS, but you do not want to take responsibility for it, because you know that risks are involved with it. At the same time, you are well paid, high ranked and can buy you out of the responsibility. Is that a fair scenario? I do not think so. Secondly, you have a job at the military office. You work hard, but you are still financially burdened and need to ensure food for your family and save money for your children's college fund. Then someone gives you the opportunity for a promotion, higher payments which will result in financial security. The money reward is high, and maybe you do not have the sufficient knowledge of the topic to know the risks of taking the responsibility. Champagne et al. (2015) argue that "social prestige in the occupation of a given office will come as a price of signing away part of one's freedom." However, I think they neglect the background from where people come from. It is not all about the prestige, but money also plays a significant role. If we think of the scenario before, would it really be fair to blame the man who signed in order to give his family security? I don't believe this is the right approach. I think, there is a big risk, that people with power and money, buy people, who are struggling in their life and maybe don't have the

sufficient knowledge of the topic, to take on the responsibility. Perhaps, these people also don't see another way out of their financial problems, then signing this contract. I definitely don't think it is a satisfying set up if those people get blamed if something fails with the AWS.

The second idea in the view that humans should take responsibility, argues that the important question is not whether humans can be held responsible. The important issue is how to distribute the responsibility among humans (Noorman & Johnson, 2014). They argue that we should not view the AWS as a single idea. They also believe that we can include the programmers, the designers, those who decide how the robotic system will be deployed and those who validate and verify how the AWS behave. They claim, that commander officers still provide robots with a goal of how they should act, and the developers can program the range of their behavior. Basically, what our goal should be, is to clarify the questions of how and at what time the responsibility can be allocated to the people, who are part of the development of AWS. The word 'Development' refers to the technical, financial and governmental participation. The authors propose several strategies on how to allocate the responsibility between the developers, such as backward-looking and forward-looking responsibility. Forward-looking responsibility includes the task of deciding which tasks and duties should be assigned to which humans involved in the development process. Backward-looking, on the other hand, involves locating where the error occurred, when something went wrong. Whether it was the software, the way humans behaved when dealing with the AWS and the interface between human and hardware. Thus, they conclude that if we understand how the tasks were assigned in the AWS, we know what went wrong and we can allocate responsibility accordingly.

I agree that the developers, policy-makers and military commanders are part of the responsibility. I think the idea, to distribute the responsibility of those who have contributed to building an AWS is justified. I believe this, because I think it is essential that we are aware of the consequences that follow with developing new machines. I think that developers often recognize the positive effect that their new programs offer. However, they often forget how other people might exploit their programs. I will illustrate my point with Facebook. The original idea, to be able to bring friends and family together, even if they live on a different continent, is very admiral. However, in the past year, we have faced several Facebook scandals, such as manipulation and cyber mobbing. Obviously, just because a product can have a side effect, which it most often will have, it still sometimes makes sense to produce it. I just want to point out, that often, we do not think about the side effects. With machines, such as AWS, it is even more evident they can have tremendous side effects and are in

considerable risk of exploitation. Therefore, I think it is essential to make people who develop AWS think about the side effects. I think if we distribute the responsibility, it might help people to inquire about the full spectrum of possibilities of what positive and negative consequences the development of AWS could have, instead of only contemplate the positive outcomes.

On the other hand, there are arguments against distributing the responsibility. Firstly, the more autonomous a machine becomes, the more difficult is it for the programmer to predict the choices it will make. A lot of autonomous weapons are based on algorithms that are learning over time. Thereby, they are becoming more and more precise and better throughout time. This, however, makes it even more difficult to give the responsibility to a programmer. He might have been part of developing the machine in the preface, but as the machine will continue to learn, it will change. Secondly, part of the definition of autonomy in the context of weapons was "weapons…will be able to operate without meaningful human control over the critical function". So, the definition clearly states, that the machine will make a decision on its own. When giving the responsibility to the military commanders, they will end up being held responsible for actions they couldn't have held responsible for. When the machine is able to learn, it is even more difficult for the commanders to keep track of this development and actions they will take. An argument is that it, therefore, would not be fair to give the responsibility to the commanders.

Nevertheless, I still argue that these ideas of distributing the responsible are currently the best we have, if we cannot prevent the development of killer AWS. Programmer, military commander and policymaker, as mentioned before, should think about the negative consequences. Negative consequences could be war crimes were AWS kill civilians in warfare or AWS could get hacked to kill innocent people in their own country. If the people, who are involved in the development, reflect on these negative consequences, which easily could happen, and afterward still decide to tribute to development, I think it is more than eligible to divide the responsibility between the people that were part of the development of the AWS. Furthermore, I believe that this only could work, if the two positions for "humans will hold the responsibility" will be combined. I think that the responsibility for the consequences of AWS should be distributed to the people being part of the project and that they should voucher for the responsibility in advance. I think it is vital to voucher for the responsibility, because the task of distributing the responsibility if something goes wrong, might be arduous and challenging. We have seen, that there are a lot of arguments for and against responsibility for the different people or machines involved in an AWS crime. These

arguments could be used in court, and thereby it will make the decision of whose fault it will be more difficult if people don't voucher for their responsibility from the beginning. Yet, I think there is a better approach to AWS, where we can prevent the negative scenarios that can result in developing AWS. In the next paragraphs, I will share and discuss my ideas towards a solution.

A different aspect on responsibility

The above discussion on responsibility within AWS has shown us that the consideration of the topic is a difficult discussion. I believe, at the present time, there will always be someone who will be unhappy, unsatisfied and disagreeing if one of the previous solutions towards responsibility in AWS would be implemented. Maybe, we should think out of the box and start thinking of whether we can find a solution which will satisfy more people. Maybe, there is a way, to develop AWS, where the crucial question might not be who should be held responsible. Maybe, if we start thinking differently about how to create new AWS and think about what the technology could offer us to prevent deaths and war crimes. Maybe, the crucial question should rather be how to design and implement the system morally and ethically?

I will explain my thoughts towards the topic, by including ideas from the Just War Theory. I will highlight three criteria mentioned in the Just War Theory, which I believe make clear that we are misusing modern technology in a wrong way in contemporary warfare. Firstly, I want to highlight the point of 'just cause' in a just war. The 'just cause' implies that we can create war when we need to protect ourselves. In modern views, the term 'just' also refers to self-defense. The main idea is not that we want to kill people. Instead, we want to defend ourselves, and war intervention is in theory made to protect lives. It is difficult in a hot war to defend yourself without spreading blood. However, this is where AWS could play a potential revolutionary role. Secondly, I want to explain the meaning of Jus in Bello, the right to go into war. This doctrine states among other, that a war needs a military necessity. The idea is that an attack should help to defeat the enemy, but the offense should also limit excessive and unnecessary death and destruction. Lastly, I want to present the idea behind "no means" in Jus in Bello. This criterion states that we should not use weapons which cannot be controlled (Crawford, 2003).

The development of the current AWS, which I presented at the beginning of the paper, are in theory in line with the criteria I mentioned in the Just War Theory. However, they can easily be violated. The significant risk of having weapons which are autonomous is if they

start acting on their own or get hacked by anonymous people. If we do not have them under control, they can kill millions of innocent people. This also directly violates the “no means” criteria. I believe, if there is a risk that we will not be able to control the AWS, then this criterion is violated. Thereby, we do not have the right to conduct war in theory. So, how can we create weapons that are aligned with the criteria of Just War Theory? What if, we develop autonomous weapons, who are not able to kill? Well, how could we then defend ourselves? The current technology has never before been able to create such precise machines as today. We are at a time, where we are able to build machines, that can aim so precisely as never before, which are able to give electroshocks, make people unconscious and even point at several people in seconds. Let us imagine warfare: Today we are able to send weapons that can easily aim at hundreds and thousands of people and make them unconscious or paralyze them. An example is the mosquito killer machine. As mentioned in the beginning, this machine can point at 30-40 mosquitos per second. Instead of mosquitos, it could point at humans, and instead of killing, it could paralyze. With our technological improvements, it is in theory outdated even to use killing machines anymore. Humans have never been so bright and well developed as today. So why not use this new technology advantages to create AWS, that don’t kill. In this way, we are much closer to the Just War Theory. With these non-killing AWS, war will be just, as it undoubtedly will be to defend ourselves and protect lives. The non-killing AWS will also be in line with the military necessity. With these weapons, we can defeat our enemies without spilling blood. Lastly, the machines are obviously not perfect either. Non-killing machines could also attack civilians, or they could get hacked. However, it still will not have such fatal results as with the current killing-AWS. Non-killing AWS are simply not designed to kill humans. Therefore, the risk of war crimes will decrease. Furthermore, the current crucial question of who is holding the responsibility, might not be as hard, as the consequences of these non-killing AWS aren't as dramatic. In creating these machines, I would imagine, that developer, military commanders, and policy-makers will be more willing to distribute the responsibility between them. I believe this because the risk for taking the responsibility on is far lower than with killing AWS.

To conclude my idea, the Just War Theory, which the international laws of war are based on, could be better respected by AWS than it has been in the past from human weapon systems. AWS could thereby prevent deaths but still make it possible to conduct war and help countries defend themselves. Therefore, I argue that researchers, philosophers, engineers, and policy-makers should stop focusing so much on the question of the responsibility in the

context of AWS. Attention should be focused on how to program and develop the AWS ethically and morally.

Conclusion

I have argued that developing autonomous weapon systems in war could be used for military purposes if we start focusing on making moral and ethical weapon designs. In the preceding, I have presented modern autonomous weapons and discussed the problems within and outside warfare while using them. Furthermore, I talked about the challenges we are facing when we develop AWS. A significant challenge, which I discussed, was the issue of responsibility. I then presented current ideas on who could be hold responsible for the actions from the AWS. I debated the assumptions and arguments that underlie the view that human versus non-human should be held responsible. I argue that we haven't reached the level of giving the machine, the non-human, the responsibility yet. I claimed that it might be the right approach in the future when machines have human-like treats such as consciousness. Despite giving specific merits to the views and ideas which current philosophers and researchers are discussing in the course of providing the humans the responsibility for the machines they have built, I still comprehend some issues with these ideas. Nonetheless, I have argued for a combination of the current ideas mentioned in the paragraph of humans holding the responsibility. A combination of "voucher for" and distribute responsibility between the parties involved in the production, would be the most justified solution in my view. This would however only be the case if we cannot stop the development of killer machines. An idea that makes much more sense for me is rethinking the way we build the machines. I argued for the development of non-killing AWS and highlighted the use and benefits of these moral and ethical weapon systems. I prelude to current researchers and policy-makers in this field, to rethink their current course of technological development.

References

Asaro, P. (2007). Robots and responsibility from a legal perspective. Proceedings of the IEEE Conference on Robotics and Automation, Workshop on Roboethics, April 14, 2007, Rome.

Champagne, M., & Tonkens, R. (2015). Bridging the Responsibility Gap in Automated Warfare. *Philosophy and Technology*, *28*(1), 125–137. https://doi.org/10.1007/s13347-013-0138-3

Crootoft, R. (2015). Citation: Rebecca Crootoft, The Killer Robots Are Here : Legal and Policy Implications , 36 Cardozo L . Rev. 1837 (2015)

Crawford, N. C. (2003). Just War Theory and the U.S. Counterterror War. *Perspectives on Politics*, *1*(1), 5–25. https://doi.org/10.1017/S1537592703000021

Noorman, M., & Johnson, D. G. (2014). Negotiating autonomy and responsibility in military robots. *Ethics and Information Technology*, *16*(1), 51–62. https://doi.org/10.1007/s10676-013-9335-0

Slijper, F. (2017). Where to Draw the Line, *27*(15), 220.

Un.org. (2018). *Treaty on the Non-Proliferation of Nuclear Weapons (NPT) – UNODA*. https://www.un.org/disarmament/wmd/nuclear/npt/ [Accessed 17 Jul. 2018].

Wareham, M. (2018). [online] Stopkillerrobots.org. Available at: http://www.stopkillerrobots.org/wpcontent/uploads/2013/03/KRC_Moscow_9Sept2016.pdf