The Maastricht Journal of Liberal Arts
2021, Volume Thirteen, pp. 78-89

KILLER ROBOTS IN CONFLICT

The Morality of Artificial Intelligence in Warfare

Author: Finja Westerhove

Abstract In light of the fast pace of technological advancement in warfare, this paper is concerned about the moral implications of the use of artificial intelligence in the weaponry industry. Specifically, it provides an interdisciplinary perspective on the application of lethal autonomous weapon systems (LAWS) in conflict. The concepts of techno-moral implications of Swierstra (2015) and techno-moral boundaries of Kamphof (2017) are applied to the case of LAWS in warfare and provide insights into future changes of morals in war. The key results of this method suggests that LAWS in warfare threaten to erase moral virtues and cause a shift to a less humane reality of war.

1. Introduction

Artificial intelligence (AI) in warfare is a reality in many military test operations already (UNSESCO, 2019). More specifically, lethal autonomous weapon systems (LAWS), which are fully autonomous robots that are programmed to kill a certain target with no human intervention whatsoever, are being developed by influential political powers such as Israel, Russia, China, the United States and South Korea (Campaign to Stop Killer Robots, 2020). Since these robots' algorithms are commanding them to kill, this paper will refer to LAWS also as killer robots. These robots are the next level of technological warfare, after the currently leading unmanned drones. An example is HARPY, which is

able to detect and attack targets and was sold to Israel (PAX, 2020). Moreover, South Korea is planning to implement the SGR-A1, another autonomous weapons system, on its border to North Korea in order to detect human intruders (PAX, 2020).

Despite the progress in development, regulating restrictions for these lethal technologies are not in place yet. An international debate started in 2010 and is supported by organisations, that have called out for internationally agreed limits to AI in warfare (Sychev, 2018). The importance of this debate cannot be understated given the possible wide-scale implications the technology could have on the future of human kind. For example, in 2020, the Asian Times even referred to the urgency of multilateral action on the topic as equally important as action on climate change. The introduction of killer robots threatens to increase the destructive potential of warfare, which naturally brings up questions about ethics and justice. Academic literature on the topic currently concentrates on technical aspects and geopolitical implications but comparatively, little research has been done on possible moral implications. This paper contributes to the filling of this gap by expanding knowledge on the possible moral implications that LAWS might have on society.

As a result, this paper asks what are the techno-moral implications of the use of killer robots in warfare? In order to address the question within a theoretical framework, I first refer to Swierstra's (2015) argument on the social implications of technology. I relate Swierstra's research to the techno-moral implications of LAWS, which are the implications on the transmission of moral values through individual narratives, the morals in governmental decision making, the moral agency of perpetrators, and the morality in the international community. Second, I introduce Kamphof's (2017) case study on the application of AI in human relations and apply it to the implications on future techno-moral boundaries. The interdisciplinary character of the relevant areas of security studies, artificial intelligence, and technology studies meets the complexity of the topic, and enables a well-considered judgment on future developments. Finally, I conclude that if technology is entrusted with a task as big as deciding over life and death, serious impacts on techno-moral changes will take place. The affected areas of society range from the perception of war by individuals and legal questions of accountability, to the facilitation of governmental decision making. In all of these social spheres, killer robots in conflict threaten to alter important moral boundaries that make up our sense of self and govern our actions.

2. Theoretical Framework

2.1 Techno-Moral Implications

In his paper, Swierstra (2015) calls for a reflective engagement with techno-moral changes in order to handle moral implications of technological progress. Swierstra describes how emerging technologies and our morals mutually shape each other and can cause, beside quantifiable hard impacts, qualitative soft impacts on society. Soft impacts are co-produced by the consumer, rather than by the technology alone, and are subjective and hard to grasp. They come into existence through the application of technology in the social sphere and take the form of complex and difficult to measure changes in attitude, perception, or mindset (Swierstra, 2015). Since soft impacts' roots are multidimensional, it is difficult to agree on concrete causes, as well as to point out responsible actors. For this reason, soft impacts have been ignored in public discourse for a long time. For more moral innovation and reflection on the topic, their normatively charged anticipation requires the understanding of morals as a lived practise. According to Swierstra, a learning attitude that includes the negotiation of the coexistence with technology will help humanity to cope with these normative challenges by putting our current morals, as well as emerging ones, into question. In the following, I assess several techno-moral changes that might occur in relation to killer robots in warfare in order to analyse the soft impacts of LAWS and spark a discussion about the future of morals.

The use of killer robots in warfare has implications on the transmission of moral values through individual narratives. Through the utilisation of killer robots, fewer soldiers will have to risk their lives on battlefields and suffer from the consequences of physical and psychological harm. Instead, emotionless algorithms will take over the most disturbing tasks without any traumatic consequences. Compared to the current latest technology, LAWS would not need someone to give commands, since they operate on their own for several hours (Antebi, 2014). In his book, *Army of None*, Scharre (2019) highlights how this absence of pain could make us lose humanity in the process. Since a device has no emotions, it will not think back to the moment when it killed a person or witnessed a death. This lack of human memory in the process also prevents a conversation on the horrors of war in the broader society. For example, the dark narratives of soldiers after their return usually become part of a collective memory on the

familiar or cultural level, as described by Jan Assmann (2008). Once the individual memories have been transmitted to others, they enable ideas about what war is like and thus awareness of its cruelty. Once society has reached a consensus that war is bad and should therefore be avoided, it might become part of the cultural canon of beliefs (J. Assmann, 2008). Besides the negative experiences of soldiers, war conducted by killer robots also prevents soldiers from having the chance to see the humanity in the enemy and thus feeling compassion (Scharre, 2019). In this case as well, the narratives will not be transmitted to society and an increase in hostilities might follow as a result of the objectification of the enemy. For these reasons, the use of killer robots in warfare trigger the techno-moral change of decreasing awareness on the horrors of war and a stronger polarization of the conflicts' parties. This change is enabled through the lack of compassion, which LAWS do not possess. The impacts on society that arise from this absence of human emotion are the increased apathy or even willingness to go to war in society and governments.

In addition, the use of killer robots in warfare has implications for governmental decision-making. Building on the observation that citizens will not be confronted with the horrors of war anymore, the decision to go to war is made easier for governments, since public opinion is more likely to be in favour. This statement is supported by the fact that the use of LAWS puts fewer of the own soldiers at risk, which is why a negative public opinion on the decision is not at stake. This phenomena was described by Mueller (2005) as the *Iraq Syndrome*. He observed the public support for military interventions to have decreased simultaneously to an increase of deaths of soldiers. This lack of risk also facilitates a pro-war positioning of individual public officials since they do not possess responsibility for human lives and can avoid the consequences of endangering these. Moreover, no expenses have to be calculated for the wellbeing of the killer robots. Since they do not have human needs such as sleep, health insurance, nutrition or pension, much commitment for the soldiers can be saved (Contratto, 2012). For a government or military, these new conveniences through LAWS are highly beneficial. Consequently, the waging of a war will be less of a financial, as well as moral burden for public officials in the government and therefore an easier decision to make.

Justice in war is a further area where the use of killer robots has moral implications. In order to act morally in war, the principles of just war theory, *jus ad bellum* (having a justified reason to go to war) and *jus in bello* (fair rules during war),

must be observed and include the viewpoint of the enemy (Contratto, 2012). Included in these criteria is that both sides must have a realistic chance to win, which is not a given if one side owns a robot army while the other mobilizes humans (Gilli, Pellegrine & Kelly, 2019). Moreover, the killer robots revolutionize the kinds of operations some nations could carry out since no human has to be put in danger (Antebi, 2014). According to Antebi (2014), such hybrid armies have grown fast, from one robot per fifty soldiers in the US army, located in the Arab sphere in 2005, to one robot in thirty soldiers in 2012. Even though these were not killer robots yet, the asymmetry of the opponents, with the purely human army being the weaker one, was already visible here. This new balance does not conform to the concept *jus in bello* either, which states that for a war to be just, proportionality has to be assured (Contratto, 2012). This change of justice in war is a moral implication of the use of killer robots. It is contrary to current military ethics and thus, a possible change of our morals through technology.

The use of killer robots in warfare also has implications on the moral agency of perpetrators. All actions of a killer robot are steered by its algorithm that has been programmed by a human. The robot interprets the algorithm literally and therefore processes the information differently to humans (Antebi, 2014). If, against this background, a malfunction occurs or the robot takes an action that is illegal and must be punished as a war crime, the question of responsibility arises. According to Scharre (2019), the chaos of war can never be predicted in advance, which is why the moral agency behind the crime is difficult to judge. This loss of agency is closely connected to the infringement of human dignity, since it is possible no one will be held accountable for atrocities (International Committee of the Red Cross, 2018). Moreover, the lack of transparency of the algorithms and their responsible programmers might lead to the unpredictability of the use of LAWS and thus, a blind spot in moral agency (Gilli, Pellegrine & Kelly, 2019). In order to enable an ethical framework, Contratto (2012) proposed implementing rules of accountability in the production of LAWS and possibilities to trace human agents behind the technology. If no one is held responsible from the beginning, there will be too many possible culprits to bring about a punishment. This creates a blind spot in the legal system (Sychev, 2018). Not being able to trace back the agency behind crimes leads to a lack of adequate punishments and victims without justice. This scenario is thus another demonstration of the implication of

the technology on our current and future morals, the concrete effects of which are a soft impact of a loss of control over the rule of law in connection to killer robots.

Finally, the use of killer robots in warfare has implications on the morals of the international community. None of the currently existing conventions on the use of arms, such as the Geneva Conventions or the Laws and Customs of War on Land, have managed to include international regulations on the use of LAWS (Human Rights Watch, 2019). Several attempts have been made by multinational organizations, such as the United Nations, but a consensus could not be reached. This delay was due to big military powers that judge any decision as premature (Human Rights Watch, 2020). In light of the unwillingness to implement common rules, the technological progress in LAWS is continuing and nations keep working on the enhancement of their military capacities. This could cause anxiety about the other's progress and might look similar to an arms race (Sychev, 2018). For example, in 2017, Putin declared that whichever nation becomes a leader in AI could also become the leader of the world. This statement reminds Synchev (2018) of Cold War-like tensions, with the difference that this time a multipolar rather than bipolar structure is visible. Drawing from the lessons of previous arms races, these tensions have often led to irresponsible decision making and an absence of control, followed by wide-scale lethal consequences (Sychev, 2018). Situations like this would pose a threat to our common security, which Gilli et al. (2019) describe to be one of the highest moral goods. While threats on the level of national security endanger a certain group of people only, threats on the level of common security are powerful enough to endanger countless numbers of people including future generations, which locates LAWS as a top security threat (Gilli et al., 2019). Since politicians are well aware of the destructive consequences of previous arms races, but still engage in the production of LAWS, it can be argued that techno-moral changes have taken place in the international community. More specifically, it seems that the competitive character of the development of LAWS is sustainably lowering the threshold to take action that risks peaceful coexistence with other major powers and thus humanity's common security.

2.2 Implications on Moral Boundaries

Since LAWS have not been implemented in large scale projects yet, there is no data on the effects on morality in society available for analysis. However, AI has been used in other areas of society, and studies on the consequences were conducted simutaniously. In order to convey an idea of the impacts of AI in warfare, I assess Kamphof's (2017) ethnographic study on the use of AI in the sector of elderly care for reference. She observed the effects of a video monitoring system on patients and care workers, as well as the change of their relation to one another and the device itself. The application of the technology aimed at solving privacy problems in elderly care and at securing good relationships. Whereas initially the monitoring of the private sphere was uncomfortable and strange to both, the participants eventually adapted to the new situation. The technology retreated, as the patient did not feel watched anymore and the care workers learned to process the information beneficially (Kamphof, 2017). As a consequence, Kamphof argues, privacy has been redefined while the application of the technology has triggered techno-moral changes on the individual level of the participants. The relationship has undergone a creative process of co-shaping by the means of technology, which she describes as a dynamic and artful process, conducted by relational beings and their quest for good relationships. On the one hand, change took place through the gradual building of trust in the device, and on the other hand through the extension of subjectivity by the produced data. Kamphof highlights that the caretakers have found it convenient to let the monitoring system speak on behalf of the patients to secure dignity about vulnerabilities. Kamphof gives the example of patients with problems with basic bodily functions that they are ashamed to address. With the monitoring system, transparency over such issues can be maintained without the burden of having to talk about them. In this way, initial concerns about a loss of privacy were replaced by a perceived gain in privacy. Overall, a modification of the manner of communication between both instances has taken place and sustainably changed patients' and care workers' understanding of the self and the other. Considering the broader implications of this experiment, Kamphof argues society as a whole can be predicted to have flexible boundaries of morality in the application of technology in various aspects of life.

The use of killer robots in warfare has the implication of the future redefinition of moral boundaries. Applying Kamphof's (2017) observations to the use of AI in warfare, it

can be assumed that similar changes in the acceptance of the application of LAWS will occur. The issue of privacy in Kamphof's case translates into the transferring of human control to machines in the case of the killer robots. At the current stage of the matter, the autonomous decision making of robots over life and death is a strong techno-moral boundary that leads people to question the technology. However, drawing from Kamphof, these techno-moral boundaries are objects of change and might shift further away as soon as the boundary has been crossed. The intensity of this shift is dependent on the culture it is taking place in. Alesich & Rigby (2017) state that culture plays a big role in what people tolerate robots to do, since there are different concepts of what counts as a living creature. In Japan, for example, a fairly high tolerance was observed, due to the underlying belief that objects inherit a soul, whereas secular Europeans tend to be sceptical. Moreover, Kamphof's study showcases that the data drawn from the monitoring system co-shaped interpersonal relationships. In the case of killer robots this might lead to the perception of technology as an extension of our human skills. With the frequent and broad use of the technology, the notion of being a part of human subjects might form and strengthen with further application. Since this humanisation of technology brings them closer to our identities, trust might be gained in the decision making and functioning of the robots and cause a decrease of the current demands for stricter regulations. In light of this comparison, the use of killer robots in warfare could cause a shift in boundaries of societies' morals. Consequently, fundamental techno-moral changes in the acceptance and willingness of the application of LAWS can be expected to take place.

3. Conclusion

Summarizing the findings, it can be stated that if technology is entrusted with a task as big as deciding over life and death without human intervention, it is to be expected that large impacts on the morals of human relation to technology take place and cause techno-moral changes in diverse areas. First, the transmission of moral values through individual narratives is impacted by the lack of human emotions on the battlefield. This might lead to a lower threshold to go to war, since the public opinion is not shaped by the collective memory of the horrors of war anymore. Second, the morals in governmental decision making might shift, since no human soldiers of the own army are

at risk and do not have to be cared for during and after conflict. This dehumanisation of armies also impacts the fairness of war, since in many cases the opposition army could consist largely of humans. Third, the moral agency of perpetrators is impacted by the question of responsibility for the killer robots' actions. Malfunctioning that leads to war crimes might not be accountable to a human actor, reducing the human dignity of victims as well as the possibility of recourse through law. Finally, the morals of the international community possibly experience a shift towards tolerating a destabilization of international security through an AI arms race. Thus, in order to safeguard the moral decision making of nations, common regulations should be agreed on.

These moral implications are applicable to Swierstra's (2015) concept of techno-moral changes, since the analysed moral implications are triggered by the technological progress in LAWS; they mutually shape each other since a change in morals impacts the progress of the technology and vice versa. The developments are followed by the soft impacts of national apathy regarding the topic of war, paying less attention to the fairness of war, not being able to seek justice in war crimes, and lowering the threshold to take actions that risk security. Moreover, the case study of Kamphof (2017) suggests that techno-moral changes are shifting moral boundaries. Kamphof has observed this process to be a creative interplay of technology and human relationships that has transformative potential. While in elderly care technology was proven to adapt the boundaries of privacy protection, the application of LAWS in warfare might increase the general tolerance towards handing over destructive tasks to machines.

In order to prevent the application of LAWS in warfare, activists are arguing in favour of international regulations. However, national interests make it difficult to find consensus. For example, authors from 2007 such as Sparrow already thematized the same issues as more recent literature from 2019. This shows how the diplomatic progress has been significanly slower than the technological one.

Nevertheless, killer robots in conflict threaten to erase important moral virtues that currently govern our actions. War is already an immoral issue but has potential to evolve in an even more cruel direction through the use of autonomous weapons. As Gilli et al. (2019) point out, machines might be better at predictions, but humans' are still better at moral decision making, which is why we would give up a part of our humanity when handing over this skill to artificial intelligence. If the techno-moral implications of

the use of killer robots in warfare shall stay within a humane framework, then the development of AI must stay human-centred.

References

Alesich, S. & Rigby, M. (2017). Gendered robots: Implications for our Humanoid future. *IEEE Technology and Society Magazine, 36*(2), 50-59. https://doi.org/10.1109/MTS.2017.2696598

Antebi, L. (2014). *Controlling robots: It's not science fiction. Arms control and national security: New Horizons*. Institute for National Security Studies (pp. 65-80). http://www.jstor.org/stable/resrep08938.7

Assmann, A. (2008). Canon and Archive. In: A. Erll & A. Nünning (Eds.), *Cultural Memory Studies* (pp. 97-107). Walter de Gruyter.

Assmann, J. (2008). Communicative and cultural memory. In: A. Erll & A. Nünning (Eds.), *A companion to cultural memory studies* (pp. 109-118). Springer. https://doi.org/10.1007/978-90-481-8945-8_2

Campaign to Stop Killer Robots (2020). Protect civilians: Stop killer robots. https://www.stopkillerrobots.org/2020/05/protect-civilians-stop-killer-robots/

Campaign to Stop Killer Robots (2020). The threat of fully autonomous weapons. https://www.stopkillerrobots.org/learn/

Contratto, M. (2012). *The decline of the military ethos and profession of arms: An argument against autonomous lethal engagements*. Air University Press.

Gilli, A., Pellegrine, M., & Kelly, R. (2019). Intelligent machines and the growing importance of ethics. In A. Gilli (Ed.), The brain and the processor: Unpacking the challenges of human-machine interaction (pp. 45–54). NATO Defense College. http://www.jstor.org/stable/resrep19966.11

Human Rights Watch (2019). Killer robots: Ban treaty is the only credible solution. https://www.hrw.org/news/2019/09/26/killer-robots-ban-treaty-only-credible-solution

Human Rights Watch (2020). *Stopping killer robots, country positions on banning fully autonomous weapons and retaining human control*. https://www.hrw.org/sites/default/files/media_2020/08/arms0820_web.pdf

International Committee of the Red Cross (2018). Ethics and autonomous weapon systems: An ethical basis for human control? *Arms Control Today, 48*(6), 38-39. https://www.jstor.org/stable/90022990

Kamphof, I. (2017). A modest art: Securing privacy in technologically mediated homecare. *Foundation of Science, 22*(1), 411-419. https://doi.org/10.1007/s10699-015-9448-5

Mueller, J. (2005). The Iraq syndrome. *Foreign Affairs, 84*(6), 44-54. https://doi.org/10.2307/20031775

PAX (2020). Dutch scientists call for ban on killer robots. https://www.paxforpeace.nl/stay-informed/news/dutch-scientists-call-for-ban-on-killer-robots

Scharre, P. (2019). *Army of none*. WW Norton & Co.

Swierstra, T. (2015). Identifying the normative challenges posed by technology's 'soft' impacts. Ethics in practice. *Nordic Journal of Applied Ethics, 9*(1), 5–20. https://doi.org/10.5324/eip.v9i1.1838

Sychev, V. (2018). The threat of killer robots. UNESCO. https://en.unesco.org/courier/2018-3/threat-killer-robots

UNESCO (2019). The UN versus LAWS. https://en.unesco.org/news/versus-laws