Autonomous Weapons Systems and the Ethics of Unmanned Warfare

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

William Orser Renken

Spring 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Pedro A. P. Francisco, Department of Engineering and Society

Introduction

Humanity has been at war for thousands of years, and weapons technology has always been a factor for whichever party wins or loses. From a simple rock to the sword, from the sword to the longbow, from the longbow to the crossbow to black powder firearms and beyond, contemporary weapons have always defined military tactics and, ultimately, the course of battle. However, a new problem arises when the weapons start to operate themselves - who is to blame for the destruction they cause or when they malfunction? Certainly the weapons themselves (or the algorithms that run them) can be blamed, but court marshaling a missile or a drone sounds counterproductive, to say the least.

In the sociotechnical lens of international naval relations and the principle of just war, the novel technology of autonomous weapons systems (AWS) is posing unique issues regarding new legislation, relative ease of waging war, and an imbalance of the risk of loss of personnel between two warring nations or groups. With the rapidly emerging technologies of artificial intelligence, AWS, and the integration of the two, organizations such as the United Nations and the Department of Defense are facing a problem of reaction rather than anticipation (Hall, 2017, 86).

The justification of military violence under the umbrella of AWS by the United States and its allies also brings about a sense of urgency for the development of necessary legislation – the "procedural-organizational appropriateness" school of thought brought about by AWS is deeply affecting the sociotechnical landscape and its relation to the law (Bode, Huelss, 2018, 413). Additionally, the existing Laws of War (LOW) and Rules of Engagement (ROE) language can and has been used as a method of circumnavigating the ethical issues brought about by AWS through an approximation (or in other words, a placeholder) for morality so that relevant ethical

concerns can be disregarded (Lin et al, 2008). The affected parties extend beyond the United States; as the price of these novel technologies is driven lower and lower, more and more players are beginning to acquire AWS. (Coyne, Alshamy, 2022, 191) This price change is also affected by the engineering companies working on these systems, such as Lockheed Martin and Northrop Grumman in the case of the United States.

As the United States is at the forefront of military technology on the world front, the issue of the ethics of AWS falls primarily on US soil. However, other global players such as China and Russia must also be considered, as the setting of autonomous weapons technology is quickly becoming international.

With the rapid, largely unregulated development of autonomous weapons systems on the forefront of military technology, who is to blame for the loss of life due to malfunction or even successful implementation? This research paper will explore the ramifications of leaving weapons control to algorithms, how these algorithms are currently regulated (or, in this case, unregulated), and which party is to blame for the destruction that they cause. AWS, while a revolutionary technology, should hold both their operators and designers accountable as they are created and implemented in the world military theater.

Background and Significance

From a technical perspective, autonomous vehicles in the armed forces is a leap forward when compared to manned craft (maritime or airborne). This means that threats can be identified and neutralized quicker and with more precision. Additionally, this has been implemented in the airborne munitions sector in the form of “lingering” missiles (primarily developed by Anduril Industries), which circle an area until they are given a target. (Horowitz, 2016) AWS also promise to be more inexpensive and easier to operate than traditional vessels, munitions, or

countermeasures. “Algorithms of violence,” particularly because they are so simple and inexpensive to implement, can present unprecedented opportunity for intense acts of violence by both small minorities and vast majorities on various populations – such destructive power on tap “might bring forth, or destroy, democracies just as easily as they replace tyrants.” (Asaro, 2019, 547) The technical ability of AWS, then, presents a novel concern for governmental stability and lowers the price of entry of highly destructive warfare.

The rapid advancement of complexity of artificial intelligence (AI) also poses an interesting challenge for the governing of such systems. Language models such as ChatGPT are based on “next word prediction” models, and thus are only as reliable as the sources they gather data from (which is not considering the fact that they can distribute incorrect information even from correct sources from their various models). Artificial intelligence implemented in weapons systems will thus be vastly different from these conventional models. Data fed into these models will primarily be military combat data, and will soon give AI the ability to direct swarms or fleets of AWS for specific targets. (Wilson, 2020, 128) Another key question that arises from this rapid, largely unregulated development is: who is responsible for destruction of life or property with the discharge of such weapons? Is it the fault of the designer of the system if the AI model makes a mistake and, for instance, prematurely discharges the weapon, injuring or killing friendly personnel? There seems to be no consensus on this matter. (Horowitz, Asaro, Heins) However, it does seem that Murphy’s Law still applies in this situation: “anything that can go wrong will go wrong.” In the case of deadly weapons of war, this principle must be taken under very careful consideration in both the design process as well as the ethical and legal processes.

Socially, this emerging technology poses an important question about the ethics and legislative systems of warfare – international legislation is currently leagues behind the actual

technology, which raises an important question of accountability. “Machine malfunction may cause great harm, but no human may be accountable.” (Coyne et al, 2021, 11) As AWS become more and more advanced, there will eventually be no human interaction with the system at all, which raises the question of what a justly declared war will look like in the near future. With autonomous technology, strategies and practices like Kamikaze will become more common due to the lack of the risk of loss of personnel. AWS poses a unique threat to society as a whole as well – governments may fall, large majorities can be silenced, and international war can be started by the literal push of a button.

Methodology

I will use the method of Actor-Network Theory (ANT) to analyze the importance of AWS in the current sociopolitical and sociotechnical landscape. ANT is an appropriate system of analysis for such a global issue due to the complexity that it can accommodate – each nation is an actor, with government defense contractors and lawyers of international law as subsets of each actor. The relationships between the actors via the connections of AWS and their morality connects the network. This analysis is important because autonomous craft and weapons systems are the cutting edge of defense and should be taken under careful consideration in practical application, drawing of new language in legislature, and practice of international affairs.

The question of the ethicality of autonomous weapons systems must be approached carefully and through several different lenses and thought frameworks. This analysis will primarily utilize the concept of Responsible Research and Innovation (RRI), which employs the notion of engineers anticipating future issues and responding to social realities. (Stilgoe, Owen, Macnaghten, 2013, 1570) RRI is an applicable lens through which to look at AWS because of the idea of response on the part of the engineering community. The idea that “just because we can,

doesn't mean we should" comes to mind when examining AWS through this concept. There is also the idea of adaptability in design, which is particularly important for AWS. As the technology improves at its current pace with the continued lack of legislature, the technology can (and will) outpace the ethical frameworks put in place by the Geneva Convention. Thus, both the research conducted on the topic here as well as that done by engineers working on such projects should heavily employ RRI to ensure that the technology does not develop at an uncontrollable pace.

The literature review for this question was also approached, as often as possible, from a bipartisan perspective. Because the United States is a primary actor in the AWS ANT framework, it is paramount to remain as impartial as possible. For this impartiality to be palpable, diverse sources from actors in the federal government, civilian scholars, and specialists in international law must be considered with similar weight. Questions relevant to this data set include the following: what is currently in service, what is in development, and how is the development occurring? This data analysis will be carried out in two ways: policy analysis and ethical analysis. Policy analysis for this topic will consist primarily of international law (such as that agreed upon in the Geneva Conventions), military strategies for NATO and the US, etc. while ethical analysis will be conducted through review of philosophical thought experiments and other abstract schools of thought: Is the dehumanization of war unethical? If soldiers' lives are no longer on the line on the offensive side and only on the defensive, what is preventing war breaking out more frequently? The frameworks of RRI and ANT will guide the logic flow that addresses the question of accountability with the implementation of new AWS, with sufficient literature to examine each actor, the research responsibilities for each (if any), and the influence each actor has over the decisions relevant to AWS and the accountabilities therein.

Literature Review

Autonomous weapons systems technology is years ahead of legislation and protocol, which creates the challenging precedent of continued development. (Hall, 2017) Furthermore, Bode and Huelss suggest that AWS are creating new norms instead of abiding by them - thus, the United States is justifying "appropriateness" for ending human life at the hands of AWS. (Bode, Huelss, 2018) This is supported by suggested norms arising from "procedural-organizational appropriateness," which is the idea that new technologies create the pretense for new methods of warfare and not vice versa.

There is also a consensus that AWS, as a technology, should be banned altogether. This is supported, according to Asaro (2019), by the fact that autonomous systems (particularly loitering munitions, which though they have been in development since the 1980s, have seen recent rapid development) have the potential to grant hitherto unseen amounts of power to smaller organizations that have violent intentions. "Algorithms of violence," as Asaro notes, are malevolent forces that are difficult to control by governing bodies due to their relative open-source nature. Critical democratic and international institutions can be critically attacked at any point from a source that can be invisible. Furthermore, Zhao et al suggest a distinction for algorithms for autonomous technology - moral algorithms and algorithms that yield moral results are two completely different ideas, and moral algorithms should be the ideal.

Some also suggest that there is currently no way to concisely create legal language for governing AWS with the state of the technology as it stands. Lin et al (2008) assert that in programming an AWS, either from bottom up or from top to bottom, moral and ethical questions can be avoided through Laws of War (LOW) or Rules of Engagement (ROE). These law sets can be used as a general approximation of morality (or at least, a set of predetermined moral

guidelines) so that ethical concerns can be disregarded. This is supported through the idea that humanity does not yet have the technology to create a fully autonomous vessel and must require some degree of interaction. Thus, ethical concerns can be postponed until that becomes a reality. They support this assertion through the idea that the technology is so undefined and unexplored in practical applications that it is nearly impossible to create frameworks that are universally applicable. This school of thought is dangerous - if AWS are allowed to develop with no guidelines, or if guidelines are postponed until the technology has progressed to some predetermined point, there may be no way to slow down or stop development for this process of rulemaking to occur. Contractors such as Northrop Grumman, Lockheed Martin, Boeing, etc. are not in business by waiting; they are in business to make money.

Heins (2018) recommends that the consequentialist approach to the morality of AWS (that is, the ends justify the means), while useful in some military applications thus far, is detrimental in the long term. As the human element of autonomous systems diminishes, the methods of the systems themselves (thus, *the means*), become exponentially more important. The grounds for this statement are that while distance between the enemy has been the primary focus of military technology up to the present day, there has always been a human controller on the operation end. If that element is taken away, sole trust rests in the systems themselves, which calls into question the consequentialist approach. This reasoning is sound, on the basis that the LOW and ROE guidelines are not sufficient for the governing of such weapons, and that further development of moral algorithms (as suggested by Zhao et al) is necessary for appropriate legislation and deployment of AWS.

J.M. Beard, in the Georgetown Journal of International Law, offers perspectives on the potential legal issues regarding the responsibility of individuals as well as states with respect to

the use of autonomous weapons. In his words, the "diminishing level of human control will continue to raise increasingly difficult questions about both state and individual accountability for the actions of autonomous weapon systems." (Beard, 5) In the 1950s, heat seeking missile technology governed changes to legal framework with regards to fault. Personnel had the final decision to launch, but once the munition was fired there was little to nothing an individual could do to stop it from discharging. Thus, the algorithms used for these weapons came under a very close eye from the US government as well as in international law. Newer, further autonomous systems are, essentially, entirely unregulated, sans the requirement under US law to "receive a legal review 'to determine whether the weapons or weapon systems or their intended use in combat are consistent with the obligations assumed by the United States Government under all applicable treaties and with customary international law.'" As the United States is not a party of Additional Protocol I of the Geneva Conventions of 1949 (which constitutes review of the convention for each new weapons technology developed), the internal review of the US is the only current safeguard.

Discussion/Results

The primary question of this investigation, that of responsibility for casualties at the hands of AWS, has an answer that is relatively straightforward: the responsibility falls on the operators, but more importantly the designers, of these systems. There are numerous actors governing these two roles (such as the US government, NATO, the UN, among others) that should serve as governing bodies as well as checks and balances, but the designers of these systems should be held accountable to the highest extent. This is supported by the fact that the legal frameworks for AWS are years behind the development of the technology (Beard, Asaro,

Heins, Bode, Huells) and the LOW and ROE are the only existing ethical codes that the systems must exhibit. Therefore, the designers and engineers of AWS set the precedent for the decisions that the munitions make. Until international law can be drawn to govern these technologies, there are two methods to maintain ethical algorithms: one is to ban their development altogether and the other is to allow development but assign accountability to the designers of the system. Banning the development altogether (as suggested by Asaro) seems unreasonable - even if the US and its allies comply with the ban, there are certainly others that will jump at the opportunity to obtain an edge over competition.

Procedural-organizational appropriateness, that is, the norms that arise through existing procedures in response to a need in organizations such as the armed forces, is a strong driver of the continued use of AWS, especially with the precedent of use in the war on terror. (Bode, Huells, 2018, 412) It is this force that is a justification for the U.S. military to continue development and deployment of AWS - it simply has been a solution in the past, so will continue to be in the future. For example, the U.S. drone program "has set novel precedents for how ending human life is 'appropriate'," and the U.S. military operates under these precedents because they now fall under the umbrella of procedural-organizational appropriateness. (Bode, Huells, 2018, 413) Therefore, at least to a large extent, the organization as a whole is to blame for the use and destruction of AWS simply because the justification for continued use is simply that the organization deems it appropriate based on previous action. Furthermore, the designers of the algorithms in these organizations in particular are accountable due to participation in such practice at the most, and complacency at the least.

Additionally, the precautionary principle approach is a system which should be used in AWS' development, but is not. (Coyne, Aishamy, 2021, 197) There are always unintended

consequences seen with the use of new technologies, and the high risk nature of AWS amplifies the implications of such consequences. The concept of precautionary principle is also in agreement with the Responsible Research and Innovation (RRI) approach outlined by Stilgoe, Owen, and Macnaghten. Indeed, those who have the most agency over the design and capabilities of AWS, the designers, are the most applicable users of RRI and precautionary principle. With these two frameworks' focus on the development stage and not (per se) the actual governing of such technologies, it follows to conclude that those primarily active in the development stage are to use these principles most actively.

Further, the rising trust in the algorithms that guide AWS should weigh heavily on the shoulders of the designers of the systems. The U.S. government uses a consequentialist approach (e.g. the ends justify the means) for the use of these weapons systems. (Heins, 2018, 50) As the human element in the targeting and deployment of AWS diminishes, the algorithms of the system (that is, *the means*) become more and more liable for the damage caused. Algorithmic bias, which is derived from the designer of the algorithm, is then extremely important to mitigate. The designers of the algorithms, then, hold responsibility for the bias of the system and the behavior of AWS.

Conclusion

Autonomous weapons systems are irreversibly changing the course of warfare forever. With the deadly combination of rising reliance on algorithms, higher destructive capability, plausible deniability of individual responsibility, and lack of regulation, AWS poses a threat to the world theater on a similar scale and timeline as the rapid introduction and integration of nuclear warheads. Therefore, the approach for appropriate ethical frameworks and legal regulations should be deliberate, thorough, and if possible, swift. However, until appropriate

legislation is developed, the accountability for the destruction and decisions made by autonomous systems should rest on the designers and engineers of the military contracting companies developing them as well as the operators of the munitions with the "big red button."

Steven Umbrello addresses the complication of the legislation of AWS by suggesting the use of LOW and ROE to develop an interim framework; while it is possible for this approach to be implemented effectively, there is also the ever-present risk of military contractors finding loopholes in the system or otherwise skirting the regulations via lobbying. To prevent this, the engineering contracting corporations associated with the development of autonomous weapons should be held directly accountable in the event of malfunction or unethical behavior of the systems through increased responsibility given to the designers. While there are also methods of avoiding this responsibility through disclosure of shortcomings to avoid, for example, a class action lawsuit, the combination of these two suggested frameworks along with the principles of responsible research and innovation create an ethical framework of responsibility resilient enough to hold until appropriate legislation is developed.

References

Asaro, P. (2019). Algorithms of Violence: Critical Social Perspectives on Autonomous Weapons. Social Research: An International Quarterly 86(2), 537-555. https://doi.org/10.1353/sor.2019.0026.

Beard, J. M. (2014). Autonomous weapons and human responsibilities. Georgetown Journal of International Law, 45(3), 617-682. https://heinonline.org/HOL/LandingPage?handle=hein.journals/geojintl45&div=25&id=&p age=

Bode, I., & Huelss, H. (2018). Autonomous weapons systems and changing norms in international relations. Review of International Studies, 44(3), 393-413. doi:10.1017/S0260210517000614

Christopher Coyne & Yahya A. Alshamy (2021) Perverse Consequences of Lethal Autonomous Weapons Systems, Peace Review, 33:2, 190-198, DOI: 10.1080/10402659.2021.1998747

Hall, B. K. (2017). Autonomous Weapons Systems Safety. JFQ, 86, 86–93. https://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-86/jfq-86_86-93_Hall.pdf

Heins, J. C. (2018). (tech.). Airpower: The Ethical Consequences of Autonomous Military Aviation (pp. 1–92). Naval War College. Retrieved October 23, 2023, from https://apps.dtic.mil/sti/citations/AD1079772.

Horowitz, M. C. (2016). The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons. Daedalus. https://www.amacad.org/publication/ethics-morality-robotic-warfare-assessing-debate-ove r-autonomous-weapons#:~:text=Autonomous%20planes%2C%20for%20example%2C%2 0flying,with%20the%20law%20of%20war

Lin, P., Beckley, G., & Abney, K. (2008). (thesis). Autonomous Military Robotics: Risk, Ethics, and Design. U.S. Department of Navy, Office of Naval Research. Retrieved October 23, 2023, from https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac.

Stilgoe, J., Owen, R., & Macnaghten, P. (2020). Developing a Framework for Responsible Innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.4324/9781003075028-22

Umbrello, S. (2019). Lethal autonomous weapons: designing war machines with values. Delphi Interdisciplinary Review of Emerging Technologies, 2(1), 30-34. https://heinonline.org/HOL/LandingPage?handle=hein.journals/delphi2&div=9&id=&page =

Zhao, H., Dimovitz, K., Staveland, B., & Medsker, L. (2016). (tech.). Responding to the Challenges in the Design of Moral Autonomous Vehicles (pp. 169–173). Washington, DC: Association for the Advancement of Artificial Intelligence.