THE ETHICAL AND STRATEGIC DEVELOPMENTS IN WARFARE THROUGH AUTONOMOUS COMBAT TECHNOLOGY

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

By

Lilian Price

March 30, 2023

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
Catherine D. Baritaud, Department of Engineering and Society

INTRODUCTION TO THE EXPANSION OF LETHAL AUTONOMOUS WEAPONS SYSTEMS (LAWS) WITHIN THE U.S. GOVERNMENT AGENDA

One of the main objectives of the United States revolves around national security, whether that be within the United States or globally. With the rise in concerns for national security, the government has increased military funding in order to improve technology that can assist with internal and external threats, and create an advantage to other world powers such as Russia, China, and North Korea (Fiott, 2018, p. 41). This rise in funding has generated a new generation of technology, specifically focused on autonomous weapons, meaning devices that use "theoretical methods and techniques for simulating and expanding human intelligence", and are essentially self-taught (Cao, 2017, p. 701). These devices have been implemented as missiles, aircrafts, drones, and other pre-existing weapons involved in armed combat. Given the rapid expansion of artificial intelligence capabilities, there is a void in national policy in regards to the extent of the usage of these systems, which still contain a multitude of flaws. In more recent years, there has been a push to not only study the ethical implications of implementing such technology into warfare before being put into practice, but also creating targeted policy in order to limit the scope of its use.

The STS research paper seeks to evaluate the previous expansion of policies regarding autonomous weapons in addition to current efforts, all in response to the growing government funding of such weapons by the United States. Previous to artificially intelligent weapons, there has long been a component of technology in weapons systems in order to perform more objective tasks such as "identifying and tracking incoming targets and engaging them" in addition to "reacting to incoming missiles or mortar shells in cases in which the timing does not allow for human decision-making" (Sauer, 2016, para. 3). Such automatic systems are static, and perform functions against objects within a well constrained set of parameters and within a given environment (Sauer, 2016, para. 7). However, for the purposes of this research paper, such automated systems create a useful comparison for newly autonomous systems. In direct contrast to previously automated systems, autonomous weapons are more dynamic in nature, and operate outside of a constrained environment in order to perform "advanced capabilities in navigation, aerial refueling, reconnaissance, maneuvering, swarming with multiple platforms, electronic warfare, and, perhaps, some aspects relating to the application of force" (Horowitz, 2014). Although their anticipated use is to expand on typical automated features such as autopilot and maneuvering, there has been a race to implement policies in order to address the potential trajectory of autonomous weapons into combat weapons with the capability to apply direct force.

By analyzing the increase in autonomous weapons in conjunction with Actor-Network Theory (ANT), the varying degrees of influence between different groups that directly correlate to the production and study of autonomous weapons can help discern the barriers to policy development. An educated understanding of the relationships between key stakeholders in the model will influence the development of potential areas of improvement within government policy in order to mitigate the potential future consequences of implementing unregulated weapons systems with the potential for inflicting direct force on a large scale.

The technical paper involves the development and fabrication of a game that interfaces between a human and computer, and involves guessing various words chosen by either the player or the computer. In broad terms, the game has three modes of play, two of which involve human to computer interaction. For each mode of play, the communication between the computer and player goes two ways, where the player can choose the word or the computer can choose the word. The objective of each round is for the player to guess the word within a given allotment of guesses.

Linking both the STS and technical papers together creates an avenue for understanding in regards to how computers learn and function. Although in this case on a simplistic scale, one potential failure to legal precedent for autonomous technology is a lack of understanding by not only policymakers but civilians and military members. By creating a game that uses computer technology to try and successfully fool the player and vice versa, the public is becoming more aware and comfortable around interacting with and understanding computer technology that is meant to interface and learn off of people. Despite the technical project being loosely coupled with the STS research paper in terms of device equality in terms of function, the technical paper provides an elementary step in increasing the understanding of computer systems by the general public in the hopes of encouraging future education on more complex computer systems such as autonomous weapons and their potential impact in various functions.

AN OVERVIEW TO THE EXPANSION OF GOVERNMENT FUNDED PROGRAMS FOR AUTONOMOUS WEAPONS

The United States government, over a five-year period, allocated $2 billion to the development of technology such as lethal autonomous weapons systems (LAWS) in order to be implemented during warfare (Di Corpo, 2021, pp. 260). Since this technology is in its first stages of development, many ethical concerns arise from the potential misuse of these weapons. While some government and defense contracting officials argue that developments in autonomous weapons protect national security and military personnel, there is a significant reason for speculation that implementing autonomous weapons into armed combat creates potential situations of mass destruction and casualties. In addition, these fears of misuse are echoed by the American people due to their mistrust in government transparency when it comes to national security. This public perception of government abuse in relation to national security and times of war stems from the 1960s with both the Vietnam War and scandals such as Watergate occurring within the same decade. Although slightly recovered, this mistrust has since continued to decline in the 2000s from the 9/11 terrorist attacks in addition to the government's involvement in the wars in Iraq (Beyond Distrust: How Americans View Their Government, 2020).

Despite this mistrust, the government intends to expand the use of LAWS under their supervision, which was made clear by Bob Work, the US Deputy Secretary of Defense in his introduction of the Third Offset Initiative:

> So, DOD is -- we are going to leverage AI technology, particularly in things like cyber defense, electronic warfare defense, missile defense. But what's also clear to us is that we need to go to huge new levels of human-machine symbiosis, allowing each to do what the other does -- which is to do what they do best (2016, para. 10).

The public perception of government control surrounding technology such as LAWS begets public concerns surrounding the potential misuse of these systems bending the rules of ethics in order to push an agenda under the guise of national security. These systems contain an extensive potential for abuse, which in the hands of the government is far from the control of the public. Nevertheless, the United States continues to push forward in the autonomous arms race with the expected funding to increase from 11 billion USD to 52 billion USB between 2016 and 2025 (Fakhreddine et al., 2019, p. 14).

AN INTRODUCTION TO THE INFLUENCE OF ANTI-AUTONOMOUS WEAPONS GROUPS ON POLICY DEVELOPMENT IN THE USAGE OF LETHAL AUTONOMOUS WEAPONS SYSTEMS

As of the present, there are six types of weapons that are banned from warfare internationally which are comprised of “poison gas, biological weapons, chemical weapons, blinding lasers, antipersonnel landmines, and cluster munitions” (Goose & Wareham, 2016), all of which were banned in the last 20 years. Each of the listed forms of weaponry were banned in an attempt to preserve civilian life as well as encourage the preservation of international humanitarian law. With the current list of banned weapons, their function greatly contrasts that of autonomous weapons, given that their outcomes are fixed, meaning that those weapons perform one function through a specific method or medium. In the case of autonomous weapons, their predictability is almost unknown, even after extensive testing.

Within the past decade, a greater call has been made internationally to study the effects of autonomous weapons systems in order to prevent the death and/or destruction of innocent civilians. According to the most recent initiative outlined by the Department of Defense Directive (2022), the implementation of autonomous weapons is allowed only after rigorous testing to verify that they:

> Function as anticipated in realistic operational environments against adaptive adversaries; complete engagements in a timeframe consistent with commander and operator intentions and, if unable to do so, terminate engagements or seek additional human operator input before continuing the engagement; and are sufficiently robust to minimize failures that could lead to unintended engagements or to loss of control of the system to unauthorized parties (Sayler, 2019).

In addition to these terms, any modifications or updates to the system require the testing process to be repeated in its entirety. Although this set of design verifications appears rather robust, each checkpoint is verified through senior officials at the Department of Defense which has no legal standard to support the approval of these systems. Despite this directive being a step toward regulation and the protection of humanitarian efforts, the rate of development and anticipated scale of implementation is unknown and not explicitly regulated in legislation. On one hand, these systems could become very useful in instances that require large data mining or quick decision making, however their unpredictability in a crisis situation is a cause for concern. These concerns have in turn lead to the creation of a vast array of non-government organizations such as the International Network on Explosive Weapons (INEW), International Campaign to Abolish Nuclear Weapons (ICAN), and the Campaign to Stop Killer Robots (Goose & Wareham, 2016). Each of these groups in affiliation with one another, seek to ban the production and implementation of autonomous weapons.

The pressure to enforce government regulation of these systems has sparked a dialogue amongst the United Nations Human Rights Council, however this push for legislation is in direct conflict with the outside pressures to innovate. Despite the call to ban the production of autonomous weapons, the United States is met with an equal pressure to maintain the upper hand against opposing world powers who are also in the production of autonomous weapons "such as China, Israel, South Korea, Russia, and the United Kingdom" (Goose & Wareham, 2016). In practice, the United States could lose any military advantage if they choose to put restrictions on autonomous weapons while other well-funded and well-established militaries choose to push for the production of autonomous weapons.

ANALYSIS OF POTENTIAL FAILURES TO AUTONOMOUS SYSTEMS THROUGH A RISK ANALYSIS

The STS paper seeks to highlight the potential failures modes in relation to recent examples of autonomous system integration. Typically, one stance argues for the usage of autonomous weapons given that most of them are preventative, and perform somewhat simplistic and repetitive tasks with a small potential for failure. However, this analysis calls the past failures in relation to the small-scale implementation of autonomous weapons, which provides insight into the risk factors in authorizing these weapons on a large scale.

RISK ASSESSMENT OF COLLISION AVOIDANCE IN UNMANNED AERIAL VEHICLE (UAV)

This risk analysis will specifically characterize the current faults in unmanned aerial drones in order to suggest avenues for improvement as well as understand the necessary developments in order for these systems to be safely used in combat. An unmanned aircraft parallels that of an autonomous system given that it is characterized as being unmanned, meaning there is no pilot on board controlling the aircraft. This, in turn, leaves most of the system functionality to the drone itself, which is controlled by various sensors in connection to artificially intelligent software. Essentially, this case study involves a form of an autonomous system that exhibits more human control in order to suggest future failures for systems that lack any human element.

Methodology

For the sake of this risk analysis assessment, both civilian and military factors are considered through the usage of a fault tree. In the typical function of an autonomous system, there are two phases which are the “strategic phase, where the mission is planned and initial requirements are completed and the tactical phase which is the mission flight phase where monitoring happens” (Khan, 2021, pg. 2). In this case, the primary risk analysis is conducted in terms of the tactical phase, given the larger potential for system fault and failure. The analysis itself will involve not only identifying the key failure modes within the tree, but also iterating through each failure mode in order to determine the scale of potential failure. Each one of the failure modes is then considered in tandem with one another in order to create a holistic view of the scope of potential failure, and the direct impact it has on not only the military but civilians.

The goal of the risk assessment is to use the findings in order to theorize potential directions of improvement in policy and methods of development for the systems themselves. The overall risk assessment consists of failure analysis for collisions which assist in predicting the scope of risk in order to inform mitigation techniques. By identifying the scale of progress needed in order for these systems to function in a way that protects both the environment and human lives, the opposing sides to either implementing or banning/limiting the use of autonomous weapons is assessed.

Results

The regards to the development of unmanned systems, the largest potential categories of fault can be summarized through Shama Ams:

> These systems carry the risk of algorithmic bias due to flaws in underlying training data and its interpretation, difficulty in maintaining meaningful human control, the potential for more conflict due to fewer barriers to military engagement, and uncertainty in accountability for machine error (2021, para. 1).

Although broad, these four key issues each contain subsets of specific problems that must amount as a result of the unencumbered development of unmanned systems. Without regulation, these issues quickly amass making it more difficult to place limitations on this technology. This rapidly growing ethical dilemma is demonstrated in Figure 1 below, showing just a small portion of highly debated roadblocks stemming from autonomous technology.

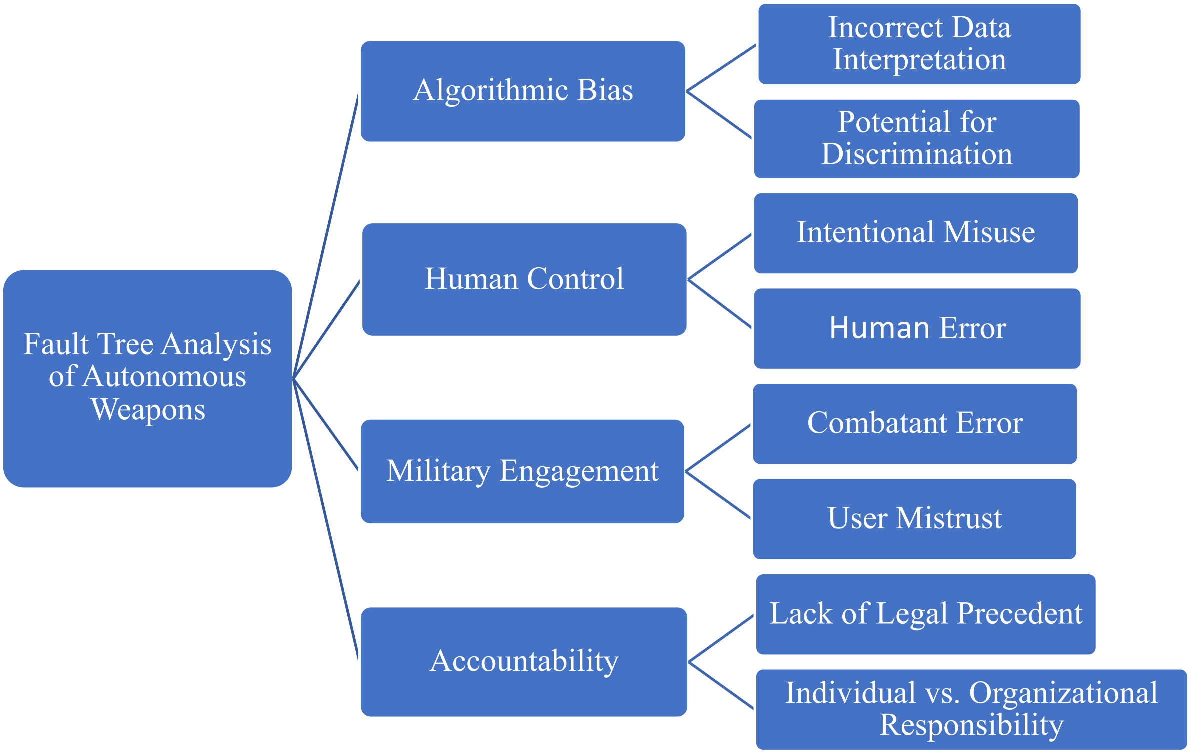
Fault Tree Analysis of Autonomous Weapons
Algorithmic Bias
Incorrect Data Interpretation
Potential for Discrimination
Human Control
Intentional Misuse
Human Error
Military Engagement
Combatant Error
User Mistrust
Accountability
Lack of Legal Precedent
Individual vs. Organizational Responsibility

Figure 1: Fault Tree Analysis of the Risk Factors Regarding Unmanned Weapons. Each potential risk category of unmanned weapons contains a multitude of smaller potential risk factors, demonstrating the numerous ethical dilemmas associated with autonomous weapon systems (Ames, 2021).

The potential fault pathways can be seen above in Figure 1, which provide a basis for the understanding of the risk scale associated with unmanned systems. Although this list is not exhaustive, it contains the most significant failure modes, and provides an accurate basis for the assessment of this risk analysis case study. When analyzing the fault-tree, the same iterative process can be repeated for each branch, but for the sake of this analysis only a few of these paths will be explored.

One of the primary fault modes involves a lack of accountability that can be traced back to the lack of legal precedent and regulation. Currently, these systems lack an "ethical benchmark" which would "establish rules for armed combat," giving these systems no ceiling for innovation and use (Zacharias, Schmitt, 2021, pp. 2). This "lack of a coherent regulatory regime" creates scenarios for legal uncertainty, making it difficult to hold any individual or organization accountable in the case of human misuse or system error (Hartmann et al., 2022, p. 2). So, the apparent solution comes through government regulation of the development of these systems, however therein lies the conflict. Not only is the government the sole organization that can regulate this technology, but they are also the largest beneficiaries and benefactor of autonomous weapon systems. This conflict of interest, alongside the public concern of "government transparency" poses the largest threat to the development of these systems (Pohle & Audenhove, 2017, p. 3).

Another example is demonstrated regarding algorithmic bias within artificially intelligent software. Both the design and function of artificially intelligent systems rely on the "vital role humans play in molding the constraints and implications" in order to mimic and serve a human entity (Gavili, 2022). However, these systems can inherit the implicit biases and prejudices of their creators, in addition to learned bias from the training data. Given that artificial intelligence is primarily for the imitation of human intelligence, it is essentially impossible to create an AI system that is without bias towards various individuals or groups. This impact may be negligible for more objective tasks performed by AI systems; however, the consequences are potentially fatal when using this software in combat weapons. When using an autonomous weapon for the sole purpose of detecting and eliminating targets, algorithmic bias could potentially contribute to the misidentification of a target versus a civilian.

Each subsection of the fault-tree poses a new angle to qualitatively identify the largest areas of risk in order to inform policy regarding the breadth of function that autonomous systems can have if fully implemented into combat. This risk analysis serves as a basis for understanding the potential consequences to prematurely approving autonomous systems that contain a multitude of potential failure modes.

MODELING THE NETWORK OF AUTONOMOUS WEAPON TECHNOLOGY DEVELOPMENT THROUGH ACTOR-NETWORK THEORY (ANT)

In order to study the implications of autonomous technology, Shi and Zheng suggest making joint research between basic theory and the technology of intelligence the primary goal (2006, p. 811). This would be accomplished through studying the relationship between these systems, being AI, and the end goal, which is replicating human intelligence artificially. In order for this to be accomplished, a sociotechnical system must be developed, which can be seen below in Figure 2, in order to form a more holistic view of these systems.

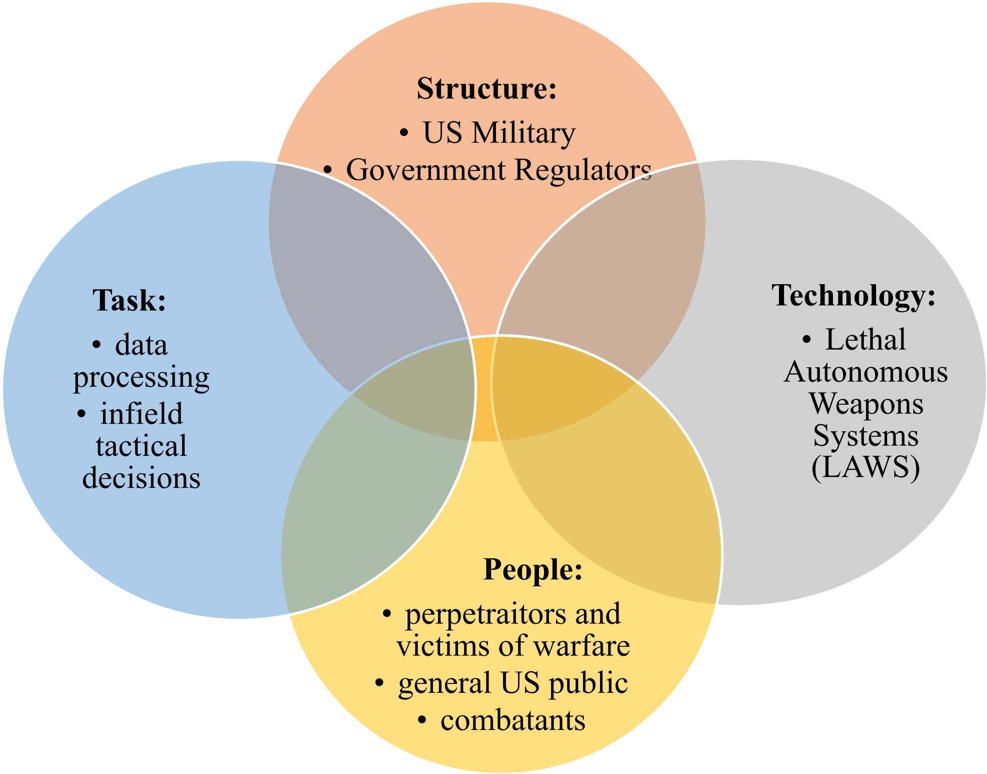

Figure 2: Autonomous Weapons Sociotechnical Model. The development of autonomous weapon systems heavily relies on the interaction between not only the technology itself but also the contributors and environment (Price, 2022).

This model relies on the relationship between four key components: the structure, technology, task and people. The sociotechnical framework provides a system for "modelling and analyzing complex systems" through "humans applying technology to perform work through a process within a social structure" (Oosthuizen & Pretorius, 2016, p. 17). In this case, the technology , which is LAWS, grows and forms within a multifaceted system that consists of both the controlling structure and the people in addition to the task. Within this model, the controlling structure is made up of key users and developers, which consists of both government regulators who influence the trajectory of innovation and the US military, who are the primary users of weapon technology. Closely related are the people, who in the case of the military are the combatants who make up the primary users of autonomous weapon technology. In opposition are the victims of these complex weapons systems in addition to the US public who will be informed on the outcomes that the introduction of such weapons causes. Finally, there is the influence of the task, which consists of the design specifications of the autonomous system such as target identification, data processing, and data analysis. In order for this system to develop in tandem into a fully functioning basis for LAWS, there needs to be a reliable sense of sociotechnical trust, which stems from an agreement between the actor's models and the actor's trust of the architecture of the system (Paja et al., 2013, p. 342). This means that each of these actors/components of the model have to work equally, without one component drawing too much of the development responsibility. The weight and function of each of these actors can be seen in Figure 3, demonstrating how each component influences the other.

The social context for the usage of the sociotechnical model seen in Figure 2 is demonstrated through Actor-Network Theory (ANT), which can be seen below in Figure 3. Actor-Network Theory provides a larger context for the development of complex technical systems and their corresponding interactions with humans and society (Crawford, 2020). Further emphasizing the usefulness of the sociotechnical model, ANT highlights the complexity of the larger working system associated with LAWS. This framework conceptualizes the mutual shaping that occurs between the different actors within the network. In terms of LAWS, figure 3 on the following page provides a visualization of the four larger societal aspects that play a role in the network (seen in orange) in addition to the various networks within each larger actor (seen in light blue, green, and red). These four broad actors within the network are: the users, designers, policymakers and industry development. Within these actors there are interconnected networks that influence each other, which can be seen through examples of engineers who not only make up the primary designers of LAWS, but also push innovation in Machine Learning which influences the development of that field. Each of these factors directly influence not only the development of autonomous combat but each other, demonstrating that there is “no longer separation between science and society, as various social actors can influence the course of science and technology” (Crawford, 2020). The development of models such as ANT and the sociotechnical model provide a basis for the development of legislation in order to regulate and provide a scope for the usage of LAWS.

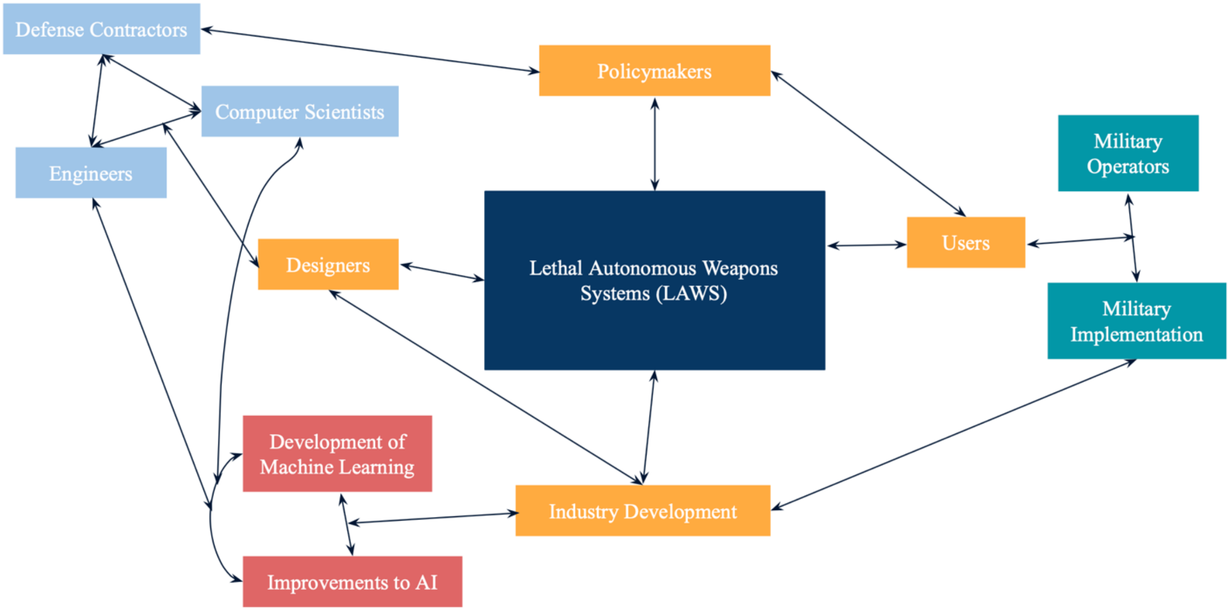

Figure 3: Autonomous Weapons Actor-Network Theory Model. The social context of autonomous weapon systems is provided through the lens of an interconnected web through various actors ranging from human to technical (Price, 2022).

FUTURE IMPLICATIONS OF AUTONOMOUS SYSTEM INTEGRATION

The goal of this research into the implications of introducing autonomous technology into warfare is to curb the potential negative consequences of using this technology that result from system malfunction and user misuse. Before analyzing potential failures of such technology, this research first identified the prevalence of addressing these ethical concerns due to the rising demand for by the United States government to fund projects involving artificial intelligence in combat weapons. In parallel with the increase in funding is the lack of legal precedent in relation to not only the research and development of LAWS, but also their function and capabilities when put into practice. From the initial background of identifying prevalence and gaps in policy, this research paper identified the significant modes of failure through a risk analysis case study involving the usage of fault-tree analysis. Through acknowledging the potential ethical conflicts that arise from this technology, there is an opportunity to develop alternative pathways of innovation in order to avoid these negative consequences.

In this paper, a qualitative analysis gave insight into the scope of influence in order to inspire future research into areas of fault within each step in the process from initial research and design to the integration of autonomous weapons into military environments. This analysis was further built-on through the use of Actor-Network Theory in order to better understand the various factors that influence each mode of failure identified in the fault-tree analysis.

With regards to future implications, there is still a growing need to further explore the breadth of influence that the failure of these systems can have on not only humans but the environment. Although not done in this research paper, the modes of failure can be analyzed through statistical modeling in order to quantify the extent of influence. The intent of conducting quantitative research is to provide a more specific set of regulations to the development of autonomous systems, given that the current legal regulations are vague and subjective. The integration of a quantitative analysis alongside future assessments of ethics and risk can in combination influence future policy in the space of artificial intelligence, even outside of the scope of combat weaponry.

REFERENCES

Ams, S. (n.d.). Blurred lines: The convergence of military and civilian uses of AI & data use and its impact on liberal democracy. *International Politics*. https://doi.org/10.1057/s41311-021-00351-y

Cao, Z. (2017). *Development and Application of Artificial Intelligence* (Z. You, Ed.; WOS:000426730000126; Vol. 70, pp. 701–704).

Crawford, T. Actor-Network Theory. Oxford Research Encyclopedia of Literature. https://oxfordre.com/literature/view/10.1093/acrefore/9780190201098.001.0001/acrefore-9780190201098-e-965.

Di Corpo, R. (2021). Autonomous Technology and the ethics of Non-Power. *Peace Review*, *33*(2), 256–262. https://doi.org/10.1080/10402659.2021.1998858

Fakhreddine, A., Bettstetter, C., Hayat, S., Muzaffar, R., & Emini, D. (2019). Handover challenges for cellular-connected drones (p. 14). https://doi.org/10.1145/3325421.3329770

Fiott, D. (2018). America first, third offset second? *Rusi Journal*, *163*(4), 40–48. https://doi.org/10.1080/03071847.2018.1529893

Fiott, D. (2017). A revolution too far? US defence innovation, europe and NATO’s military-technological gap. *JOURNAL OF STRATEGIC STUDIES*, *40*(3), 417–437. https://doi.org/10.1080/01402390.2016.1176565

Gavili, A. (2022). Towards Semantic Search in Building Metadata: A System Exploration; The Impact of Cognitive Bias and Algorithmic Formalism in Enabling the Creation of Discriminatory AI Technologies [University of Virginia]. https://doi.org/10.18130/F5QF-9277

G. Badakis, M. Koutsoubelias and S. Lalis, "Robust precision landing for autonomous drones combining vision-based and infrared sensors," 2021 IEEE Sensors Applications Symposium (SAS), 2021, pp. 1-6, doi: 10.1109/SAS51076.2021.9530091.

Goose, S. D., & Wareham, M. (2016). The growing international movement against killer robots. Harvard International Review, 37(4). Master FILE Premier. https://proxy01.its.virginia.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=f5h&AN=119802555

Hartmann, J., Jueptner, E., Matalonga, S., Riordan, J., & White, S. (2022). Artificial intelligence, autonomous drones and legal uncertainties. *European Journal of Risk Regulation,* 1-18.

Horowitz, M. C. (2014). Coming next in military tech. Bulletin of the Atomic Scientists, 70(1). Academic Search Complete. https://doi.org/10.1177/0096340213516743

Jaeger, K., & Bers, K.-H. (2001). image-based 3D scene analysis for navigation of autonomous airborne systems. *SPIE Proceedings*. https://doi.org/10.1117/12.444197

Khan, A. (2021). Risk assessment, prediction, and avoidance of collision in autonomous drones (arXiv:2108.12770). arXiv. http://arxiv.org/abs/2108.12770

K. Zacharias and K. Schmitt, "Canada's policy approach to "killer robots" and the ethics of autonomous weapons systems," 2021 IEEE International Symposium on Technology and Society (ISTAS), 2021, pp. 1-4.

Melancon, A.-A. (2020). What's wrong with drones? automatization and target selection. *Robotics, Autonomous Systems and Contemporary International Security*, 111–131. https://doi.org/10.4324/9781003109150-6

Oosthuizen, R., & Pretorius, L. (2016). Assessing the Impact of New Technology on Complex Sociotechnical Systems. *South African Journal of Industrial Engineering*, *27*(2), 15–29. https://doi.org/10.7166/27-2-1144

Paja, E., Chopra, A. K., & Giorgini, P. (2013). Trust-based specification of sociotechnical systems. *Data & Knowledge Engineering*, *87*, 339–353. https://doi.org/10.1016/j.datak.2012.12.005

Pew Research Center. (2020, May 30). *Beyond distrust: How Americans view their government*. Pew Research Center - U.S. Politics & Policy. Retrieved from https://www.pewresearch.org/politics/2015/11/23/beyond-distrust-how-americans-view-their-government/

Pohle, J., & Audenhove, L. V. (2017). Post-snowden internet policy: Between public outrage, resistance and policy change. *Media and Communication*, *5*(1), 1–6. https://doi.org/10.17645/mac.v5i1.932

Sauer, F. (2016). Stopping "Killer Robots": why now is the time to ban autonomous weapons systems. Arms Control Today, 46(8). Military & Government Collection. https://proxy01.its.virginia.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=mth&AN=119075691

Sayler, K. M. (2019). Defense primer: U.S. policy on Lethal Autonomous Weapon Systems (Internet materials; In Focus (Library of Congress. Congressional Research Service)). Congressional Research Service. https://purl.fdlp.gov/GPO/gpo134420

Shi, Z., & Zheng, N. (2006). Progress and challenge of artificial intelligence. *Journal of Computer Science And Technology*, *21*(5), 810–822. https://doi.org/10.1007/s11390-006-0810-5

Work, B. (n.d.). *Remarks by deputy secretary work on third offset strategy*. U.S. Department of Defense. Retrieved from https://www.defense.gov/News/Speeches/Speech/Article/753482/remarks-by-deputy-secretary-work-on-third-offset-strategy/