ESSAI

Volume 15 Article 27

Spring 2017

Command and Control: A Moral Framework for the Development of Artificial Intelligence

Tiffany Montgomery
College of DuPage

Command and Control: A Moral Framework for the Development of Artificial Intelligence

by Tiffany Montgomery

(Philosophy 1840)

I. Introduction

Militaries around the world currently depend on robotics to give them an advantage on the battlefield. On the border of North and South Korea robots are used to maintain and defend against the extreme conflict between the two countries. One South Korean company, DoDAAM, has taken advantage of the near constant fighting on the border to manufacture the Super aEgis 2, an automated gun turret capable of locking onto and engaging targets from nearly two miles away (Blain). China, Russia, and Israel are also key players in robotic warfare, with all three countries suspected to be developing robotic weapons of their own (Muller 4). The United States Department of Defense (DoD) believes that robotics will be a key piece in the war of the future stating “autonomy in unmanned systems will be critical to future conflicts that will be fought and won with technology” in their Unmanned Systems Integrated Roadmap (US Department of Defense). This roadmap also mentions the use of autonomous systems and software to make use of the vast amounts of data that robots collect through sensors and other electronics on the battlefield.

Autonomous systems are also being integrated into other parts of society, including the development of autonomous cars, the algorithms that control the news people get on their social media timelines, and the robots that clean their floors. For example, algorithmic trading is very attractive to Wall Street due to its speed, but it has some associated drawbacks. One day in 2010, the stock market dropped 9% on account of a breakdown in algorithmic trading. In his talk, “How Algorithms Shape Our World”, Kevin Slavin discusses the implications of Black Box or Algo (Algorithmic) Trading. The Flash Crash of 2010 lasted approximately 36 minutes but it will take years before there will be a consensus on what happened that day. Because much of the stock market consists of algorithms that order trades without any human supervision or any ability to adapt to turmoil, our economy has become vulnerable to dramatic price swings as algorithms fight over how to deal with the sudden change.

Humans increasingly rely on autonomous systems that do not require any conscious effort from an operator. The prospect of losing control have even more frightening consequences in a physical space. For example, Section 1.4.3 of the U.S DoD “Unmanned Systems Integrated Roadmap” provides an operational vignette to illustrate the DoD’s vision for the use of unmanned systems in the year 2020. The vignette involves ground, aerial, and underwater unmanned systems working together to investigate and report activity regarding the possible creation of a weapon of mass destruction in the fictional country of Norachi. The unmanned systems in this story are networked together to allow them to quickly make decisions about how to execute a mission independent of a human operator. What happens when these systems have their own flash crash? How can people engineer a stop button for robots falling out of the sky? How can people prevent a bullet from shooting once an automated gun turret has already pulled the trigger?

II. Thesis

The more reliant human beings are on autonomous systems, the more vulnerable they become to the systems shaping our world without their consent. The advancement of autonomous systems jeopardizes human autonomy by replacing conscious thought and fostering an overreliance on autonomous systems. The novelty of the threat that autonomous systems pose to humanity's current way of life requires a new way to think about how and why humans create autonomous systems. In this paper I argue that the development of autonomous systems require that a moral framework include control as a value. Developers of autonomous systems ought to create solutions to maintaining control over autonomous systems in parallel to efforts to increase the degree of autonomy a system has. While many of the current moral frameworks that surround the use of technology revolve solely around their technological capability and any physical harm they may cause, this moral framework seeks to place value on human consciousness and moral judgement over pure calculations. Currently, there is no evidence that it is possible to create a human-level artificial intelligence. Regardless, any creation of a human-level artificial intelligence would still not be human and would not be able to parallel the moral judgements of a human. An autonomous system that is capable of making true moral judgements risks losing its utility should the actions of such an artificial intelligence not align with the original goals that its creators has for its behavior.

III. Elements of a Possible Framework

Much of the fear associated with automated systems comes from a lack of understanding of the varying degrees of control that an autonomous system has over its assigned task. The effects that an autonomous system has on the ability of a human to control a situation depends on both the level of autonomy and how the autonomous system is used. ALFUS, a framework for Autonomy Levels for Unmanned Systems attempts to unify various definitions, classifies autonomous systems, and provides a method of modelling the risk associated with the rise of an autonomous system. In its first workshop, held on July 18, 2003 at National Institute of Standards and Technology (NIST), the Ad Hoc ALFUS Working Group determined that its objectives were to define metrics for autonomy and to develop a framework for autonomy levels for unmanned systems (ALFUS). ALFUS defines the level of autonomy as the ability of an unmanned system to sense, perceive, analyze, communicate, plan, make decisions, and act to achieve its goals as assigned independent of any human interaction. ALFUS then defines three markers of an unmanned systems' autonomous capabilities: human independence, mission complexity, and environmental complexity. The ALFUS framework uses these metrics to then allow an operator to evaluate the risk in using an autonomous system for any given task, stating that "When the unmanned system's contextual autonomous capability is equal to or higher than the required levels, lowered risk can be anticipated, although risk may exist even when all the requirements are all fulfilled." (ALFUS) The use of this framework then not only requires a knowledge of the capabilities of the unmanned system but an assessment of the complexity of the task at hand and clear requirements from which you can evaluate the potential performance of the task by an automated system.

As the requirements of a task become more complex, an autonomous system will be required to have more generalized knowledge and autonomy over itself. Ezio Di Nucci and Fillippo Santoni de Sio claim in *Who's Afraid of Robots: Fear of Automation and Direct Control* that an AI having control over itself, another agent, or the task presented to it is not the cause for concern. In fact according to Nucci and de Sio, delegation of tasks is a marker of progress for humanity. Di Nucci and de Sio, make a distinction between the capacity to control and the activity of controlling. The act of controlling is having any direct influence on some task. As a human becomes more proficient in a task, they do not have to make an effort to execute the task, such as riding a bike or reading a paper. Di Nucci and de Sio argue that this proficiency extends beyond the body. When autonomous systems are used as an extension of humans, they exemplify our ability to not only understand a task but "teach" another agent how to do it. Whereas the capacity to control reflects one's ability to have control over itself and its motive, consciousness is not required for *controlling*, only proficiency of the task at hand. Having the capacity to control however requires a moral judgement of what tasks are valuable enough to be made. For example, when a soldier is in control of a non-automated weapon it is because they are conscious of what they are doing and have the ability to stop the weapon. They execute the activity of controlling a bullet when they fire the weapon. A soldier does not have the capacity to control the war however, since one soldier cannot begin or end a war.

There is no guarantee that a high-level artificial intelligence will continue to abide by the wishes of the operator. An autonomous system does not have to be conscious in order for disobedience to occur. The more complex the tasks we allow AI to complete become, the more general their programming must become. Human beings must trust a high-level AI to make complex decisions on its own while upholding any values associated with the task at hand.

In *The Human Condition*, philosopher Hannah Arendt concerns herself with how humanity's growing ability to control itself and our environment through scientific instrumentation could have an adverse effect on how we view our world. Galileo Galilei, famed astronomer, played a major role in the scientific revolution of the seventeenth century by using the telescope to observe the planets and confirm that the Earth gravitates around the Sun. Usually the acceptance of the heliocentric view of the solar system gains attention, but it is the use of the telescope that captures the mind of Arendt. Using the telescope, Galilei was able to take on the Archimedean perspective—or at least a point similar to it—a point Archimedes hoped would allow human beings to view the world in relation to all other things, while simultaneously untethering ourselves from the Earth in an effort to not obscure our view of the world. The telescope proved that the desire to reach for the Archimedean point comes at a cost. The failure of our senses to accurately perceive the world and our ability to view the world from the Archimedean point go hand-in-hand. Arendt claims that exploration in physics from this point on requires taking the perspective of the Archimedean point (263). We are still bound to the Earth, however, and any changes that we make to the Earth from the Archimedean Point will still affect our well-being. Just as Arendt is concerned with the use of data gathered through scientific instrumentation, this paper concerns itself with the "data" gathered through and manipulated by *artificially* intelligent machines. The discovery of the Archimedean point was a source of great despair or part of the scientists and mathematicians during the 17th century, the paradox of being able to fully observe the world without being consciously present still affects how we think about technology today. When we begin using artificial intelligence as a tool we alienate ourselves not only from all sense experience, but also conscious thought. In summary, taking Arendt, Di Nucci and de Sio, and ALFUS into account, it is important to be aware of an autonomous system's purpose, make sure that it can be trusted to perform its own task and not have control over itself or any influence on the larger goal. It is required that a human being's sense experience is incorporated into these decisions of how the autonomous system is used and that we are constantly monitoring it.

IV. Case Study: Warfare

Much of the debate around whether we can trust robots in warfare has revolved around the use of lethal autonomous weapons systems (LAWS), robots that can identify and attack targets without any direct human supervision. The DoD has listed three driving factors for producing unmanned systems (including LAWS): 1) the success of unmanned systems during combat operations in Southwest Asia, 2) the need for cost effectiveness caused by a predicted reduction in the military budget, and 3) a shift of national security to the Asia-Pacific theater of war many require that unmanned systems replace human soldiers where freedom to operate is contested. Other advantages of LAWS include the reduction of risk to human soldiers, increase in the ability to prosecute war crimes, increase in accuracy, precision, and speed, and the ability to operate in places that are dangerous for soldiers to fight in.

As of April 15th, 2016, there have been three informal meetings of artificial intelligence, ethics, and policy experts on LAWS hosted by the United Nations Convention on Certain Conventional Weapons (CCCW) (The United Nations). The meetings have consisted of a discussion on whether fully autonomous weapons systems truly exist, a need for a clear definition of LAWS, and the ethical concerns that the presence of such systems present. While there is doubt whether fully autonomous weapons systems exist, semi-autonomous weapons systems currently are in use and in development. The super aEgis 2 is currently being used in many locations in the Middle East, including three airbases in the United Arab Emirates, the Royal Palace in Abu Dhabi, an armory in Qatar and numerous other unspecified airports, power plants, pipelines and military air bases elsewhere in the world (Parkin). According to DoDAAM's CEO Myung Kwang Chang, "Need is the mother of invention," and it is expected that both North and South Korea will want to use similar gun turrets in the demilitarized zone on the border of both countries. The United States Joint Forces Command's Project Alpha has been working on the development of Tactical Autonomous Combatants, robots that are meant to be largely autonomous with little human supervision. Gordon Johnson, the Unmanned Effects Team leader for Project Alpha, claims that these robots will be able to save human soldiers from having to go into chemically, biologically, or radiologically contaminated environments. Semi-autonomous LAWS have even been used for domestic conflict. Texas police officers have also made use of an autonomous robot sniper after the shooting of Dallas police officers took advantage of the opportunity in 2015. Dallas police Chief David Brown said, "the robot was used in the building where Johnson was still sniping, intent on continuing to kill." (CNN).

Despite all these advantages, the development of these systems has not gone without opposition. The majority of the ethical debates surrounding military robotics have been in favor of a preemptive ban on the development and wartime use of LAWS. More than one-thousand leaders and experts in the field of Artificial Intelligence signed an open letter urging the United Nations to ban the development and use of autonomous weapons to be announced at the opening of the International Joint Conference on Artificial Intelligence in Buenos Aires, Argentina, July 28, 2015 (Del Prado). AI researchers fear that the proliferation of LAWS could discourage research in the field and inhibit the benefits that could come from developing non-lethal artificially intelligent systems. The letter also warns that LAWS have the potential to become the "Kalashnikovs of tomorrow"—in reference to the designer of the popular AK machine gun—and notes that LAWS have been described as the third revolution in warfare after gunpowder and nuclear arms. The writers of the letter fear that the lost-cost of materials required to mass-produce these weapons will lead to an international arms race. The Stop Killer Robots Campaign shares the same fear of such a slippery-slope. The organization consists of five international NGOs, a regional NGO Network, and four national NGOs, that believe that the ease of robot warfare could lead to a lack of constraint by army officials and could make the decision to go to war more feasible, putting a greater number of civilians at risk in the long term. People supporting the Stop Killer Robots Campaign also argue that the inability of LAWS to reason and make decisions like humans do presents an ethical concern in itself (Stop Killer Robots).

V. Analysis

In asking whether or not the use of LAWS is morally permissible, we utilize the elements of a possible moral framework outlined in the former half of this paper to evaluate the trustworthiness of LAWS. First we must define the task, by definition LAWS are created to be used as lethal weapons. The weapons under the scope of this paper are meant to be used during battles of war. As a weapon to be used during war, LAWS must follow the rules laid out by International Humanitarian Law (IHL). A large part of IHL is contained in the four Geneva Conventions of 1949. Since then there have been other conventions called to prohibit the use of weapons such as chemical weapons or land mines and to extend protections of noncombatants. The restrictions placed on weapons of war mirror the principles of just war. Just war theory aims to provide an ethical framework for deciding when to go to war (*jus ad bellum*) and how to fight during war (*jus in bello*). The use of LAWS can be evaluated under "*jus in bello*" according to three principles: necessity (whether it is appropriate to take an action), proportionality (making sure that the action taken is suited to the reason), and discrimination (being able to distinguish between combatants and noncombatants). After defining the task, we answer the question of whether we can trust LAWS to uphold the principles of IHL. In *Robots, Trust, and War*, Muller defines predictive trust as reliance on an agent to do some job like a mountaineer trusting a rope not to break.

So far much of the effort of those who want to support a ban on the development of LAWS has centered around forming arguments against LAWS' technological capabilities to uphold the principles of just war and IHL. These analyses can be used to determine if LAWS are deserving of predictive trust. Noel E. Sharkey a professor of Artificial Intelligence and Robotics and a professor of Public Engagement in the Department of Computer Science at the University of Sheffield, UK, addresses the principles of proportionality and distinction in his paper "The evitability of autonomous robot warfare". First, Sharkey notes that there are not currently LAWS in use that can separate civilians from combatants (especially in insurgent warfare) or that can recognize combatants that are wounded or surrendering, thereby unsatisfying the principle of distinction. Sharkey explains that current sensing technologies can not can tell humans from one another, nor can a programmer ever be able to give the robot the proper instructions on how to tell civilians apart since there is no clear definition of what a civilian is (787-99).
Machine learning, can be defined as follows:

> "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P improves with experience E." (Mitchell 2)

In other words, if a program is fed a lot of data, and asked to do a task with that data, one can tell that it "learned" if it gets better at that task. The data that is fed through the algorithm is broken down into features and labels that describe what the data means. In a visual recognition system, an image is treated as one example of a solution to the task of recognizing something in the image. Its features consist of all the values of each pixel it contains. After "learning" about the pixel representations of visual data a computer vision algorithm can then predict if any of the visuals it has learned about will be in another image. Because the number of features of an image are very large, it is common to use neural networks, a network of the usual calculations that would be made in a machine learning problem, to compute on this data. This software becomes even more complicated when developers use deep learning (the layering of many neural networks) to train their software to complete a task. (Knight). Despite its efficiency, the mathematical complexity of computer vision algorithms makes it difficult to work backwards and understand why an algorithm made the prediction it did. As machine learning algorithms become more common, programmers must be able to design software that can provide many different reasoning metrics about itself, instead of just the amount of confidence an algorithm has in its prediction. (Knight)

Because of the nature of machine learning algorithms, LAWS will also have a hard time understanding context and being able to associate any meaning with the calculations it makes. While a human can easily tell whether an object in the road is a piece of trash or an animal, LAWS might not have such an easy time telling the difference if they learn that small animals and trash are around the same size and color. Even when the algorithms work correctly, issues can arise if they are applied poorly. In *Morality and Machines: Perspectives on Computer Ethics*, Stacey Edgar highlights three types of human error: syntax error (a programmer's code does not follow the rules of the language that they are using), logic errors (when there are no syntax errors but the code still doesn't perform the task you want it to), and application errors (when there are no syntax or logic errors, but your code still outputs incorrect results because your program model was not being used in the correct way). Edgar notes that these errors can be remedied by using program verification and extensive testing. In the case of autonomous vehicles which use similar technology as LAWS, according to Fortune, thorough testing of computer vision systems could take hundreds of years. The RAND research project notes that for at least fatalities and injuries during car accidents, it is impossible for autonomous vehicles to provide enough information through test driving (Knight). There is no reason to believe that a similar implementation on LAWS would yield different results.

Even if LAWS had the ability to discern between different actors in the theater of war, the robots themselves still lack the common sense needed to make decisions about what actions to take after they have made such a distinction. The second conflict with international humanitarian law is that of "situational awareness or agency to make proportionality decisions". Sharkey concedes that LAWS might be better at solving the easy proportionality decisions of choosing and deploying munitions that minimize collateral damage, but LAWS cannot solve the hard proportionality problem of being able to decide whether it is necessary to apply force in the first place. This decision is a moral one that requires intuition that inanimate object do not possess. While a human soldier's intuition is of course not always reliable, it is the understanding of the soldiers' actions and our ability to communicate that warrants one's forgiveness of any wrongdoing. Knowing that one will be able to understand and communicate changes to any potential wrongdoings instills normative trust. Overreliance on learning algorithms may cause soldiers to forget that examples of human behavior are needed to guide the program in the first place. Where current artificial intelligence lacks in judgement, it is the experience of a human being that will fill in any moral gap.

Muller also argues against Sharkey in claiming that there are LAWS that can also satisfy the principle of proportionality, by targeting other weapons systems and not human beings (6). Muller adds to this claim by noting that there are situations in which a robot could determine that no collateral damage was likely. Then under the principles of IHL, it would be morally permissible for LAWS to act. Muller also defines normative trust as reliance on an agent after they have taken the needs of trustor into account. Muller spends the latter half of his paper analyzing the ability of LAWS to be trustworthy to civilian conflict during asymmetric war. While not every war LAWS can be used for will be asymmetric, it is useful to apply Muller's arguments to trust between soldiers, commanding officers and LAWS. This paper has established that there is reason to believe that current LAWS are not reliable. Even if LAWS were advanced enough to become reliable weapons and earn the predictive trust of their fellow combatants, it is impossible for LAWS to have any human in mind, because it has no mind at all—just calculations. The responsibility of acknowledging how one's actions may affect those around them before taking the action may be shifted to any human being that is responsible for the operation of the LAWS, there is debate on who or what that may be.

In *"Killer Robots"*, Robert Sparrow, Professor of Philosophy at Monash University in Australia, considers three possible parties that could be held responsible for war crimes: the programmer of the LAWS, the commanding officer or operator of the LAWS, and the LAWS itself. First, he rules out the programmer since any error in the system can put in a warning by the manufacturer, shifting the blame to the commanding officer that chose to use the system despite its faults. A programmer also should not be held responsible for any transgressions on behalf of the LAWS because a successful automation engineer would be required to create an entity that can adapt to the unpredictable theaters of war in a way that was unpredictable to both the programmer and the commanding officer. Not designing LAWS to this level of automation forgoes many of the utilities of having a robot fill in for human soldiers in more dangerous environments. Sparrow defines full autonomy and moral responsibility as going hand in hand. Sparrow writes, "to say of an agent that they are autonomous is to say that their actions originate in them and reflect their ends" (Sparrow

65). Because such an agent would be the only one capable of foreseeing and preventing its own actions, it would be the only one that could fairly be held responsible for itself. Sparrow extends this reasoning to the commanding officer as well. Even if the commanding officer was aware of all the risks of deploying such an agent as told by the manufacturer beforehand, there are still risks that the manufacturer and therefore the commanding officer could not foresee. Finally, Sparrow rules out the robot itself, because it does not have the capability to be punished or held accountable for its actions since it, by Sparrow's definition, is not capable of feeling emotions and responding to pain like humans do (69-73).

Sparrow creates an analogy between the robot and child soldiers. A child soldier possesses a similar autonomy to the type of LAWS that Sparrow describes since they can make unpredictable choices that satisfy ends that begin and end in themselves, but they are still not mature enough for anyone to reasonably trust them with the task of choosing when to take another human's life. Sparrow writes, "It is the prospect of intelligent actors without any moral responsibility that makes child soldiers especially terrifying." (74) Sparrow argues that no one ought to be held responsible in the case of the child soldier since their parents, their commanding officer, nor themselves held enough prior knowledge to stay in control of the child soldier. In *Just War and Killer Robots, Simpson and Muller* offer a counter argument by extending this analogy to the example of Joseph Kony's use of child soldiers in the 'Lord's Resistance Army' operating in northern Uganda, the D. R. Congo and South Sudan. Simpson and Muller argue that it is Joseph Kony who is clearly responsible for the actions of the child soldiers in the Lord's Resistance Army. Simpson and Muller claim that Kony is, in fact, responsible for any war crime committed by these child soldiers because the incapability of children to understand the complicated ethics of war is common knowledge. It is the decision to use child soldiers violates the principles of just war, not the actions of the child soldiers themselves. (5)

If we accept child soldiers to exemplify Simpson and Muller's idea of near-autonomous robots, then we can also say that the robot would not be able to make decisions about who to target during a war because "they lack the technical and moral competence" to use their artillery appropriately. Therefore, it is possible to hold commanding officers responsible for using weapons that are known to violate the principles of just war and international humanitarian law. While the previous analogy ignores that LAWS can have varying degrees of autonomy, much of the reasoning still applies to a wide variety of LAWS. Regardless of autonomy level if a combatant chooses to wrongfully use LAWS then any harm caused by the weapon would be a direct result of their action, rendering them responsible. This line of reasoning, however, still supports that LAWS cannot gain predictive trust and should not be used exclusively lest soldiers lose control of our ability to operate on the battlefield.

Any attempts to create a LAWS capable of executing actions that reflect their own ends would be in vain. The advantages of using LAWS in place of human beings on the battlefield is the likelihood of having fewer of one's own soldiers die. A robot that was "conscious" enough to be acceptable to the public would require either the robot itself to mimic human emotion or have human beings feel emotion toward it as well as flexibility to adopt to potentially unpredictable complex contexts. Going back to the example of child soldiers, creating children for the sake of doing labor is slavery. The existence of LAWS of any autonomous system that was worthy of being compared to a child or animal, would prompt more discussion on the rights of such machine, including the right to freedom from the control of a human operator. Such machines might have their own rights to control, jeopardizing the current control humans have over our own autonomy. Such an intelligence would also challenge our control over ourselves and the planet seeing as our superiority over other animals allows us to "rule the Earth" now.

While it is important to consider the presence of a fully autonomous machine, the possibility of humans wrongfully assigning anthropomorphic qualities to LAWS or autonomous systems, is more realistic than the creation of a sentient artificial intelligence. It is a popular belief that a human's brain works like a computer: the senses collecting data and the brain operating on that data and saving memories on a sort of "wet drive", opposed to a computer's hard drive. However, the notion that our brains are complex computers could not be any less true. Research in artificial intelligence began as a way to understand the brain through reverse engineering, however we do not yet understand how many of our brain functions work, let alone having the ability to recreate them with electronics. (Aeon) The result of such a project has only product a reduction of reality: a true *artificial* intelligence lacking context and consciousness. Since the 1940s, artificial intelligence research has shifted from creating a general purpose intelligence to creating algorithms to do very complicated, yet specific tasks. Some of the most infamous examples of this include: the defeat of Chess Grand Master Garry Kasparov by IBMs DeepBlue in 1997 (Wired), the success of IBMs Watson on Jeopardy in 2011 (NYTimes), and the defeat of legendary Go player Lee Se-dol by AlphaGo, a program developed by Google's Deep Mind (The Verge). As AI continues to surpass humans in activities once thought to require human intelligence, the technology continues to gain a significant amount of hype regarding is true abilities.

According to Sharkey, soldiers have already begun to treat robots like animals. Soldiers have been reported taking the robots they work with fishing and letting them hold the fishing rod in their hand. One researcher had to stop testing on a mine clearing robot because he thought letting the robot's legs get blown off after detecting a mine was inhumane. Repeated use of autonomous systems can also cause humans to over trust them. During testing of a driver's capability to stop an autonomous car from getting in an accident most of the drivers weren't paying attention to the road or what the car was doing (Borowsky, Oron-Gilad). Despite the lack of understanding of the capabilities of autonomous vehicles, drivers were still willing to relinquish control.

VI. Conclusion

In conclusion, LAWS are not able to gain predictive or normative trust because there is no weapon currently in existence that its technologically capable of upholding the principles of IHL. Even if we could build such a weapon, combatants risk the weapon losing its utility because the qualities that are required to uphold IHL involve the robot being able to have a significant amount of control over its own functions, increasing the possibility that it can disobey commands from a human operator. By allowing LAWS that cannot uphold IHL, humans risk allowing ourselves to trust LAWS regardless of their technological capability. In the case of LAWS being used during times of war, developers ought to stay away from the creation of LAWS that can select and attack targets without a command from a human. This conclusion does not mean that autonomous systems are not useful in the battlefield. There are still numerous research opportunities in creating LAWS and other autonomous systems that emphasize human-robot interaction and assist a soldier in executing some task. An example includes robots created to gather intel during reconnaissance missions. The need for trustworthiness under the moral framework outlined in this paper creates an opportunity for researchers to focus on robots that can not only gather intel but explain why their intel is important as well.

Works Cited

Arendt, Hannah. *The Human Condition*. Chicago: U of Chicago, 2012. Print.

Blain, Loz. "South Korea's Autonomous Robot Gun Turrets: Deadly from Kilometers Away."*New Atlas*. New Atlas, 7 Dec. 2010. Web. 11 Oct. 2016.

Borowsky, A., and T. Oron-Gilad. "The Effects Of Automation Failure And Secondary Task On Drivers' Ability To Mitigate Hazards In Highly Or Semi-Automated Vehicles." *Advances In Transportation Studies* 1 (2016): 59-70. *Academic Search Complete*. Web. 18 July 2016.

Byford, Sam. "Google's DeepMind Defeats Legendary Go Player Lee Se-dol." *The Verge*. The Verge, 09 Mar. 2016. Web. 16 May 2017.

Di Nucci, Ezio and Santoni de Sio, Filippo, Who's Afraid of Robots? Fear of Automation and the Ideal of Direct Control (May 6, 2014). Forthcoming in: Battaglia F & Weidenfeld N (eds.), Roboethics in Film. RoboLaw Series. Pisa: Pisa University Press. Available at SSRN: https://ssrn.com/abstract=2433433

Edgar, Stacey L. *Morality and Machines: Perspectives on Computer Ethics*. Boston: Jones and Bartlett, 1997. Print.

Epstein, Robert. "Your Brain Does Not Process Information and It Is Not a Computer." *Aeon*. Aeon, 18 May 2016. Web. 16 May 2017.

Finley, Klint. "Did a Computer Bug Help Deep Blue Beat Kasparov?" *Wired*. Conde Nast, 28 Sept. 2012. Web. 16 May 2017.

How Algorithms Shape Our World. By Kevin Slavin. Perf. Kevin Slavin. *TED*. TED, July 2011. Web. 02 Aug. 2016.

Knight, Will. "If a Driverless Car Goes Bad We May Never Know Why." *MIT Technology Review*. MIT Technology Review, 07 July 2016. Web. 29 July 2016.

Lackey, Douglas P. *The Ethics of War and Peace*. Englewood Cliffs, N.J.: Prentice Hall, 1989.Print.

Mitchell, Tom M. *Machine Learning*. New York: McGraw-Hill, 1997. Print.

Muller, Vincent C., and Thomas W. Simpson. "Killer Robots: Regulate, Don't Ban." *Blavatnik School of Government Policy Memo* (2014): 1-4. *Philpapers*. Web. 10 Oct. 2016.

Müller, Vincent C. (2016). Autonomous killer robots are probably good news. In Ezio Di Nucci&Filippo Santonio de Sio (eds.), *Drones and responsibility: Legal, philosophical and sociotechnical perspectives on the use of remotely controlled weapons*. Ashgate 67-81.

Open Letter on Autonomous Weapons - Future of Life Institute." *Future of Life Institute*. N.p.,28 July 2015. Web. 11 Oct. 2016.

Parkin 16 July 2015, Simon. "Killer Robots: The Soldiers That Never Sleep."*BBC*. BBC, 16 July2015. Web. 11 Oct. 2016.

Prado, Guia Marie Del. "Stephen Hawking, Elon Musk, Steve Wozniak and over 1,000 AIResearchers Co-signed an Open Letter to Ban Killer Robots." *Business Insider*. BusinessInsider, Inc, 27 July 2015. Web. 11 Oct. 2016.

"Recommendations to the 2016 Review Conference Submitted by the Chairperson of theInformal Meeting of Experts." (n.d.): n. pag. *The United Nations Office at Geneva*. The United Nations, 2016. Web.

"Review of the 2012 US Policy on Autonomy in Weapons Systems." *Human Rights Watch*.Human Rights Watch, 17 Apr. 2015. Web. 11 Oct. 2016.

Russell, Debra L. F. "Autonomy Levels For Unmanned Systems." *NIST*. N.p., 21 Sept. 2016. Web. 16 May 2017.

Sharkey, Noel E. "The Evitability of Autonomous Robot Warfare." *International Review of the Red Cross* 94.886 (2012): 787-99. Web. 10 Oct. 2016.

Sherman, Erik. "It's Impossible to Find Out If Self-Driving Cars Are Safe: Report." *Fortune*. Fortune, 16 Apr. 2016. Web. 3 Aug. 2016.

Sidner, Sara, and Mallory Simon. "How Robot, Explosives Took out Dallas Sniper." *CNN*. Cable News Network, 12 July 2016. Web. 16 May 2017.

Simpson, Thomas W., and Vincent C. Muller. "Just War and Robot's Killings." *Philosophical Quarterly* (2015): 1-22. *Http://www.sophia.de*. University of Oxford & Anatolia College/ACT, 7 Aug. 2015. Web.

Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy* 24.1 (2007): 62-77. Print.

"The Problem." *Campaign to Stop Killer Robots*. N.p., n.d. Web. 11 Oct. 2016. <http://www.stopkillerrobots.org/>.