LIBERAL ARTS AND HUMAN SCIENCES
CENTER FOR HUMANITIES
VIRGINIA TECH.

Mellon Foundation

Tech for Humanity Case Studies

Automating Death: AI and Battlefield Weapons Decisions

Nowhere is artificial intelligence (AI) more feared than on the battlefield. While AI is feared as a component of modern weapon systems it is simultaneously being increasingly implemented as a feature of modern military systems in most developed nations. The inclusion of AI into weapon systems leads most people to think of the Terminator movies and the rise of some future artificial intelligence in the form of SkyNet that takes over the world and kills people. Yet the implementation of AI in weapon systems is far from the world of SkyNet. At present all AI systems are constrained to what is commonly referred to as narrow AI as opposed general AI. Breaking down these concepts is important to understanding the complexities of AI inclusion in military systems. First, narrow AI is focus on a specific task or limited set of tasks. All present forms of AI through 2021 constitute some form of narrow AI. They are good at a number of different tasks such as image recognition, personalization, chatbots, predictive text, translation and more. Narrow AIs are trained typically trained on large quantities of data and use a variety of statistical methods to derive answers to specific problems. Narrow AIs do not have self-awareness, consciousness, or the ability to think and are therefore unlikely to gain sentience and become killer robots. An example of narrow AI is an image recognition platform that can differentiate between dogs and cats in photographs. While it can differentiate between the two species, this differentiation in and of itself has no meaning and the AI will not be a dog or a cat "person" or AI. By contrast general AI is thought to constitute machines that can process a range of cognitive tasks with little oversight, learn, generalize, apply knowledge in new and abstract manners, and plan for the future. General AI if it comes into existence is expected to possess intelligent qualities such as common sense, creativity, and emotive preferences.

Narrow AI and general AI are often confused by the general public and applications of narrow AI can often resemble general AI in many ways. For instance, Deep Blue was able to defeat Gary Kasparov the world chess champion, and AlphaGo was able to beat world Go champion Lee Seedol. Both of these activities appear to meet many of the requirements associated with general AI, yet upon closer inspection each AI was largely task specific to playing chess or Go. Similarly, many narrow AIs seemingly defeat the well-known Turing test. AIs such as Google Duplex make it increasingly difficult to differentiate between humans and machines in certain text and voice-based interactions. Yet, the success of Google Duplex might be more of a commentary on human communication practices than on the reality of an AI achieving general level capabilities.

Such is the fear of AI enabled weapon systems that there are calls from the United Nations, Human Rights Watch, Amnesty International and others to preemptively ban any weapon system in which the decision to initiate a lethal action is not undertaken by a human operator. Dozens of nations have joined the call for a prohibition against lethal autonomous weapons (LAWS). Yet there are notable exceptions to this list, including the United States (US), China, and the Russian Federation. There is presently an ongoing United Nations Governmental Group of Experts (UN

GGE) devoted to investigating the topic of LAWS that has met annually since 2019 to discuss and establish the guiding principles for the utilization of LAWS. At the core of UN GGE discussions on LAWS are the guiding principles that international humanitarian law applies fully to all weapons systems including the LAWS and establishes grounds for responsibility and accountability of the use of such systems. Establishing basic definitions about LAWS is complex and a point of significant discussion.

Among the nations with no prohibition on the development and use of LAWS is the US. The US Department of Defense Directive (DODD) 3000.09 defines LAWS as "weapon system[s] that, once activated, can select and engage targets without further intervention by a human operator." This definition removes the human from the decision-making loop for such autonomous systems and further expands the category to include systems that are semi-autonomous with a human in the decision loop. Weapons in this category are more commonly known and constitute fire and forget weapons such as air-to-air and surface-to-air missile systems. It is little secret that the US is actively pursuing a range of different LAWS with variable levels of human interaction. These efforts are partly in response to perceived efforts of adversary nations efforts to utilize AI in warfare. Yet despite all the movement forward on research and development DODD 3000.09 regulations on LAWS also states weapon systems should:

> *"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."*

What is clear is that LAWS are going to be part of the future battlefield experience faced by civilians and soldiers alike. This reality is spurring a number of firms to get a head start on the development of LAWS.

One firm, American Innovation Defense (AI-D) - a small defense contracting firm, began in 2016 to read the shift in the global defense market towards increasingly sophisticated weapons systems. AI-D specialized in perimeter defense systems, often with an emphasis on barriers and other physical infrastructures designed to keep soldiers at forward operating bases (FOBs) safe from attacking militants or terrorists in Iraq and Afghanistan. Their primary emphasis was on keeping land-based attacks from breaching the perimeters of FOBs and over-running the often-smaller sized units of 20-100 soldiers on the inside. FOBs confront a number of unique challenges that are often not faced by larger well-established military bases and are often in remote regions that are difficult to reach over land. Resupplying FOBs or surging in additional military personnel in times of crisis is generally not possible. FOBs can frequently form a daisy chain for supply vehicles into increasingly remote areas of a country. In an emergency FOBs can be resupplied by helicopters or parachute supply drops from cargo planes. Yet on most days FOBs are isolated forward operating outposts with little callback support if attacked. Often more remote FOBs can only receive air support after substantial delay of 2-4 hours after an attack has already begun. And this air support in the form of specially equipped attack helicopters is constrained by weather conditions which at many locations can be quite bad, particularly in the winter months.

To address all these problems AI-D initially developed an FOB construction kit that enabled a layered defense system for soldiers that allowed them to seek structural sanctuary and provide cover from inner redoubt. These structural barriers provided soldiers with substantial cover from which to defend their positions from approaching enemy land forces. Soldiers positioned throughout the FOB behind various predesigned structures could minimize exposure to small-arms and Rocket Propelled Grenade (RPG) fire while still maintaining reasonable visibility to counter incoming assaults. Inner redoubts (overlooks) provided cover fire for perimeter positions and provided a place to fall back to in a worst-case scenario.

Initially the AI-D FOB structures were extremely successful in providing a safe and resilient structure in which to base forward positioned soldiers. Yet by 2016 adversary forces were adapting to the AI-D FOB structures and using a variety of rockets, mortar, and unmanned aerial vehicles (UAVs) to target the inner areas of the FOB structures. This shift in technology and tactics was working with deadly effect and making it more difficult for US forces to maintain FOBs in contested areas. This was pressuring US forces from these areas and resulting in the loss of territory to forces that undermine human rights and security local populations. As casualties began to mount AI-D was tasked with providing soldiers with a enhanced security features for their FOBs that could provide protection against incoming low-altitude (30-1,000 feet) munitions.

AI-D immediately set to work on the problem. Initial testing demonstrated that US servicemen were ill-suited to manually shoot down UAVs, rockets and mortars. Initially signal jamming technology and devices such as the DroneDefender were used with moderate success against UAVs but provided little protection against rockets and mortars. Jamming technology also degraded area communications and increased the FOBs electromagnetic spectrum signature making it both more difficult to communicate for help and more visible for automated targeting. Soldiers also complained of the difficulty of identifying UAVs at altitudes higher than a couple hundred feet. When adversary forces discovered the difficultly posed by flying at slightly higher altitudes, they began to fly UAVs higher and drop various munitions including small incendiary devices from just outside of the operational range of the soldiers' defenses.

When none of the initial systems proposed served to remediate the challenges soldiers faced. AI-D turned from non-lethal structural solutions such as barriers and non-lethal electronic warfare solutions such as DroneDefender to an emphasis on lethal means of countering the threats posted. They began work on what they called the Automated Integrated Munitions Response System (AIM-R). AIM-R was based on the US Navy's Phalanx Close-In Weapons System (CIWS). The Phalanx CIWS was designed to protect against small incoming threats including boats, torpedoes, anti-ship missiles, and helicopters. When learning about the Phalanx CIWS system the AI-D engineers and developers began to recognize many of the problems that the system would face in on-land environments where projectiles would have a significantly increased probability of collateral damage.

Engineers noted several principal challenges associated with adaptation of the Phalanx CIWS system to land environments including topographical features such as mountains, hills and buildings, civilian proximity, descending munitions fired from an automated system resulting in collateral damage and target identification challenges within constrained distances. AIM-R engineers needed to rethink the K-band radar and Forward-Looking Infrared (FLIR) target identification systems and algorithms, attitude orientations, and munition types used by the system. Starting with radar identification algorithms they had to adjust the system to differentiate between fast- and slow-moving objects (drones vs. rockets and mortars). They also had to account for friendly incoming aircraft such as attack helicopters, supply plans and friendly drones. To do this they both programmed in friendly aircraft signatures and coded the AIM-R system not to fire at those airframes. They also built in a failsafe that used friendly aircraft transponder signals as a system kill switch that would disable the AIM-R system when friendly aircraft entered within range of the system.

Next the AIM-R engineers and programmers began working on the civilian proximity challenges. To do this they worked with smart munition manufacturers to develop explosive rounds that self-detonated at a pre-determined range and altitude above FOB ground level. Because FOBs are often located at different altitudes the system auto-adjusts itself upon initial set up to set a floor (lower firing and detonation limit) for its munitions. Any munition that might contact land would detonate at a safe distance from any land-based targets. This meant the system could not be used to directly engage incoming adversary forces human or machine. If a land-based assault were in bound with automobiles or other mechanized devices soldiers would have to manually respond with Javelin shoulder launched missiles or small arms fire.

Next because of the varied topography of FOB locations and concerns about unintentionally harming civilians in buildings or on mountain or hillsides upon initial setup the AIM-R used a Laser Detection and Ranging sensor to create a three-dimensional topographical map of the surrounding terrain. Any munition fired from the AIM-R system would automatically alter its detonation time based on speed and range relative to the terrain to prevent detonation anywhere near fixed land-based objects.

In 2018 the system was tested at a mock range. Dummy US soldiers were placed in a mockup of an FOB and fired upon with rockets, mortars, and UAVs. Initial tests began slowly. Single rockets or mortar fire were lobbed at the FOB and the AIM-R system identified these incoming munitions and immediately responded by firing its Vulcan cannon with special explosive rounds. Initial tests against single munitions were all successful with not a single incoming adversary munition entering the FOB. The tempo of incoming adversary munitions was increased and again the AIM-R responded by putting up an veritable wall of explosive rounds to protect the FOB. Combinations of adversary munitions were tried and again not a single adversary munition entered the FOB.

By early 2019 AI-D began testing the AIM-R system to ensure it did not injure land based civilian or adversary targets. It began testing by launching complex volleys of adversary munitions at the FOB while concurrently having Unmanned Ground Vehicles (UGVs) advance on the FOB. With no soldiers inside the FOB the UGVs easily made it to the outside walls of the base. Next the system was tested in areas with different topography to ensure that it did not harm ground-based targets in elevated positions on hills, mountains, or buildings. The system passed all tests. Not a single test resulted in collateral damage. The system even automatically shut down when it identified a friendly signature in its airspace. The system was found to have one flaw. If an advancing adversarial force was able to get within 300 yards of an FOB and launch a low trajectory rocket, mortar, or UAV the system would not target the adversary to prevent collateral damage. The system was tested in all weather conditions, and it was found to be accurate and effective at protecting FOBs against incoming volleys of adversary munitions while simultaneously preventing collateral damage.

Army personnel who participated in the tests noted that from the outside looking in the wall of explosive munitions fired out formed what looked like a blinding mushroom of molten metal that hovered above the about 35 meters off the ground in a half spherical shape over the FOB. Testers who viewed the system from inside the FOB commented that the noise of the system at full activity was deafening and that it formed the inside of a sphere of molten metal and that once accustomed to the sound soldiers reoriented and focused on land-based adversaries. By early 2020 the AIM-R system had passed all DoD safety checks and was being deployed to select forward bases. Soldiers who relied on the system referred to it as the Uncle Sam's halo or Uncle Sam's hat. When firing at full capacity small hot fragments of metal rained down on the FOB necessitating soldiers to ensure that all exposed skin was covered to prevent burns. After implementation not a single soldier was lost to an incoming aerial projectile from a drone, rocket, or mortar. The success of the system in combination with the FOB structures virtually eliminated all challenges fixed position soldiers in bases.

The AIM-R system once turned on was fully automated and selected and responded to all incoming projectiles independent of human intervention. It was governed by a narrow AI algorithm that controlled where it fired and set the range for the detonation of each munition it fired. The system although not directed at humans is by all measures a LAWS. It is governed by rules, but learns and adapts to changing topographical and conflict conditions. The system was not designed and is not intended for use against living adversaries. The system cannot think in the way that a General AI based system might one day think but is able to adapt and adjust in real-time. They system has had a net positive effect on the safety of US soldiers and adheres to all aspects of DODD 3000.09.

Discussion Question #1
There is understandable concern pertaining to the introduction of LAWS into militaries around the world. Beyond concerns about the potential uncontrollable nature of future general intelligence AI systems, there are more immediate and pressing concerns about the creation of narrow AI systems that possess lethal capabilities even if those capabilities are used in non-lethal functions. What limits should be placed on the development of LAWS? Can these systems be used only for defensive purposes? What if those defensive purposes result in a loss of life (i.e. the shooting down of an enemy pilot in the case of a surface-to-air missile)?

Discussion Question #2
LAWS have been used in combat for quite some time to protect ships (Phalanx CIWS), to protect territory (Patriot missiles and surface-to-air missile systems). Each of these systems has lethal intent. Is a system like AIM-R somehow more justifiable because it has no lethal intent? What if the AI-D were to enable it to defend a perimeter against incoming enemy forces as well as airborne munitions? Would the efficiency defense and the further reduction of risk to those inside an FOB be justifiable? What if a civilian were to be caught in the mix of an armed attack on an FOB and were killed?

Discussion Question #3
LAWS are based on code and code is highly susceptible to cyber-attacks. Should soldiers depend on a system that has the potential to be hacked? At what level is the potential for hacking a system too great. What if adversary forces were to deploy chaff in the form of small reflective metal streamers to confuse an AI targeting system into potentially harming civilians?

Discussion Question #3
What if there are conditions that the AI-D engineers and developers didn't consider, and the AIM-R system malfunctions or has biases built in and the system harms an unintended target? In the first Iraq war a rounding error in the Patriot Missile system batteries resulted in missiles missing their targets by miles.

Discussion Question #4
Although the present limits the development of LAWS to narrow AI – AI systems are increasingly being bundled together. The increasing use of bundles of AI and ML in applications increases their

complexity and reduces their auditability. Might this pose a problem specifically to the development of LAWS? If so, how?

Discussion Question #5
What issues should be addressed in AI prior to the implementation of AI in LAWS? Is there a need to establish legal, ethical, and moral boundaries for AI? Much progress has been made in the laws governing the conduct of war, specifically in the form of the Geneva Conventions and more broadly norms and laws associated with international humanitarian law. Where does AI fit in the applications of these laws?

Discussion Question #6
Many major adversarial powers to the United States are developing LAWS without regard to international law. These weapons are likely to pose both a strategic and a tactical threat to the United States in the coming decades. How should the United States address this threat? Should we build equivalent weapon systems to those of our adversaries, or should we ensure that our weapon systems adhere to international laws and norms? How far is too far to go in countering potential adversary LAWS?

Reflecting on Automating Death

The US and her allies face multiple challenges in the coming decades related to the incorporation of AI in weapon systems. While there are clear moral and ethical reasons why the inclusion of AI into weapons poses problems their implementation might be an inevitable feature of security competition between opposing powers. The overriding need to match adversaries' capabilities in combat is likely force the issue and make it necessary for the US to establish a trajectory on the automation of weapon systems from which it will be hard to change course. The US will be increasingly pressured to develop LAWS to maintain its interests in a hyper competitive global military environment.

There is little need for concern about the possibility of general AI and the rise of a Skynet replete with Terminators, but the possibility of weapon systems operating independently and without human intervention should be of concern. DoD policy makes retaining control over weapon systems a priority, but as the complexity of systems increases, the comprehension of system decisions will become more difficult. Understanding why AIs make decisions under various conditions and being able to audit those decisions and subsequently update learning functions to adhere to moral, ethical and even legal constraints will be critical to ensuring the proper use of LAWS.

The AIM-R weapon system is hypothetical, but it is based on a real system known as the Centurion C-RAM (Counter – Rocket, Artillery, and Mortar) system that has many of the same features. The

Centurion C-RAM is a truck mounted system that is actively in use and protects large military bases. It was adapted from the Navy's Phalanx CIWS. Knowing that these systems are real and in use surprises many who often think that automated weapon systems are something reserved for science fiction. However, these types of systems are real and in active service today. When a Centurion C-RAM fires in response to an incoming threat it streams out high explosive rounds at a rate of 4,500 rounds per minute and that looks like something out of a Star Wars film.

Implementing AI on the battlefield poses a number of challenges and ethical concerns ranging from issues of accountability and transparency to more abstract debates on whether decisions over life and death can be devolved to a machine. This third concern is best referred to as the dehumanization of killing. This list is not comprehensive, and it is important to discuss and identify other ethical concerns raised by both the development and use of LAWS.

Accountability

LAWS are not prohibited by US law and are expressly permitted in US Department of Defense policy. LAWS however raise key questions related to accountability. When LAWS are activated who is responsible for their function and perhaps malfunction. If LAWS kill civilians, is it a war crime and if so, who is responsible, the soldier(s) who activated the system, the developer or designers of the system, or the company that sold the system? One of the central tenets of controlling military forces is a legal and procedural adherence to the Geneva Conventions and international law. US soldiers are told they may only follow lawful orders. Despite the lethality of the US military its hierarchy controls its use of force in many ways. The intent of existing LAWS is to transfer this accountability into the chain of command. A commander who activates a LAWS must do so at a time and place in accordance with the legal conduct of military activities. If a commander activates a LAWS in a situation where it is likely to violate humanitarian or US legal constraints that commander is accountable for the activation of the weapon system and its subsequent effects. What, however, happens if a LAWS system malfunctions? If in the case of the AIM-R system who is responsible if the AI misidentifies a target, is hacked, or has some other malfunction that results in a violation of US law? Establishing accountability around LAWS is challenging imparts a number of assumptions on both commanders and the systems they are using. Most commonly there is an assumption that systems will and should function as designed. Yet, experience has shown that there are unforeseen events that arise as system complexity increases. Addressing the concept of accountability in the use of LAWS is a first step in establishing an ethical framework in which to utilize them.

Transparency

As AI complexity increases the decision engines (algorithms) of AIs can obscure the logic behind decisions or findings. Often AIs tasked with identifying patterns will skip over entire swaths of potential programming and find efficiencies that might lead to negative externalities. In the case of LAWS it is critical to ensure that the AIs being used in systems are auditable. That their decision structures are understandable to the soldiers using them. Having an AI behind the AIM-R system come to a decision about the distance a munition may travel or when it is unable to engage or engage targets must be clearly understood by operators and programmers. If it engages in actions that are outside of desired parameters these actions must be visible and correctable. It sounds strange to refer to a weapon as transparent, but the code behind LAWS must be transparent to prevent or correct undesirable outcomes.

Dehumanization

Of all the concerns surrounding the integration of AI into weapon systems the most pertinent remains the inherent dehumanization of its targets. The act of removing life or death decisions from a human operator and devolving those to a machine seemingly devalues the intrinsic nature of human life. When machines are the arbiters of life and death the moral and ethical implications for the future of humanity are in some way altered. Can a non-living entity programmed to "think" and set in motion ever be a valid executioner? Life and all the messiness that comes with it is somehow distinct from non-life. The loss of a human life can often inspire a range of emotions in both those who knew the deceased and those who took the diseased life. Yet a machine with an embedded AI has no emotion good or bad associated with the taking of life. It just calculates. In taking a life the machine while at risk of destruction feels no fear, suffers no real risk, expects no consequences from its actions. It is an analytical engine with lethal capabilities, but it is not human. Therefore, its ability to end life by its very nature removes or distances those attributes of war that make war a horrible human endeavor from one side and imparts and makes the target of its lethality the consequence of a decision modeled in a coding schema non-human. There is no doubting that LAWS will be used. But the question is more than will the AI follow orders, will it act within its defined code, and will we understand the decisions it made. LAWS pose a more prescient and critical question. What does it mean to be human if lives are valued so little that their destruction can be turned over to a machine that operates independently of human control.