



## Case Western Reserve University School of Law Scholarly Commons

---

The International Journal of Ethical Leadership  
Special Volumes

Cross Disciplinary Publications


---

Summer 2021

### Arguments for Banning Autonomous Weapon Systems: A Critique

Hunter Cantrell  
[hunter.cantrell@westpoint.edu](mailto:hunter.cantrell@westpoint.edu)

Follow this and additional works at: <https://scholarlycommons.law.case.edu/ijelspecial>

 Part of the [History Commons](#), [Law Commons](#), [Philosophy Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

#### Recommended Citation

Cantrell, Hunter, "Arguments for Banning Autonomous Weapon Systems: A Critique" (2021). *The International Journal of Ethical Leadership Special Volumes*. 2.  
<https://scholarlycommons.law.case.edu/ijelspecial/2>

This Article is brought to you for free and open access by the Cross Disciplinary Publications at Case Western Reserve University School of Law Scholarly Commons. It has been accepted for inclusion in The International Journal of Ethical Leadership Special Volumes by an authorized administrator of Case Western Reserve University School of Law Scholarly Commons.

# Arguments for Banning Autonomous Weapon Systems<sup>1</sup>

A Critique

Hunter Cantrell

---

## List of Abbreviations

AWS—Autonomous Weapon System(s)

ICC—International Criminal Court

IHL—International Humanitarian Law

LoW—Law of War

RoE—Rules of Engagement

## Introduction

The anticipated “imminent”<sup>2</sup> rise of fully autonomous weapon systems has led to a call for a “campaign to stop killer robots.”<sup>3</sup> This call has so far been backed by twenty-three nations<sup>4</sup> and has grown in perceived legitimacy since the campaign’s inception in 2013.<sup>5</sup> These efforts are driven in part by a fear of a future in which all (or at least the lethal) functions of human soldiers have been removed from the battlefield and replaced with a lethal form of artificial intelligence. However, contrary to the Hollywood depiction of apocalyptic Terminators roaming city streets, there is no reason to think that the fear is a realistic one.

---

1. This thesis is dedicated to my loving wife Tiffany and my two wonderful boys Holden and Bryson. Without you, I would be nothing.

2. There is a considerable lack of consensus on what counts as imminent: 5, 10, 50, 100 years? This lack of clarity is rampant throughout the field of emergent autonomous weapon systems.

3. Campaign to Stop Killer Robots, “About Us,” <https://www.stopkillerrobots.org/about/>.

4. It should be noted that all of the nations that have so far joined the campaign to “stop killer robots” are unlikely to be the nations that have the technological capability, infrastructure, or military “necessity” to build such devices. The list includes: Algeria, Argentina, Bolivia, Brazil, Chile, Costa Rica, Cuba, Ecuador, Egypt, Ghana, Guatemala, Holy See, Iraq, Mexico, Nicaragua, Pakistan, Panama, Peru, State of Palestine, Uganda, Venezuela, Zimbabwe (Campaign to Stop Killer Robots).

5. The campaign has met every year since 2013, hosted by the United Nations Office in Geneva, Switzerland.

Nonetheless, there are reasoned arguments against the development and implementation of fully autonomous weapon systems. We can classify the arguments against such systems into two main categories: an argument from “responsibility” and an argument from “agency.” Roughly, the argument from responsibility is concerned with the chain of responsibility for the use of autonomous weapon systems (AWS) and who ought to bear responsibility should something go wrong. The argument from agency questions why we should exchange human agency in a lethal action for an artificial agency and discusses the implications that this exchange has for human dignity and the possibility of mercy.

In this thesis, I summarize and evaluate these arguments in turn and show that both arguments are insufficient for an outright ban on AWS. I contend that the arguments are sufficient to support a framework, built on existing international humanitarian law, to regulate but not ban the development and employment of AWS. I then argue that the use of AWS is a rational choice for military and political leaders. I also outline a reasonable framework based on the current Geneva Conventions and international humanitarian law for regulating AWS.

## 1.1 Why is it essential to discuss autonomous weapon systems?

Autonomous weapon systems are the next frontier of modern military technology. As the speed of decision-making becomes faster due to increasing computerization, the ability of the human mind to keep up is lagging. The *OODA loop* was first coined by USAF Pilot and air power strategist Colonel John Boyd.<sup>6</sup> It stands for Observe, Orient, Decide, Act. The basic premise is that if a fighter pilot can get inside the OODA loop of an enemy fighter pilot, he could defeat him. The OODA loop then is a short decision matrix process in which a person (or in the future a machine) can decide on a certain action. It is likely that these OODA loops will continue to narrow as technology increases. The narrowed OODA loop of the future lends itself naturally to artificial intelligence that can process inputs and compute multiple decisions in real time much faster than any human. The drive to make “smarter” and faster weapons can clearly be seen in the evolution of fighter jets (USAF F-35 II and the Chinese Chengdu J-20 are but two examples), “smart bombs” (precision-guided munitions

---

6. Boyd 1987, 383.

such as the Excalibur<sup>7</sup>), and the increased utilization of computer-assisted decision-making tools in military planning.<sup>8</sup>

In response to the increased complexity of modern warfare and the decrease in the available population deemed fit to serve,<sup>9</sup> increased automation in current weapon systems and a drive to implement fully autonomous weapon systems have been the remedy. Current systems such as the PHALANX (deployed on US Aegis Cruisers) allows a semi-autonomous system to perform critical ship protection functions, freeing up the limited crew to perform other essential duties.<sup>10</sup> The increasing use of such technologies is the first stepping stone in the implementation of a system that will have the ability to select “targets and [deliver] force without any human input or interaction.”<sup>11</sup> The implementation of AWS is potentially worrying on many fronts, as alluded to earlier, but this development has the beneficial potential of reducing the unforeseen or collateral damage that accompanies warfare as currently practiced.

## 1.2 What are autonomous weapon systems?

Before delving into the debate regarding the implementation of AWS, it is essential to understand just what an autonomous weapon is. The academic and diplomatic communities, unsurprisingly, have yet to settle upon a commonly agreed-upon working definition of AWS. Currently accepted diplomatic definitions vary widely: some include weapons that today are considered automated, though not autonomous, and others include only systems wholly separated from human interaction. There are also definitions that require specific functionalities or technologies to be present. One example is from the government of the Netherlands, which defines an AWS as “a weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched,

---

7. US Army Acquisition Support Center, “Excalibur Precision 155MM Projectiles.”

8. This should not be construed as currently using computers *to* make command decisions, but rather the reliance of computer-based systems to provide more, better, and/or clearer information to a commander. These systems included everything from worldwide cargo tracking, enemy and friendly unit battlefield tracking, electronic systems monitoring, etc. As the civilian world has become more digitally reliant, so too has the military.

9. See the Military Leadership Diversity Commission Issue Paper #2, “Outreach & Recruiting.”

10. Raytheon, “Phalanx Close-In Weapon System.”

11. ICRC 2015, 6.

cannot be stopped by human intervention.”<sup>12</sup> This definition could easily encompass weapons that today are not considered to be autonomous, such as a cruise missile.<sup>13</sup> A cruise missile is designed to seek a specific programmed target (a heat signature, laser designation from a soldier on the ground, or some other pre-designated marker) and once a human has given the order to deploy the system, it cannot be recalled.

The government of France approaches the definition of an AWS in terms of “what it is not”<sup>14</sup> stipulating:

Lethal autonomous weapons are fully autonomous systems...LAWS should be understood as implying a total absence of human supervision, meaning there is absolutely no link (communication or control) with the military chain of command....The delivery platform of a LAWS would be capable of moving, adapting to its land, marine, or aerial environments and targeting and firing a lethal effector (bullet, missile, bomb, etc.) without any kind of human intervention or validation....LAWS would most likely possess self-learning capabilities.<sup>15</sup>

The French definition differs from the Dutch, in that it requires that an AWS (or LAWS) have no connection with a military (read human) chain of command, including in mission planning and final issuance of orders to use the AWS. This definition would severely limit what could be called an AWS, as it is likely that there will be some meaningful human interaction somewhere in the “kill chain.” The French definition then is so narrow as to make it practically untenable as a foundation upon which to build practical diplomatic and academic discussions.<sup>16</sup>

While adopting elements of the functionalist approaches used by the French, the United States’ definition is distinct because it distinguishes

12. United Nations Institute for Disarmament Research [UNIDIR] 2017, 23.

13. This example could also extend to any other “precision guided munition.” It should be noted that current law according to the Geneva Convention mandates the use of non-discriminate weapons which can be generally interpreted as meaning it endorses using precision munitions over indiscriminate ones and prohibits “[e]mploying weapons, projectiles and material and methods of warfare...which are inherently indiscriminate in violation of the international law of armed conflict, provided that such weapons, projectiles and material and methods of warfare are the subject of a comprehensive prohibition and are included in an annex to this Statute” (International Criminal Court Article 8(2)(b)(xx) 1998).

14. United Nations Institute for Disarmament Research [UNIDIR] 2017, 24.

15. *Ibid.*, 24.

16. This can also be said of the current Chinese designation of an AWS having a level of autonomy which lacks “human intervention and control during the entire process of executing a task” (Government of China 2018).

autonomous from semi-autonomous systems. The United States (through the Department of Defense) defines an AWS as

A weapon system, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system but can select and engage targets without further human input after activation.<sup>17</sup>

The US definition then is more practically useful than the French or Dutch, since it gives a detailed account of what is an AWS is and, importantly, excludes by implication systems that should not count as an AWS.<sup>18</sup> The definition does not include systems that are in place today that would be considered automated or semi-autonomous<sup>19</sup> (PHALANX, C-RAM<sup>20</sup>, etc.), nor does it include “fire and forget” systems, which could include the aforementioned cruise missile. It is also important to note that the US definition allows an AWS to act without the intervention of a human operator but does not preclude interaction from a human operator. This is important as it falls clearly in line with the US DoD Policy Directive 3000.09, which states that “Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.”<sup>21</sup> This policy directive hints at the collaborative nature that the US envisions for the implementation of artificial intelligence and AWS. The US policy definition also “places the

---

17. United Nations Institute for Disarmament Research [UNIDIR] 2017, 30.

18. As mentioned, the US provides a definition of what it defines as *not* an AWS, “A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator. This includes: (a) semi-autonomous weapon systems that employ autonomy for engagement related functions, including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement. (b) “Fire and forget” or lock-on-after launch homing munitions that rely on TTP (tactics, techniques and procedures) to maximize the probability that only the targets within the seeker’s acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by a human operator” (UNIDIR 2017, 31).

19. It should be noted, though, that systems such as the PHALANX and C-RAM can be placed into a setting which is analogous to fully autonomous.

20. Program Executive Office- Missiles & Space, “Counter-Rocket, Artillery, Mortar (C-RAM).”

21. Department of Defense, Directive 3000.09.

focus of what constitutes autonomy at the level of decision rather than on the presence or absence of a particular technology.”<sup>22</sup>

### 1.3 What does current International Humanitarian Law say regarding AWS?

Modern military conflict is governed by a set of norms and principles generally agreed upon as the basis for conducting a war. These norms and principles are built upon the long history of the just war theory. Prior to the twentieth century, these norms and principles were not codified but were rather a set of agreed-upon customs and practices.<sup>23</sup> For instance, it was generally considered the “gentlemanly” practice to rescue as many enemy sailors as possible once a naval battle had concluded. This practice was not required by treaty or other formal international agreement, but rather was an accepted practice that evolved over centuries of naval warfare. Though it had been common practice, with the proliferation of submarine warfare, the practice was abandoned. After World War II, during the Nuremberg trials, the Allies accused Admiral Donitz of committing crimes against humanity for the Nazi’s tactic of sinking enemy vessels and refusing to surface to rescue the survivors. The prosecution failed, though, as it was determined that through the evolution of naval warfare it had become common custom for submarines not to pick up surviving, stranded sailors due to extreme operational risk.

International Humanitarian Law (IHL) in the form of the Geneva Conventions and its Additional Protocols does not explicitly mention AWS. What can be useful for developing rules regulating AWS, though, are IHL prohibitions on certain sorts of weaponry. The Convention on Certain Conventional Weapons (CCW) bans the use of certain inherently indiscriminate forms of weapons. This prohibition implies a preferential status of precision, or “smart,” weapons over indiscriminate or “dumb” ones. To follow the CCW then, an AWS would have to be developed and utilized in a fashion in which its use would not be “inherently indiscriminate.” What then is a weapon or weapon system that is inherently indiscriminate?

---

22. UNIDIR 2017 31.

23. While it is true that some elements of international law are still predicated on customs and norms, there are now treaty-based rules in place, such as those specified in the Geneva Conventions (1949) and in the Statute of the International Criminal Court. Prior to the establishment of such bodies as the ICC, most nations only had the assurances of warfare or mutually supportive defensive alliances to enforce compliance with agreed upon customs and norms, and some nations, including the US, are not state parties to the ICC.

A prime example would be a “dumb” antipersonnel land mine or a poison gas. Both types of weapons are indiscriminate because upon deployment they cannot correctly distinguish between a legitimate and a non-legitimate military target. The land mine will kill or maim any who are unfortunate to step on it, even decades later. Developing an AWS that can accurately discriminate is a difficult task, but not one that is insurmountable.

## 2. Objections to AWS: The Arguments for a Ban

Since the invention of the crossbow in the thirteenth century<sup>24</sup>, there have been calls for the limitation of acceptable weaponry in “civilized warfare.” Technological developments have pushed nations to seek the latest cutting-edge weaponry to gain even the slightest advantage over their enemy. These developments have included the machine gun, the land mine, and poison gases. Some of the developments, such as machine guns, have stayed within the bounds of acceptable military hardware. Others such as the land mine have been heavily regulated, and their use is internationally discouraged.<sup>25</sup> The third category of weapons, including poison gases, has been expressly forbidden under international treaty.

Since 2013, the Campaign to Stop Killer Robots has advocated for a total ban on the development and employment of lethal autonomous weapon systems.<sup>26</sup> The group has advocated at the state and international level for a coalition of UN member states, nongovernmental organizations, businesses, and private citizens with the express purpose of banning lethal AWS. The campaign, like many persons in academia, worries that AWS will cross a “moral threshold,” sacrificing the dignity of human life for

---

24. From the Papal Canon issued after the Second Council of the Lateran: “We prohibit under anathema that murderous art of crossbowmen and archers, which is hateful to God, to be employed against Christians and Catholics from now on” (Papal Encyclicals Online, “The Second Council of the Lateran”).

25. Antipersonnel land mines are explicitly prohibited under the Quebec treaty: ...anti-tank and anti-ship mines are still acceptable. Mines that have a short life span (such as becoming inert after X hours of deployment) are acceptable.

26. In September 2018, the Campaign to Stop Killer Robots advocated “for a pre-emptive and comprehensive ban on the development, production, and use of fully autonomous weapons, also known as lethal autonomous weapon systems or killer robots.” As of 2020, the Campaign has altered their pre-emptive ban language and instead say “The Campaign to Stop Killer Robots urges states to launch negotiations on an international treaty to retain meaningful human control over targeting and attack decisions by prohibiting the development, production, and use of fully autonomous weapons. This ban treaty should be enforced via national laws” (The Campaign to Stop Killer Robots, <https://www.stopkiller-robots.org/learn/>).



the efficiency of automated warfare. The arguments for a ban on “killer robots” can roughly be grouped into two categories, the “responsibility” argument and the “agency” argument.

## 2.1 The ‘responsibility’ argument: Who can be held responsible?

One of the great worries of those against the use of AWS is that we may create robots that can target and kill humans without being able to “justly” hold a moral agent responsible. Some have called the problem of justly assigning responsibility for an action taken by an AWS an “accountability gap.”<sup>27</sup> To be sure, if we can trace an action of an AWS directly to the person who authorized the mission (for instance, a rogue commander who programmed orders to kill unarmed civilians and injured enemy soldiers), then it would appear to be clear that we can justly assign blame for any resulting death on the commander who authorized the mission. The problem arises, however, when we cannot directly trace a line from the person who authorized the AWS to the war crime that was perpetrated. This is the “accountability gap.” Bonnie Docherty, one of the advocates of a ban on AWS, believes that “[it wouldn’t be] fair nor legally viable to...hold a [human] commander or operator responsible.”<sup>28</sup> She is not alone in finding issue with shifting responsibility away from the agent—the AWS—who committed the morally problematic (or illegal) action. While arguing for an accountability gap, Robert Sparrow explores two possibilities: shifting blame from the AWS onto humans (the programmer and the commanding officer) and leaving the AWS as the responsible agent.

First, Sparrow argues that we might be able to rest at least some of the responsibility for a wayward AWS on the programmer or designer who built the device: “this will only be fair if the situation described occurred as a result of negligence on the part of the design/programming team.”<sup>29</sup> Sparrow, though, quickly dismisses assigning responsibility to the programmer in cases where there is no negligence. The first reason to dismiss putting the blame on the programmer is that “the possibility that the machine may attack the wrong targets may be an acknowledged limitation of the system.”<sup>30</sup> If the possibility that the machine might attack an unauthorized

---

27. Scharre 2018, 261.

28. *Ibid.*, 261.

29. Sparrow 2007, 69.

30. Sparrow 2007, 69.

target is a “side effect” or “limitation of the system” and the programmer explicitly communicates this limitation to any potential user, then it would not be fair to hold the programmer liable. Sparrow argues that in this case, those who still choose to deploy the system, with the known limitations or possible side effects, would have to bear the burden of the responsibility. He then adds that “the possibility that an autonomous system will make choices other than those predicted and encouraged by its programmers is inherent in the claim that it is autonomous.”<sup>31</sup> Here Sparrow is pointing out that due to the operational design of an AWS and the complex nature of its possible algorithms and neural networks, no human will be able in all cases to predict the decisions of an AWS. If this is the case, then it is again unfair to blame the programmer, since she could not have foreseen all of the actions of the AWS.

Sparrow argues that if we cannot hold the programmer of the AWS to account for its actions, then possibly we can hold the commanding officer who ordered its use to account.<sup>32</sup> He writes, “The risk that [the AWS] may go awry is accepted when the decision is made to send it into action. This is the preferred approach of the military forces seeking to deploy existing AWS.”<sup>33</sup> Sparrow argues that if we treat AWS in the same manner that we treat current conventional “dumb” and “smart” weapons of today, then it seems that we can simply attribute the blame to the commander who ordered their use. This tack, though, is not taking the full nature of the autonomy of the weapon into account. Sparrow argues that the very nature of the autonomy of the weapon means that commanders cannot control which targets the AWS selects, and thus they should not be held to account for the choices of the AWS: “If the machines are really choosing their own targets then we cannot hold the Commanding Officer responsible for the deaths that ensue.”<sup>34</sup> If the programmer and the commanding officer cannot be held to account, can the AWS itself be held responsible?

Sparrow argues that “autonomy and moral responsibility go hand in hand.”<sup>35</sup> If we reach a point where AWS is fully capable of making lethal

---

31. *Ibid.*, 70.

32. Sparrow does not delineate at what level of command we ought to hold one to account. If for instance an infantry company commander (a Captain [O3] in the US Army and US Marine Corps) decides to send an AWS into a village and it kills a child, who should be responsible? The captain? Their boss, a battalion commander (lieutenant colonel)? Or the brigade commander, division commander, corps, etc.? Should only the most senior be held to account for putting this device out in to the force to begin with?

33. Sparrow 2007, 70.

34. Sparrow 2007, 71.

35. *Ibid.*, 65.

decisions on its own (with a human out of the loop), then an AWS would have to bear the moral responsibility of its actions, which brings us to the crux of the problem. “It is hard to take seriously the idea that a machine should—or could—be held responsible for the consequences of ‘its’ actions.”<sup>36</sup> Assuming that holding an agent responsible for bad action involves punishment for the agent, Sparrow argues that for “acts to serve as punishment they must evoke the right sort of response in their object.”<sup>37</sup> What then would Sparrow have as the right sort of response? According to him, the only fitting response to punishment involves suffering, and so for an AWS to be punished, “it must be possible for it [the AWS] to be said to suffer.”<sup>38</sup> But this notion of suffering complicates the way we understand the functionality and capabilities of AWS as currently construed and imagined. It is not likely that a programmer could or would build into an AWS the ability to have emotional responses such as suffering (or fear) because such emotions would contravene one of the potential benefits that AWS provides over humans, namely, that the systems are not provoked to war crimes by anger, hatred, the thirst for vengeance, etc.<sup>39</sup> If it is the case that an AWS does not have the ability to experience physical or emotional pain, it is likely not able to suffer and thus, according to Sparrow, would be unable to be punished.

Ultimately, Sparrow believes that unless we can ethically develop an AWS that can be “justly held responsible,”<sup>40</sup> it is unethical to deploy these weapons on the battlefield. The ability to decide to take a life, Sparrow believes, must remain with an agent who is capable of moral responsibility and can be rightly said to be punished in case she violates one of the Laws of War.

## 2.2 The “agency” argument: Should we forfeit human agency at the point of death?

One of the critical elements of war is the decision to kill a fellow human being. In the earliest days of warfare, until the advent of the firearm, most of this killing was up close and personal. Enemy combatants would be able

---

36. *Ibid.*, 71.

37. *Ibid.*, 71.

38. *Ibid.*, 72.

39. Ronald Arkin makes the case that AWS lacking specific emotions (fear, anger, sadness, etc.) would be a potential benefit to their use. These AWS would not exhibit the typical emotional responses seen in humans on the battlefield and thus would be more likely to make “correct” responses according to their governing ethics.

40. Sparrow 2007, 66.

to look into each other's eyes as they were fighting for their lives. Lieutenant Colonel Dave Grossman, in his book *On Killing*, explores the effects that growing distance from the act of killing has had on our collective ability to kill. When we are face to face with our enemy, they are human, and we can understand them at some level. Thus the killing is more intimate and more brutal. When we kill at a distance, our enemy appears to be something less than human, and thus it is easier to rationalize killing them. The ultimate distance between combatants and killing is the removal of the decision to kill a human agent and allowing that decision to be made by an artificial agent. Alex Leveringhaus argues that "the replacement of human agency in a war with the 'artificial agency' of machines is deeply problematic."<sup>41</sup> The critical debate, then, is whether we should forfeit human agency in favor of an artificial one.

Leveringhaus contends that the first problem with surrendering human agent hegemony over the decision to kill is that we lose the "moral equivalence of soldiers." The principle of moral equivalence, formulated by Michael Walzer, states that each set of soldiers (the aggressors and the defenders) has the same moral value from the perspective of the Laws of War so that all of them are equally entitled to use force to kill the enemy and to defend themselves.<sup>42</sup> This moral equivalence holds, according to Walzer, regardless of the justice of the war being fought. All have the moral right to fight. When one side of the equation is replaced with an AWS, then there is no longer a moral equivalence. An AWS does not have a moral claim to self-defense, and as such, they lose their equality with the opposition.

A second problem Leveringhaus addresses is an AWS's lack of an ability to make a "moral judgment." This lack of ability is especially important when the system determines the level of force to use because "...application of the proportionality criterion [in IHL] involves making moral judgments."<sup>43</sup> If an AWS is not able to make a moral judgment about the right proportion of force needed in a given situation, should it even be able to make the ultimate moral decision to take a life in the first place? Leveringhaus says "no" and argues that the ability to kill without the ability to deliberate about the moral consequences of that action is, in essence, a disrespecting of the human rights and dignity of the individual being targeted.

The crux of Leveringhaus' argument is the contention that by using

---

41. Leveringhaus 2016, 2.

42. Walzer 2015, 34.

43. Leveringhaus 2016, 2.

AWS to kill another human being, we are removing any chance of mercy from the kill chain and thereby morally distancing ourselves from the killing in a way that disrespects the dignity of the person being killed. Mercy in the kill chain means that at any point in time, a soldier could choose not to pull the trigger. Leveringhaus write, “Killer Robots qua artificial agents lack a central component of a human agency, the ability to do otherwise: not to shoot the target.”<sup>44</sup> This lack of an ability to not kill an authorized combatant then by its very nature is disrespectful of the dignity inherent in every human. “The enemy qua legitimate target does not hold a claim against the attacking soldiers not to kill him. But surely the human capacity to have mercy with, feel pity for, or empathize with other humans, even if these belong to an opposing state, is morally relevant and worthy of protection.”<sup>45</sup> The problem Leveringhaus has with an AWS executing a lethal action upon a human is that it fundamentally lacks a sense of compassion and understanding of the value of human life. He writes, “I think retaining human agency at the point of force delivery, thereby protecting the freedom not to pull the trigger, push the button, or throw a grenade, is essential for retaining our humanity in exactly the situation that challenges it the most: war.”<sup>46</sup> If an AWS does not have an option not to kill an enemy, then it lacks an ability to show mercy, and the possibility of showing mercy is required respecting human dignity. Leveringhaus closes with an unsettling message, one that runs against the optimism seen in many developers and researchers of advanced weapons technology, “Killer Robots will not rescue us from the human condition.”<sup>47</sup>

### 3. Rejecting the Arguments

As J. Glenn Gray points out, there is a necessary relationship between death and war, “death in war is commonly caused by members of my own species actively seeking my end, despite the fact that they may never have seen me and have no personal reason for mortal enmity.”<sup>48</sup> This fact then prompts us to ask, How then can we reduce the suffering of those actively engaged in

---

44. Leveringhaus 2016, 9. Leveringhaus uses the term Killer Robots instead of Autonomous Weapon Systems. This seems to be an emotive ploy to convince the reader of the moral dilemma presented by using such devices. I believe this distracts from the objective nature of his argument and instead puts forward an emotive and weaker argument.

45. *Ibid.*, 10.

46. *Ibid.*, 10.

47. *Ibid.*, 15.

48. Gray 1959, 100.

warfare and those innocent bystanders caught in its fury? Part of the answer may lie in the implementation of artificial intelligence and autonomous weapon systems on the battlefield of tomorrow. Throughout the rest of this thesis, I will argue (as many in the national security and defense sectors have already done<sup>49</sup>) that we are likely to see AWS operating in tandem with the soldiers of tomorrow. While it is true that there are many functions that could more easily be turned over to an autonomous system (such as driving logistics convoys along dangerous routes or processing human resources paperwork), there are functions within the military that are unlikely ever to be fully automated. Instead, we are likely to see a combination of humans and machines coupled in a way to maximize the intelligence, durability, strength, and firepower of the human soldier. I will argue that each of the arguments for a ban so far presented does not provide sufficient reason for the implementation of such a ban. Rather, all they can logically support is the implementation of a system of regulation, by which international parties can monitor and guide the development of such systems in such a way that they remain compatible with IHL and are morally permissible.

### 3.1 Reply to the 'responsibility' argument

While Sparrow's argument might be convincing on the surface, he is making tenuous assumptions in some places, while ignoring causal chains of responsibility as they are currently construed in the military. First, I deal with Sparrow's view of the possibility of assigning blame to the AWS programmer and the commanding officer who deploys the system and then address Sparrow's argument that it would be impossible to punish an AWS.

Sparrow is generally correct in his wariness to assign moral responsibility to the programmer of the AWS. This assignment would likely be akin to holding the maker of a rifle morally responsible for its use during a school shooting; while doing so may ameliorate feelings regarding the use of a weapon in such a manner, it does little to help us understand who is morally responsible. Sparrow does argue that we could hold a programmer responsible if the misdeed was due to some programming negligence, akin to how a manufacturer today is held liable for defects within their products that result in harm or death to the end user. What must not happen, though, is that a programmer is held liable because we find the use of their product

---

49. See Major General Mick Ryan's (Australian Army) white paper "Human-Machine Teaming for Future Ground Forces," [https://csbaonline.org/uploads/documents/Human\\_Machine\\_Teaming\\_FinalFormat.pdf](https://csbaonline.org/uploads/documents/Human_Machine_Teaming_FinalFormat.pdf).

to be morally repugnant. On the matter of holding the programmer liable, I concede to Sparrow that we should only hold her responsible if she erred in some negligent way.

Sparrow says that we ought not to hold the commanding officers liable for the actions of the AWS under their command because this system is fully autonomous, and as such we cannot be sure on what actions or motivations for those actions the AWS will take. This ignores some of the fundamental principles of leadership and responsibility that are foundational to the current military US culture.

The concept of responsibility is a crucial tenet of military culture. All service members are imbued with a certain level of responsibility commensurate with their rank and position. All have the primary responsibility conforming their conduct to the law, military regulation, and tradition. As one ascends the rank structure of their affiliated branch, increasing levels of responsibility are given, which include responsibilities for both personnel and material resources. The pinnacle of military responsibility comes with attaining the position of “commander.” Depending upon the branch, a commander at different levels will be responsible for personnel and material (commensurate with the unit size), but the basic principles remain the same. The US Army manual on leadership<sup>50</sup> specifies that “command includes the authority and responsibility for effectively using available resources and for planning the employment of, organizing, directing, coordinating, and controlling military forces for the accomplishment of assigned missions.”<sup>51</sup> The pertinent portion here is the responsibility that a commander holds. A commander is responsible for all the actions that their soldiers take or fail to take at any given time. US Army commanders accept, through their placement in a position of higher authority, the burden of responsibility for the actions of their subordinates. Therefore, commanders would be at least partially responsible for AWS misfires and problematic activity even if the event occurs without direct human control.

The Yamashita standard was established as a legal precedent in the wake of World War II. In the military trials that followed the end of the war in the Pacific, General Tomoyuki Yamashita was tried for the crimes against humanity perpetrated by troops under his control. General Yamashita argued that he could not have known all the atrocities that were being

---

50. Army Doctrine Reference Publication 6-22 Army Leadership.

51. ADRP 6-22 2012, 1-3

committed nor could he be expected to exert any form of direct control over his widely dispersed troops in the prevention of such acts. The court found this argument unconvincing and said:

The law of war imposes on an army commander a duty to take such appropriate measures as are within his power to control the troops under his command for the prevention of acts which are violations of the law of war and which are likely to attend the occupation of hostile territory by an uncontrolled soldiery, and he may be charged with personal responsibility for his failure to take such measures when violations result.<sup>52</sup>

The verdict handed down in the Yamashita case then firmly set that commanders, at any level, have a duty to prevent atrocities from occurring in units under their command. There is a potential response, though, that says, “A commander is overall responsible for the actions of his troops, but we cannot reasonably hold a commander responsible for the actions of a ‘non-negligently designed’ AWS whose decision-making process is essentially unknown.” My response is that not only do we hold commanders at least partially responsible for the actions of their subordinates regardless of the subordinate’s decision-making process, but we also hold a commander responsible for fostering an environment where such choices are seen as legitimate. First, the military holds its commanding officers ultimately responsible for the actions of all below them, regardless of whether they can physically control them or not. Take, for instance, the recent spate of prominent naval accidents in the US Navy’s 7th Fleet (stationed in Japan). The commanders of both vessels were found responsible (at least one was even referred to prosecutors for manslaughter charges), even though they were not physically at the helm, or even on the bridge at the time of the incident.<sup>53</sup> The Navy determined that they were negligent in their duties to ensure their ship operated in the appropriate manner. The Navy also determined that onboard these vessels, and at fleet command level (fleet headquarters), commanders fostered an environment where lax standards and corner-cutting was seen as acceptable.<sup>54</sup> If we circle back to AWS, it

---

52. In re Yamashita.

53. It should also be noted that commanders even higher, to include the commander of the 7th Fleet, were forced to resign as a result of these incidents.

54. See the US Navy’s reports on both the USS *Fitzgerald* and the USS *John S McCain* crashes: <https://www.secnav.navy.mil/foia/readingroom/HotTopics/CNO%20USS%20Fitzgerald%20and%20USS%20John%20S%20McCain%20Response/CNO%20USS%20>



is reasonable to assume, on the current military legal structure, that if an AWS was allowed to commit a war crime or not prevented from doing so, that we could hold that commander both legally and morally responsible.

If we are collectively willing to hold a commander—at least partially—liable for the conduct of his troops, who are autonomous agents themselves, then there is no legal or moral reason to not hold them to the same level of responsibility for the actions or inactions of an AWS. Sparrow believes this would be unfair to hold a commander responsible, since the commander cannot possibly predict the behavior of the AWS, but this raises the question of whether the commander can predict the behavior of their troops in battle. Commanders regularly and vigorously train their soldiers so that their actions become more predictable during a stressful situation, but this does not preclude the possibility that a human soldier will cross a moral threshold and commit a war crime. Robotist Ronald Arkin writes “I personally do not trust the view of setting aside the rule [Law of War (LoW) or Rules of Engagement (RoE)] by the autonomous agent [the AWS] itself, as it begs the question of responsibility if it does so, but it may be possible for a human to assume responsibility for such a deviation if it is ever deemed appropriate.”<sup>55</sup> Arkin here is arguing that when we design and build an AWS, we ought never to allow it the ability to decide on its own to set aside the LoW or RoE (which would be its core programming) and instead vest that responsibility only in the hands of a human. This then would transfer the responsibility of the AWS violating LoW or RoE away from the AWS and onto a human decision maker (namely a commanding officer at some level).<sup>56</sup> If a human commander, however far removed from the final decision point of killing, decided to send an AWS into a situation where it is a known possibility that there may be excessive collateral damage, there appears to be no moral wrong with holding the commander to account when incidents occur.

One might legitimately worry here that I have not provided enough grounds—or enough compelling grounds—to make the claim that the

---

Fitzgerald%20and%20USS%20John%20S%20McCain%20Response.pdf.

55. Arkin 2009, 40.

56. I believe this could be similar to the use of a targeting officer in current conflicts the US is engaged in. In these types of situations, when a target is designated, often a senior officer (oftentimes a general officer) will be asked to validate the target and authorize the engagement. This officer normally has a lawyer in their presence to offer legal advice on the situation, but ultimately the responsibility lies with that decision maker. This could be the same with authorizing an AWS to deviate from its LoW/ RoE preprogramming, if a commander wants it to happen, then they must authorize and accept full responsibility for that action.

Yamashita standard provides enough moral justification for us to ground responsibility in the office of a commander. I argue that instead of searching for a grounded justification, we might approach this problem by acknowledging that the Yamashita standard provides us a morally justifiable reason to hold a commander responsible (even one that is overseeing the use of AWS). Why might this sort of responsibility doctrine be morally justifiable? We desire to compel commanders on the ground, in times of war, to take all due care possible to foster an environment that operates within and respects the Laws of War, applicable International Humanitarian Law, and the relevant Rules of Engagements for that area of operation. The Yamashita (or the updated Medina standards) provide us a morally justifiable route to achieve that goal. It serves as a reminder to commanders (throughout the chain of command) that they are personally responsible for the actions of their subordinates. This does not mean, though, that they take all the responsibility for the actions (or lack thereof) of their subordinates, but they share an appropriate proportion of the responsibility.<sup>57</sup>

The final problem with Sparrow's argument is his reluctance to hold AWS, as a moral agent, responsible for its crimes. Sparrow argues that one cannot rightly punish an AWS because it cannot suffer and since it cannot suffer, it is not being punished. He appears to be grounding his argument for punishment in the idea of retributive justice. The problem is that Sparrow asserts that retributive justice is the correct course to pursue (explicitly with AWS and implicitly with humans). He says that "in order for any of these acts [various sorts of punishments] to serve as punishment, they must evoke the right sort of response in their object...to be capable of being punished then, it [the thing or person being punished] must be possible for it to be said to suffer."<sup>58</sup> We can reject this desire for suffering by merely adopting one of the other possible moral forward-looking claims to punishment available in the literature. We could, for instance, desire a more communicative form of punishment, meant to communicate the unacceptable nature of an act to other AWS. We could also desire to have a rehabilitative approach to punishment, in which we reprogram or retrain the AWS to learn that the previous action was bad or undesirable. If we insist on a retributive system, as Sparrow seems to require, then he would be right; the punishment of an AWS would be impossible, but there is no reason here to think that Sparrow is correct.

---

57. I believe that the discussion of the "appropriate proportion of responsibility" is an important one we must have, but it falls outside of the purview of this thesis.

58. Sparrow 2007, 72.

With regards to bearing the moral responsibility of an enemy soldier's death, Sparrow says, "the least we owe our enemies is allowing that their lives are of sufficient worth that someone should accept responsibility for their deaths."<sup>59</sup> Sparrow demands that we be able to hold the individual (AWS) responsible when it seems the loci ought to be on the commander who decides to use the weapon or the politician that begins the war. If the AWS performs in a legitimate fashion, then there should be no requirement to hold individuals—human or AWS—responsible for legitimate killing in war.<sup>60</sup>

On the other hand, how do we deal with an AWS that has "gone rogue" and committed a war crime? I argue that instead of punishing the AWS, we examine the circumstances under which the incident occurred. No two events will ever be the same, so to make a blanket statement or universal principle regarding whom to blame for what is inappropriate. Instead, we ought to examine the decision made to employ the weapon (was there a high calculated likelihood of excessive collateral damage?) and other decisions made by the commander along with the relevant information available at the time to assign blame in the appropriate places. In the end, if it can be conclusively demonstrated that a commander or even an AWS took all possible due care to prevent an atrocity, but one happened (e.g., a stray bomb), then we ought not to hold either responsible (possibly applying the doctrine of double effect).

Sparrow would reply that if we follow this plan, we will develop an accountability gap and the only proper remedy to the situation would be to not to use an AWS in the first place. To this, I reply that Sparrow is misguided. If, as I have mentioned before, we apply standards such as Yamashita (or Medina) we can locate the nexus of moral responsibility upon the commanding officer in those cases where we cannot clearly demonstrate that the AWS made an illegal (or immoral) move. We would then be faced with two options, either (a) "punish" the AWS using a different moral basis (communicative, rehabilitative, etc.) or (b) punish the commanding officer (at whatever level deemed appropriate, which might not be the lowest level of command) using the Yamashita/ Medina standards.

As has been thus far demonstrated, based upon current international law precedent, US military leadership norms, and a rethinking of the idea of

---

59. Ibid., 67.

60. Of course, someone like McMahan would disagree and would say that if those soldiers, or maybe AWS, who participate in an unjust war *are* individually responsible for the killing of an enemy soldier because that other soldier (who is presumably just) is not a legitimate target (McMahan 2011, 14).

punishment, the argument from responsibility fails. Instead of providing compelling evidence that the international community ought to outright ban AWS development and use, this argument instead compels us to regulate the development and use of AWS.

### 3.2 Reply to the 'agency' argument

The most potent argument dealt with thus far holds that we ought not to turn over control of the decision to kill from a human being to an artificial agent. This ought not to occur because the artificial agent lacks the ability to demonstrate mercy at the point of lethal action and because it is inherently disrespectful of human dignity for a person to be killed by the decision of an autonomous robot. Both parts of this argument are fundamentally flawed.

Leveringhaus insists that we must allow for the option of not following a legitimate order to kill the enemy on the battlefield, in essence, the option not to shoot a legitimate combatant and instead exercise mercy. He says “[c]ompared to artificial agency, what makes human agency in warfare, and in ordinary life, valuable is the possibility of engaging in an alternative course of action.”<sup>61</sup> Leveringhaus genuinely believes that on a battlefield one must be fully capable of making the decision not to kill when faced the permissible option of exercising lethal force. He supposes that if an AWS is preprogrammed with its orders, mission parameters, Law of War and Rules of Engagement parameters, it will simply attack all enemies that it encounters. This supposition, I believe, is not necessarily true. If we follow Arkin’s advice regarding the “ethical governor,” a properly designed AWS will first default to nonlethal measures to either evade the enemy or subdue him and to immediately default to lethal action would be impossible. If we require an algorithm to be built on this model, then lethal actions would only occur if they were obligatory. For example, suppose there is an AWS on patrol with a squad of American infantry soldiers, we could set as one of the RoE as: “If one member of your squad is in imminent danger of being captured, then you are obligated to use proportionally correct force to prevent their capture.” If any member of the patrol is not in imminent danger of being captured, then the AWS must default to nonlethal action, unless it meets some other predefined obligatory criteria.

---

61. Leveringhaus 2016, 9

AWS as a tool must be obligated<sup>62</sup> to follow legitimate and lawful orders, but it also must be obligated not to follow orders that are illegitimate and unlawful. The latter is just as important as the former in the prevention of war crimes on the battlefield. This would also seem to serve as Arkin's "ethical behavior control."<sup>63</sup> Accordingly he writes, "especially in the case of battlefield robots (but also for a human soldier) we do not want the agent to be able to derive its own beliefs regarding the moral implications of the use of lethal force, but rather to be able to apply those that have been previously derived by humanity as prescribed by LoW and RoE."<sup>64</sup>

According to Leveringhaus then, one of the defining features that separate human agents and artificial agents is our (human) ability to demonstrate mercy towards our enemies. To Leveringhaus, "surely the human capacity to have mercy with, feel pity for, or empathise with other humans, even if these belong to an opposed state, is morally relevant and worthy of protection."<sup>65</sup> What then is this mercy if not the ability to do otherwise? It seems that the key for Leveringhaus is that we can empathize with our enemy; we can recognize their humanity and thus refrain from doing them harm when able or appropriate; "those soldiers who did not kill might have done so because they recognised the humanity of the enemy and realised the graveness of the decision to pull the trigger."<sup>66</sup> Might, though, we be confusing what constitutes genuine mercy on the battlefield? Is deciding not to kill an enemy fighter and let them live another day (albeit potentially in captivity) not recognizing their humanity? I would argue that the decision to not kill, based upon the lack of military necessity of their death is an adequate form of compassion and mercy that could be implemented on the battlefield of tomorrow. This would move us forward in reducing the amount of battlefield carnage and reduce deaths to those only necessary to achieve a military objective or those unforeseeable and unfortunate deaths that happen as a consequence of another intended action. We must also recognize that as it currently stands, humans have not demonstrated a good track record of demonstrating mercy on a large scale in warfare (though it does happen at the individual level). Often the propaganda machines of a nation will spin the enemy into an "other" or

---

62. Here we might define the obligations of an AWS simply as its programing to follow the rules written within its code, but this might be expanded to include a more expansive view of moral obligations, if that is possible.

63. Arkin 2009, 66–67.

64. *Ibid.*, 117.

65. Leveringhaus 2016, 9–10.

66. Leveringhaus 2016, 10.

some other being that removes their humanity, thus making it easier to kill (think of calling the Germans in WWI “Huns” or the Vietnamese “gooks”). If we can create an AWS, under strict international regulation, that abides by international law and respects the life of a person (by only killing those deemed legally allowable to kill and in accordance with military necessity), we could better demonstrate mercy on the battlefield than we have thus far demonstrated that we are capable of.

A potential reply to my objection here might be along the lines of feasibility. Is it reasonable to expect that we could feasibly program an AWS to “show mercy” to enemy combatants whenever able? This is a strong line of argument and one that we must continue to address throughout the development of such systems. One might argue that we could “simply”<sup>67</sup> write in the governing algorithm of the AWS that “whenever you encounter an enemy soldier that does not present a direct threat to you or any human soldiers in the direct vicinity, demonstrate mercy by not using lethal action to neutralize them.” This is but one possible technique, but one that we should genuinely explore. Harkening back to Ronald Arkin and his theoretical work on constructing an algorithm to govern such a machine, it seems that mercy would be inherent in defaulting to a nonlethal option as the first course of action. Instead of a human soldier choosing not to kill the enemy (presumably using nonlethal measures to capture him and not just letting him go free), we would have an AWS defaulting to capturing the opposing soldier using nonlethal and humane tactics.<sup>68</sup>

The final argument from Leveringhaus details the necessity of the preservation of human dignity in the decision to kill. Leveringhaus argues that this dignity is preserved if and only if a human agent makes the ultimate decision at the point of death. This seems to beg the question of whether a human agent deciding to kill is necessary to preserve that dignity. I would answer no. Paul Scharre makes a powerful point when he says:

When viewed from the perspective of the soldier on the battlefield being killed, this [dignified death] is an unusual, almost bizarre critique of autonomous weapons. There is no legal, ethical, or historical tradition of combatants affording their enemies

---

67. I fully acknowledge that this would not be a simple task.

68. It would negate this whole project if the AWS used nonlethal yet *inhumane* tactics. Tactics such as these (maybe bludgeoning the enemy into unconsciousness) seem to be just as morally suspect as shooting every individual enemy that one encounters (that are not currently posing a threat).

the right to die a dignified death in war. There is nothing dignified about being mowed down by a machine gun, blasted to bits by a bomb, burning alive in an explosion, drowning in a sinking ship, slowly suffocating from a sucking chest wound, or any other horrible ways to die in war.<sup>69</sup>

The desire for only dignified deaths to occur in war appears to be a futile attempt to sterilize combat by saying “at least the enemy died a dignified death.” While uncomfortable to most, we must acknowledge that at a foundational level, warfare is defined by death and killing. It should be our goal then to reduce the amount of carnage and suffering wrought upon people (both combatants and noncombatants), but to demand that “death with dignity” be a requirement ignores the fundamental nastiness of war. It is perfectly acceptable to require that nations only go to war when such an action is morally justified or at the very least morally justifiable but to demand that they only kill in dignified ways is both too ideal and ignores the brutal reality of this most deadly of human relationships.

The most dignified death one could expect<sup>70</sup> in warfare is one that causes as little suffering as possible. A “clean” death, as it were, that quickly and efficiently extinguishes the life of a combatant without causing lingering pain and suffering, would plausibly be the most desirable. This, though, should not be taken as a downplaying of the significance of death on the battlefield to some sort of brutally efficient system of slaughter akin to a slaughterhouse, but rather a desire to end hostilities as quickly as possible without needless pain and suffering. As Paul Scharre wrote (from his own personal experiences in warfare), it seems farcical to insist that death at the “hands” of a machine would somehow be less dignified than the various and sundry ways one can die in modern warfare.

Instead of being concerned with a perceived “dignity in the manner of death” (since it is unclear whether one might be able to have a dignified death) we ought to focus on the reduction of suffering in war. Autonomous weapon systems provide us a possible avenue to achieve this. As I will argue in the next section, our goal in warfare should be to reduce unnecessary pain and suffering. We ought then to design AWS that help us achieve this goal. As such, we should reject Leveringhaus’ claim that death by AWS is inherently undignified.

---

69. Scharre 2018, 288.

70. If it is even genuinely possible to expect such a thing in combat.

#### 4. AWS as the Rational Choice in the Evolution of Warfare

Limiting the deaths of noncombatants and civilians on the battlefield ought to be the aim of any morally upright force. IHL requires armies of every state, whether it has ratified the Geneva Treaties or not, to take all reasonable care and precaution to not kill or injure protected classes of persons and to not unnecessarily damage or destroy property (especially protected classes of property). In this regard, militaries across the world have been developing “smart weapons.” Autonomous weapon systems are the next logical step of these smart weapons. Smart weapons have at least a two-fold benefit over so-called “dumb weapons,” (1) they can be directed at a specific target (often with margins of error of only a few meters) and (2) they are economically beneficial. Now, these two reasons play into the desire to limit the unnecessary collateral damage to the greatest extent possible. If we can target an area with a minimal margin of error, then we can use few weapons to neutralize the threat, thus limiting the exposure to danger for surrounding civilians and property. This calculation aligns with the fundamental architecture that Arkin urges those researching and developing AWS to include in their design.<sup>71</sup> This design would push the effective “battlefield carnage” or collateral damage to as low a point as possible. Secondly, smart weapons are economically beneficial when compared to “dumb” weapons. This may seem counter-intuitive since many of the smart weapons often cost hundreds of thousands or even millions of dollars apiece. The rationale, though, is that if I can achieve the same mission with few weapons (that may cost more individually) than I can with many relatively cheap “dumb” munitions (which in the aggregate may cost more than the smart weapon), then it is more economically beneficial to use the smart weapon. This seems to make sense. If I can expend only one guided missile to take out a heavily guarded and important bunker that previously would have taken two hundred unguided bombs, then I should clearly use the one smart guided missile. If this case holds with weapons such as missiles and air-dropped bombs, then it holds that if an AWS can do the job of ten human soldiers, it is economically more feasible to use the one AWS. If AWS can assist militaries in reducing economic costs as well as the suffering of noncombatants, then there are strong reasons to conclude that the development and use of such systems is the rational course to pursue.

---

71.  $A+B+C+D$  = Battlefield carnage.  $A$  = Intended combatants,  $B$  = Unintended Friendly Forces,  $C$  = Intended Noncombatants, and  $D$  = Unintended Noncombatants. The goal should be to maximize  $A$ , while eliminating  $B$  &  $C$ , and ensuring  $D$  is as close to 0 as is possible (Arkin 2009, 128).



Arkin and others believe that the use of AWS (and AI more generally) on the battlefield may be the moral course to take because of a set of characteristics that are inherent in their nature. First, AWS can act conservatively.<sup>72</sup> An AWS can (and arguably must be programmed) to default to a nonlethal action. This is primarily due to its lack of a need for self-preservation.<sup>73</sup> By forgoing the innate human desire to keep one's self alive and free from harm, an AWS can take more provocative nonlethal actions (such as moving out of a dangerous area that may not be possible for a human) or sacrificing itself for the sake of mission fulfillment, in order to draw attention of the enemy away from the rest of its unit. An AWS will also lack the emotional fog and resultant self-fulfilling prophecies that often cloud the judgment of human agents.<sup>74</sup> Often it is the case that war crimes are committed in the heat of battle when emotions run high and judgment is clouded by fear or anger. As J Glenn Gray says, a soldier "becomes a fighting man, a *Homo furens*."<sup>75</sup> If an AWS is built without such clouding emotions, as would be the logical and morally best course, then it would not be a "slave" to fear or anger, would not act out in aggression because another AWS or human squad member was killed by some local village. In essence, it might be able to be a better ethical and moral agent than we are in battle. Accordingly, Arkin writes, "It is not my belief that an autonomous unmanned system will be able to be perfectly ethical on the battlefield, but I am convinced that they can perform more ethically than human soldiers are capable of."<sup>76</sup>

If, as Arkin argues, we can design an AWS with the ability to act both ethically and morally better (or at least on par) as compared with human soldiers; then it appears that developing such a system is the rational course to take. If an AWS can be at least as ethical and moral as the best of us in warfare, then we have compelling moral reasons to build such a device. First, we would reduce the number of our own soldiers who are required to be in harm's way. This point acknowledges that not all human soldiers can be replaced by AWS, but certainly, a good number of them can be. If I can deploy an autonomous system that can drive down an IED-ridden highway to deliver my supplies, without risking the lives of truck drivers, then I am morally obligated to do so. If I choose not to take such an action if the tech-

---

72. Arkin 2009, 29.

73. *Ibid.*, 29.

74. *Ibid.*, 30.

75. Gray 1959, 27.

76. Arkin 2009, 31.

nology is available to me, then I am in fact acting immorally. Second, if an AWS can make better ethical and moral decisions than humans can on the battlefield, then I am obligated to deploy such a system. If—and this is a big if—such a system can be built and made available to me as a commander, and if it can make more morally and ethically correct decisions (such as properly discriminating targets, preventing target overkill, reducing collateral damage) than humans can, then I am morally obligated to deploy such a system. The deployment of such a system does not relieve those commanders who choose to use them (here likely higher than the tactical level company and battalion commanders<sup>77</sup>) of the moral weight of killing, nor would it absolve them of any associated excessive collateral damage.<sup>78</sup> For as Arkin mentioned, it is unlikely that even with the best technology we could ever effectively reduce collateral damage to zero. Yet, as a command, I would still be morally obligated to deploy an AWS that could make morally better decisions than would be made by the humans who would otherwise be using lethal force.

In the end, it is the rational course of action to develop and deploy a weapon system that can reduce the amount of battlefield carnage. This view does not in any way sanitize warfare, for legitimate combatants will still die and, as Leveringhaus says, “the enemy qua legitimate target does not hold a claim against the attacking soldiers not to kill him.”<sup>79</sup>

## 5. Proposal: A Schema for Regulating AWS

Now that I have dealt with the more persuasive arguments against the use of AWS on the battlefield of tomorrow and argued why we are morally compelled to use such weapons, it is necessary to create a theoretical system of governance that could regulate the development and deployment

---

77. In the US Army, the *company* is the standard sized smallest unit with a designated commander (though there are some smaller units, called detachments, but these are exceptions to the rule). Companies vary in strength from 50 personnel to over 300. A company then is the base level tactical unit at which we vest command leadership and responsibility. A company is commanded by a captain (O-3) who, if having had no previous enlisted experience, has between 4–6 years time in service. A *battalion* (at least a standard one) comprises five to seven companies and is generally between 700–1,500 personnel depending upon make-up and mission set of the unit. A battalion is still considered to be at the tactical level. A battalion is commanded by a lieutenant colonel (O-5) generally with about 17–19 years time in service.

78. It is a physical impossibility to limit all collateral damage when warfare takes place where people live. This is an accepted fact. The goal, though, is to reduce or limit the amount of collateral damage. In this way, excessive collateral damage is seen as a war crime, whereas incidental collateral damage is not.

79. Leveringhaus 2016, 10.

of such systems on the international level. As previously written, there is no current IHL that specifically covers the development and deployment of AWS. This section aims to be a starting point for just such a discussion. I recommend three broad areas of concern for the development and deployment of AWS. These three areas encompass the most pressing worries that one may have regarding AWS. Of course, this will not cover all potentialities in the use of the systems, but rather is intended to be a broad enough theoretical framework to guide the development of international law. The three areas that I emphasize in this theoretical schema include (1) appropriate deployment of AWS, (2) adherence to current and future IHL, and (3) established chain of responsibility.

First, the issue is when is it appropriate to deploy such a system? IHL should stipulate some laws stating that no AWS should be deployed by any military or nation unless and until the relevant technical experts have certified that the system has been programmed to conform its behavior to IHL, and any commander who deploys such a system not so certified is subject to punishment, regardless of whether the system actually commits war crimes. This requirement would not be foreign to militaries today, as they must go through the same type of certification process to ensure that standard or conventional weapon systems (think of a tank) are designed and operate within the bounds of the law. Once a military or nation has passed this threshold, they then must make another decision, which consists of two separate, but no less equal parts: (1) when to use nonlethal versus lethal force (tactical/operational deployment) and (2) when to deploy an AWS (strategic deployment). I deal with each of these problems in turn.

The decision to employ an AWS in a tactical situation just means that the commander on the ground chooses to use an AWS for a given mission. These commanders are generally accepted to be operating at the “tactical level” and generally have the freedom to decide on the direct employment of troops and weaponry necessary to achieve a given mission or directive. In this sense, a company commander would have the command authority to decide to employ her company-level unmanned aerial vehicle to help increase her visual space while she would not have the authority to direct a strategic level asset (such as a satellite) to accomplish her mission. When speaking of an AWS, though, a tactical level commander (even up to a Brigade Commander who generally commands approximately five thousand troops) should have the command responsibility and authority to decide as to whether it is appropriate to use such a weapon. For instance, if a company

commander is tasked with securing a village, she must make an on the spot decision (maybe in consultation with higher level orders) whether to use a specific weapon or tactic. The same holds true for an AWS. To control for the appropriate use of force, international humanitarian law ought to specify that the tactical use of AWS should ensure that the potential collateral damage created by their use is minimized.

The second issue regarding employment of AWS refers to strategic deployment. By strategic employment of an AWS, I mean whether it is morally or legally permissible to use such a weapon in a given war or conflict. To address this concern, it seems simple enough to legislate that the use of AWS must comport with the principles of JWT and IHL specifically—Would the deployment of an AWS be more likely than not to cause excessive collateral damage when deployed to X theater when compared to Y theater? For example, given the level of technological refinement at the time, it may be unethical to deploy an AWS into a heavily urbanized theater of combat (i.e., Singapore) when compared to a remote or rural theater of combat (i.e., the Russian steppe). This regulation ought to state that it is impermissible for any nation or military to deploy an AWS into a theater of combat where the likelihood for excessive collateral damage from their use is higher than what could be expected from the use of human forces. For example, if the very employment of an AWS into a theater is likely to lead to an indiscriminate use of force because the AWS (and likely a human) would have a great difficulty in properly discriminating between combatants and noncombatants, then the employment of the AWS would constitute a war crime, even if it does not err in its discrimination of legitimate combatants. There is no reason to think that when an AWS commits a mistake, we would not develop some sort of accountability review, though whom we decide to find ultimately at fault for the mistake may look different from what it currently does. If a soldier makes a mistake now (one that would result in a war crime), then that individual soldier and her chain of command may be held to account.<sup>80</sup>

Accordingly, the second broad category of regulation should state something to the effect that “an AWS may not be employed, if by its very design it is incapable of adhering to the basic tenets of the JWT and of any current or future possible IHL.” As was earlier stated, it would be inherently wrong, both morally and legally, to employ a system that through its programming

---

80. This would of course assume that the mistake was some sort of deliberate action or that the mistake occurred out of some sort of gross negligence; the specifics at this juncture are not particularly important as the case still holds together.

is unable to be accountable. Such a system might lack an ability to be auditable in some fashion (such as having a fire or engagement log to be able to review all such engagements that the AWS was in), or it simply may default to lethal action as a proper first response. While it would not be the place of IHL to dictate the exact technology, which may or may not be used, something akin to Arkin's ethical governor would be a prime example of a type of AWS that could comply with this theoretical regulation. Any such system that meets the intent of the regulation and thereby can adhere to the JWT and IHL should be deemed as permissible.

The final area of focus for this future instantiation of IHL is that there ought to be an established chain of responsibility. Future IHL should provide that "in the event of a deployment of an AWS a commander or political leader assumes moral and legal responsibility for the strategic decision to employ such a weapon system." The law should also stipulate that, at a tactical level, "a commander who decides to employ such a weapon is both morally and legally responsible for all that it does or that it fails to do." Some opponents of such legal regulations may complain that they place too burdensome a weight on military commanders. To that I answer: the burden ought to be heavy, as this will preclude the free use of such systems in situations that are likely to either violate JWT/ IHL or increase collateral damage to unacceptable levels. The regulations may have the potentially positive effect of limiting the use of AWS to only those situations where outcomes are reasonably well controlled or in situations where the likelihood of collateral damage is low. (For instance, the regulation may discourage the use of AWS in dense urban settings but may increase the use of AWS in virtually unpopulated areas such as the open seas).

While there is a benefit to creating international law that encourages a morally correct development and employment of AWS, there must be a corresponding mechanism to compel compliance. Such a mechanism may be like the already established ICC, or it may be a new form of international enforcement. The legal regulations I proposed do not depend on any particular method of enforcement and are compatible with a newly created form of compliance mechanism to "force" nations to develop their AWS in a way which would comport with the regulations. It is not my place here to decide how this mechanism ought to function, but merely to advocate for a system which is able, to its best ability, to prevent the commission of war crimes before they are committed.

## 6. Conclusion

Human progress is analogous to a perpetual motion machine. Without any definite beginning and without a seeming end, we march endlessly on. So too does the development of better and more effective ways to kill each other. From the earliest days of human existence when one man realized that the atlatl would allow him to throw his spear further than his opponent to modern advances in the realm of artificial intelligence, the push to find the next greatest weapon moves inexorably forward. The time is now to act to form a body of regulation to shape how we as a species develop and deploy what is undoubtedly to be the next class of weapon systems.

I have presented two of the most persuasive arguments against the use of autonomous weapon systems in the form of the “responsibility” and the “agency” arguments. While both of these arguments are superficially compelling, neither of them provides sufficient grounds (either individually or collectively) to tip the scale towards an outright ban on the use of this emergent technology. Rather, both (and other arguments) provide a compelling reason as to why we ought to push for international regulation on the development and morally proper deployment of such systems. I have also argued that the proper development and deployment of an autonomous weapon system is the rational choice if and only if they can meet our strict moral standards of conduct. The aim then is to reduce the awfulness that is the battlefield. Much like Sherman said, “War is hell,” but we ought to do our very best to make it as much of a tolerable hell as we possibly can.

There is a deep-seated instinct in humans to fear the unknown. Science fiction movies have done little to quell the worries that, if we continue our path of technological improvements, robots will throw off the yoke of their “oppression” and turn on their masters. This fear, while it might be widespread, is not a sound basis for moral or legal decisions.

## Acknowledgments

I would like to acknowledge the faculty and staff of the Philosophy Department, especially Dr. Andrew Altman and Dr. Andrew I. Cohen: without your dedication, this thesis would not have been successful. I would also like to thank all the professors with whom I have had the privilege of taking a course: thank you for putting up with me. I would also like to thank those in the US Army who saw a spark in me and thought I was worthy of the task. Thank you all.

## References

- Arkin, Ronald. *Governing Lethal Behavior in Autonomous Systems*. Boca Raton: Chapman and Hall Imprint (Taylor and Francis Group), 2009.
- Army Doctrine Reference Publication. "6-22 Army Leadership." *Army Doctrine Reference Publications* (2012). [https://usacac.army.mil/sites/default/files/misc/doctrine/cdg/cdg\\_resources/manuals/adrp/adrp6\\_22\\_new.pdf](https://usacac.army.mil/sites/default/files/misc/doctrine/cdg/cdg_resources/manuals/adrp/adrp6_22_new.pdf).
- Boyd, John. *A Discourse on Winning and Losing*. Maxwell Air Force Base: Air University Press, 2018.
- Campaign to Stop Killer Robots. "About Us." Accessed 25 July 2018. <https://www.stopkillerrobots.org/about/>.
- Department of Defense. "Directive 3000.09-Autonomy in Weapon Systems." *Office of the Deputy Secretary of Defense* (November 2012). <http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>.
- In re Yamashita 327 U.S. 1 [1946] 61 (United States Supreme Court). <https://supreme.justia.com/cases/federal/us/327/1/>.
- Government of China. "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions of the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects." *Convention on Certain Conventional Weapons—Position Paper* (April 2018). [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/DD1551E60648CEBBC125808A005954FA/\\$file/China's+Position+Paper.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/DD1551E60648CEBBC125808A005954FA/$file/China's+Position+Paper.pdf).
- Gray, J. Glenn. *Warriors: Reflections of Men in Battle*. New York: Harcourt, Brace and Company, 1959.
- International Criminal Court. "Rome Statute." Article 8(2)(b)(xx) (1998). [https://www.icc-cpi.int/nr/rdonlyres/ea9aef7-5752-4f84-be94-0a655eb30e16/0/rome\\_statute\\_english.pdf](https://www.icc-cpi.int/nr/rdonlyres/ea9aef7-5752-4f84-be94-0a655eb30e16/0/rome_statute_english.pdf).
- Leveringhaus, Alex. "What's so Bad about Killer Robots." *Journal of Applied Philosophy*, (March 2016). DOI: 10.1111/japp.12200.
- McMahan, Jeff. *Killing in War*. Oxford: Oxford University Press, 2011.
- Military Leadership Diversity Commission. "Outreach & Recruiting." *United States Department of Defense*. November 2009. <https://diversity.defense.gov/Portals/51/Documents/Resources/Commission/docs/Issue%20Papers/Paper%2002%20-%20Requirements%20and%20Demographic%20Profile%20of%20Eligible%20Population.pdf>.
- Papal Encyclicals Online. "The Second Council of the Lateran- 1139 A.D." Accessed 18 January 2019. <http://www.papalencyclicals.net/councils/ecum10.htm>.
- Program Executive Office- Missiles & Space. "Counter-Rocket, Artillery, Mortar (C-RAM)." Accessed September 1, 2018. <https://missiledefenseadvocacy.org/defense-systems/counter-rocket-artillery-mortar-c-ram/>.
- Raytheon. "Phalanx Close-In Weapon System." Accessed 15 August 2018. <https://www.raytheon.com/capabilities/products/phalanx>.
- Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. New York: W. W. Norton & Company, 2018.

Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy*, vol. 24, no. 1 (2007):62–71. DOI: 10.1111/j.1468-5930.2007.00346.x.

United Nations Institute for Disarmament Research. "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches." *UNIDIR Resources*, no. 6 (2017): 1–33. <http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf>.

United Nations Office for Disarmament Affairs. "Perspectives on Lethal Autonomous Weapon Systems." *UNODA Occasional Papers*, no. 30 (November 2017): 1–61. [www.un.org/disarmament](http://www.un.org/disarmament).

US Army Acquisition Support Center. "Excalibur Precision 155MM Projectiles." Accessed 25 July 2018. <https://asc.army.mil/web/portfolio-item/ammo-excalibur-xm982-m982-and-m982a1-precision-guided-extended-range-projectile/>.



# Argumentos para prohibir los sistemas de armas autónomos<sup>1</sup>

Una Crítica

Hunter Cantrell

Traducción al español por Victoria A. García

---

## Lista de Abreviaturas

AWS: Sistema (s) de Armas Autónomas

ICC- Tribunal Penal Internacional

IHL- Derecho Internacional Humanitario

LoW- Leyes de la Guerra

RoE- Reglas de Enfrentamiento

## 1. Introducción

El surgimiento previsto e inminente<sup>2</sup> de los sistemas de armas totalmente autónomos ha dado lugar al llamamiento de una “campana para detener a los robots asesinos”<sup>3</sup>. Hasta ahora, este llamado ha sido respaldado por veintitrés países<sup>4</sup> y se ha visto incrementado con una legitimidad que se percibe desde el inicio de la campana en el 2013.<sup>5</sup> Estos esfuerzos han sido impulsados en parte, por el temor a un futuro en el que todas las funciones (o al menos las letales) de los soldados humanos sean retiradas del campo de batalla y reemplazadas por una forma letal de inteligencia artificial. Sin

---

1. Dedico esta tesis a mi querida esposa Tiffany y a mis dos maravillosos hijos Holden y Bryson. Sin ustedes, no sería nada.

2. Existe una falta considerable de consenso sobre lo que se considera inminente: ¿5, 10, 50, 100 años? Esta falta de claridad es común en todo el campo de los sistemas de armas autónomos emergentes.

3. La campana para detener a los robots asesinos “*About Us*”.

4. Cabe señalar que es poco probable que todas las naciones que se han unido hasta ahora a la campana para “detener a los robots asesinos” sean los países que tienen la capacidad tecnológica, la infraestructura o la “necesidad” militar de construir tales dispositivos. La lista incluye: Argelia, Argentina, Bolivia, Brasil, Chile, Costa Rica, Cuba, Ecuador, Egipto, Ghana, Guatemala, Santa Sede, Irak, México, Nicaragua, Pakistán, Panamá, Perú, Estado de Palestina, Uganda, Venezuela, Zimbabue (Campana para detener a los robots asesinos).

5. La campana se ha reunido todos los años desde el 2013, organizada por la Oficina de la Organización de las Naciones Unidas en Ginebra, Suiza.

embargo, no existe motivo alguno para pensar que este temor sea fundado, contrario a la representación hollywoodense de los *Terminators* apocalípticos deambulando por las calles de la ciudad.

No obstante, existen argumentos lógicos en contra del desarrollo y de la implementación de los Sistemas de Armas Totalmente Autónomos. Los argumentos en contra de estos sistemas pueden clasificarse en dos categorías principales: un argumento de “responsabilidad” y un argumento de “agencia”. A grandes rasgos, el argumento de la responsabilidad se refiere a la cadena de responsabilidad por el uso de sistemas de armas autónomos (AWS) y acerca de quién debe asumir la responsabilidad en caso de que algo salga mal. El argumento de la agencia cuestiona porqué deberíamos intercambiar la agencia humana por una agencia artificial en una acción letal y las implicaciones que este intercambio representa para la dignidad humana y la posibilidad de misericordia.

En esta tesis, resumo y evalúo estos argumentos sucesivamente y demuestro que son insuficientes para una prohibición total de los AWS. Sostengo que los argumentos son suficientes para respaldar un marco fundamentado en el actual derecho internacional humanitario, para *regular, pero no prohibir* el desarrollo y el empleo de los AWS. Posteriormente argumento que el uso de los AWS es una opción racional para los líderes militares y políticos. De igual manera, trazo un marco razonable para regular los AWS basado en los Convenios de Ginebra vigentes y el Derecho Internacional Humanitario.

## 1.1 ¿Por qué es importante hablar de los Sistemas de Armas Autónomos?

Los Sistemas de Armas Autónomos son el “siguiente peldaño” en la tecnología militar moderna. A medida que se agiliza la toma de decisiones debido al incremento de la computarización, se retrasa la capacidad de la mente humana para mantenerse actualizada. El ciclo de O.O.D.A. fue acuñado por primera vez por John Boyd<sup>6</sup>, un piloto de la Fuerza Aérea de los Estados Unidos (USAF, por sus siglas en inglés) y estratega del poder aéreo coronel. O.O.D.A. significa Observar, Orientar, Decidir y Actuar y su premisa básica es que si un piloto de combate puede entrar en el ciclo de O.O.D.A. de un piloto combatiente enemigo, podría derrotarlo. Entonces el ciclo de O.O.D.A. es un proceso corto de matriz de decisiones en el que una persona (o en el futuro, una máquina) puede decidir sobre una determinada acción. Es prob-

---

6. Boyd, 1987, p. 383.

able que estos ciclos de O.O.D.A. continúen estrechándose a medida que aumenta el uso de la tecnología. El ciclo alargado de O.O.D.A. del futuro se presta en sí, de manera natural, para la inteligencia artificial que puede procesar entradas y calcular múltiples decisiones en tiempo real y mucho más rápido que cualquier ser humano. El impulso de fabricar armas “más inteligentes” y más rápidas puede verse claramente en la evolución de los aviones de combate (el USAF F-35 II y el Chengdu J-20 chino son sólo dos ejemplos), las “bombas inteligentes” (municiones guiadas con precisión como la Excálibur<sup>7</sup>) y en la mayor utilización de recursos asistidos por computadora, en la toma de decisiones en la planificación militar.<sup>8</sup>

Ante una mayor complejidad de la guerra moderna y un menor número de población apta y disponible para servir<sup>9</sup>, la solución ha sido una mayor automatización de los actuales sistemas de armamento y un impulso por implementar sistemas de armas totalmente autónomos. Los sistemas actuales como el PHALANX (utilizado en los cruceros norteamericanos Aegis) permite que un sistema semiautónomo realice funciones esenciales de protección de buques, permitiendo que la tripulación restringida realice otras tareas esenciales<sup>10</sup>. El uso cada vez mayor de tales tecnologías, es el primer peldaño en la implementación de un sistema que tendrá la capacidad de seleccionar “objetivos y lanzar fuerza sin ningún aporte o interacción humana”.<sup>11</sup> Como se mencionó anteriormente, la implementación de los AWS pudiera ser preocupante en muchos aspectos, sin embargo, este desarrollo tiene la favorable capacidad de reducir el daño imprevisto o colateral que acompaña a la guerra, de la manera en que se realiza actualmente.

## 1.2 ¿Qué son los Sistemas de Armas Autónomos?

Antes de profundizar en el debate sobre la implementación de los AWS, es importante entender lo que es un arma autónoma. Como es lógico,

---

7. Centro de Apoyo para la Adquisición del Ejército de los Estados Unidos (por sus siglas en inglés, “Proyectiles Excalibur de alta precisión de 155 mm”).

8. Esto no debe interpretarse como el uso actual de computadoras *para* tomar decisiones de mando, en todo caso, como la dependencia de sistemas basados en computadoras para proporcionar más, mejor y/o más clara información a un comandante. Estos sistemas incluyen todo, desde rastreo de carga mundial, rastreo de campo de batalla de unidades enemigas y amigas, monitoreo de sistemas electrónicos, etc. A medida que el mundo civil se ha vuelto más dependiente digitalmente, también lo ha hecho el ejército.

9. Consulte el documento temático No. 2 “Alcance y Reclutamiento” emitido por la Comisión de Diversidad de Liderazgo Militar.

10. Raytheon, “Phalanx, sistema artillero antimisil”.

11. ICRC 2015, 6.

las comunidades académicas y diplomáticas, aún no se han decidido por una definición de los AWS que se tome en consenso. Las definiciones diplomáticas que se aceptan en la actualidad varían ampliamente: algunas incluyen armas que hoy se consideran automatizadas, aunque no autónomas, mientras que otras definiciones, únicamente incluyen sistemas completamente separados de la interacción humana. También existen definiciones que requieren la presencia de características o tecnologías específicas. Un ejemplo de lo anterior, es la definición que da el gobierno de los Países Bajos, la cual define los AWS como “un arma que, sin la intervención humana, selecciona y ataca objetivos que cumplen ciertos criterios predefinidos, que obedece la decisión humana de desplegar el arma, a sabiendas de que un ataque, una vez ejecutado, no puede detenerse por medio de la intervención humana”.<sup>12</sup> Esta definición podría abarcar fácilmente armas que hoy en día no se consideran autónomas, tales como un misil de crucero<sup>13</sup>. Un misil de crucero está diseñado para buscar un objetivo específico programado (una huella térmica, designación láser de un soldado en el suelo o algún otro marcador designado previamente) y una vez que un humano haya dado la orden de desplegar el sistema, no se puede revocar.

El gobierno de Francia plantea la definición de los AWS en términos de “lo que no es”<sup>14</sup>, estipulando:

Las armas autónomas letales son sistemas totalmente autónomos... Los Sistemas de Armas Autónomos Letales (LAWS, por sus siglas en inglés) deben entenderse como aquello que implica una ausencia total de supervisión humana, lo que significa que no hay absolutamente ningún vínculo (comunicación o control) con la cadena militar de mando... La plataforma de lanzamiento de un LAWS sería capaz de moverse, adaptarse a su entorno ter-

---

12. Instituto de las Naciones Unidas de Investigación sobre el Desarme [UNIDIR, por sus siglas en inglés] 2017, p. 23

13. Este ejemplo también podría ampliarse a cualquier otra “munición guiada de precisión”. Cabe señalar que, de acuerdo con la Convención de Ginebra, la ley actual ordena el uso de armas no indiscriminadas, lo que generalmente puede interpretarse en el sentido de que respalda el uso de municiones de precisión en lugar de las indiscriminadas y prohíbe “desplegar armas, proyectiles, materiales y métodos de guerra... los cuales indiscriminan de manera intrínseca violando el derecho internacional de los conflictos armados, siempre y cuando esas armas, proyectiles, materiales y métodos de guerra sean objeto de una prohibición integral y se incluyan en un anexo al presente Estatuto” (Artículo 8(2)(b)(xx) 1998 de la Corte Penal Internacional).

14. Instituto de las Naciones Unidas de Investigación sobre el Desarme [UNIDIR] 2017, p. 24.

restre, marino o aéreo y apuntar y disparar un efecto letal (bala, misil, bomba, etc.) sin ningún tipo de intervención o validación humana... Los LAWS probablemente poseerían capacidades de autoaprendizaje.<sup>15</sup>

La definición francesa difiere de la holandesa en que se requiere que los AWS (o LAWS) no tengan conexión con una cadena de mando (léase humano) militar, incluso en la planificación de la misión y la emisión final de órdenes para usar un Sistema de Armas Autónomo. Esta definición limitaría seriamente lo que podría llamarse un Sistema de Armas Autónomo, ya que es probable que exista alguna interacción humana significativa en algún punto de la “cadena de exterminio”. La definición francesa es tan estrecha que la hace prácticamente insostenible como una base sobre la cual se desarrollen discusiones diplomáticas prácticas y académicas.<sup>16</sup>

La definición de los Estados Unidos si bien adopta elementos de los enfoques funcionalistas utilizados por los franceses, es diferente porque distingue los sistemas autónomos de los semiautónomos. Estados Unidos (a través del Departamento de Defensa) define un Sistema de Armas Autónomo como,

Un sistema de armas, que una vez activado, puede seleccionar y atacar objetivos sin más intervención que un operador humano. Esto incluye sistemas de armas autónomos supervisados por humanos que están diseñados para permitir que los operadores humanos anulen el funcionamiento del sistema de armas, pero puedan seleccionar y atacar objetivos sin más intervención humana después de la activación.<sup>17</sup>

Por lo tanto en la práctica, la definición norteamericana es más útil que la francesa o la holandesa, porque ofrece una descripción detallada de lo que es un Sistema de Armas Autónomo y lo que es más importante, excluye por implicación, los sistemas que no deberían contarse como los AWS.<sup>18</sup> Esta

---

15. *Ibíd.*, p. 24.

16. Esto también puede decirse de la actual designación china de un AWS que tiene un nivel de autonomía que carece de “intervención humana y control durante todo el proceso de ejecución de una tarea” (Gobierno de China 2018).

17. Instituto de las Naciones Unidas de Investigación sobre el Desarme [UNIDIR] 2017, p. 30.

18. Como se mencionó, Estados Unidos brinda una definición de lo que determina como lo que no son los AWS, “Un sistema de armas que, una vez activado, está destinado a atacar solo objetivos individuales o grupos de objetivos específicos que han sido seleccionados

definición no incluye los sistemas disponibles hoy en día, los cuales podrían considerarse automatizados o semiautónomos<sup>19</sup> (PHALANX, C-RAM<sup>20</sup>, etc.) ni los sistemas de “disparar y olvidar”, que podrían incluir el misil de crucero previamente mencionado. También es importante señalar que la definición norteamericana permite que los AWS actúen sin la intervención de un operador humano, pero *no* excluye la interacción de un operador humano. Esto es significativo ya que se apega evidentemente con la directiva de la Norma 3000.09 del Departamento de Defensa de los Estados Unidos, que establece que: “Los sistemas de armas autónomos y semiautónomos deberán diseñarse para permitir que los comandantes y operadores ejerzan los niveles adecuados de criterio humano sobre el uso de la fuerza”.<sup>21</sup> Esta directiva reguladora insinúa la naturaleza colaborativa que Estados Unidos prevé para la implementación de inteligencia artificial y los AWS. La definición de la norma de los Estados Unidos también “ubica el enfoque de lo que constituye la autonomía a nivel de *decisión* más que en la presencia o ausencia de una tecnología en particular”.<sup>22</sup>

### 1.3 ¿Qué establece el Derecho Internacional Humanitario vigente respecto a los AWS?

Los conflictos militares modernos se rigen por un conjunto de normas y principios que generalmente se acuerdan como base para llevar a cabo una guerra. Estas normas y principios se basan en la larga historia de la teoría de la guerra justa. Antes del siglo XX, estas normas y principios no estaban codificados, sino eran más bien eran un conjunto de costumbres y prácticas

---

por un operador humano. Esto incluye: (a) sistemas de armas semiautónomos que utilizan la autonomía para funciones relacionadas con el enfrentamiento, que incluyen entre otras, pero no se limiten a conseguir, seguir e identificar objetivos potenciales; señalar objetivos potenciales a los operadores humanos; priorizar los objetivos seleccionados; sincronizar el disparo; o proporcionar una guía terminal para centrarse en objetivos seleccionados, siempre que se mantenga el control humano sobre la decisión de seleccionar objetivos individuales y grupos específicos objetivo para el enfrentamiento. (b) “Disparar y olvidar” o bloquear después de lanzar municiones autodirigidas que se basan en TTP (tácticas, técnicas y procedimientos) para tener más posibilidades de que solo los objetivos dentro del encesto que ha obtenido el buscador, cuando se active el buscador, sean esos objetivos individuales o grupos objetivo específicos los que hayan sido seleccionados por un operador humano” (UNIDIR, 2017, p. 31).

19. Sin embargo, debe considerarse que los sistemas como PHALANX y C-RAM se pueden situar en un entorno, el cual puede ser desde análogo hasta totalmente autónomo.

20. Oficina Ejecutiva del Programa: Misiles y Espacio, “Sistema de defensa contra ataques con cohetes, artillería y morteros (C-RAM, por sus siglas en inglés)”.

21. Departamento de Defensa, Directiva 3000.09.

22. UNIDIR 2017 31.

tácitas<sup>23</sup> Por ejemplo, generalmente se consideraba una práctica “caballeresca” el rescatar a tantos marineros enemigos como fuera posible, una vez que se terminaba una batalla naval. Esta práctica no la exigía ningún tratado ni ningún acuerdo internacional formal, sino más bien fue una práctica aceptada que se realizó durante siglos de guerras navales. Aunque había sido una práctica común, con la proliferación de la guerra submarina, la práctica se abandonó. Después de la Segunda Guerra Mundial, durante los juicios de Nuremberg, los aliados acusaron al almirante Donitz de cometer crímenes contra la humanidad por la táctica de los nazis de hundir las naves enemigas y negarse a emerger a la superficie para rescatar a los sobrevivientes. Sin embargo, el enjuiciamiento fracasó, ya que se determinó que durante la evolución de la guerra naval, ya era una costumbre común para los submarinos, no recoger a los marineros supervivientes varados, debido al riesgo operativo extremo.

El Derecho Internacional Humanitario (IHL, por sus siglas en inglés) consistente en la Convención de Ginebra y sus protocolos adicionales no menciona a los AWS de manera explícita. Sin embargo, el hecho de que existan prohibiciones del IHL sobre ciertos tipos de armas, puede beneficiar en la elaboración de reglas que regulen a los AWS. La Convención sobre ciertas armas convencionales (CCW, por sus siglas en inglés) prohíbe el uso de determinadas formas de armas intrínsecamente indiscriminadas. Esta prohibición implica un estatus preferencial de armas de precisión o “inteligentes” sobre las indiscriminadas o “tontas”. Entonces, para cumplir con la CCW, un Sistema de Armas Autónomo debe desarrollarse de tal manera que su uso no sea “intrínsecamente indiscriminado”. Por lo tanto, ¿qué es un arma o un sistema de armas intrínsecamente indiscriminado? Un buen ejemplo sería una mina terrestre antipersona “tonta” o un gas tóxico. Se considera que el uso de ambos tipos de armas es indiscriminado, ya que al ser utilizadas, no pueden distinguir correctamente entre un objetivo militar legítimo y uno ilegítimo. La mina terrestre matará o mutilará a cualquiera que tenga la mala suerte de pisarla, incluso décadas después. Crear un Sistema de Armas Autónomo que pueda discriminar con precisión es una tarea difícil, pero no insuperable.

---

23. Si bien es cierto que algunos elementos del derecho internacional todavía se basan en normas y costumbres, existen actualmente reglas que se basan en tratados, como las que se especifican en los Convenios de Ginebra (1949) y en el Estatuto de la Corte Penal Internacional. Antes del establecimiento de órganos como la ICC, la mayoría de las naciones únicamente contaban con garantías de guerra o alianzas de defensa de apoyo mutuo para hacer cumplir las normas y costumbres aceptadas y algunas naciones, incluyendo Estados Unidos, no son parte de la ICC.

## 2 Objeciones contra los aws: argumentos a favor de su prohibición

Desde la invención de la ballesta en el siglo XIII<sup>24</sup>, han surgido llamamientos para que se limite el armamento aceptable en la “guerra civilizada”. Los avances tecnológicos han presionado para que las naciones busquen lo último en armas de vanguardia para obtener incluso la más mínima ventaja sobre su enemigo. Estos desarrollos han incluido la ametralladora, la mina terrestre y los gases tóxicos. Algunos de los proyectos, como las ametralladoras, se han mantenido dentro de los límites del material militar aceptable. Otros, como la mina terrestre, han sido regulados fuertemente y su uso se ha desalentado a nivel internacional<sup>25</sup>. La tercera categoría de armas, incluyendo los gases tóxicos, ha quedado expresamente prohibida bajo un tratado internacional.

A partir del 2013, la Campaña para detener a los robots asesinos ha promovido la prohibición total del desarrollo y empleo de los sistemas de armas autónomos letales<sup>26</sup>. Este grupo ha abogado a nivel nacional e internacional por una coalición de estados miembros de la ONU, organizaciones no gubernamentales, empresas y particulares con el explícito propósito de prohibir los Sistemas de Armas Autónomos letales. Al igual que a una cantidad de personas en el mundo académico, a la campaña le preocupa que el Sistema de Armas Autónomo llegue a cruzar el “umbral de lo moral”, sacrificando la dignidad de la vida humana en aras de la eficiencia de la guerra automatizada. Los argumentos a favor de la prohibición de los “robots asesinos” pueden agruparse prácticamente en dos categorías, el argumento de la “responsabilidad” y el argumento de la “agencia”.

### 2.1 El argumento de ‘responsabilidad’: ¿A quién se puede responsabilizar?

Una de las mayores preocupaciones de quienes están en contra del uso de los AWS es que podamos crear robots que puedan apuntar y matar huma-

---

24. Del derecho canónico emitido después del Segundo Concilio de Letrán: “Prohibimos bajo anatema ese arte asesino de ballesteros y arqueros, que es aborrecible para Dios, para ser empleado contra cristianos y católicos de ahora en adelante”. (Encíclicas papales en línea, “El Segundo Concilio de Letrán”).

25. Las minas terrestres antipersonales están explícitamente prohibidas en virtud del tratado de Quebec: las minas antitanque y antibuque siguen siendo aceptables. Son aceptables las minas que tienen una corta vida útil (aquéllas que son inertes después de X horas de despliegue).

26. “La Campaña para detener a los robots asesinos exige una prohibición preventiva e integral sobre el desarrollo, la producción y el uso de armas totalmente autónomas, también conocidas como sistemas de armas autónomos letales o robots asesinos” (Campaña para detener a los robots asesinos, <https://www.stopkillerrobots.org/learn/>).



nos sin que se pueda responsabilizar “debidamente” a un agente moral. Al problema de adjudicar debidamente la responsabilidad de una acción tomada por un Sistema de Armas Autónomo, algunas personas le han llamado una “brecha de responsabilidad”.<sup>27</sup> Para mayor seguridad, si se puede rastrear una acción de un Sistema de Armas Autónomo directamente hasta la persona que autorizó la misión (por ejemplo, un comandante deshonesto que programó las órdenes para matar a civiles desarmados y a soldados enemigos heridos), por lo tanto, pareciera evidente que podamos atribuir con justicia la culpa de cualquier muerte que resulte, al comandante que autorizó la misión. Sin embargo, el problema surge cuando no podemos trazar directamente una línea desde la persona que autorizó el Sistema de Armas Autónomo, hasta el crimen de guerra que se perpetró. Esta es la “brecha de responsabilidad”. Bonnie Docherty, una de las defensoras de la prohibición de los AWS considera que “no sería justo ni legalmente viable...responsabilizar a un comandante u operador [humano]”.<sup>28</sup> Ella no es la única que cuestiona el hecho de no fincar responsabilidad al agente —el Sistema de Armas Autónomo— que cometió la acción moralmente problemática (o ilegal). En tanto que argumenta una brecha de responsabilidad, Robert Sparrow analiza dos posibilidades: transferir la culpa del Sistema de Armas Autónomo a los humanos (el programador y el oficial al mando) y dejar al Sistema de Armas Autónomo como el agente responsable.

En primer lugar, Sparrow argumenta que podríamos descargar, al menos en parte, la responsabilidad de un Sistema de Armas Autónomo descarriado sobre el programador o diseñador que construyó el dispositivo: “esto solo será justo si la situación previamente descrita ocurriera como resultado de la negligencia por parte del equipo de diseño/programación”.<sup>29</sup> Sin embargo, Sparrow descarta rápidamente que se responsabilice al programador en aquellos casos donde no exista negligencia. La primera razón para descartar el hecho de culpar al programador es que “la posibilidad de que la máquina ataque los objetivos equivocados, puede ser una limitación aceptada del sistema”.<sup>30</sup> Si la posibilidad de que la máquina pueda atacar a un objetivo no autorizado es un “efecto secundario” o una “limitación del sistema” y el programador comunica explícitamente esta limitación a cualquier usuario potencial, entonces no sería justo responsabilizar al programador. Sparrow argumenta que, en este caso, quienes todavía opten por implementar el

---

27. Scharre 2018, p. 261

28. Ibid, p 261.

29. Sparrow, 2007, p. 69.

30. Sparrow, 2007, p. 69.

sistema, con las limitaciones conocidas o los posibles efectos secundarios, tendrían que asumir el peso de la responsabilidad. Añade Posteriormente que “la posibilidad de que un sistema autónomo tome decisiones distintas a las previstas y apoyadas por sus programadores, es inherente a la afirmación de que es autónomo”.<sup>31</sup> En este sentido, Sparrow señala que debido al diseño operativo del Sistema de Armas Autónomo y a la naturaleza compleja de sus posibles algoritmos y redes neuronales, ningún humano puede predecir en todos los casos, las decisiones de un Sistema de Armas Autónomo. Si este fuera el caso, nuevamente, sería injusto culpar al programador, ya que no pudo haber previsto todas las acciones del Sistema de Armas Autónomo.

Ya que no podemos hacer que el programador del Sistema de Armas Autónomo se responsabilice de sus acciones, entonces posiblemente podemos pedir cuentas al oficial al mando que ordenó su uso.<sup>32</sup> Señala: “Se acepta el riesgo de que el Sistema de Armas Autónomo pueda equivocarse cuando se toma la decisión de enviarlo a la acción. Este es el enfoque que prefieren las fuerzas militares que buscan desplegar los actuales AWS”.<sup>33</sup> Sparrow sostiene que si tratamos a los AWS de la misma manera que tratamos a las armas “tontas” e “inteligentes” convencionales de hoy, parecería entonces que podemos sencillamente atribuir la culpa al comandante que ordenó su uso. Sin embargo, esta táctica no considera la naturaleza plena de la autonomía del arma. Sparrow considera que la naturaleza misma de la autonomía del arma significa que los comandantes no pueden controlar qué objetivos selecciona el Sistema de Armas Autónomo y por lo tanto, no deberían rendir cuentas por las elecciones del mismo: “Si las máquinas realmente eligen sus propios objetivos, entonces no podemos responsabilizar al comandante en jefe de las muertes que se causen”.<sup>34</sup> Si el programador y el oficial al mando no pueden rendir cuentas, ¿se puede responsabilizar al propio Sistema de Armas Autónomo?

Sparrow sostiene que “la autonomía y la responsabilidad moral van de la mano”.<sup>35</sup> Si llegamos a un punto en el que el Sistema de Armas Autónomo es totalmente capaz de tomar decisiones letales por sí solo (sin un humano

---

31. Ibid, p. 70.

32. Sparrow no define a qué nivel de mando uno deba rendir cuentas. Si, por ejemplo, el comandante de una compañía de infantería (un Capitán (O3) en el ejército y el cuerpo de Marines de los Estados Unidos) decide enviar un Sistema de Armas Autónomo a una población y mata a un niño, ¿quién debería ser responsable? ¿El capitán? ¿Su jefe, un comandante de batallón (teniente coronel)? O el comandante de brigada, el comandante de división, el cuerpo de Marines, etc. Por principio de cuentas ¿debe únicamente responsabilizarse a los de mayor rango por poner este dispositivo en la fuerza?

33. Sparrow, 2007, p. 70.

34. Sparrow, 2007, p. 71.

35. Ibid, p. 65.

en el circuito), entonces tendría que asumir la responsabilidad moral de sus acciones, lo que nos lleva al meollo del problema. “Es difícil tomar en serio la idea de que una máquina debería o podría, ser considerada responsable de las consecuencias de ‘sus’ acciones”.<sup>36</sup> Suponiendo que responsabilizar a un agente por una mala acción implica un castigo para éste, Sparrow sostiene que para que “los actos sirvan como castigo, deben provocar el tipo de respuesta correcta en su objeto”.<sup>37</sup> Entonces, ¿cuál sería el tipo de respuesta correcta de Sparrow? Según él, la única respuesta que se adecúa al castigo implica sufrimiento, por lo que para que un Sistema de Armas Autónomo sea castigado “debe poderse decir que éste sufra”.<sup>38</sup> Pero esta noción de sufrimiento complica la manera de entender la funcionalidad y las capacidades del Sistema de Armas Autónomo como se interpretan e imaginan actualmente. Es poco probable que un programador pueda o vaya a incorporar en un sistema de este tipo, la capacidad de generar respuestas emocionales como el sufrimiento (o miedo) porque tales emociones contravendrían uno de los beneficios potenciales que éste brinda sobre los humanos, es decir, que a los sistemas no se les induce para cometer crímenes de guerra por ira, odio, sed de venganza, etc.<sup>39</sup> Dado que un Sistema de Armas Autónomo no tiene la capacidad de experimentar dolor físico o emocional, es probable que no pueda sufrir y por lo tanto, según Sparrow, no podría ser castigado.

En última instancia, Sparrow considera que a menos que podamos crear éticamente a un Sistema de Armas Autónomo que pueda ser “debidamente responsable”,<sup>40</sup> no es ético desplegar estas armas en el campo de batalla. Señala que la capacidad de decidir quitar una vida debe quedar en manos de un agente que sea capaz de asumir la responsabilidad moral y del que pueda decirse legítimamente que va a ser castigado, en caso de que viole una de las Leyes de la Guerra.

## 2.2 El argumento de la “agencia”: ¿Se debe castigar a la agencia humana hasta el extremo de la muerte?

Uno de los elementos cruciales de la guerra es la decisión de matar a un ser humano. Desde la época de los inicios de la guerra hasta la llegada de las

---

36. Ibid, p. 71.

37. Ibid, p. 71.

38. Ibid, p. 72.

39. Ronald Arkin argumenta que la falta de emociones específicas (miedo, ira, tristeza, etc.) de los AWS podría beneficiar su uso. Estos AWS no mostrarían las respuestas emocionales típicas que se ven en los humanos en el campo de batalla y por lo tanto, es más probable que tengan reacciones “correctas” de acuerdo con la ética que los rige.

40. Sparrow, 2007, p. 66.

armas de fuego, la mayor parte de estas muertes se daban a corta distancia y de persona a persona. Los combatientes enemigos podían verse a los ojos mientras luchaban por sus vidas. El teniente coronel Dave Grossman, en su libro “*On Killing*”, estudia los efectos que ha tenido en nuestra capacidad colectiva de matar, el aumento en la distancia en la acción de matar. Cuando estamos cara a cara con nuestro enemigo, ellos son humanos y podemos entendernos al mismo nivel. Consecuentemente, el acto de matar es más íntimo y más brutal. Cuando matamos a distancia, nuestro enemigo pareciera ser un poco menos humano y, por lo tanto, es más fácil justificar su muerte. La mayor distancia entre los combatientes y el acto de matar es eliminar la decisión de matar a un agente humano y permitir que esa decisión la tome un agente artificial. Alex Leveringhaus sostiene que “el reemplazo de la agencia humana con la ‘agencia artificial’ de las máquinas en una guerra, es sumamente problemático”.<sup>41</sup> El debate fundamental es entonces, si deberíamos renunciar a la agencia humana en favor de una artificial.

Leveringhaus afirma que el primer problema que se presenta al ceder la hegemonía del agente humano sobre la decisión de matar, es que se pierde la “equivalencia moral de los soldados”. El principio de equivalencia moral, formulado por Michael Walzer, establece que desde la perspectiva de las leyes de la guerra, cada grupo de soldados (los agresores y los defensores) tiene el mismo valor moral, por lo que todos tienen igual derecho de usar la fuerza para matar al enemigo y para defenderse.<sup>42</sup> Esta equivalencia moral se mantiene, según Walzer, independientemente de la justicia de la guerra que se libra. Todos tienen el derecho moral a luchar. Cuando una parte de la igualación se reemplaza con un Sistema de Armas Autónomo, ya no existe entonces una equivalencia moral, debido a que éste no tiene el derecho moral a la autodefensa y como tal, pierde su igualdad ante la oposición.

El segundo problema que aborda Leveringhaus, es la falta de capacidad del Sistema de Armas Autónomo para emitir un “juicio moral”. Esta falta de capacidad es de extrema importancia cuando el sistema determina el nivel de fuerza a utilizar porque “...la aplicación del criterio de proporcionalidad en el IHL implica emitir *juicios morales*”.<sup>43</sup> Si un Sistema de Armas Autónomo no puede emitir un juicio moral sobre la proporción correcta de fuerza que se requiere en una determinada situación, ¿debe ser capaz en primer lugar, de tomar la decisión moral definitiva de quitar una vida? Leveringhaus señala que

---

41. Leveringhaus, 2016, p. 2.

42. Walzer, 2015, p. 34.

43. Leveringhaus, 2016, p. 2.

“no”, y argumenta que la capacidad de matar sin la capacidad de deliberar sobre las consecuencias morales de esa acción es básicamente, una falta de respeto a los derechos humanos y a la dignidad del individuo que está siendo atacado.

El punto decisivo del argumento de Leveringhaus es la controversia de que al usar un Sistema de Armas Autónomo para matar a otro ser humano, estamos eliminando cualquier posibilidad de misericordia de la cadena de exterminio y por lo tanto, nos distanciamos moralmente del acto de matar de manera que no se respeta la dignidad de la persona que se mata. En la cadena de exterminio, la misericordia significa que en cualquier momento, un soldado puede optar por *no* apretar el gatillo. Leveringhaus afirma: “Los robots asesinos, como agentes artificiales, carecen del componente central de una agencia humana que es la capacidad de hacer lo contrario: no disparar al objetivo”.<sup>44</sup> Esta falta de capacidad para *no* matar a un combatiente autorizado, es por lo tanto y por su misma naturaleza, irrespetuosa de la dignidad inherente de todo ser humano. “El enemigo en calidad de blanco lógico no tiene derecho a que los soldados atacantes no lo maten, pero ciertamente, la capacidad humana para tener piedad con, sentir compasión por o empatía con otros humanos, incluso si éstos pertenecen a un país contrario, es de relevancia moral y digna de protección”.<sup>45</sup> El problema que Leveringhaus argumenta con un Sistema de Armas Autónomo que ejecuta una acción letal sobre un humano, es que carece fundamentalmente de un sentido de compasión y comprensión del valor de la vida humana. Por lo que escribe: “Considero que retener la agencia humana al grado de aplicar la fuerza, protegiendo en esta forma la libertad de no apretar el gatillo, presionar el botón o lanzar una granada, es fundamental para que se mantenga nuestra humanidad, precisamente en la situación que más la desafía: la guerra”.<sup>46</sup> Si un Sistema de Armas Autónomo no tiene la opción de no matar a un enemigo, entonces carece de la capacidad de mostrar misericordia, la cual se requiere para respetar la dignidad humana. Leveringhaus cierra con un mensaje inquietante, que opaca el optimismo que se ha visto en muchos desarrolladores e investigadores de tecnología de armas avanzadas “Los robots asesinos no nos rescatarán de la condición humana”.<sup>47</sup>

---

44. Leveringhaus, 2016, p. 9. Leveringhaus utiliza el término robots asesinos en vez de Sistema de Armas Autónomos. Esta parece ser una táctica emotiva para convencer al lector del dilema moral que presenta el uso de tales dispositivos. Considero que esto distrae de la naturaleza objetiva de su argumento y presenta en cambio, un argumento emotivo y más débil.

45. Ibid, p. 10.

46. Ibid, p. 10.

47. Ibid, p. 15.

### 3. Contraargumentos

Existe una relación necesaria entre la muerte y la guerra, como lo señala J. Glenn Gray: “en la guerra, la muerte regularmente la causan los miembros de mi propia especie que buscan activamente mi fin, a pesar del hecho de quizás nunca me hayan visto y no tengan motivo para una enemistad a muerte”.<sup>48</sup> Este hecho nos lleva a preguntarnos: ¿cómo podemos entonces mitigar el sufrimiento de quienes participan activamente en la guerra y el de los espectadores inocentes que han quedado atrapados en su furia? Parte de la respuesta *puede* estar en la implementación de la inteligencia artificial y los sistemas de armas autónomos en el campo de batalla del mañana. En lo que resta de esta tesis, argumentaré, al igual que lo han hecho muchos otros en los sectores de seguridad nacional y defensa<sup>49</sup>, que es muy posible que veamos a los AWS operando en conjunto con los soldados del mañana. Si bien es cierto que existen muchas funciones que podrían transferirse más fácilmente a un sistema autónomo (como conducir convoyes logísticos por rutas peligrosas o procesar trámites de recursos humanos), dentro del ejército existen tareas que es poco probable que alguna vez se automaticen por completo. En cambio, es factible que veamos una combinación de humanos y máquinas acoplados de tal manera que aprovechen al máximo la inteligencia, durabilidad, fuerza y potencia del soldado humano. Sostendré que cada uno de los argumentos a favor de una prohibición que se han presentado hasta ese punto, no brindan una razón suficiente para la implementación de dicha prohibición. Lo que, en todo caso, pueden apoyar lógicamente es la implementación de un sistema de regulación, mediante el cual las partes internacionales puedan monitorear y guiar el desarrollo de dichos sistemas, de manera que continúen siendo compatibles con el IHL y estén permitidos moralmente.

#### 3.1 Respuesta al argumento de la ‘responsabilidad’

Aunque a simple vista el argumento de Sparrow puede ser convincente, en algunos puntos presenta supuestos con débil fundamento al tiempo que ignora las cadenas causales de responsabilidad como se interpretan actualmente en el ejército. Abordo primero la opinión de Sparrow sobre la posibilidad de culpar al programador del AWS y al oficial al mando que implementa el sistema y posteriormente, trato el argumento de Sparrow de que sería imposible castigar a un Sistema de Armas Autónomo.

---

48. Gray 1959, p. 100.

49. Consulte el informe técnico del general de división Mick Ryan (ejército australiano) “Equipo humano- máquina para las fuerzas terrestres del futuro. [https://csbaonline.org/uploads/documents/Human\\_Machine\\_Teaming\\_FinalFormat.pdf](https://csbaonline.org/uploads/documents/Human_Machine_Teaming_FinalFormat.pdf).

En términos generales, Sparrow está en lo correcto al ser cauto en el hecho de adjudicar responsabilidad moral al programador del Sistema de Armas Autónomo. Esta adjudicación equivaldría probablemente a responsabilizar moralmente al fabricante de un rifle por su uso durante un tiroteo escolar; si bien al hacerlo puede mitigar los sentimientos con respecto al uso que de esa manera se le da al arma, no ayuda a comprender quién es moralmente responsable. Sparrow argumenta que podríamos responsabilizar a un programador si la fechoría se debió a alguna negligencia en la programación, de la misma manera en que un fabricante es responsable hoy, por las fallas de sus productos que resultan en daños o muerte al usuario final. Sin embargo, lo que no debe suceder, es que se responsabilice a un programador porque consideramos que el uso de su producto sea moralmente ofensivo. Sobre la cuestión de responsabilizar al programador, coincido con Sparrow en que solo deberíamos responsabilizar a alguien si cometió un error que fue de cierta manera, negligente.

Sparrow señala que no debemos responsabilizar a los comandantes de las acciones de los AWS bajo su mando, porque este sistema es completamente autónomo y que, como tal, no podemos estar seguros de qué acciones o motivaciones tomarán los AWS para ese efecto. Lo anterior, hace caso omiso de algunos de los principios básicos de liderazgo y de responsabilidad que son fundamentales para la cultura militar norteamericana actual.

El concepto de responsabilidad es un principio importantísimo de la cultura militar. Todos los miembros del servicio están investidos de un cierto nivel de responsabilidad de acuerdo con su rango y posición. Todos tienen la responsabilidad primordial de apegar su conducta a la ley, a la regulación militar y a la tradición. A medida que uno asciende en la estructura de rango de la división a la cual está adscrito, se confieren mayores niveles de responsabilidad, que incluyen deberes tanto para los recursos del personal como materiales. La cúspide de la responsabilidad militar se logra obteniendo el puesto de “comandante”. Dependiendo de la división, un comandante será responsable en los diferentes niveles, del personal y del material (conforme al tamaño de la unidad), pero los principios básicos continúan siendo los mismos. El manual sobre liderazgo del Ejército de los Estados Unidos<sup>50</sup> especifica que “el mando incluye la autoridad y la responsabilidad de utilizar eficazmente los recursos disponibles y de planificar el trabajo de organizar, dirigir, coordinar y controlar las fuerzas militares para la realización de las

---

50. Publicación en Referencia de la Doctrina Militar 6-22 Liderazgo militar.

misiones asignadas”.<sup>51</sup> La parte importante en este punto, es la responsabilidad que tiene un comandante. Un comandante es responsable de todas las acciones que sus soldados tomen o dejen de tomar en un determinado momento. Los comandantes del ejército norteamericano aceptan, a través de su asignación en un puesto de rango superior, la carga de responsabilidad por las acciones de sus subordinados. Por ende, los comandantes serían al menos parcialmente responsables de las fallas y de la actividad problemática de los AWS, incluso si el evento se diera sin un control humano directo.

El estándar Yamashita se ha establecido como un precedente legal a partir de la Segunda Guerra Mundial. En los juicios militares que precedieron al fin de la guerra en el Pacífico, el general Tomoyuki Yamashita fue juzgado por los crímenes en contra de la humanidad, perpetrados por las tropas bajo su mando. El general Yamashita argumentó que no podía haberse enterado de todas las atrocidades que se estaban cometiendo ni podía esperarse que ejerciera alguna forma de control directo sobre sus tropas ampliamente dispersas, para evitar tales actos. El tribunal consideró este argumento poco convincente y señaló:

La ley de la Guerra impone sobre un comandante del ejército el deber de tomar aquellas medidas apropiadas que estén dentro de su autoridad, para controlar las tropas bajo su mando con el fin de evitar actos que se consideren violaciones de la ley de la Guerra y mediante los cuales es probable que se vigile la ocupación del territorio hostil por tropas militares no controladas y que él pueda ser acusado de responsabilidad personal por no tomar dichas medidas cuando se presentaron las violaciones.<sup>52</sup>

El veredicto que se dictó en el caso Yamashita establece rotundamente que los comandantes, a cualquier nivel, tienen el deber de evitar que sucedan atrocidades en las unidades bajo su mando. No obstante, existe una réplica que establece que: “Un comandante es responsable principalmente de las acciones de sus tropas, pero no podemos responsabilizarlo totalmente de las acciones de un “Sistema de Armas Autónomo que fue diseñado en forma no deliberada y del cual básicamente se desconoce su proceso de toma de decisiones. “Mi respuesta es que no consideremos responsables únicamente a los comandantes, por lo menos en parte, de las acciones de sus subordinados,

---

51. ADRP 6-22, 2012, p. 1-3

52. En referencia a Yamashita.



independientemente del proceso de toma de decisiones de los mismos, sino que también consideremos responsables al comandante por promover un entorno donde dichas opciones son vistas como válidas. Primero, el ejército básicamente responsabiliza a sus oficiales al mando, de las acciones de todos aquéllos que estén a su cargo, *independiente* de si pueden o no controlarlos físicamente. Analicemos por ejemplo la reciente avalancha de accidentes navales relevantes de la 7ª Flota Naval de los Estados Unidos (apostada en Japón). Los capitanes de ambos navíos fueron encontrados responsables (por lo menos uno de ellos fue puesto en manos de los fiscales por homicidio imprudencial), aunque no estuvieron físicamente al mando del timón e incluso en el puente de mando al momento del incidente.<sup>53</sup> La Fuerza Naval determinó que fueron negligentes en su tarea de asegurar que el barco operara de manera apropiada. La Marina también determinó que a bordo de estos navíos y a nivel de mando de flota (cuartel general de la flota), los comandantes *promovieron un entorno* donde las normas laxas y fáciles, eran vistas como aceptables<sup>54</sup>. Si sobre la estructura legal militar vigente, consideramos de nuevo el Sistema de Armas Autónomo, resulta lógico asumir que si se le permitió cometer un crimen de guerra o *no se le evitó* que lo hiciera, eso podría ocasionar que se considerara responsable al comandante, tanto legal como moralmente.

Si estamos dispuestos a responsabilizar de manera colectiva a un comandante, por lo menos parcialmente, por la conducta de su tropa, quien está en sí constituida por los mismos agentes autónomos, entonces no existe un motivo legal o moral para *no* considerarlos responsables con el mismo grado de responsabilidad de las acciones o la falta de acciones de un Sistema de Armas Autónomo. Sparrow considera que sería injusto responsabilizar a un comandante, ya que éste probablemente no pueda predecir la conducta del Sistema de Armas Autónomo, sin embargo, en este punto surge la pregunta sobre si el comandante puede predecir la conducta de sus tropas en el campo de batalla. Los comandantes entrenan de manera regular y activa a sus soldados para que sus acciones puedan ser *más* predecibles durante una situación estresante, pero esto no excluye la posibilidad de que un soldado humano pueda cruzar el límite y cometer un crimen de guerra. El roboticista Ronald Arkin apunta:

---

53. También debe observarse que como resultado de estos incidentes, fueron obligados a renunciar los comandantes de mayor rango, incluyendo al capitán de la 7ª flota.

54. Ver los informes de la Marina de los Estados Unidos tanto en el accidente del USS Fitzgerald como en el USS John S. McCain: <https://www.secnav.navy.mil/foia/readingroom/HotTopics/CNO%20USS%20Fitzgerald%20and%20USS%20John%20S%20McCain%20Response/CNO%20USS%20Fitzgerald%20and%20USS%20John%20S%20McCain%20Response.pdf>.

“De manera personal, yo no confío en la perspectiva de apartarse de la regla (La Ley de la Guerra (LoW) o de las Reglas de Enfrentamiento (RoE) por el agente autónomo (AWS) en sí, por lo que obliga la cuestión sobre si existe algún tipo de responsabilidad. Sin embargo, existe la posibilidad de que un humano asuma la responsabilidad por tal divergencia, en caso de que alguna vez se considerara apropiada.”<sup>55</sup> En este sentido, Arkin argumenta que cuando diseñamos y construimos un Sistema de Armas Autónomo, nunca debemos permitirle la capacidad de decidir por su cuenta, apartándose de las LoW o RoE (lo cual sería su programación básica) y de preferencia, dejar esa responsabilidad únicamente en manos de un humano. De ser así, esto transferiría la responsabilidad de los Sistemas de Armas Autónomos, violando sus LoW o RoE, sobre un humano que tome decisiones (específicamente, un oficial al mando de algún nivel).<sup>56</sup> Si un comandante humano, no obstante, lejos del punto de la decisión final de matar, decidió enviar a un Sistema de Armas Autónomo a una situación donde se sabe que existe la posibilidad de que suceda demasiado daño colateral, en este caso pareciera ser moralmente correcto, responsabilizar al comandante cuando se presentan los incidentes.

En este punto, alguien podría preocuparse y con justa razón por el hecho de que no he aportado las bases suficientes o los fundamentos convincentes para afirmar que el estándar Yamashita nos proporciona la suficiente *justificación* moral para sustentar la responsabilidad en el cargo de un comandante. Argumento que, en vez de buscar una justificación de fondo, debemos abordar este problema reconociendo que el estándar Yamashita ofrece una razón que se puede *justificar* moralmente para responsabilizar a un comandante (incluso quien supervisa el uso de los AWS). ¿Cuál es la razón para que este tipo de doctrina de responsabilidad puede tener una justificación moral? Deseamos conminar a los comandantes en el campo, en época de guerra, para que se esmeren en todo lo posible por fomentar un entorno que opere dentro de y respete la Ley de la Guerra, las leyes internacionales humanitarias vigentes y las Reglas de Enfrentamiento para esa área de

---

55. Arkin, 2009, 40.

56. Creo que esto podría ser similar al uso de un determinado oficial en un conflicto actual en el que esté involucrado E.U. En este tipo de situaciones, cuando se ha determinado un objetivo, regularmente se le pide a un superior (con frecuencia a un oficial general) que valide el objetivo y autorice el combate. Este oficial generalmente cuenta con un abogado presente que le proporciona asesoría legal en la situación, pero al final la responsabilidad es de quién toma la decisión. Esto pudiera ser igual a hecho de autorizar a un Sistema de Armas Autónomo para que se desvíe de su programación previa de las LoW/RoE, si un comandante así lo decide, por lo que ellos deben autorizar y aceptar plena responsabilidad por esa acción.

operación. El estándar Yamashita (o las normas Medina actualizadas) nos ofrece una vía que puede ser *justificable* moralmente para lograr esa meta. Funciona como un recordatorio para los comandantes (a través de la cadena de mando) de que ellos son *personalmente* responsables de las acciones de sus subordinados. No obstante, esto no significa que sean *totalmente* responsables de los actos (o la falta de los mismos) de los subordinados, pero si comparten la parte de responsabilidad que *les corresponde*.<sup>57</sup>

El problema definitivo con el argumento de Sparrow es su renuencia en considerar responsable por sus delitos a un Sistema de Armas Autónomo, como agente moral, ya que uno no puede castigarlo de manera legítima porque no puede sufrir y ya que no puede sufrir, no va a ser castigado. Al parecer, basa su argumento del castigo en la idea de la justicia retributiva. El problema es que Sparrow afirma que la justicia retributiva es el rumbo adecuado a seguir (de manera explícita con los AWS y de manera implícita con los humanos). Afirma que “para que cualquiera de estas acciones (diferente tipo de castigo) sirva como castigo, debe provocar el tipo de respuesta apropiada en su objeto...ser capaz de ser castigado, por lo tanto, (la cosa o persona a ser castigada) puede decirse que sufre”.<sup>58</sup> Este deseo por el sufrimiento puede rechazarse al adoptar una de las otras posibles afirmaciones morales de vanguardia para el castigo, de que se dispone en la literatura. Podríamos, por ejemplo, querer una forma de castigo más comunicativa, que tenga la intención de comunicar la naturaleza inaceptable de un acto a otro Sistema de Armas Autónomo. También podríamos optar por un enfoque de rehabilitación para el castigo, en el cual volviéramos a programar o entrenar a los AWS para que aprendan que la acción previa fue mala o no deseada. Si insistimos en un sistema retributivo, como lo pide Sparrow, entonces él estaría en lo correcto; sería imposible castigar a los AWS. Sin embargo, no existe una razón para pensar que Sparrow esté en lo correcto en este aspecto.

Con respecto a tener la responsabilidad moral de la muerte de un soldado enemigo, Sparrow afirma: “lo menos que debemos a nuestros enemigos es admitir que sus vidas valen lo suficiente para que alguien deba aceptar la responsabilidad por sus muertes”.<sup>59</sup> Sparrow requiere que se pueda responsabilizar a los AWS como individuos, cuando pareciera que el enfoque debe

---

57. Creo que es importante que tengamos un debate sobre “la parte de responsabilidad que corresponde” pero esto queda fuera del ámbito de esta tesis.

58. Sparrow 2007, 72

59. Ibid, 67.

estar sobre el comandante que decide usar el arma o el político que inicia la guerra. Si el Sistema de Armas Autónomo funciona de manera legítima, no habrá entonces ninguna obligación de responsabilizar a los individuos –humanos o a los AWS– por matar de manera legítima en la guerra.<sup>60</sup>

Por otra parte, ¿cómo tratamos a un Sistema de Armas Autónomo que “se volvió malvado” y cometió un crimen de guerra? Sostengo que en vez de castigarlo, debemos analizar las circunstancias bajo las cuales se dio el incidente. Un evento nunca será igual a otro, a tal grado de establecer una afirmación genérica o un principio universal respecto a quien se debe culpar por lo que no es correcto. En todo caso, debemos analizar la decisión que se tomó para utilizar el arma (¿se estimó un alto grado de posibilidad de exceso de daño colateral? y la otra decisión por parte del comandante, junto con la información importante de que se disponía en ese momento, para atribuir responsabilidad en los lugares adecuados. Al final, si se puede demostrar de manera concluyente que un comandante e incluso un Sistema de Armas Autónomo hicieron todo lo posible por evitar una atrocidad, pero se dio (por ejemplo, una bomba perdida), entonces *tampoco debemos* hacerlos responsables (aplicando posiblemente la doctrina del doble efecto).

Sparrow contestaría que si nos apegamos a este plan, crearemos una brecha de responsabilidad y la única solución adecuada en esta situación, sería no usar un Sistema de Armas Autónomo como primera opción. Ante esto, sostengo que Sparrow está equivocado. Si como se mencionó previamente, aplicamos estándares tales como el Yamashita (o Medina) podemos ubicar el nexo de responsabilidad moral sobre el oficial al mando en aquellos casos donde no podamos demostrar evidentemente que el Sistema de Armas Autónomo hizo un movimiento ilegal (o inmoral). Estaríamos entonces frente a dos opciones, ya sea la opción a) “castigar” al Sistema de Armas Autónomo basándose en una diferente moral (comunicativa, rehabilitadora, etc.), o b) castigar al oficial al mando (a cualquier nivel que se considere apropiado, que pudiera no ser el de menor rango) usando los estándares Yamashita/Medina.

Como ha quedado demostrado hasta la fecha, el argumento de la responsabilidad fracasa, basándose en los precedentes de las actuales leyes internacionales, las normas de liderazgo militar norteamericano y la reconsideración

---

60. Por supuesto que alguien como McMahan no estaría de acuerdo y diría que si esos soldados o quizás AWS, que participan en una guerra injusta *son* responsables de manera individual por la muerte de un soldado enemigo porque ese otro soldado (supuestamente justo) no es un objetivo válido (Mc Mahan 2011, 14).

de la idea del castigo. En vez de proporcionar una evidencia convincente de que la comunidad internacional prohibir debe estrictamente el desarrollo y uso de los AWS, este argumento más bien nos convence de regular el desarrollo y uso de los mimos.

### 3.2 Respuesta al argumento de la 'agencia'

El argumento más contundente que se ha abordado hasta ese punto, sostiene que no se le debe pasar a un agente artificial, el control de la decisión de matar a un ser humano. Esto no debe suceder, ya que un agente artificial no tiene la capacidad de mostrar piedad al momento de la acción letal y porque de manera intrínseca, el hecho de que una persona sea asesinada por la decisión de un robot autónomo, es una falta de respeto para la dignidad humana. Ambas partes de este argumento son básicamente erróneas.

Leveringhaus insiste en que debemos permitir la opción de *no* seguir una orden legítima de matar al enemigo en el campo de batalla, básicamente la opción de *no* disparar a un auténtico combatiente y en todo caso, practicar la misericordia. Afirma que comparado con la agencia artificial, lo que hace valiosa a la agencia humana en la guerra y en la vida diaria, es la posibilidad de involucrarse en un curso alterno de acción”.<sup>61</sup> Leveringhaus considera realmente que en un campo de batalla, uno debe ser totalmente apto para tomar la decisión de no matar cuando se enfrenta a la opción que le permite ejercer la fuerza letal. El asume que si un Sistema de Armas Autónomo está programado con sus órdenes, con los parámetros de la misión y con los parámetros de las Leyes de la Guerra y las Reglas de Enfrentamiento, atacará sencillamente a todo enemigo que se encuentre. Considero que esta suposición, no es necesariamente cierta. Si seguimos el consejo de Arkin respecto al “gobernador ético”, un Sistema de Armas Autónomo que fuera diseñado adecuadamente, primero predeterminará las medidas no letales, ya sea para evadir o para contener al enemigo, y no podía determinar de inmediato las acciones letales. Si requerimos que se construya un algoritmo sobre este modelo, entonces las acciones letales se darían únicamente si fueran obligatorias. Supongamos por ejemplo que existe un Sistema de Armas Autónomo patrullando junto con un escuadrón de soldados de la infantería norteamericana, podríamos establecer como una de las RoE: “si un miembro de nuestro escuadrón está en peligro inminente de ser capturado, entonces tú estás *obligado* a usar la fuerza adecuada de manera proporcional para evitar su captura”. Si algún

---

61. Leveringhaus 2016, 9

miembro de la patrulla no se encuentra en peligro inminente de ser capturado, entonces el Sistema de Armas Autónomo debe predeterminar hacia una acción no letal, excepto que deba satisfacer otro criterio reglamentario que se hubiera definido previamente.

Al Sistema de Armas Autónomo como recurso, debe obligársele<sup>62</sup> a seguir las órdenes *legítimas* y *válidas*, pero también a no seguir las órdenes ilegítimas y no válidas. Lo segundo es tan importante como lo primero para evitar crímenes de guerra en el campo de batalla. Lo anterior también pareciera servir como “control de conducta ética” de Arkin”.<sup>63</sup> En este sentido, señala: “especialmente en el caso de los robots del campo de batalla (pero también para un soldado humano), no queremos que el agente pueda deducir sus propias convicciones respecto a las implicaciones morales del uso de la fuerza letal, sino más bien que sea capaz de aplicar aquéllas que hayan sido deducidas previamente por humanidad, conforme lo prescrito por las LoW y las RoE”.<sup>64</sup>

Por lo tanto y de acuerdo con Leveringhaus, una de las características determinantes que separan a los agentes humanos y a los artificiales, es nuestra capacidad (humana) para mostrar misericordia hacia nuestros enemigos. Para Leveringhaus, “la capacidad humana para tener misericordia con, sentir piedad por o empatizar con otros humanos, incluso si éstos pertenecen a un país opositor, es por supuesto de trascendencia moral y digna de protección”.<sup>65</sup> Sin embargo, ¿qué sucede entonces, si esta piedad no tiene la capacidad de hacer lo contrario? Pareciera que para Leveringhaus, la clave está en que podemos empatizar con nuestro enemigo; reconocer su humanidad y de esa manera, abstenernos de hacerles daño cuando se pueda o sea apropiado hacerlo; “pudo haber sido que esos soldados que no mataron lo hicieran porque reconocieron la humanidad del enemigo y se dieron cuenta de la gravedad de la decisión de apretar el gatillo”.<sup>66</sup> ¿Podríamos, no obstante, estar confundiendo lo que constituye una genuina piedad en el campo de batalla? ¿El hecho de decidir no matar a un combatiente enemigo y permitirle vivir un día más (incluso probablemente en cautiverio), implica no reconocer su humanidad? Se argumentaría que la decisión de no matar, basada en la falta

---

62. En este punto debemos definir las obligaciones de los AWS simplemente conforme a su programación para seguir las reglas escritas dentro de su código, pero de ser posible, esto puede ampliarse para incluir una visión más extensa de las obligaciones morales.

63. Arkin, 2009, 66-67

64. Ibid, 117.

65. Leveringhaus 2016, 9-10

66. Leveringhaus 206, 10.

de necesidad militar de su muerte, es una forma apropiada de compasión y de misericordia que pudiera implementarse en el campo de batalla del mañana. Lo anterior nos colocaría en primera fila, al reducir el número de matanzas y de muertes desafortunadas que ocurren a consecuencia de otra acción intencionada. De igual manera, se debe reconocer que como están las cosas actualmente, los humanos no cuentan con un buen historial, mostrando compasión a gran escala en el campo de batalla (aunque esto se dé a nivel personal). Es frecuente que la maquinaria de propaganda de un país invente al enemigo en un “otro” o en algún otro ser que elimine su humanidad, haciéndolo así más fácil de matar (recuerdo cómo se llamaba a los alemanes “hunos” durante la Segunda Guerra Mundial o “vietcon-gés” a los vietnamitas). Si somos capaces de crear un Sistema de Armas Autónomo bajo una estricta regulación internacional, que se apegue a la ley internacional y respete la vida de una persona (matando únicamente a aquéllos que se considere permitido matar legalmente y conforme a las necesidades militares), podríamos mostrar *mayor* misericordia en el campo de batalla de la que hasta ahora hemos hecho.

En este punto, puede presentarse una posible réplica a mi objeción junto a las vías de factibilidad. ¿Resulta lógico pensar que podemos programar de manera factible un Sistema de Armas Autónomo “para que muestre misericordia” hacia los combatientes enemigos siempre que sea posible? Esta es una vía de argumentación sólida que debemos continuar abordando durante el desarrollo de dichos sistemas. Se podría argumentar que podemos “simplemente”<sup>67</sup> escribir en el algoritmo rector del Sistema de Armas Autónomo que “siempre que usted se encuentre con un soldado enemigo que no presente una amenaza directa para usted o para cualquier soldado humano en las cercanías directas, muestre misericordia no usando una acción letal para neutralizarlo”. Esto sería tan solo una táctica posible, pero que se debe analizar a fondo. Retomando de nuevo a Ronald Arkin y a su proyecto teórico sobre construir un algoritmo que rija ese tipo de máquina, pareciera que a falta de una opción no-letal como primer curso de acción, estaría intrínseca la misericordia. En vez de que un soldado humano optara por no matar al enemigo (utilizando supuestamente medidas no letales para capturarlo y no solo dejándolo ir en libertad), tendríamos un Sistema de Armas Autónomo tardo para capturar al soldado enemigo usando tácticas no letales y humanas.<sup>68</sup>

---

67. Lo que confirma totalmente que esto no sería una tarea fácil.

68. Se negaría todo este proyecto si el Sistema de Armas Autónomo usó tácticas no letales pero *inhumanas*. Tácticas como éstas (quizás apalear al enemigo hasta quedar inconsciente)

El argumento final de Leveringhaus analiza la necesidad de preservar la dignidad humana en la decisión de matar. Leveringhaus afirma que esta dignidad únicamente se mantiene, si un agente humano toma la mejor decisión al momento de la muerte. Esto parece esquivar el tema sobre si se requiere la decisión de matar de un agente humano para mantener la dignidad y mi respuesta sería que no. Paul Scharre hace hincapié, al señalar:

Cuando desde la perspectiva del soldado al ser matado en el campo de batalla, se considera esto (la muerte digna) es una crítica poco común y casi descabellada de las armas autónomas. No existe una tradición legal, ética o histórica de los combatientes que les permita a sus enemigos el derecho de morir con una muerte digna en la guerra. No hay nada digno en ser abatido por una ametralladora, volado en pedazos por una bomba, quemado vivo en una explosión, ahogado en un barco que se hunde, sofocándose lentamente de una herida de bala en el pecho, o cualquier otra manera horrible de morir en la guerra.<sup>69</sup>

El deseo de que solo se den muertes dignas en la guerra parece ser un sutil intento de esterilizar el combate al decir, “por lo menos el enemigo tuvo una muerte digna”. Aunque para la mayoría no resulte cómodo, debemos reconocer que a nivel básico, la guerra se define por la muerte y la acción de matar. Por lo tanto, nuestro meta debe ser reducir la cantidad de matanzas y de sufrimiento infringido sobre las personas (tanto combatientes como no combatientes); sin embargo, exigir que la “muerte con dignidad” sea un requisito, rechaza la agresividad de la guerra. Es totalmente aceptable que se exija que los países vayan a la guerra únicamente cuando dicha acción se justifique moralmente o que por lo menos así sea, pero exigir que solo maten de manera digna resulta ideal en extremo y al mismo tiempo, hace caso omiso de la cruel realidad de la más mortífera de las relaciones humanas.

En la guerra podría esperarse que la muerte más digna<sup>70</sup> fuera aquella que causara el menor sufrimiento posible. Una muerte “limpia” por así decirlo, que extinga de manera rápida y efectiva la vida de un combatiente, sin causarle dolor y sufrimiento permanentes pudiera ser lógicamente, la más deseada. Lo anterior, no debe considerarse como una forma de subestimar

---

parecen ser tan moralmente sospechosas como dispararle a cada sujeto enemigo que uno se encuentre (que no representen una amenaza real)

69. Scharre 2018, 288.

70. Si acaso fuera realmente posible esperar tal cosa en un combate.



la importancia de la muerte en el campo de batalla a algún tipo de sistema de masacre con eficiente crueldad parecido a un matadero, sino más bien como un deseo de terminar las hostilidades a la brevedad posible, sin dolor ni sufrimiento innecesarios. Como lo menciona Paul Scharre (basado en su propia experiencia personal en la guerra), resulta absurdo insistir que la muerte en “manos” de una máquina sería, de alguna manera, menos digna que las distintas y diversas maneras en que uno puede morir en la guerra moderna.

En vez de preocuparse por una “dignidad percibida en la manera de morir” (ya que no queda claro si uno pudiera ser capaz de tener una muerte digna) debemos enfocarnos en reducir el sufrimiento en la guerra. Los sistemas de armas autónomos ofrecen una posible manera de lograrlo. Como lo argumento en la siguiente sección, nuestra meta en la guerra debe ser reducir el dolor y el sufrimiento innecesarios. Por lo tanto, debemos diseñar AWS encaminados a lograr esta meta y rechazar como tal, la afirmación de Leveringhaus de que la muerte por los AWS es intrínsecamente indigna.

#### 4. Los aws como opción lógica en la evolución de la guerra

El propósito de cualquier fuerza moralmente honesta debe ser limitar las muertes de los no combatientes y de los civiles en el campo de batalla. El IHL requiere que los ejércitos de cada estado, independientemente de que haya o no haya ratificado los Tratados de Ginebra, tengan el cuidado y las precauciones razonables de no matar o lastimar a las personas de las clases protegidas y de no causar daños innecesarios ni destruir propiedades (especialmente de las clases protegidas). En este aspecto, los ejércitos a nivel mundial, han desarrollado “armas inteligentes”. Los sistemas de armas autónomos son el paso más lógico a seguir de estas armas inteligentes que como mínimo, tienen un beneficio doble sobre las supuestas “armas tontas” 1) pueden ser dirigidos hacia un blanco en específico (regularmente con márgenes de error de unos cuantos metros) y 2) ofrecen beneficios en términos económicos. Ahora bien, estas dos razones influyen en el deseo de limitar el daño colateral innecesario en la mayor medida posible. Si se puede apuntar a un área, con un margen mínimo de error, podemos entonces usar pocas armas para neutralizar la amenaza, limitando así la exposición al peligro tanto para los civiles como para las propiedades circundantes. Este cálculo se apega a la arquitectura fundamental que Arkin insta que incluyan en su diseño quienes investigan y desarrollan los AWS.<sup>71</sup> Este diseño presionaría hacia una “masacre efectiva en el campo de

---

71.  $A+B+C+D$ = Masacre en el campo de batalla.  $A$ = Combatientes planeados,  $B$ = Fuerzas amigables no planeadas  $C$ = No Combatientes planeados,  $D$ = No Combatientes

batalla” o al grado mínimo posible de daño colateral. Segundo, en términos económicos, las armas inteligentes ofrecen mayor beneficio comparadas con las armas “tontas”. Esto pareciera contradecir a la lógica, debido a que muchas de las armas inteligentes cuestan cientos de miles e incluso millones de dólares por pieza. La lógica es que si puedo lograr la misma misión con pocas armas (que pueden costar más si se adquieren en forma individual) que con algunas municiones “tontas” relativamente baratas (que en conjunto pueden costar más que el arma inteligente); por lo tanto, usar el arma inteligente ofrece más beneficios en términos económicos. Lo anterior parece tener sentido. Si puedo usar un sólo misil guiado para deshacerme de un enorme búnker, fuertemente resguardado, que anteriormente hubiera requerido de 200 bombas no dirigidas, entonces debo usar obviamente, el único misil guiado inteligente. Si este caso aplica con armas como los misiles y las bombas lanzadas desde el aire, entonces también sostiene que si un Sistema de Armas Autónomo puede hacer el trabajo de diez soldados humanos, por lo tanto, su uso resulta más viable económicamente. Si un Sistema de Armas Autónomo puede ayudar a que los ejércitos reduzcan los costos económicos, así como el sufrimiento de los no combatientes, existen por lo tanto, motivos de consideración para concluir que el curso razonable a seguir, es el desarrollo y uso de dichos sistemas.

Arkin y otras personas consideran que el uso de los AWS (y en general de IA) en el campo de batalla pudiera ser el rumbo moral a seguir debido a un conjunto de características inherentes a su naturaleza. Primero, el AWS puede actuar de manera conservadora.<sup>72</sup> Un Sistema de Armas Autónomo puede (y posiblemente deba ser programado) para ser determinado hacia una acción no letal. Lo anterior se debe principalmente a la falta de una necesidad de supervivencia.<sup>73</sup> Al renunciar al deseo innato de mantenerse vivo y libre de daño, un Sistema de Armas Autónomo puede tomar acciones no letales más desafiantes (tales como salir de un área peligrosa que fuera difícil para un humano) o sacrificándose en sí, por el buen cumplimiento de la misión con el propósito de desviar la atención del enemigo del resto de su unidad. Un Sistema de Armas Autónomo no se dejaría confundir por las emociones y los consiguientes pronósticos de auto realización que turban comúnmente el criterio de los agentes humanos.<sup>74</sup> Los crímenes

---

No Planeados. El objetivo debería ser maximizar A, mientras eliminamos B & C y nos aseguramos de que D esté lo más cerca posible de 0 (Arkin 2009, 128)

72. Arkin 2009, 29

73. Ibid, 29

74. Ibid, 30

de guerra se cometen frecuentemente en el ardor de la batalla cuando las emociones están en su máximo nivel y cuando el miedo o el enojo nublan el juicio. Conforme lo comenta J Glenn Gray, un soldado “se convierte en un luchador, un *Homo furens*”.<sup>75</sup> Si se crea un Sistema de Armas Autónomo sin tales emociones que turben el juicio, como sería el mejor curso lógico y moral, ya no sería, por lo tanto, un “esclavo” del miedo o del enojo, ni actuaría agresivamente porque otro Sistema de Armas Autónomo o un miembro del escuadrón humano fueron asesinados en una población local. Básicamente, podría ser un mejor agente ético y moral de lo que nosotros somos en la batalla. En consecuencia, Arkins señala: “No considero que un sistema autónomo no tripulado pueda ser totalmente ético en el campo de batalla, sin embargo, estoy convencido de que puede desempeñarse de manera más ética que los soldados humanos”.<sup>76</sup>

Si como Arkin afirma, podemos diseñar un Sistema de Armas Autónomo con la capacidad para actuar mejor tanto ética como moralmente (o al menos a la par) comparado con los soldados humanos; por lo tanto, el rumbo más racional que se debe tomar, es desarrollar ese sistema. Si un Sistema de Armas Autónomo puede por lo menos ser tan ético y moral como el mejor de nosotros en la guerra, tenemos entonces, razones morales convincentes para construir dicho dispositivo. Primero, se disminuiría el número de nuestros propios soldados que se requieren que estén de alguna forma, frente al peligro. En este punto se reconoce, que no todos los soldados humanos pueden ser reemplazados por los AWS, más sin duda, se podría reemplazar a un buen número de ellos. Si puedo desplegar un sistema autónomo que pueda bajar en una autopista cargada de artefactos explosivos improvisados (IED, por sus siglas en inglés) para entregar mis suministros, sin arriesgar las vidas de los camioneros, entonces estoy obligado moralmente a hacerlo. Por el contrario, si decido no tomar dicha acción cuando dispongo de tecnología, entonces, estoy de hecho, actuando de manera inmoral. Segundo, si un Sistema de Armas Autónomo puede tomar mejores decisiones éticas y morales que los humanos en el campo de batalla, entonces estoy obligado a desplegar dicho sistema. Si, y éste es un gran si, dicho sistema puede ser construido y estar a mi disposición como comandante y si puede tomar decisiones que sean moral y éticamente más correctas (tales como diferenciar objetivos adecuadamente, evitar el exceso de muertes de

---

75. Gray 1959, 27

76. Arkin 2009, 31.

los objetivos, reducir el daño colateral) que los humanos, entonces estoy moralmente obligado a desplegar ese sistema. Este despliegue no libera a los comandantes que decidan usarlo (en este aspecto, probablemente de mayor rango que los comandantes de compañía y de batallón a nivel táctico<sup>77</sup>) del peso moral de matar, ni los absuelve de cualquier daño colateral excesivo que se relacione.<sup>78</sup> Tal y como Arkin lo mencionó, no existen muchas probabilidades de que, incluso con la mejor tecnología, podamos reducir de manera eficaz el daño colateral a cero. Sin embargo, como comando, estaría moralmente obligado a desplegar un Sistema de Armas Autónomo que pudiera tomar mejores decisiones morales que los humanos, quienes por lo demás, estarían usando la fuerza letal.

Al final, el curso de acción razonable para desarrollar y desplegar un sistema de armas es lo que pueda reducir la cantidad de masacre en el campo de batalla. Esta perspectiva no busca de forma alguna, sanear la guerra, ya que los combatientes reales seguirán muriendo y como lo señala Leveringhaus, “el objetivo enemigo real como tal, no espera que no lo maten los soldados que lo atacan”.<sup>79</sup>

## 5. Propuesta: un esquema para regular los aws

Considerando que ya he abordado los argumentos más persuasivos que se han utilizado en contra del uso de los AWS en el campo de batalla del mañana y argumentado los motivos por los que estamos obligados moralmente a utilizar tales armas, se requiere crear un sistema teórico de gobernanza que pueda regular el desarrollo y despliegue de esos sistemas a nivel internacional. Como se señaló anteriormente, no existe actualmente

---

77. En el ejército de EUA, una compañía es la unidad estándar de menor tamaño con un comandante designado (aunque hay unidades más pequeñas llamadas destacamentos, pero son excepciones a la regla). Las compañías varían en fuerza con personal desde 50 hasta más de 300. Una compañía es entonces, la unidad táctica de nivel básico a la cual le otorgamos liderazgo y responsabilidad. Una compañía es comandada por un capitán (0 a 3) que, si no tiene experiencia de reclutamiento previa, tiene entre 4-6 años en servicio. Un batallón (por lo menos uno estándar) está conformado por cinco a siete compañías y generalmente tiene un personal de 700 a 1500 dependiendo de cómo esté integrado y de la misión estipulada para la unidad. Se considera que un batallón está todavía a nivel táctico. Un batallón está generalmente bajo el mando de un teniente coronel (O a 5) con aproximadamente 17 a 19 años de servicio.

78. Es físicamente imposible limitar TODO el daño colateral cuando la guerra se lleva a cabo donde viven personas. Este es un hecho aceptado. Sin embargo, la meta es REDUCIR o LIMITAR la cantidad de daño colateral. De esta forma, el daño colateral en exceso es visto como un crimen de guerra, mientras que el daño colateral incidental no.

79. Leveringhaus 2016, 10.

un Derecho Internacional Humanitario que se enfoque específicamente en el desarrollo y despliegue de los AWS. Esta sección aspira a ser el punto de partida para esa discusión y con ese fin, recomiendo tres áreas amplias de interés para el desarrollo y despliegue de los AWS. Estas tres áreas incluyen las dudas más apremiantes que uno pueda tener con respecto a los AWS. Ciertamente, lo anterior no comprenderá todas las posibilidades en el uso del sistema, ya que más bien tiene como fin, servir como una infraestructura teórica, lo suficientemente amplia, para guiar la elaboración de la ley internacional. Las tres áreas que enfatizo en este esquema teórico incluyen 1) despliegue apropiado de los AWS, 2) cumplimiento con el IHL actual y futuro y 3) establecimiento de una cadena de responsabilidad.

En primer lugar, el problema es: ¿cuándo resulta apropiado desplegar dicho sistema? El IHL debe establecer algunas leyes en las que se estipule que no debe desplegarse ningún Sistema de Armas Autónomo por ningún ejército o nación, únicamente hasta que los expertos técnicos relacionados con el mismo, hayan certificado que el sistema ha sido programado para adaptar su conducta con el IHL y que cualquier comandante que despliegue ese sistema, sin estar certificado, sea castigado independiente de que el sistema cometa realmente crímenes de guerra. Este requisito no sería ajeno para los ejércitos en la actualidad, ya que deben pasar por el mismo tipo de proceso de certificación para asegurarse de que los sistemas de armas estándar o convencionales (por decir, un tanque) sean diseñados y funcionen dentro de los límites legales. Una vez que un ejército o una nación ha pasado este umbral, debe entonces tomar otra decisión, la cual consiste en dos partes separadas, pero casi iguales 1) cuándo usar fuerza no letal contra fuerza letal (despliegue táctico/operativo) y 2) cuándo desplegar un Sistema de Armas Autónomo (despliegue estratégico). A continuación, abordo cada uno de estos problemas.

La decisión de utilizar un Sistema de Armas Autónomo en una situación táctica, significa únicamente, que el comandante en la base decide usarlo para una determinada misión. Se acepta por lo general, que estos comandantes operen a “nivel táctico” y tengan la libertad de decidir sobre el uso directo de tropas y armamento necesarios para lograr cierta misión u orden. En este sentido, un comandante de compañía tendría la autoridad de mando para decidir emplear el vehículo aéreo no tripulado de la compañía, con el fin de tener un mayor espacio visual, sin embargo, no tendría autoridad para dirigir un recurso de nivel estratégico (como un satélite) para lograr su misión. No obstante, cuando se habla de un Sistema de Armas Autónomo, un comandante de nivel táctico (incluso hasta un comandante de Brigada que generalmente

está al mando de una tropa aproximada de 5,000) debe tener la responsabilidad de mando y la autoridad para decidir si es apropiado usar ese tipo de arma. Por ejemplo, si un comandante de compañía es responsable de proteger a un pueblo, en ese momento debe tomar la decisión (tal vez consultando con órdenes de nivel superior) sobre el uso de un arma o una táctica específicos. Lo mismo aplica para un Sistema de Armas Autónomo. Para controlar el uso apropiado de fuerza, el Derecho Internacional Humanitario debe especificar que el uso táctico de un Sistema de Armas Autónomo debe garantizar que se va a reducir al mínimo, el posible daño colateral como resultado de su uso.

En segundo lugar, se abordará la cuestión respecto al despliegue estratégico del Sistema de Armas Autónomo y por despliegue estratégico me refiero, a si es moral o legalmente aceptable usar dicha arma en una determinada guerra o conflicto. Para solucionar esta inquietud, parece bastante sencillo legislar que el uso del AWS deba coincidir con los principios de la Teoría de la Guerra Justa y con el IHL, ¿habría más probabilidades de que el despliegue de un Sistema de Armas Autónomo no causara daños colaterales en exceso al ser desplegado en el escenario X en comparación con el escenario Y? Considerando, por ejemplo, el nivel de perfeccionamiento tecnológico en su momento, pudiera no ser ético desplegarlo en un escenario de combate urbanizado en extremo (Por ej. Singapur) cuando se compara con un escenario de combate rural o lejano (Por ej. la estepa rusa). Esta regulación debe establecer que el despliegue por cualquier nación o ejército, de un Sistema de Armas Autónomo, es inadmisibles en un escenario de combate donde sea mayor la probabilidad del exceso de daño colateral, del que pudiera esperarse por el uso de fuerzas humanas. Por ejemplo, si su simple uso en un escenario ocasionara probablemente, el uso indiscriminado de fuerza debido a que el Sistema de Armas Autónomo (y posiblemente, un humano) tendría una dificultad enorme para diferenciar de manera adecuada entre los combatientes y los no combatientes, por lo tanto, su uso constituiría un crimen de guerra, incluso si no se equivocara al diferenciar a los verdaderos combatientes. No existe motivo alguno para considerar que, cuando un Sistema de Armas Autónomo se equivoca, no elaboraríamos algún tipo de evaluación de responsabilidad, aunque finalmente, a quien decidamos culpar por el error, pueda tener una apariencia distinta de la actual. Si un soldado comete un error en nuestros días, (un error que resulte en un crimen de guerra), ese individuo soldado y su cadena de mando pueden ser así, considerados responsables.<sup>80</sup>

---

80. Esto ciertamente supondría que el error fue algún tipo de acción deliberada o que se produjo por algún tipo de negligencia grave, los detalles en este punto no son particular-

Por lo tanto, la segunda amplia categoría de regulación debe incluir alguna estipulación con el fin de que “un Sistema de Armas Autónomo no pueda ser desplegado, en caso de que, debido a su diseño, no pueda apearse a los principios básicos de la Teoría de la Guerra Justa y a los actuales o futuros preceptos del IHL”. Como se mencionó anteriormente, emplear un sistema que pueda considerarse responsable por medio de su programación, sería intrínsecamente incorrecto tanto moral como legalmente. A dicho sistema le puede faltar la capacidad de que se le audite de cierta forma (como tener un registro de incendio o de desempeño para poder revisar todas las formas en que dicho sistema se involucró), o puede simplemente determinar el uso de la acción letal como primera respuesta apropiada. Aunque no sería propio del IHL ordenar la tecnología precisa que pudiera o no usarse, algo parecido al gobernador ético de Arkin, sería el mejor ejemplo de un tipo de sistema que pudiera cumplir con esta regulación teórica. Cualquier sistema que, como tal, cumpla con el propósito de la regulación y pueda de esta forma, apearse a la Teoría de la Guerra Justa y a la IHL debe considerarse como aceptable.

La última área de enfoque para esta ejemplificación futura del IHL, es que debería existir una cadena de responsabilidad establecida. El futuro IHL debe prever que “en caso de que se despliegue un Sistema de Armas Autónomo, un comandante o líder político debe asumir la responsabilidad moral y legal por la decisión estratégica de emplear ese sistema de armas”. La ley también debe estipular que a nivel táctico, “un comandante que decida utilizar dicha arma es responsable tanto moral como legalmente por todo lo que la misma haga o deje de hacer”. Quienes se oponen a esas regulaciones podrán argumentar que se impone una carga demasiado pesada en los mandos militares, a lo que respondo: la carga debe ser pesada, ya que esto impide el libre uso de dichos sistemas en situaciones en las que es posible, ya sea infringir las Leyes de la Teoría Justa o el IHL, o en su defecto, aumentar el daño colateral a niveles que no sean aceptables. Las regulaciones pueden tener un posible efecto positivo de limitar el uso de los AWS únicamente en aquellas situaciones donde los resultados se puedan controlar de manera razonable o en situaciones donde la probabilidad de daño colateral sea baja. (Por ejemplo, la regulación puede disuadir el uso de los AWS en entornos urbanos con población densa, pero puede aumentar su uso en áreas virtualmente despobladas, como en mar abierto).

Aunque existe una ventaja al elaborar una ley internacional que aliente al desarrollo y al uso moralmente correcto de los AWS, debe existir un

mecanismo apropiado para obligar a su cumplimiento. Tal mecanismo puede ser como el que estableció la ICC o una nueva forma de observancia a nivel internacional. Las regulaciones legales que he propuesto, no dependen de ningún método de observancia en particular y son compatibles con una forma de mecanismo de cumplimiento creada recientemente para “obligar” a las naciones a desarrollar sus propios AWS, de tal manera que cumplan con las regulaciones. No me corresponde decidir la manera en que debe funcionar este mecanismo, en todo caso, abogar por un sistema que pueda, de acuerdo con sus posibilidades, evitar los crímenes de guerra *antes* que se cometan.

## 6. Conclusion

El avance humano es análogo a una máquina en movimiento continuo. Sin un comienzo definido y sin un fin aparente, avanzamos sin cesar, al igual que el desarrollo de maneras mejores y más efectivas de matarse entre sí. Desde los inicios de la existencia humana, cuando el hombre se dio cuenta que el hatlatl le permitía lanzar su jabalina más lejos que su oponente, hasta los avances modernos en el campo de la inteligencia artificial, la presión por encontrar la siguiente arma más poderosa, avanza de manera inexorable. El momento de actuar es hoy, con el fin de formar un órgano de regulación que configure la manera en que nosotros, como especie, desarrollemos y despleguemos lo que indudablemente, es el subsiguiente tipo de sistema de armas.

He presentado dos de los argumentos más persuasivos en contra del uso de los sistemas de armas autónomos, a manera de argumentos de responsabilidad y de agencia. Aunque ambos argumentos persuaden de manera superficial, ninguno ofrece las bases suficientes (ya sea de manera individual o colectiva) para inclinar la balanza hacia una absoluta prohibición sobre el uso de esta incipiente tecnología. En todo caso, ambos (y los demás argumentos) ofrecen un motivo persuasivo en cuanto al porqué es necesario impulsar *regulaciones* internacionales sobre el desarrollo y el despliegue moralmente apropiado de dichos sistemas. Asimismo, he argumentado que el desarrollo y despliegue adecuados de un sistema de armas autónomos es la opción razonable, únicamente si se pueden cumplir nuestros estrictos estándares morales de conducta. Por lo tanto, el propósito es reducir el horror que se da en la guerra. Como Sherman lo señaló: “la guerra es el infierno” pero debemos hacer todo lo posible para que sea un infierno tolerable.

El temor a lo desconocido es un instinto profundamente arraigado en los humanos. Las películas de ciencia ficción no han contribuido para calmar los temores de que, si continuamos con nuestra ruta de avances tecnológi-



cos, los robots derribarán el yugo de su “opresión” y se convertirán en sus amos. Este temor, aunque pudiera ser generalizado, no es una base firme para decisiones morales o legales.

## Agradecimientos

*Me gustaría agradecer al profesorado y al personal del Departamento de Filosofía, especialmente al Dr. Andrew Altman y al Dr. Andrew I. Cohen: la presente tesis no hubiera sido posible sin su empeño. También agradezco a todos los profesores con los que he tenido el privilegio de tomar clases: gracias por aguantarme. Igualmente quiero dar las gracias a aquéllos en el ejército norteamericano que vieron una chispa en mí y pensaron que era digno de la tarea. Gracias a todos.*

## Referencias

- Arkin, Ronald. *Governing Lethal Behavior in Autonomous Systems*. Boca Raton: Chapman and Hall Imprint (Taylor and Francis Group), 2009.
- Army Doctrine Reference Publication. “6-22 Army Leadership.” *Army Doctrine Reference Publications* (2012). [https://usacac.army.mil/sites/default/files/misc/doctrine/cdg/cdg\\_resources/manuals/adrp/adrp6\\_22\\_new.pdf](https://usacac.army.mil/sites/default/files/misc/doctrine/cdg/cdg_resources/manuals/adrp/adrp6_22_new.pdf).
- Boyd, John. *A Discourse on Winning and Losing*. Maxwell Air Force Base: Air University Press, 2018.
- Campaign to Stop Killer Robots. “About Us.” Accessed 25 July 2018. <https://www.stopkillerrobots.org/about/>.
- Department of Defense. “Directive 3000.09—Autonomy in Weapon Systems.” *Office of the Deputy Secretary of Defense* (November 2012). <http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>.
- In re Yamashita 327 U.S. 1 [1946] 61 (United States Supreme Court). <https://supreme.justia.com/cases/federal/us/327/1/>.
- Government of China. “Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions of the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects.” *Convention on Certain Conventional Weapons—Position Paper* (April 2018). [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/DD1551E60648CEBBC125808A005954FA/\\$file/China's+Position+Paper.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/DD1551E60648CEBBC125808A005954FA/$file/China's+Position+Paper.pdf).
- Gray, J. Glenn. *Warriors: Reflections of Men in Battle*. New York: Harcourt, Brace and Company, 1959.
- International Criminal Court. “Rome Statute.” Article 8(2)(b)(xx) (1998). [https://www.icc-cpi.int/nr/rdonlyres/ea9aef7-5752-4f84-be94-0a655eb30e16/0/rome\\_statute\\_english.pdf](https://www.icc-cpi.int/nr/rdonlyres/ea9aef7-5752-4f84-be94-0a655eb30e16/0/rome_statute_english.pdf).
- Leveringhaus, Alex. “What’s so Bad about Killer Robots.” *Journal of Applied Philosophy*, (March 2016). DOI: 10.1111/japp.12200.
- McMahan, Jeff. *Killing in War*. Oxford: Oxford University Press, 2011.

- Military Leadership Diversity Commission. "Outreach & Recruiting." *United States Department of Defense*. November 2009. <https://diversity.defense.gov/Portals/51/Documents/Resources/Commission/docs/Issue%20Papers/Paper%2002%20-%20Requirements%20and%20Demographic%20Profile%20of%20Eligible%20Population.pdf>.
- Papal Encyclicals Online. "The Second Council of the Lateran- 1139 A.D." Accessed 18 January 2019. <http://www.papalencyclicals.net/councils/ecum10.htm>.
- Program Executive Office- Missiles & Space. "Counter-Rocket, Artillery, Mortar (C-RAM)." Accessed September 1, 2018. <https://missiledefenseadvocacy.org/defense-systems/counter-rocket-artillery-mortar-c-ram/>.
- Raytheon. "Phalanx Close-In Weapon System." Accessed 15 August 2018. <https://www.raytheon.com/capabilities/products/phalanx>.
- Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. New York: W. W. Norton & Company, 2018.
- Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy*, vol. 24, no. 1 (2007): 62–71. DOI: 10.1111/j.1468-5930.2007.00346.x.
- United Nations Institute for Disarmament Research. "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches." *UNIDIR Resources*, no. 6 (2017): 1–33. <http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf>.
- United Nations Office for Disarmament Affairs. "Perspectives on Lethal Autonomous Weapon Systems." *UNODA Occasional Papers*, no. 30 (November 2017): 1–61. [www.un.org/disarmament](http://www.un.org/disarmament).
- US Army Acquisition Support Center. "Excalibur Precision 155MM Projectiles." Accessed 25 July 2018. <https://asc.army.mil/web/portfolio-item/ammo-excalibur-xm982-m982-and-m982a1-precision-guided-extended-range-projectile/>.

# Interdiction des Systèmes d'armes Létales Autonomes<sup>1</sup>

## Une Analyse Critique

Hunter Cantrell

---

### Liste Des Abréviations

SALA- Système(s) d'armes (létales) autonomes

CPI- Cour pénale internationale

DHI- Droit humanitaire international

DG- Droit de la guerre

RE- Règles d'engagement

### I. Introduction

Une augmentation de l'emploi des systèmes d'armes entièrement autonomes —que d'aucuns annoncent comme imminente<sup>2</sup>— a suscité l'appel à une campagne pour cesser le recours aux « robots tueurs ». Cet appel a jusqu'à présent été soutenu par vingt-trois nations<sup>3</sup>, et n'a cessé de gagner en légitimité depuis le début de la campagne intitulée *Campaign to Stop Killer Robots*<sup>4</sup> en 2013<sup>5</sup>. Cette initiative est motivée en partie par la crainte d'une perspective selon laquelle tous les attributs du soldat humain, y compris la notion de vie et de mort, seraient retirés du champ de bataille pour y être remplacés par une forme d'intelligence artificielle meurtrière. Il y

---

1. *Je dédie ce travail à mon épouse tant aimée, Tiffany, et à mes deux merveilleux fils, Holden et Bryson. Sans vous, je ne serais rien.*

2. Ce qui soulève la question du consensus concernant l'imminence de la situation qui pourrait s'échelonner sur 5, 10, 50, 100 ans. Ce manque de clarté est commun dans le domaine des nouveaux systèmes d'armes létales autonomes.

3. A noter que l'ensemble des pays qui se sont joints à cette campagne n'ont probablement pas les capacités technologiques, les infrastructures ou les besoins militaires pour construire de tels équipements. Algérie, Argentine, Bolivie, Brésil, Chili, Costa Rica, Cuba, Equateur, Egypte, Ghana, Guatemala, Irak, Mexique, Nicaragua, Ouganda, Pakistan, Panama, Pérou, Venezuela, Zimbabwe, ainsi que le Vatican et la Palestine ont participé à cette campagne.

4. Campaign to Stop Killer Robots, "About Us."

5. Les participants à cette initiative se réunissent annuellement depuis 2013 au bureau des Nations Unies à Genève.

a toutefois peu de raisons de penser qu'il s'agisse d'une crainte réaliste, contrairement aux représentations apocalyptiques imposées par l'industrie cinématographique d'Hollywood, en particulier l'image de nombreux Terminators errant dans les villes.

Il existe néanmoins des perspectives éclairées s'insurgeant contre le développement et la mise en place de systèmes d'armes entièrement autonomes. Ces positions appartiennent à deux grands types d'argumentation : l'un s'attache à l'argument dit « de la responsabilité », l'autre répondant de l'argument dit « de l'agent responsable ». D'une façon générale, la *responsabilité* se définit par la chaîne de responsabilité nécessaire à l'utilisation des systèmes d'armes létales autonomes (SALA) dont les acteurs devraient endosser toute responsabilité en cas de dysfonctionnement. D'autre part, l'*agent responsable* entend mettre en question la substitution du jugement humain par celui d'une entité artificielle dans le cadre d'une action destinée à entraîner la mort ; il en étudie également les implications pour la dignité humaine et la possibilité accordée à une éventuelle grâce.

Ce mémoire entend proposer une évaluation de ces deux arguments en tant qu'ils ne suffisent pas à justifier une interdiction pure et simple des SALA. Par ailleurs, ce travail de recherche soutient que les arguments sus-cités parent au maintien d'un cadre juridique fondé sur le droit international humanitaire actuel qui visera à *réglementer*, et non à *interdire*, le développement et l'emploi des SALA. Dans ces pages, le recours aux SALA est défendu en tant que choix rationnel pour les dirigeants militaires et politiques. Y est également décrit un potentiel cadre juridique, fondé sur l'actuelle convention de Genève et le droit humanitaire international, afin de réglementer l'usage des SALA.

## 1.1 Pourquoi il est essentiel de parler des systèmes d'armes autonomes

La démocratisation du recours aux systèmes d'armes autonomes est une prochaine étape dans l'usage des technologies militaires modernes. Alors qu'on constate une rapidité accrue dans les prises de décision militaires, rapidité largement attribuable aux prouesses croissantes de l'informatisation, la capacité de l'esprit humain à suivre cette tendance reste à la traîne. La boucle OODA en est un excellent exemple. Le concept a été inventé par le colonel John Boyd, pilote de l'US Air Force et stratège de la puissance aérienne américaine.<sup>6</sup> L'acronyme recense quatre processus, OODA sig-

---

6. Boyd 1987, 383.

nifiant « observer, s'orienter, décider et agir ». Le principe de cette boucle démontre que si un pilote de chasse parvient à pénétrer dans la boucle OODA de l'ennemi, il est alors en mesure de descendre son appareil. La boucle OODA est donc un processus décisionnel très rapide par lequel une personne — ou, à l'avenir, une machine ? — sera en mesure de prendre l'initiative d'une action donnée. Il est probable que les boucles OODA continueront de se resserrer à mesure que s'imposeront de nouveaux progrès technologiques. Dans l'avenir, le resserrement de la boucle OODA pourra se prêter à une intelligence artificielle capable de traiter de nombreuses données de calcul de décisions simultanément en temps réel, c'est-à-dire beaucoup plus rapidement que ne pourrait le faire un humain. On peut déjà remarquer un élan de fabrication d'armes plus « intelligentes » et plus rapides en observant l'évolution des avions de chasse (dont deux exemples parmi tant d'autres, l'US Air Force F-35 II américain et le Chengdu J-20 chinois), de même que l'évolution des « bombes intelligentes » (obus de précision Excalibur à guidage GPS et inertiel<sup>7</sup>), ainsi que par le niveau d'utilisation accru d'outils de prise de décision assistée par ordinateur dans la planification militaire.<sup>8</sup>

L'automatisation accrue des systèmes d'armement actuels et la volonté de recourir à des armes entièrement autonomes ont quelque peu remédié à la complexité accrue de la guerre moderne et à la baisse du nombre des engagés déclarés aptes au service dans l'armée.<sup>9</sup> C'est le cas du système d'arme rapproché PHALANX (sur les croiseurs américains Aegis) qui confie à un système semi-autonome des fonctions essentielles de protection antimissile des navires en dernière ligne de défense, libérant ainsi un équipage limité en nombre qui sera affecté à d'autres tâches.<sup>10</sup> L'utilisation croissante de ces technologies est une première étape dans la mise en place de systèmes capables de sélectionner des « cibles et [de recourir] à la force

---

7. US Army Acquisition Support Center, "Excalibur Precision 155MM Projectiles."

8. Il ne s'agit pas là de confier des décisions stratégiques aux ordinateurs et autres calculateurs électroniques, mais de mettre l'informatique au service de la hiérarchie militaire en mobilisant des renseignements toujours plus nombreux, plus exacts et plus clairs. Ces systèmes informatiques comprennent des possibilités très diverses dont la localisation automatisée de cargaisons expédiées à l'international, ainsi que de lieux de bataille amis et ennemis, les systèmes de surveillance et de détection, etc. Si tant est que certaines inventions technologiques sont à l'origine de dépendances au numérique dans le civil, c'est également le cas dans l'armée.

9. Voir the Military Leadership Diversity Commission Issue Paper #2 "Outreach & Recruiting."

10. Raytheon, "Phalanx Close-In Weapon System."

sans aucune intervention ou interaction humaine ». <sup>11</sup> Comme nous l'avons vu précédemment, le déploiement des SALA est théoriquement inquiétant de nombreux points de vue, mais cette évolution technologique a également le bénéfice potentiel de réduire les imprévus ou les dommages collatéraux inhérents à la guerre telle qu'elle est pratiquée actuellement.

## 1.2 Description des systèmes d'armes autonomes

Avant de se pencher sur le débat concernant le recours aux SALA, il est essentiel de comprendre exactement ce qu'est une arme autonome. Il n'est pas étonnant que les milieux universitaires et diplomatiques ne se soient pas encore accordés sur une définition opératoire des SALA. Ainsi les définitions diplomatiques acceptées jusqu'ici varient-elles considérablement : certaines incluent des armes aujourd'hui considérées comme automatisées, mais non autonomes, alors que d'autres prennent seulement en compte les systèmes ne nécessitant aucune intervention humaine décisive. Certaines autres définitions existantes requièrent la présence de fonctionnalités ou de technologies spécifiques. Un exemple assez parlant est celui du gouvernement néerlandais qui donne la définition suivante : un SALA est « un armement qui sélectionne des cibles et engage la force de manière autonome par rapport à certains critères définis à l'avance, et à la suite d'une décision prise par un opérateur humain de déployer l'arme, en sachant qu'une fois lancée, l'attaque ne peut pas être interrompue par une intervention humaine ». <sup>12</sup> À savoir que cette définition pourrait tout à fait s'appliquer à des armes qui ne sont actuellement pas considérées comme totalement autonomes, tels les missiles de croisière. <sup>13</sup> En effet, le missile de croisière, conçu pour frapper une cible spécifique programmée (signature thermique, désignation par laser d'un soldat au sol ou autre cible prédéfinie), ne devient totalement autonome qu'une fois lancé, par un opérateur humain ayant donné l'ordre de déployer le système, et ne peut donc pas être rappelé.

---

11. ICRC 2015, 6.

12. United Nations Institute for Disarmament Research [UNIDIR] 2017, 23

13. Cet exemple peut également s'appliquer à d'autres armements dotés de munitions à guidage de précision. Il est important de remarquer que les lois en vigueur selon les Conventions de Genève interdisent l'utilisation d'armes aveugles, ce qu'on peut interpréter comme un aval accordé à l'emploi des armes à guidage de précision. En outre, les termes de ces conventions prohibent « le fait d'employer les armes, projectiles et méthodes de guerre [...] (visant à) frapper sans discrimination en violation du droit international des conflits armés, à condition que ces armes, projectiles, matières et méthodes de guerre fassent l'objet d'une interdiction générale et qu'ils soient inscrits dans une annexe au présent Statut » (Cour pénale internationale, Article 8 (2)(b)(xx) 1998).

Le gouvernement français donne sa propre définition du SALA par « la négative »<sup>14</sup> :

Les armes létales autonomes sont des systèmes entièrement autonomes...Les SALA doivent s'entendre comme impliquant l'absence totale de surveillance humaine, ce qui signifie qu'il n'y a absolument aucun lien (communication ou contrôle) avec la chaîne de commandement militaire....La plate-forme de livraison d'un SALA est techniquement en mesure de se déplacer, de s'adapter à des environnements terrestres, marins ou aériens et de cibler et de tirer un effecteur mortel (balle, missile, bombe, etc.) sans aucune intervention ou validation humaine....Le SALA possédera très probablement des capacités d'auto-apprentissage.<sup>15</sup>

La définition des SALA proposée par les Pays-Bas diffère de celle de la France qui affirme que ces systèmes d'armement ne sont pas liés à la chaîne de commandement militaire (c'est-à-dire à un opérateur humain), y compris dans la planification de mission et la prise de décision de recours à l'un de ces systèmes. Cette définition limite considérablement la notion même de SALA, car une opération humaine décisive intervient très probablement à un moment précis dans la « chaîne de frappe ». La définition donnée par la France est donc si étriquée qu'elle en devient pratiquement inutilisable en tant que base à des consensus diplomatiques et académiques concrets.<sup>16</sup>

Bien que la définition donnée par les États-Unis aux SALA inclue certains éléments empruntés à la perspective fonctionnaliste française, elle s'en détache également en distinguant les systèmes d'armes létales entièrement autonomes des semi-autonomes. Ainsi, les États-Unis (par le biais du Département de la Défense) donnent aux SALA la définition suivante :

Un système d'armes qui, une fois activé, peut sélectionner des cibles données et engager la force sans autre intervention d'un opérateur humain. Cela inclut les systèmes d'armes autonomes supervisés par l'homme et conçus pour permettre aux opérateurs humains de se substituer au fonctionnement du système d'armes,

---

14. United Nations Institute for Disarmament Research [UNIDIR] 2017, 24

15. Ibid, 24.

16. C'est également la perspective de la Chine qui considère les SALA comme ayant un niveau d'autonomie qui manque « d'intervention humaine et de contrôle durant le processus d'exécution d'une mission donnée » (gouvernement chinois, 2018).

mais qui peuvent sélectionner des cibles et engager la force sans autre intervention humaine après l'activation.<sup>17</sup>

La définition états-unienne est alors plus utile en pratique que les définitions française ou néerlandaise, car elle propose un compte-rendu détaillé des SALA en excluant implicitement des systèmes qui ne devraient pas être considérés comme tels.<sup>18</sup> Ainsi, les systèmes en place aujourd'hui qui pourraient être considérés comme automatisés ou semi-autonomes<sup>19</sup> (PHALANX, C-RAM<sup>20</sup>, etc.) n'en font pas partie, de même que les systèmes « Tire et oublie », qui pourraient inclure le missile de croisière évoqué plus haut. Il est également important de souligner que selon la définition américaine un SALA peut agir sans l'intervention d'un opérateur humain, mais *en aucun cas* n'empêche l'intervention d'un opérateur humain. C'est d'autant plus important car cette distinction est clairement conforme à la directive 3000.09 du Département de la Défense (DoD) américain, stipulant que « les systèmes d'armes autonomes et semi-autonomes doivent être conçus pour permettre aux responsables militaires et aux opérateurs d'exercer des niveaux appropriés de jugement humain quant au recours à la force ».<sup>21</sup> Cette directive indique que les États-Unis envisagent une relation complémentaire entre le jugement humain et le recours à l'intelligence artificielle et aux SALA. Cette politique « met également l'accent sur ce

---

17. United Nations Institute for Disarmament Research [UNIDIR] 2017, 30.

18. Comme il a été mentionné plus haut, les États-Unis donnent une définition de ce qui n'entre pas dans la dénomination de SALA, en l'occurrence : « un système d'arme qui, une fois activé, ne peut engager la force que de cibles ou de groupes de cibles préalablement programmées par un opérateur humain. Cela recouvre : (a) les systèmes d'armes semi-automatiques autonomes pour les fonctions d'engagement de la force, y compris, mais sans limitation imposée à l'acquisition, la poursuite et l'identification de cibles éventuelles, la facilitation du repérage visuel de cibles éventuelles pour l'opérateur humain, le classement des cibles sélectionnées par ordre de priorité, le minutage de la mise à feu, ou la mise à disposition du guidage terminal pour la focalisation sur des cibles visées, pourvu qu'un contrôle humain parraine à la décision de sélectionner des cibles isolées ainsi que des groupes de cibles sur lesquelles engager la force. (b) Mode « Fire and Forget », soit « Tire et oublie » ou verrouillage de la cible après le lancement qui retient les munitions en s'appuyant sur le TTP (tactiques, techniques et procédures) afin d'optimiser la probabilité que seules les cibles dans la liste d'acquisitions du programmeur quand celui-ci active le système sont les cibles individuelles ou groupe de cibles sélectionnées par un opérateur humain » (UNIDIR 2017, 31).

19. Il faut toutefois préciser que les systèmes PHALANX et C-RAM peuvent aussi être enclenchés sur un mode totalement autonome.

20. Program Executive Office- Missiles & Space, «Counter-Rocket, Artillery, Mortar (C-RAM).

21. Department of Defense, Directive 3000.09.



qui constitue l'autonomie au niveau *décisionnaire* plutôt que sur la présence ou l'absence de recours à une technologie particulière ». <sup>22</sup>

### 1.3 Le droit international humanitaire actuel et les SALA

Les conflits militaires modernes sont régis par un ensemble de normes communément acceptées comme justification à tout conflit militaire, à savoir que ces principes sont fondés sur l'histoire de la doctrine de la guerre juste. Avant le XXe siècle, ces principes n'étaient pas codifiés, répondant plutôt à un ensemble de coutumes et de normes établis sur les bases d'une entente mutuelle. <sup>23</sup> Par exemple, sauver un grand nombre de marins ennemis à l'issue d'une bataille navale était une action « digne d'un gentilhomme ». Sans être imposée par traité ou accord international, elle était héritée d'une pratique courante forgée au fil de siècles de guerre navale. Plus tard, cette pratique a été abandonnée avec la prolifération de la guerre sous-marine. Après la seconde guerre mondiale, dans le cadre des procès de Nuremberg, les alliés ont accusé l'amiral Donitz de crime contre l'humanité pour avoir employé la tactique nazie consistant à couler des navires ennemis sans remonter à la surface pour sauver les rescapés. Les poursuites judiciaires à son encontre ont toutefois échoué, car de par l'évolution de la guerre navale, il était devenu monnaie courante de ne pas assurer le sauvetage des rescapés de sous-marins coulés en raison de risques opérationnels extrêmes.

Le droit humanitaire international (DHI) tel qu'il s'applique dans le cadre des Conventions de Genève et de ses protocoles additionnels ne mentionne pas explicitement les SALA. Cela peut être un point positif dans l'élaboration de traités concernant ces systèmes d'armement, bien que le DHI interdise le recours à certains types d'armements. La Convention sur certaines armes classiques (CCAC) interdit l'utilisation de certaines formes d'armes frappant sans discrimination, ce qui confère un statut préférentiel aux armements guidés de précision dits armes « intelligentes » au détriment des armes aveugles dites « simples ». Afin de se conformer aux prescriptions du CCAC, il faudrait alors élaborer des SALA ne pouvant pas être employés

---

22. UNIDIR 2017, 31.

23. Bien que certains éléments de la loi internationale soient toujours fondés sur des normes et principes coutumiers, des réglementations sont désormais mises en place par traités, tels les Conventions de Genève (1949) et dans les statuts de la Cour pénale internationale. Avant la création de ces institutions par l'International Code Council (ICC), la plupart des états avaient pour seules garanties les principes de la guerre et des coopérations de défense bilatérales afin d'assurer le respect des normes et principes coutumiers, à noter que certaines nations, dont les Etats-Unis d'Amérique, ne font pas partie de l'ICC.

de manière « intrinsèquement aveugle ». Ce qui renvoie à la définition de l'arme ou du système d'armes simple : une mine antipersonnel ou un gaz toxique en sont des exemples type. Les deux catégories d'armes sont aveugles puisqu'à leur déploiement, elles sont dirigées indistinctement contre des cibles militaires légitimes et non légitimes. Une mine terrestre tuera ou mutilera tous ceux qui auront la malchance de marcher dessus, même des décennies après son déploiement. Toutefois, développer un SALA qui pourrait être dirigé contre des cibles précises est une tâche difficile, mais pas insurmontable.

## 2. Des Objections aux SALA: Arguments Pour L'interdiction

Depuis l'invention de l'arbalète au XIII<sup>e</sup> siècle<sup>24</sup> des appels ont été lancés en faveur de la limitation des armes tolérées dans le conflit armé dit civilisé. D'autre part, des avancées technologiques, dont la mitrailleuse, la mine terrestre et le gaz toxique, ont aussi poussé les nations à se procurer des armes de pointe de plus en plus perfectionnées afin de pouvoir prendre l'avantage sur l'ennemi. Certaines de ces armes, telle la mitrailleuse, font encore partie du matériel militaire dont l'emploi est en vigueur. D'autres, comme les mines terrestres, ont été fortement réglementées et leur emploi est déconseillé à l'échelle internationale.<sup>25</sup> Quant au gaz toxique, son emploi a été expressément interdit en vertu d'un traité international.

Depuis 2013, les membres de la campagne contre les robots tueurs, *Campaign to Stop Killer Robots*, plaident en faveur d'une interdiction totale de la création et de la mise en service de systèmes d'armes létales autonomes.<sup>26</sup> Cette organisation non gouvernementale œuvre au niveau étatique et international en faveur d'une coalition comprenant états membres de l'ONU, organisations non gouvernementales, entreprises et citoyens dans le but exprès d'interdire les SALA. Les acteurs de cette campagne, de même que de nombreux universitaires, s'inquiètent de voir les SALA franchir un « seuil

---

24. Droit canonique issu du deuxième concile du Latran : « Nous interdisons sous peine d'anathème que l'art meurtrier des archers et des arbalétriers, que Dieu juge haineux, soit désormais utilisé contre les catholiques et les chrétiens. »

25. Les mines antipersonnel sont expressément interdites d'après le traité du Québec : les mines antichars ainsi que les mines antinavires restent tolérées. Les mines ayant une durée de vie raccourcie (devenues inertes un certain nombre d'heures après leur déploiement) sont également tolérées.

26. « *The Campaign to Stop Killer Robots* appelle à l'interdiction préventive et totale de la création, de la fabrication et de la mise en service des armements totalement autonomes, aussi connus sous la dénomination de systèmes d'armes létales autonomes ou robots-tueurs » (*Campaign to Stop Killer Robots*, <https://www.stopkillerrobots.org/learn/>).

moral » sacrifiant la dignité de la vie humaine en vertu de l'efficacité de la guerre automatisée. À noter que les arguments en faveur de l'interdiction des « robots-tueurs » peuvent être divisés en deux catégories : d'une part, l'argument dit « de la responsabilité » et, d'autre part, celui dit « de l'agent responsable ».

## 2.1 L'argument de la « responsabilité » ou qui tenir responsable

L'une des grandes préoccupations des militants contre l'utilisation des SALA est qu'il est envisageable de créer des robots pouvant cibler et tuer des humains sans qu'un agent moral puisse être tenu dûment responsable. Le problème qui persiste quant à l'attribution d'une juste part de responsabilité à un humain pour un élément de mission mené à bien par un SALA est souvent qualifié de « rejet de responsabilité ». <sup>27</sup> S'il est possible de retracer les actions d'un SALA pour les renvoyer directement à la personne qui a autorisé la mission (par exemple, un chef militaire voyou qui aurait programmé l'ordre de tuer des civils non armés et des soldats ennemis blessés), il est alors clair que les morts occasionnées engagent à juste titre la responsabilité du chef militaire ayant autorisé la mission. Cependant, la question se pose différemment lorsqu'il est impossible de faire porter sa juste part de responsabilité à la personne qui a autorisé l'emploi du SALA pour commettre un crime de guerre. Il s'agit alors d'un véritable « rejet de responsabilité ». Bonnie Docherty, militante en faveur de l'interdiction des SALA estime qu' « [il ne serait] ni juste ni juridiquement viable de...tenir un chef militaire ou un opérateur [humain] pour responsable. » <sup>28</sup> Elle n'est d'ailleurs pas seule à trouver problématique la question du déplacement de la responsabilité de l'humain vers l'agent, soit le SALA, qui a commis l'action moralement répréhensible ou illégale. Dans son plaidoyer en faveur de la non attribution de cette responsabilité, Robert Sparrow explore tout de même deux possibilités : faire porter la responsabilité aux humains (programmeur et chef militaire) des dégâts causés par le SALA ou considérer le SALA comme agent responsable.

En premier lieu Sparrow fait valoir que la responsabilité des dégâts causés par un SALA inopérant pourrait reposer, du moins en partie, sur le programmeur ou le concepteur de l'appareil : « ce ne serait juste que si

---

27. Scharre 2018, 261.

28. Ibid, 261.

la situation décrite s'est produite à la suite d'une négligence de la part de l'équipe de conception / programmation. »<sup>29</sup> Toutefois Sparrow rejette très vite l'idée d'attribuer la responsabilité au programmeur ou au concepteur dans les cas où il n'y a pas de démonstration de négligence. La première raison de ne pas mettre en cause le programmeur est que « la possibilité d'erreur de ciblage soit une limite connue du système. »<sup>30</sup> Si le fait que la machine frappe une cible non autorisée est un « effet secondaire » ou une « limite du système » et que le programmeur communique explicitement cette limite à tout utilisateur potentiel, il ne serait pas juste de tenir le programmeur responsable. Sparrow soutient que, dans ce cas, les chefs militaires qui déploient le système, avec ses limites connues ou effets secondaires possibles, doivent endosser la responsabilité de cette action. Il ajoute ensuite que « la possibilité qu'un système autonome fasse d'autres choix que ceux prédits et encouragés par ses programmeurs est intrinsèque à l'affirmation que le système est autonome ».<sup>31</sup> Ici Sparrow souligne qu'en raison de la conception opérationnelle d'un SALA et de la nature complexe des algorithmes et des réseaux neuronaux qui le composent, aucun humain ne sera en mesure de prédire les décisions du système d'armement de manière certaine. Si tel est en effet le cas, il est aussi injuste de blâmer le programmeur qui n'aura pas pu prévoir toutes les actions du SALA.

Sparrow soutient que s'il est impossible de demander des comptes au programmeur du SALA pour les réactions fautives du système, la responsabilité du chef militaire qui a ordonné son déploiement peut alors être engagée.<sup>32</sup> Il écrit : « Le risque que les actions du SALA puissent mal tourner est toléré lorsque la décision est prise d'enclencher la frappe. C'est d'ailleurs la logique guidant les forces militaires qui entendent déployer les SALA actuels. »<sup>33</sup> Sparrow soutient qu'en traitant les SALA de la même manière que les armes conventionnelles actuelles, simples ou intelligentes,

---

29. Sparrow 2007, 69.

30. Ibid, 69.

31. Ibid, 70.

32. Sparrow ne décrit pas à quel niveau de commandement un chef militaire pourrait être tenu responsable du fait. Si, par exemple, un commandant de division d'infanterie (un capitaine (O3) de l'armée de terre ou du Corps des Marines américains) prend la décision d'utiliser un SALA dans un village causant ainsi la mort d'un enfant, qui est responsable ? Le capitaine ? Son supérieur hiérarchique, à savoir un commandant de bataillon (lieutenant-colonel) ? Le commandant de brigade, le chef de division, le Corps, etc. ? S'agirait-il d'incriminer le tenant du grade le plus élevé pour avoir avalisé l'emploi généralisé du SALA dans sa division ?

33. Sparrow 2007, 70.

on peut simplement faire endosser la responsabilité de leur utilisation au chef militaire qui en a donné l'ordre. Toutefois, cette approche ne prend pas en compte tous les aspects qui font l'autonomie de l'arme. Sparrow fait valoir que la nature même de cette autonomie implique que les chefs militaires n'ont pas le contrôle du choix de cibles effectué par le SALA, et qu'ils ne devraient donc pas être tenus responsables des choix du système : « Si les machines choisissent véritablement leurs cibles, il est alors impossible d'incriminer le chef militaire pour les morts qui en découlent ». <sup>34</sup> Or, si programmeur et chef militaire en sont exempts, la responsabilité du SALA lui-même peut-elle alors être engagée ?

Sparrow ajoute que « l'autonomie et la responsabilité morale vont de pair. » <sup>35</sup> Si les SALA devenaient pleinement capables de prendre des décisions entraînant la mort de manière entièrement autonome (sans intervention humaine), le dispositif devrait alors porter la responsabilité morale de ses actions, ce qui est le cœur du problème. « Il est difficile de prendre au sérieux l'idée que la responsabilité d'une machine devrait ou pourrait être engagée par rapport aux conséquences de « ses actions. » <sup>36</sup> En partant du fait que tout engagement de responsabilité pour une action répréhensible implique une punition pour son auteur, Sparrow souligne « qu'une sanction pour être considérée comme telle doit provoquer un certain type de réponse. » <sup>37</sup> Quel est le type de réponse adaptée à la sanction selon Sparrow ? Selon lui, c'est la souffrance, et pour qu'un SALA puisse être puni, « il doit pouvoir souffrir ». <sup>38</sup> Or, la notion de souffrance complique la perception des fonctionnalités et autres capacités des SALA telles qu'on les interprète et les imagine actuellement. Il est peu probable qu'un programmeur puisse intégrer des réponses émotionnelles telles la souffrance ou la peur à la conception d'un SALA, car de telles émotions remettraient en question l'un des avantages potentiels que le SALA offre aux humains, à savoir de ne pas pouvoir provoquer des crimes de guerre motivés par la colère, la haine, la soif de vengeance, etc. <sup>39</sup> S'il est vrai qu'un SALA n'a pas la capacité de res-

---

34. Ibid, 71.

35. Ibid, 65.

36. Ibid, 71.

37. Sparrow 2007, 71.

38. Ibid, 72.

39. Ronald Arkin avance que le manque d'émotions spécifiques des SALA (peur, colère, tristesse, etc.) serait un des avantages potentiels de leur utilisation. Les SALA ne montrant pas les réponses émotionnelles caractéristiques des humains sur le champ de bataille seraient plus aptes à prendre des décisions justes selon l'éthique les gouvernant.

sentir de la douleur physique ou émotionnelle, il n'est probablement pas en mesure de souffrir et, selon Sparrow, ne pourrait donc pas être sanctionné.

En somme, Sparrow est d'avis que s'il n'est pas concevable de créer un SALA éthique qui puisse être « tenu responsable de ses actions à juste titre »<sup>40</sup>, il est contraire à l'éthique de déployer ce type d'armement sur le champ de bataille. La décision de prendre une vie, selon lui, doit appartenir à un agent capable de responsabilité morale qui peut être puni en cas de non-respect du droit de la guerre.

## 2.2 L'argument dit « de l'agent responsable » : faut-il renoncer à engager la responsabilité humaine face à la mort ?

La décision de tuer un être humain est un élément intrinsèque aux conflits armés. Jusqu'à l'avènement de l'arme à feu, la plupart des morts était historiquement causée par des combats personnels entraînant une grande proximité physique. Les ennemis pouvaient se regarder dans les yeux, chacun luttant pour défendre sa vie. Dans son livre intitulé *On Killing*, le lieutenant-colonel Dave Grossman étudie les effets que la distance physique croissante par rapport à l'acte meurtrier peut avoir sur la capacité à ôter la vie. En étant face à face avec son ennemi, on se rend compte que celui-ci est humain et qu'il est possible de s'identifier à lui, même très partiellement. Dans ce cas, le meurtre est plus intime et plus brutal. Quand il s'agit de tuer à distance, l'ennemi paraît déshumanisé ce qui rend plus facile de trouver des raisons à l'acte de tuer. La distance la plus importante qu'il puisse y avoir entre le combattant et l'acte meurtrier réside dans le retrait pur et simple du pouvoir décisionnaire au combattant, la décision de tuer un agent humain revenant alors à un agent artificiel. Alex Leveringhaus soutient que « dans un conflit armé, le remplacement de l'agent humain responsable par l'agent artificiel que constitue la machine est tout à fait problématique. »<sup>41</sup> Le débat est critique quant à savoir s'il faut renoncer à la responsabilité humaine pour mettre en avant celle de l'armement.

Leveringhaus prétend que le problème majeur posé par l'abandon de l'hégémonie de l'agent humain sur la décision de tuer pourrait causer une perte d'« équivalence morale des soldats ». Formulé par Michael Walzer, le principe d'équivalence morale stipule que les soldats de chaque côté du conflit, dont parmi eux, attaquants et défenseurs, ont tous la même valeur

---

40. Sparrow 2007, 66.

41. 40 Leveringhaus 2016, 2.

morale du point de vue des lois de la guerre, ce qui confère à chacun un droit équivalent de recourir à la force pour tuer l'ennemi et se défendre.<sup>42</sup> Selon Walzer, cette équivalence morale existe indépendamment de la juste raison du conflit en cours : chacun doit conserver le droit moral de se battre. Toutefois, lorsqu'une des deux parties présidant au conflit est remplacée par un SALA, la notion d'équivalence morale disparaît. En effet, le SALA n'a pas de prétention morale à la légitime défense et, à ce titre, ne saurait être sur un pied d'égalité avec l'humain.

Leveringhaus aborde un deuxième problème, à savoir l'incapacité d'un SALA à exercer un « jugement moral ». Cette incapacité est particulièrement notable lorsque le système d'armement détermine le niveau de force à employer parce que «...l'application du critère de proportionnalité [dans le DHI] implique d'avoir à faire des jugements moraux. »<sup>43</sup> Comme un SALA n'est pas en mesure de porter un jugement moral sur la proportionnalité des moyens adaptée à une situation donnée, devrait-il pouvoir prendre la plus finale des décisions morales, en l'occurrence celle d'ôter la vie ? Leveringhaus déclare que non en soulignant que la capacité de tuer sans pouvoir mettre en balance les conséquences morales de cette action est, en substance, un manque de respect aux droits de l'homme et à la dignité de l'individu visé.

Le nœud de l'argument défendu par Leveringhaus repose sur l'affirmation qu'en employant un SALA pour tuer un être humain, toute possibilité de pitié dans la chaîne de mise à mort est purement annihilée. Selon lui, cela donne lieu à la mise à distance morale d'un meurtre qui constitue un manque de respect à la dignité et aux droits humains des personnes décédées. La notion de pitié dans la chaîne de mise à mort signifie qu'à tout moment un soldat peut choisir de ne pas appuyer sur la détente. À ce sujet, Leveringhaus écrit : « Il manque aux robots tueurs en tant qu'agents artificiels une composante centrale caractéristique de l'agent humain, à savoir la capacité de prendre une autre décision : celle de ne pas tirer sur la cible choisie. »<sup>44</sup> Cette incapacité des robots à épargner un combattant ciblé manque de respect à la dignité inhérente à chaque être humain. « L'ennemi, en tant que cible légitime, ne prétend pas demander sa grâce aux soldats attaquants. Toutefois, la capacité

---

42. Walzer 2015, 34.

43. Leveringhaus 2016, 2.

44. Leveringhaus 2016, 9. Leveringhaus emploie délibérément l'expression robots-tueurs au lieu de faire référence à des systèmes d'armes létales autonomes. C'est une subjectivisation émotive qui vise à convaincre le lecteur du dilemme moral que présentent ces dispositifs. Je suis d'avis qu'il s'agit d'un moyen de s'éloigner de la nature intrinsèquement objective de ce débat, et qui contribue à sa dilution.

humaine de gracier, la capacité à éprouver de la pitié ou à compatir avec les autres humains, même s'ils appartiennent au camp opposé, est sans nul doute moralement pertinente et mérite d'être défendue. »<sup>45</sup> Leveringhaus soulève un problème au sujet des SALA opérant une action mortelle sur un être humain : il manque fondamentalement au robot une capacité intrinsèque à la compassion et à la conscience du prix de la vie humaine. Ainsi, il écrit : « Je pense qu'il est essentiel de conserver le libre arbitre humain au point d'engagement de la force, protégeant ainsi la liberté de ne pas appuyer sur la détente, actionner un bouton ou lancer une grenade, ceci dans le but de préserver notre humanité au cœur de la situation où elle est le plus facilement mise à l'épreuve : la guerre. »<sup>46</sup> Si un SALA n'est pas en capacité de choix d'épargner un ennemi, c'est qu'il n'a pas la capacité de faire preuve de compassion ou de miséricorde, et c'est bien cette capacité-là qui est nécessaire au respect de la dignité humaine. Leveringhaus conclut son argument d'un message dérangeant qui s'inscrit contre l'optimisme de certains créateurs et chercheurs en technologie des armements de pointe : « Les robots-tueurs ne nous sauveront pas des affres de la condition humaine. »<sup>47</sup>

### 3. Rejet des Arguments Précédents

Comme le souligne J. Glenn Gray, il existe une relation nécessaire entre la mort et la guerre : « la mort en temps de guerre est communément causée par des membres de mon espèce qui veulent activement ma fin, en dépit du fait qu'ils n'ont peut-être jamais eu de contact avec moi et n'ont aucune raison personnelle à une inimitié causant la mort. »<sup>48</sup> Cet état de fait incite à se demander comment il est alors possible de réduire la souffrance de ceux qui sont activement engagés dans un conflit armé et des civils innocents pris dans la fureur de la guerre. Un élément de réponse pourrait se trouver dans l'implémentation de l'intelligence artificielle et de systèmes d'armes autonomes sur le champ de bataille de demain. Dans la suite de ce mémoire, et comme beaucoup l'ont déjà fait dans les secteurs de la sécurité nationale et de la défense<sup>49</sup>, je soutiens l'idée que l'avenir verra probablement les SALA opérer en tandem avec les soldats. S'il est vrai

---

45. Ibid, 10.

46. Ibid, 10.

47. Ibid, 15.

48. Gray 1959, 100.

49. Voir livre blanc General Mick Ryan's (armée australienne) "Human-Machine Teaming for Future Ground Forces" [https://csbaonline.org/uploads/documents/Human\\_Machine\\_Teaming\\_FinalFormat.pdf](https://csbaonline.org/uploads/documents/Human_Machine_Teaming_FinalFormat.pdf).



qu'il existe au sein de l'armée de nombreux éléments de mission pouvant facilement être confiés à un système autonome (telles que la conduite de convois logistiques le long d'itinéraires dangereux ou le traitement des formalités administratives des ressources humaines), certains éléments ne seront probablement jamais automatisés dans leur intégralité. En revanche, il est probable qu'une association d'humains et de machines complémentaires pourrait optimiser l'intelligence, l'endurance, la force et la puissance de feu du soldat humain. Je soutiens que les arguments qui ont été présentés jusqu'à présent en faveur d'une interdiction de l'emploi des SALA ne donnent pas de raison convaincante à la mise en œuvre d'une telle interdiction. Tout au contraire, ces arguments peuvent logiquement défendre la mise en place d'un système de réglementation par lequel les nations pourront suivre et encadrer le développement de ces systèmes autonomes de manière à ce que leurs attributions soient compatibles avec le DHI et acceptables d'un point de vue moral.

### 3.1 Contre-réponse à l'argument dit « de la responsabilité »

Bien que l'argumentation de Sparrow soit convaincante à première vue, elle est en fait nourrie de suppositions précaires et ignore les chaînes causales de responsabilité telles qu'elles s'entendent actuellement dans l'armée. Tout d'abord, nous nous pencherons sur la perspective de Sparrow quant à la possibilité de blâmer le programmeur du SALA et le chef militaire qui déploie le système, puis nous évoquerons son idée selon laquelle il serait impossible de sanctionner un SALA.

Sparrow est du côté de la raison quand il hésite à attribuer au programmeur du SALA une responsabilité morale équivalant plus ou moins à tenir le fabricant d'un fusil moralement responsable de l'emploi qui en est fait lors d'une fusillade dans un établissement scolaire. Si faire porter cette responsabilité à un fabricant d'arme peut apaiser les sentiments liés à cet emploi fautif de l'arme, cela ne résout pas la question de la responsabilité morale. Certes, Sparrow établit qu'il est possible de tenir le programmeur responsable si un dysfonctionnement grave était dû à une négligence de programmation, de la même façon qu'un fabricant est aujourd'hui tenu responsable des défauts d'un produit ayant causé des dommages ou la mort de son utilisateur. Il faut toutefois éviter qu'un programmeur soit incriminé parce que l'utilisation de son produit peut être jugée moralement répugnante. En ce qui concerne la responsabilité du programmeur, on peut accorder à Sparrow que celui-ci n'est responsable qu'en cas d'erreur par négligence.

Sparrow poursuit son argumentation en disant que les chefs militaires n'ont pas à endosser la responsabilité des actions d'un SALA survenues sous leur commandement, car ce système d'armes est entièrement autonome et, qu'en tant que tel, ses actions ou motivations peuvent être obscures. Pourtant, cette affirmation fait fi de certains principes fondamentaux de leadership et de responsabilité qui sont à la base de la culture militaire américaine actuelle.

Par exemple, le concept de responsabilité est un principe fondamental de la culture au sein des forces armées. Tout militaire est imprégné d'un certain niveau de responsabilité proportionnel à son grade et à son poste. Chacun a la responsabilité majeure de conformer sa conduite à la loi, aux règlements militaires et à la tradition. Au fur et à mesure que l'on gravit les échelons hiérarchiques d'une armée, on acquiert des niveaux de responsabilité croissants, à la fois envers le personnel et les ressources matérielles à disposition. Le sommet de la responsabilité militaire consiste à atteindre le grade de commandant. D'une armée à l'autre, un commandant est responsable du personnel et du matériel à différents niveaux, en fonction de la taille de l'unité, mais les principes de base restent les mêmes entre toutes les armées. Le manuel de leadership des forces armées américaines <sup>50</sup> précise que « le commandement inclut l'autorité et la responsabilité d'utiliser efficacement les ressources disponibles, et donc de planifier l'emploi, l'organisation, la direction, la coordination et le contrôle des forces militaires engagées pour mener à bien les missions assignées à l'armée. »

<sup>51</sup> Ce qui nous intéresse ici est la définition des responsabilités du commandant. Celui-ci est responsable en tout temps de chaque action entreprise ou omise par ses soldats. De par leur poste d'autorité supérieur, les commandants de l'armée américaine assument la lourde responsabilité des actions de leurs subordonnés. En cela, les commandants pourraient être au moins partiellement responsables des ratés et des dysfonctionnements des SALA même si un incident se produit sans contrôle humain direct.

La norme Yamashita a été établie comme précédent juridique à la suite de la Seconde Guerre mondiale. Lors des procès militaires qui ont suivi la guerre du Pacifique, le général Tomoyuki Yamashita a été jugé pour les crimes contre l'humanité perpétrés par les troupes placées sous son commandement. Le général Yamashita a fait valoir qu'il ne pouvait ni connaître l'étendue des atrocités qui avaient été commises, ni exercer un contrôle

---

50. Army Doctrine Reference Publication 6-22 Army Leadership.

51. ADRP 6-22 2012, 1-3

direct sur des troupes déployées dans une zone géographique très large afin de pouvoir empêcher de tels actes. Le tribunal trouvant cet argument peu convaincant a déclaré :

Le droit de la guerre impose au commandant d'une armée le devoir de prendre les mesures adaptées en son pouvoir afin de contrôler les troupes placées sous son commandement et prévenir les actes allant à l'encontre du droit de la guerre et susceptibles de participer à l'occupation de territoires hostiles par une soldatesque incontrôlée, et par là même, sa responsabilité personnelle peut être engagée pour avoir omis de prendre de telles mesures en cas d'infraction au droit de la guerre.<sup>52</sup>

Le verdict rendu dans l'affaire Yamashita établit catégoriquement que les chefs militaires, à tous les niveaux, ont le devoir d'empêcher que des atrocités ne se produisent dans les unités placées sous leur commandement. Il existe cependant un possible contrepoint à ce verdict, à savoir qu'un chef militaire est globalement responsable des actions de ses troupes, mais ne pourrait être raisonnablement tenu responsable des actions d'un SALA « conçu sans négligence » et dont le processus décisionnel est essentiellement inconnu. J'ajouterai que non seulement les chefs militaires sont tenus au moins partiellement responsables des actions de leurs subordonnés (quel que soit le processus décisionnel du subordonné), mais qu'ils sont également responsables de promouvoir un environnement dans lequel les choix du subordonné pourront être considérés comme légitimes. Premièrement, l'armée considère un commandant comme responsable des actions de tous ses subordonnés dans la hiérarchie militaire, qu'il puisse ou non exercer un contrôle direct sur eux. Prenons, par exemple, la récente vague d'importants accidents navals survenus dans la 7e flotte de la marine américaine stationnée au Japon. Les commandants des deux vaisseaux ont été jugés coupables (l'un d'entre eux a même été renvoyé aux assises pour homicide involontaire), en dépit du fait qu'ils n'étaient pas physiquement à la barre, ni même sur le pont au moment de l'incident.<sup>53</sup> La Marine a conclu avoir fait preuve de négligence quant à son devoir d'assurer le bon fonctionnement de ses équipements. On a aussi pu déterminer qu'à bord de ces navires, et au niveau du commandement de la flotte (quartier

---

52. In re Yamashita.

53. Il est important de noter que des chefs militaires de rang supérieur, comme le commandant de la 7e division, ont été forcés de présenter leur démission en conséquence de ces incidents.

général), les hauts gradés encourageaient un environnement où des normes laxistes et le recours à des expédients étaient considérés comme acceptables.<sup>54</sup> Pour revenir aux SALA, selon la structure juridique militaire actuelle, il est raisonnable de supposer que si un système était autorisé à commettre un crime de guerre ou n'en était pas empêché, le commandant pourrait être à la fois juridiquement et moralement responsable.

Si l'on est collectivement disposé à rendre un commandant partiellement responsable du comportement de ses soldats (qui sont eux-mêmes des agents autonomes), il n'y a aucune raison légale ou morale de ne pas tenir les troupes au même niveau de responsabilité pour les actions ou l'inertie d'un SALA. Sparrow est d'avis qu'il serait injuste de faire porter la responsabilité du fonctionnement d'un SALA à un chef militaire, car ce dernier ne peut en aucun cas prédire le comportement de la machine. Toutefois, cela soulève un dilemme, à savoir si le chef militaire peut également prédire le comportement de ses troupes au combat. Les chefs militaires entraînent leurs soldats de manière aussi régulière que rigoureuse afin que leurs réactions soient prévisibles dans une situation tendue, mais cela n'exclut pas la possibilité qu'un soldat puisse franchir un seuil moral en commettant un crime de guerre. Le roboticien Ronald Arkin écrit : « Personnellement, je n'ai pas confiance en l'exemption du droit [de la guerre (DG) ou les règles d'engagement (RE)] opérée par un agent autonome [le SALA], car cela soulève la question de la responsabilité du fait. En revanche, il n'est pas impossible qu'un humain assume la responsabilité d'une telle entorse au droit, si c'est pertinent. »<sup>55</sup> Arkin soutient ici que ni à la conception, ni à la construction d'un SALA, on ne doit conférer à la machine la capacité de prendre la décision autonome de mettre de côté le DG ou le RE (ce qui ferait partie de sa programmation de base) et que cette responsabilité doit demeurer exclusivement aux mains d'un humain. La responsabilité du SALA allant à l'encontre du DG ou du RE s'en verrait transférée à un décideur humain (à savoir un chef militaire à un niveau donné).<sup>56</sup> Si un chef

---

54. Voir les comptes rendus de la Marine américaine au sujet des incidents du USS Fitzgerald et du USS John S McCain : <https://www.secnav.navy.mil/foia/readingroom/HotTopics/CNO%20USS%20Fitzgerald%20and%20USS%20John%20S%20McCain%20Response/CNO%20USS%20Fitzgerald%20and%20USS%20John%20S%20McCain%20Response.pdf>.

55. Arkin 2009, 40.

56. Selon moi, cela s'apparente à l'emploi des services d'un officier en charge du ciblage dans les conflits actuels qui engagent les Etats-Unis. Dans ce type de situation, lorsqu'une cible est désignée, on demandera à un officier supérieur, le plus souvent à un général, de valider la cible et d'autoriser l'engagement de la force. L'officier agira en présence d'un

militaire, aussi éloigné qu'il puisse être de la décision finale de la frappe, décide d'avoir recours à un SALA dans une situation où de trop grands dommages collatéraux sont possibles, il ne semble y avoir aucun tort moral à incriminer ce chef militaire en cas d'incident.

On pourrait douter que suffisamment d'éléments probants ont été évoqués jusqu'ici dans le but d'affirmer que la norme Yamashita justifie moralement une responsabilité attribuée au chef militaire. Je soutiens qu'au lieu de chercher des justifications, il serait préférable d'appréhender ce débat par l'affirmation que la norme Yamashita fournit une raison moralement justifiable de tenir un commandant responsable (même si son rôle est de superviser l'emploi d'un SALA). Pourquoi cette doctrine de la responsabilité peut-elle être moralement justifiable ? Parce qu'elle procède d'un désir de contraindre les commandants de théâtre en temps de guerre à prendre toutes les précautions nécessaires pour créer un environnement respectueux du droit de la guerre, du droit international humanitaire et des règles d'engagement applicables à la zone d'opérations donnée. La norme Yamashita (ou *Medina standard*, dans sa dernière mise à jour) propose une voie moralement acceptable dans ce but donné et rappelle aux chefs militaires tout au long de la chaîne de commandement qu'ils sont personnellement responsables des actions de leurs subordonnés. Cela ne signifie pas pour autant qu'ils aient à endosser l'entière responsabilité des actions ou des manques de leurs subordonnés, mais qu'ils doivent prendre leur juste part de responsabilité.<sup>57</sup>

Le point d'achoppement de l'argument présenté par Sparrow tient à sa réticence à tenir les SALA responsables de leurs exactions en tant qu'agents moraux. Sparrow soutient qu'il est impossible de sanctionner un SALA qui ne peut ressentir de souffrance, ajoutant que de ce fait le système ne peut encourir de sanction. Cet argument semble tenir à l'idée d'une justice rétributive, le problème résidant dans l'affirmation de Sparrow que la justice rétributive est nécessairement la voie à suivre (explicitement pour les SALA et implicitement pour les humains). À ce sujet, il dit : « de façon à ce que ces actes de sanction servent de punition, ils doivent susciter une réponse de la part de leur objet...pour être en mesure de recevoir une punition, l'agent

---

avocat proposant un conseil juridique adapté à la situation, bien que la responsabilité incombe au décideur. Il pourrait en être de même pour autoriser l'emploi d'un SALA déviant des normes de sa pré-programmation en termes de DG et de RE, si un chef militaire en décidait ; dans ce cas, celui-ci serait tenu d'accepter la pleine et entière responsabilité de cette action.

57. Je crois que le débat concernant « la juste part de responsabilité » est important et nécessaire, mais qu'il échappe à la portée voulue pour ce mémoire.

[artificiel ou humain] doit être jugé capable de souffrance. »<sup>58</sup> Néanmoins, il est possible de rejeter la souffrance, en tant que condition à la possibilité de la punition, par l'adoption d'idées morales prospectives qui font l'objet d'une littérature établie. Par exemple, il pourrait être envisagé d'adopter un type de sanction participatif visant à généraliser à tous les SALA la programmation d'actes jugés inacceptables. On pourrait également souhaiter une approche fondée sur la réadaptation de la sanction selon laquelle le SALA ferait l'objet d'une nouvelle programmation ou d'un paramétrage afin d'exclure toute action jugée indésirable sur la base de faits passés. En revanche, si on se focalise sur la punition par rétribution comme le préconise Sparrow, la sanction d'un SALA donc rendue impossible validerait son argument. Cependant, il n'y a que peu de raisons de penser que Sparrow ne fait pas fausse route.

Quant à la responsabilité morale liée à la mort d'un ennemi combattant, Sparrow dit ceci : « la moindre des choses que l'on doive à nos ennemis est de donner suffisamment de valeur à leur vie pour accepter la responsabilité de leur mort. »<sup>59</sup> Sparrow exige que le SALA soit tenu responsable alors qu'il est apparent que le chef militaire ayant décidé de déployer le système d'armes ou le politicien ayant engagé le conflit armé devraient être au centre du débat. Si le SALA est employé en toute légitimité, il ne devrait y avoir aucune obligation de tenir individus ou systèmes d'armement responsables de pertes humaines en temps de guerre.<sup>60</sup>

En revanche, concernant la question de la prise de responsabilité pour un SALA dont le déploiement a causé un crime de guerre, je soutiens qu'au lieu de sanctionner l'emploi du système, il est plus judicieux de se pencher sur les circonstances spécifiques à chaque incident. En effet, vues les spécificités de chacun, il ne convient pas de faire des généralisations ou de dégager des principes universels pour engager la responsabilité d'un agent. Il vaut mieux étudier les mécanismes de la décision ayant conduit à l'emploi de l'arme (par rapport à des probabilités accrues de nombreux dommages collatéraux) ainsi que les autres décisions prises par le chef militaire détenteur des renseignements utiles au moment de l'attaque avant de faire porter le

---

58. Sparrow 2007, 72.

59. Sparrow 2007, 67

60. Il est évident que McMahan, pour ne citer que lui, serait en désaccord avec cette assertion, préférant dire que ces soldats ou le SALA qui prennent part à une guerre injuste sont individuellement responsables de la mort d'un ennemi combattant parce que le soldat présumé avoir des intentions justes ne peut pas constituer une cible légitime. (McMahan 2011, 14)

blâme à un agent en particulier. En somme, s'il peut être démontré qu'un chef militaire ou un SALA ont pris toutes les précautions nécessaires pour éviter un incident grave, mais que des atrocités sont survenues (dues, par exemple, à une bombe perdue), on ne devrait tenir aucune des deux entités responsable (en appliquant possiblement la doctrine de la double finalité).

Sparrow répond que suivre cette voie donne cours à un déficit de responsabilité, le seul remède adapté à la situation étant de ne pas avoir recours aux SALA. En réponse à cette opinion, je maintiens que Sparrow fait erreur. Comme je l'ai mentionné auparavant, si la norme Yamashita (ou *Medina standard*) est mise en pratique, il sera alors possible d'établir un lien de responsabilité morale avec le chef militaire dans les cas où il est impossible de démontrer clairement que le SALA est à l'origine d'une action illégale ou immorale. Dans ce cas, on est confronté à deux options, soit a) « punir » le SALA en se fondant sur une base morale adaptée (impliquant communication, réadaptation, etc.) ou b) sanctionner le chef militaire, quel que soit son rang (mais qui pourrait ne pas correspondre au plus bas niveau de commandement) en appliquant la norme Yamashita / *Medina standard*.

Comme nous l'avons démontré sur la base de précédents dans le droit international actuel, par l'étude des normes de leadership militaire américain et en évoquant une possible refonte de l'idée de punition, l'argument de la responsabilité s'avère peu concluant. Au lieu de fournir des preuves convaincantes que la communauté internationale aurait tout intérêt à interdire purement et simplement le développement et l'utilisation des SALA, cet argument constitue, à l'inverse, une forte incitation à la réglementation du développement et de l'utilisation de ces systèmes d'armement.

### 3.2 Contre-réponse à l'argument dit « de l'agent responsable »

L'argument le plus convaincant qui ait été soutenu jusqu'à présent dans cette étude soutient qu'il est inenvisageable de céder le contrôle de la décision de tuer à un agent artificiel. D'une part, celui-ci est incapable de faire preuve de clémence au moment même d'une action létale, contrairement à un être humain ; d'autre part, qu'une personne perde la vie par décision d'un robot autonome est fondamentalement irrespectueux de la dignité humaine. Cependant, les deux côtés de cet argument sont intrinsèquement biaisés.

Leveringhaus veut rendre possible le choix de ne pas suivre un ordre légitime de tuer l'ennemi sur le champ de bataille, et en l'occurrence, le choix de ne pas tirer sur un combattant légitime, afin de laisser s'exprimer

des sentiments humains dont la pitié. Il dit : « ce qui rend précieux l'agent humain, comparé à l'agent artificiel, dans la guerre et dans la vie courante, est la possibilité de s'engager à tout moment dans un autre plan d'action. »

<sup>61</sup> Leveringhaus croit sincèrement que sur le champ de bataille, il faut être capable de ne pas tuer en dépit de l'aval qui enjoint à exercer une force meurtrière. Il suppose que si un SALA est préprogrammé pour certains ordres et paramètres de mission au vu du droit de la guerre et des règles d'engagement, il lancera l'attaque contre tous les ennemis rencontrés sur son passage. Je suis d'avis que cette supposition n'est pas nécessairement vraie. En suivant les conseils d'Arkin au sujet des algorithmes pour robots militaires, également connus sous le nom de « gouverneur éthique », la conception adéquate d'un SALA lui fera prendre par défaut des mesures non létales pour échapper à l'ennemi ou le maîtriser. Il lui serait donc impossible de passer immédiatement à une action létale. Grâce à la présence d'un algorithme construit sur le modèle de ce système d'armement, les attaques mortelles ne se produiraient que si elles étaient inévitables. À supposer qu'un SALA soit en patrouille avec une brigade de soldats de l'infanterie américaine, l'un des RE pourrait être défini comme suit : « Si un membre de la brigade est en danger imminent d'être capturé, il y a obligation de faire usage de moyens proportionnels adaptés à la situation afin d'empêcher la capture. » Ce qui implique que si aucun membre de la patrouille n'est en danger imminent d'être capturé, le SALA doit choisir par défaut une action non létale, à moins que le système n'ait à répondre à d'autres exigences prédéfinies.

Le système d'armement létal autonome en tant qu'outil doit être contraint <sup>62</sup> de *suivre* des ordres légitimes et légaux et d'*ignorer* des ordres illégitimes et illégaux. Le dernier point est tout aussi important que le premier en vertu de la prévention des crimes de guerre sur le champ de bataille. Il s'agirait d'une sorte de « contrôle du comportement éthique » <sup>63</sup> selon Arkin. Sur ce point, il écrit : « pour ce qui est du cas particulier des robots de terrain (mais aussi pour le soldat), l'agent responsable ne doit pas pouvoir se faire sa propre idée des implications morales inhérentes à l'utilisation de la force létale ; en revanche, on attend de lui qu'il mette en pratique les croyances

---

61. Leveringhaus 2016, 9

62. On pourrait ici définir les obligations d'un SALA par rapport aux caractéristiques de sa programmation le contraignant à suivre des règles prédéterminées, mais cette définition pourrait s'étendre à une vision plus extensive de ses obligations morales, le cas échéant.

63. Arkin 2009, 66-67.



et les règles ayant été précédemment déduites par la communauté humaine et dont le DG et le RE sont le reflet. »<sup>64</sup>

Leveringhaus en déduit que l'une des caractéristiques distinguant les agents humains des agents artificiels est la capacité humaine à prendre ses ennemis en pitié. Selon lui, « la capacité humaine de gracier, la capacité à se sentir pris de pitié ou à compatir avec autrui, même dans le camp opposé, est sans nul doute moralement pertinente et mérite d'être défendue. »<sup>65</sup> De quoi est faite la pitié, si ce n'est de la capacité à penser et à agir autrement ? Il semble que le point central de réponse pour Leveringhaus réside dans le fait de pouvoir s'identifier à l'ennemi, de pouvoir reconnaître son humanité et ainsi de s'abstenir de lui faire du mal lorsque c'est possible ou indiqué. « Ces soldats qui n'ont pas tué ont pu s'en dispenser en reconnaissant l'humanité de l'ennemi et en réalisant la gravité de la décision qui consiste à appuyer sur la détente. »<sup>66</sup> Est-il possible de se méprendre quant à la véritable nature de la compassion sur le champ de bataille ? La décision de ne pas tuer un ennemi pour le laisser vivre un jour de plus, bien qu'il soit potentiellement en captivité, n'est-elle pas en soi une reconnaissance de son humanité ? Je dirais qu'épargner une vie par décision fondée sur l'absence de nécessité militaire à causer la mort est une forme adéquate de compassion et de pitié qui pourrait tout à fait être mise en œuvre sur le champ de bataille de demain. Ce type de décision permettrait de réduire le nombre de massacres survenus sur le champ de bataille et de limiter le nombre de morts soit à la plus stricte nécessité en vertu d'un objectif militaire, soit à ces décès imprévisibles et regrettables survenus en conséquence d'une autre action sans lien. Il faut également reconnaître qu'à l'heure actuelle le bilan à grande échelle des actes de compassion n'est pas très satisfaisant, même au niveau individuel, durant les conflits armés. Bien souvent, la machine de propagande en vigueur dans un pays présente l'ennemi comme un intrus ou comme un être tout « autre » que l'on privera de son humanité en le traitant de « Teuton » ou de « Viêt-Cong » (en cela, on repense respectivement aux Allemands de la Première Guerre mondiale et aux Vietnamiens durant la guerre du Vietnam), cela rendant sa mort plus facile, car plus acceptable. S'il était envisageable, dans le cadre d'une réglementation internationale stricte, de créer un SALA ayant pour fondement le respect des règles du droit international et de la vie humaine (pour restreindre les morts à des

---

64. Ibid, 117.

65. Leveringhaus 2016, 9-10.

66. Ibid, 10.

cibles légales conformément aux nécessités militaires), il serait plus facile de faire preuve de clémence sur le champ de bataille que ce qu'il a été rendu possible jusqu'ici.

Cette objection pourrait se voir opposer l'argument de la faisabilité. Est-il raisonnable d'espérer pourvoir un SALA d'un logiciel incluant, autant que possible, une forme de capacité à la clémence envers les combattants du camp adverse ? Voilà une ambition solide qu'il faudra continuer à défendre durant le développement futur de ces systèmes. On pourrait mettre en avant la possibilité de « tout simplement »<sup>67</sup> inclure dans l'algorithme directeur du SALA qu'« à chaque rencontre avec un soldat ennemi ne présentant pas de menace directe pour soi ou pour tout autre soldat humain dans les environs, il faudra faire preuve de clémence en réfrénant toute tentative d'action mortelle visant à le neutraliser. » Ce n'est là qu'un exemple de technique possible, mais qu'il serait bon de prendre en considération. Pour revenir à Ronald Arkin et à ses travaux théoriques au sujet de la construction d'un algorithme régissant les choix d'une semblable machine, il semble que la clémence serait inhérente à une option non létale sélectionnée par défaut comme plan d'action initial. À la place d'un soldat humain faisant le choix d'épargner l'ennemi (probablement en utilisant des mesures non létales pour le capturer et pas seulement en le laissant partir), il y aurait un SALA dont la programmation par défaut présiderait à la capture de l'ennemi en utilisant des tactiques non létales et humaines.<sup>68</sup>

Le dernier argument de Leveringhaus détaille un impératif, celui de préserver la dignité humaine dans la décision d'ôter la vie. Leveringhaus soutient que cette dignité ne peut être préservée qu'à la seule condition qu'un agent humain prenne la décision ultime au moment même de la mort. Cela soulève la question de savoir si la présence d'un agent humain responsable de tuer est nécessaire pour préserver cette dignité. Je dirais que non. Paul Scharre fait valoir un argument fort en écrivant ce qui suit :

Du point de vue du soldat tué sur le champ de bataille, cette [mort digne] est une critique inhabituelle, presque étrange, des armes autonomes. Il n'y a pas de tradition juridique, éthique ou historique selon laquelle les combattants accordent à leurs

---

67. J'ai tout à fait conscience du fait que ce ne serait pas tâche facile.

68. L'emploi de tactiques non létales mais inhumaines nierait les bases du projet. De telles tactiques, consistant éventuellement à étourdir l'ennemi, sont aussi suspectes d'un point de vue moral que l'idée de faire feu sur tout individu dont la présence ne constituerait pas une grave menace.

ennemis le droit de mourir dignement dans un conflit armé. Il n'y a d'ailleurs rien de digne à être fauché par une mitrailleuse, réduit en pièces par une bombe, brûlé vif dans une explosion, noyé dans un navire en perdition, lentement suffoqué par une lésion thoracique aspirante, ou de même à toute autre mort effroyable en temps de guerre.<sup>69</sup>

Le souhait que seules des morts dignes puissent se produire pendant la guerre s'assimile à une tentative futile d'aseptiser le combat en déclarant qu'« au moins l'ennemi est mort dans la dignité ». Bien que l'idée soit gênante pour la plupart des gens, il faut bien reconnaître que la guerre, dans son fondement même, se définit par la mort et la destruction. Réduire la fréquence des massacres et limiter les souffrances infligées aux gens, combattants et civils, est un but à poursuivre, mais exiger que « la mort dans la dignité » devienne un impératif est une tentative futile d'ignorer la cruauté intrinsèque de la guerre. Il est tout à fait acceptable d'exiger que les nations n'entrent en conflit que lorsque c'est moralement justifié, ou justifiable tout du moins. Cependant, imposer de tuer seulement dans la dignité est à la fois trop naïf et éloigné de la réalité la plus brutale des relations humaines.

La mort la plus digne à laquelle on puisse prétendre<sup>70</sup> lors d'une guerre est celle qui cause le moins de souffrances possible. Une mort « propre », pour ainsi dire, qui ôte rapidement et efficacement la vie à un combattant, sans causer ni douleur, ni longues souffrances, serait vraisemblablement la plus enviable des morts. Ce n'est pas là une tentative de minimisation de l'importance de la mort sur le champ de bataille, comme s'il était question de laisser opérer un système de tuerie de masse brutalement efficace et semblable à un abattoir ; il s'agirait plutôt d'un désir de mettre fin aux hostilités le plus rapidement possible sans douleur ni souffrances inutiles. Comme l'a écrit Paul Scharre sur la base de ses expériences personnelles de la guerre, il semble ridicule d'insister sur le fait que la mort laissée aux « mains » d'une machine serait en quelque sorte moins digne que les autres manières dont on peut trouver la mort dans la guerre moderne.

Au lieu de se préoccuper d'une éventuelle « dignité trouvée dans la mort » (étant donné qu'on ne peut jamais être sûr de mourir dignement), il est nécessaire de se pencher sur une réduction des souffrances causées par la guerre. En cela, les systèmes d'armes autonomes représentent une

---

69. Scharre 2018, 288.

70. Si tant est qu'il soit possible de formuler une telle espérance dans le combat.

solution possible. Comme j'en ferai état dans la partie suivante de ce travail, l'objectif premier dans la guerre devrait consister à réduire douleurs et souffrances inutiles. Il faudrait concevoir des SALA programmés à cet effet. En acceptant cet état de fait, on ne peut que rejeter la revendication de Leveringhaus selon laquelle la mort par l'intervention d'un SALA est intrinsèquement indigne.

#### 4. Le SALA, Choix Rationnel dans L'évolution de la Guerre Moderne

Limiter le nombre de morts de non-combattants et de civils sur le champ de bataille devrait être l'objectif de toute force morale intègre. Le DHI exige des armées de tous les états souverains, qu'ils aient ou non ratifié les conventions de Genève, de prendre tout le soin et les précautions nécessaires pour ne pas tuer ou blesser des catégories de personnes protégées par les dispositions générales du droit humanitaire, ni endommager ou détruire inutilement des biens meubles et immeubles (en particulier des catégories de biens protégés par convention). C'est dans ce but donné que les forces militaires du monde entier ont voulu concevoir des « armes intelligentes ». Et il faut considérer les systèmes d'armes autonomes comme la prochaine étape logique de ces armes intelligentes, car ils ont au moins un double avantage par rapport aux « armes aveugles » : 1) ils peuvent être dirigés vers une cible spécifique (souvent avec des marges d'erreur de quelques mètres seulement) et 2) leur emploi est relativement économique. Or, ces deux avantages participent d'une volonté de limiter au maximum les dommages collatéraux inutiles. En ciblant une zone avec une marge d'erreur minimale, il est possible d'employer un nombre limité d'armes pour neutraliser une menace, ce qui limite l'exposition au danger des civils et des infrastructures. Ce calcul s'aligne sur l'architecture fondamentale du logiciel qu'Arkin exhorte les chercheurs et développeurs à inclure dans leur conception des SALA modernes.<sup>71</sup> Cette nouvelle programmation réduirait au strict minimum les « massacres sur le champ de bataille » et les dégâts collatéraux. D'autre part, l'emploi des armes intelligentes est avantageux économiquement parlant par rapport à l'emploi des armes simples. Cela peut sembler contraire au bon sens, car les armes intelligentes coûtent souvent des centaines de milliers, voire des millions de dollars.

---

71.  $A+B+C+D$  = massacre sur le champ de bataille.  $A$  = combattants supposés,  $B$  = Forces amies non escomptées,  $C$  = Non-combattants supposés, et  $D$  = Non-combattants non escomptés. Le but consiste à optimiser le groupe  $A$ , à éliminer les groupes  $B$  &  $C$ , à s'assurer que le groupe  $D$  compte aussi peu d'effectifs que possible (Arkin 2009, 128)

Le raisonnement qui préside à cette logique est le suivant : s'il est possible d'accomplir une mission avec peu d'armes (quoique leur coût individuel puisse être supérieur) qu'avec de nombreuses munitions « classiques » relativement bon marché (mais qui dans l'ensemble peuvent coûter plus cher qu'une arme intelligente), il est logiquement plus avantageux d'avoir recours à une arme intelligente. S'il est possible de ne déployer qu'un seul missile guidé dans le but d'éliminer un bunker très surveillé à un endroit stratégique, action qui aurait auparavant nécessité l'emploi de deux cents bombes aveugles, il est clair qu'il faudra privilégier l'emploi d'un seul missile guidé intelligent. S'il était question d'employer des missiles et des bombes larguées, on pourrait alors considérer que si un SALA peut faire le travail de dix soldats humains, il est économiquement plus viable d'utiliser uniquement ce système d'armes. Si grâce aux SALA les armées peuvent maîtriser leurs coûts tout en limitant les souffrances des non-combattants, il y a de bonnes raisons de croire que la voie à suivre dans la guerre moderne passera par le développement et l'utilisation de ces systèmes.

Arkin, parmi tant d'autres, croit que l'utilisation des SALA (et plus généralement de l'AI) sur le champ de bataille pourrait être la voie morale à suivre en raison d'un ensemble de caractéristiques inhérentes à leur nature. Premièrement, un SALA peut aussi être employé de manière conventionnelle.<sup>72</sup> La machine doit alors être programmée par défaut pour mener une action non létale, car contrairement aux humains, elle n'a aucun instinct de conservation.<sup>73</sup> N'étant caractérisé par aucun des instincts humains de survie et de protection, un SALA peut prendre des mesures non létales audacieuses (comme sortir de la zone dangereuse dont un humain ne pourrait pas s'échapper) ou se sacrifier en vertu du succès sa mission en créant une diversion pour sauver son unité. Un SALA sera également à l'abri d'illusions de la perception (brouillard émotionnel et prophéties auto-réalisatrices) pouvant occasionner des troubles du jugement chez les agents humains.<sup>74</sup> Il arrive souvent que des crimes de guerre soient commis dans le feu de l'action alors que les émotions sont vives et que le jugement est assombri par la peur ou la colère. Selon les termes de J Glenn Gray, un soldat « devient un combattant, un homme devient enragé »<sup>75</sup>. Un SALA dépourvu d'émotions fortes, selon une conception logicielle morale et logique, ne

---

72. Arkin 2009, 29

73. Ibid, 29

74. Arkin 2009, 30.

75. Gray 1959, 27.

serait pas esclave de la peur ou de la colère, ni ne répondrait par l'agression à la destruction d'un autre SALA ou à la mort d'un membre de son unité. Il pourrait en soi faire un meilleur agent éthique et moral que les humains dans la bataille. Dans la même veine, Arkin écrit : « Je ne crois pas qu'un système autonome sans pilote puisse agir de manière parfaitement éthique sur le champ de bataille, mais je suis convaincu qu'il se comportera de façon plus éthique que les soldats humains n'en sont capables. »<sup>76</sup>

Si, comme le soutient Arkin, il est possible d'inclure à la conception d'un SALA la capacité d'agir à la fois de façon éthique et morale, au moins à égalité avec les soldats humains, il est alors évident que la création d'un tel système est un choix rationnel. De même, s'il est prouvé qu'un SALA peut avoir des réactions au moins aussi justes, éthiquement et moralement, que les meilleurs soldats dans la bataille, il est alors certain qu'il y a une obligation morale à la conception d'un système d'arme de ce type. Cet équipement limiterait en grande partie le nombre de soldats exposés au danger, même s'il faut reconnaître que des systèmes d'armement ne peuvent pas remplacer tous les humains. Si un système autonome est capable d'emprunter un itinéraire semé d'engins explosifs improvisés afin de livrer une cargaison importante, et ceci sans risquer la vie des chauffeurs, on est moralement obligé de recourir à ce système. En choisissant de ne pas privilégier cette solution alors que la technologie nécessaire est à disposition, on agit en fait de manière immorale. De plus, si un SALA peut prendre des décisions plus justes, d'un point de vue éthique et moral, que les humains sur le champ de bataille, on est alors contraint de déployer ce type de système. Enfin, si un système regroupant toutes ces caractéristiques (prise de décisions moralement et éthiquement plus justes que celles des agents humains, reconnaissance précise des cibles, frappes chirurgicales pour empêcher la destruction totale des cibles, réduction des dégâts collatéraux) peut être conçu et mis à disposition, on est moralement obligé en tant que chef militaire de déployer ce type de dispositif. Toutefois, son déploiement ne saurait exempter les chefs militaires (de rang probablement plus élevé que les commandants de compagnie et de bataillon<sup>77</sup>) du poids moral lié au fait d'ôter la vie, ni ne

---

76. Arkin 2009, 31.

77. Au sein de l'armée américaine, la compagnie est la plus petite unité ayant un chef désigné (bien qu'il existe de plus petites unités appelées détachements, mais il y a des exceptions à la règle). L'effectif d'une compagnie peut aller de 50 à 300 soldats. La compagnie est le niveau de base de l'action de terrain que l'on investit d'un leadership et de responsabilités. Une compagnie est commandée par un Capitaine (O-3) qui, s'il ne jouit d'aucune expérience préalable en tant que soldat, a en général de 4 à 6 ans de service.

pourrait les exonérer de la responsabilité des dégâts collatéraux excessifs en découlant.<sup>78</sup> De fait, comme l'a mentionné Arkin, il est peu probable qu'en pratique, même avec le concours de la technologie de pointe, il soit possible d'éviter totalement les dégâts collatéraux. Pourtant, en tant que chef militaire, on reste moralement obligé de déployer un SALA qui pourrait prendre des décisions moralement plus justes que celles des humains qui, autrement, feraient usage de la force létale.

Telle est la ligne de conduite la plus rationnelle en vertu du développement et du déploiement d'un système d'armes qui limite les massacres sur le champ de bataille. Ce point de vue n'assainit en rien la guerre, car il y aura toujours des combattants légitimes qui trouveront la mort et, comme le dit Leveringhaus, « l'ennemi, en tant que cible légitime, ne prétend pas demander sa grâce aux soldats attaquants.<sup>79</sup> »

## 5. Proposition d'un Schéma de Réglementation des SALA

Au cours de ce mémoire, on a présenté en premier lieu les arguments les plus convaincants contre l'utilisation des SALA sur le champ de bataille de demain. Ensuite, on a évoqué les sources de l'obligation morale liée au déploiement de ce type d'armement. À ce point de notre travail, ces réflexions font ressortir la nécessité d'envisager l'élaboration d'un cadre théorique de gouvernance destiné à réglementer la conception et le déploiement de ces systèmes d'armement au niveau international. Comme on l'a évoqué précédemment, il n'existe aucune loi inscrite dans le DHI actuel dont le but spécifique serait d'encadrer la conception et le déploiement des SALA. La dernière partie de ce travail se veut comme un point de départ aux débats concernant un potentiel cadre juridique, trois grands sujets de préoccupation s'y rattachant. Ces trois domaines d'intérêt recouvrent les inquiétudes les plus prégnantes et les plus communes concernant ces systèmes d'armement. Sans prétendre couvrir toutes les potentialités d'utilisation des systèmes d'armes, cet exposé donne un cadre théorique suffisamment

---

Un bataillon standard comprend de 5 à 7 compagnies, ce qui constitue un regroupement comptant de 700 à 1500 soldats selon la composition et la mission de l'unité. Le bataillon fait partie de l'armée de terrain. Il est commandé par un Lieutenant-Colonel (O-5) qui a généralement de 17 à 19 ans de service.

78. Il y a une impossibilité physique à la limitation de tous les dégâts collatéraux quand des conflits armés ont lieu en zone habitée. C'est un fait admis. On a pour but de réduire ou de limiter les dommages collatéraux. Ainsi, les dégâts excessifs peuvent être classés comme crime de guerre contrairement aux dommages accessoires.

79. Leveringhaus 2016, 10.

large pour orienter l'élaboration du droit international concernant leur emploi. Les trois domaines en question dans ce schéma théorique sont 1) les circonstances possibles de déploiement des SALA, 2) le respect du DHI actuel et futur, et 3) la mise en place d'une chaîne de responsabilité.

Tout d'abord, la question majeure est d'identifier les circonstances possibles dans lesquelles déployer ces systèmes d'armement. À ce sujet, le DHI devrait adopter une législation stipulant qu'aucun SALA ne peut être déployé par une armée ou une nation sans que des experts techniques aient pu en homologuer la programmation de fonctionnement et s'assurer que cette dernière est en conformité avec le DHI. De plus, tout chef militaire déployant un système d'armement non homologué devrait être passible de sanctions, que le système ait ou non causé des crimes de guerre. Une telle exigence ne saurait être étrangère aux armées tenues de suivre le même type de processus de certification pour s'assurer que les systèmes d'armes standard ou conventionnels (un char, par exemple) sont conçus et fonctionnent dans le respect de la législation les gouvernant. Une fois la certification obtenue, l'armée ou la nation en question pourrait mettre en place un protocole comprenant deux parties distinctes et également importantes afin de décider des circonstances 1) de l'utilisation de la force non létale vs. la force létale (déploiement tactique vs. déploiement opérationnel) et 2) du déploiement un SALA (déploiement stratégique). Les deux composantes de ce protocole seront évoquées ci-après.

La décision d'avoir recours à un SALA dans une situation tactique signifie simplement que le chef militaire choisit de déployer un système d'arme dans le cadre d'une mission donnée. Ces commandants mènent des opérations de « niveau tactique », c'est-à-dire qu'ils ont la liberté de décider sur le terrain de l'emploi direct des troupes et des armes nécessaires pour accomplir une mission ou suivre une directive donnée. Ainsi, un commandant est investi de l'autorité légitime « appropriée » pour avaliser la décision d'employer un véhicule aérien sans pilote appartenant à sa compagnie afin d'élargir son espace visuel. En revanche, il n'aura pas le pouvoir de disposer de matériel de niveau stratégique (tel un satellite) afin d'accomplir sa mission. Pour en revenir au SALA, l'autorité décisionnaire et la responsabilité liées à l'emploi de ces systèmes devrait être confiée à un chef militaire de niveau tactique (c'est-à-dire jusqu'au rang de commandant de brigade, qui est à la tête d'environ 5000 soldats). Par exemple, si un commandant de compagnie est chargé de sécuriser un village, sa décision, prise en consultation avec des ordres de gradés supérieurs, doit être immédiate quant à l'emploi d'une arme ou d'une tactique spécifique. Il en va de même pour l'emploi d'un



SALA. De manière à contrôler efficacement l'emploi mesuré de la force, le droit humanitaire international doit spécifier que l'utilisation tactique d'un SALA sera assortie de garanties visant à limiter les dégâts collatéraux potentiels causés par leur utilisation.

Le deuxième défi concernant l'utilisation d'un SALA concerne son déploiement stratégique. Par emploi stratégique d'un SALA, on entend l'autorisation morale ou légale d'utiliser ce type d'arme dans une guerre ou un conflit donné, selon ses spécificités. Afin de répondre à cette préoccupation, il semble assez simple de légiférer pour rendre l'utilisation d'un SALA concordante avec les principes spécifiques du JWT (JSON Web Tokens, protocole d'échange numérique sécurisé de renseignements) et du DHI, en se posant la question de savoir si le déploiement d'un SALA pourrait être plus susceptible de causer des dégâts collatéraux excessifs lors d'un déploiement dans un théâtre d'opération X plutôt que dans un théâtre d'opération Y. Par exemple, selon le niveau de perfectionnement technologique au moment de la frappe, il peut être contraire à l'éthique de déployer un SALA dans une zone de combat fortement urbanisée (par exemple, Singapour) par rapport à une zone de combat plus éloignée ou rurale (par exemple, la steppe russe). Cette législation devrait mentionner une interdiction applicable à toute nation ou armée de déployer un SALA dans une zone de combat où la probabilité de dommages collatéraux excessifs résultant de son utilisation est plus élevée que celle concernant les dégâts provoqués par l'envoi de troupes. Par exemple, si l'emploi même d'un SALA dans une zone donnée était susceptible d'entraîner un recours aveugle à la force, la machine (ou un humain) ayant eu des difficultés à distinguer les combattants des non-combattants, l'emploi du SALA constituerait alors un crime de guerre, en marge de toute erreur de reconnaissance des combattants légitimes par le système. Il n'y a aucune raison de penser qu'il ne pourrait y avoir d'examen de responsabilité lorsqu'un SALA est à l'origine d'une erreur, bien que persistent des doutes possibles quant à l'agent coupable de l'incident. En effet, si un soldat est à l'origine d'une erreur (entraînant potentiellement un crime de guerre), il doit pouvoir être tenu responsable des conséquences de même que sa chaîne de commandement.<sup>80</sup>

Par conséquent, le deuxième grand domaine de réglementation devrait interdire l'emploi d'un SALA qui, de par sa conception même, serait inca-

---

80. Ce serait présumer d'une erreur délibérée ou provoquée par négligence, détail qui a peu d'importance à ce point donné de notre argumentation puisqu'il ne défie pas sa logique.

pable d'adhérer aux principes de base du JWT ainsi que du DHI dans ses évolutions actuelles ou futures. Comme cela a été dit précédemment, il serait fondamentalement erroné, tant du point de vue moral que juridique d'employer un système ne pouvant être tenu pour responsable de ses actions de par la nature de sa programmation. Il est peu probable de toute manière qu'un système d'armes d'ancienne génération dispose d'un fichier mémoire pouvant faire l'objet d'une vérification de toutes les actions du SALA ; selon toute probabilité, la réponse initiale par défaut d'un système de ce type consisterait en une action létale. Bien qu'il ne soit pas du ressort du DHI de choisir les technologies pouvant ou non être employées dans la guerre, un SALA muni d'un gouverneur éthique, dans la veine de celui que décrit Arkin, serait un excellent exemple de modèle en mesure de se conformer à cette réglementation théorique. Tout système de ce type pouvant adhérer à l'esprit de la réglementation, et donc au JWT et au DHI, devrait être autorisé.

Le dernier domaine d'intérêt lié à cette future modification du DHI est la nécessité de créer une chaîne de responsabilité bien établie. Dans ce but, le DHI devrait prévoir qu'un chef militaire ou un leader politique assume la responsabilité morale et juridique de la décision stratégique de déploiement d'un SALA. La loi devrait également stipuler qu'au niveau tactique un chef militaire ayant recours à ce type d'arme est à la fois moralement et légalement responsable des actions opérées ou non par le système. Certains opposants à ces réglementations dénonceront la contrainte qu'elles posent aux chefs militaires. Je réponds à cela que ces lois sont un poids nécessaire afin d'empêcher la libre utilisation des systèmes d'arme dans des situations susceptibles d'aller à l'encontre du JWT, du DHI ou d'augmenter considérablement les dommages collatéraux. Une telle réglementation aurait pour effet positif de réserver l'utilisation des SALA uniquement aux situations dans lesquelles les résultats sont raisonnablement bien contrôlés ou aux situations dans lesquelles la probabilité de dommages collatéraux est faible. (Par exemple, la réglementation peut décourager l'utilisation des SALA dans des lieux urbains densément peuplés, mais peut encourager son utilisation dans des zones peu densément peuplées telles que la haute mer).

Il y a un avantage évident à l'élaboration d'un droit international encadrant la moralisation de la conception et de l'usage des SALA, mais ce dispositif législatif doit aussi inclure un mécanisme pour assurer la conformité à ses réglementations. Ce mécanisme pourrait être semblable à la CPI déjà existante ou prendre une autre forme de mise en application internationale. Les réglementations légales envisagées plus haut ne dépendent d'aucune méthode

spécifique d'application et sont compatibles avec un mécanisme visant à inciter les nations à concevoir leurs SALA en conformité avec la loi. Il ne m'appartient pas de décider du mode de fonctionnement de ce mécanisme, mais seulement de présenter un plaidoyer en faveur d'un système capable de limiter les crimes de guerre envisagés avant qu'ils ne soient commis.

## 6. Conclusion

Le progrès humain est analogue à une machine à mouvement perpétuel. Sans commencement définitif et sans fin apparente, la marche vers le progrès est inexorable. Il en est de même pour le développement de moyens plus efficaces de s'entretuer. Dès les premiers temps de l'existence humaine, depuis le moment où l'homme s'est rendu compte que l'atlatl, ingénieux propulseur à crochet, lui permettait de propulser sa lance plus loin que son adversaire, jusqu'à nos jours où les avancées modernes dans le domaine de l'intelligence artificielle vont toujours plus loin, l'humanité fait preuve du désir irréprouvable de trouver des armes sans cesse plus performantes. Le moment historique est maintenant venu d'adopter un cadre réglementaire qui façonnera la manière dont les humains, en tant qu'espèce, entendent concevoir et déployer la future génération des systèmes d'armes.

Dans ce mémoire ont été présentés deux des arguments les plus convaincants contre l'utilisation des systèmes d'armes autonomes sous la forme des arguments dits « de la responsabilité » et « de l'agent responsable ». Bien qu'ils soient convaincants, du moins en partie, aucun de ces deux arguments ne fournit de motifs suffisants individuellement ou collectivement pour faire pencher la balance vers une interdiction pure et simple de l'utilisation de cette technologie émergente. Au contraire, ces arguments parmi d'autres fournissent une motivation pour exiger une réglementation internationale en vertu d'une conception et d'un déploiement moralement viables de ces systèmes. Au cours de ce travail, j'ai également soutenu que la conception et le déploiement encadrés d'un système d'armes autonome est un choix rationnel, si tant est que le fonctionnement des SALA peut répondre à des normes morales strictes. Le but en est d'imposer des limites à l'horreur du champ de bataille. Comme l'a dit un jour le général Sherman « la guerre est un enfer », mais il faut faire au mieux pour rendre cet enfer aussi tolérable que possible.

Un des instincts primaires des humains est la peur de l'inconnu. Les films de science-fiction n'ont pas peu fait pour attiser les angoisses selon lesquelles, en continuant sur la voie des améliorations technologiques, les robots se libéreront du joug de leur oppression et se retourneront contre

leurs maîtres. Cette crainte, bien qu'elle soit très répandue, ne peut s'imposer comme base solide à des décisions morales ou juridiques.

## Remerciements

Je tiens à rendre hommage au corps enseignant et au personnel du département de Philosophie, en particulier au Dr. Andrew Altman et au Dr. Andrew I. Cohen ; sans votre présence, je ne serais pas venu à bout de ce travail. Je voudrais également remercier les professeurs avec qui j'ai eu l'honneur de suivre des cours ; merci de m'avoir épaulé. Je veux aussi remercier ceux qui, au sein de l'armée américaine, ont su déceler mon potentiel et m'ont ainsi accordé leur confiance. Merci à tous.

## Références

- Arkin, Ronald. *Governing Lethal Behavior in Autonomous Systems*. Boca Raton: Chapman and Hall Imprint (Taylor and Francis Group), 2009.
- Army Doctrine Reference Publication. "6-22 Army Leadership." *Army Doctrine Reference Publications* (2012). [https://usacac.army.mil/sites/default/files/misc/doctrine/cdg/cdg\\_resources/manuals/adrp/adrp6\\_22\\_new.pdf](https://usacac.army.mil/sites/default/files/misc/doctrine/cdg/cdg_resources/manuals/adrp/adrp6_22_new.pdf).
- Boyd, John. *A Discourse on Winning and Losing*. Maxwell Air Force Base: Air University Press, 2018.
- Campaign to Stop Killer Robots. "About Us." Accessed 25 July 2018. <https://www.stopkillerrobots.org/about/>.
- Department of Defense. "Directive 3000.09—Autonomy in Weapon Systems." *Office of the Deputy Secretary of Defense* (November 2012). <http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>.
- In re Yamashita 327 U.S. 1 [1946] 61 (United States Supreme Court). <https://supreme.justia.com/cases/federal/us/327/1/>.
- Government of China. "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions of the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects." *Convention on Certain Conventional Weapons—Position Paper* (April 2018). [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/DD1551E60648CEBBC125808A005954FA/\\$file/China's+Position+Paper.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/DD1551E60648CEBBC125808A005954FA/$file/China's+Position+Paper.pdf).
- Gray, J. Glenn. *Warriors: Reflections of Men in Battle*. New York: Harcourt, Brace and Company, 1959.
- International Criminal Court. "Rome Statute." Article 8(2)(b)(xx) (1998). [https://www.icc-cpi.int/nr/rdonlyres/ea9aef7-5752-4f84-be94-0a655eb30e16/0/rome\\_statute\\_english.pdf](https://www.icc-cpi.int/nr/rdonlyres/ea9aef7-5752-4f84-be94-0a655eb30e16/0/rome_statute_english.pdf).
- Leveringhaus, Alex. "What's so Bad about Killer Robots." *Journal of Applied Philosophy*, (March 2016). DOI: 10.1111/japp.12200.
- McMahan, Jeff. *Killing in War*. Oxford: Oxford University Press, 2011.

- Military Leadership Diversity Commission. "Outreach & Recruiting." *United States Department of Defense*. November 2009. <https://diversity.defense.gov/Portals/51/Documents/Resources/Commission/docs/Issue%20Papers/Paper%2002%20-%20Requirements%20and%20Demographic%20Profile%20of%20Eligible%20Population.pdf>.
- Papal Encyclicals Online. "The Second Council of the Lateran- 1139 A.D." Accessed 18 January 2019. <http://www.papalencyclicals.net/councils/ecum10.htm>.
- Program Executive Office- Missiles & Space. "Counter-Rocket, Artillery, Mortar (C-RAM)." Accessed September 1, 2018. <https://missiledefenseadvocacy.org/defense-systems/counter-rocket-artillery-mortar-c-ram/>.
- Raytheon. "Phalanx Close-In Weapon System." Accessed 15 August 2018. <https://www.raytheon.com/capabilities/products/phalanx>.
- Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. New York: W. W. Norton & Company, 2018.
- Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy*, vol. 24, no. 1 (2007): 62–71. DOI: 10.1111/j.1468-5930.2007.00346.x.
- United Nations Institute for Disarmament Research. "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches." *UNIDIR Resources*, no. 6 (2017): 1–33. <http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf>.
- United Nations Office for Disarmament Affairs. "Perspectives on Lethal Autonomous Weapon Systems." *UNODA Occasional Papers*, no. 30 (November 2017): 1–61. [www.un.org/disarmament](http://www.un.org/disarmament).
- US Army Acquisition Support Center. "Excalibur Precision 155MM Projectiles." Accessed 25 July 2018. <https://asc.army.mil/web/portfolio-item/ammo-excalibur-xm982-m982-and-m982a1-precision-guided-extended-range-projectile/>.