Center on Law, Ethics and National Security

Essay Series
Number 11
August 31, 2021

Regulating the Third Frontier: The Current Unrestricted Nature of Autonomous Weapons and the Need for Regulatory Safeguards

By Angelina Bianchi

REGULATING THE THIRD FRONTIER: THE CURRENT UNRESTRICTED NATURE OF AUTONOMOUS WEAPONS AND THE NEED FOR REGULATORY SAFEGUARDS

Angelina Bianchi[*]

INTRODUCTION

The invention of gunpowder fundamentally altered the conduct of war.[1] Tactics needed to change, strategies needed to evolve, and nations that failed to adapt placed their existence in jeopardy.[2] Centuries later, the world experienced a second revolution in war with the emergence of nuclear weapons.[3] Today, the development of autonomous weapons threatens to be the next step in warfare's evolution.[4] The emergence of this third revolution has drawn criticism across the globe.[5] The absence of regulations or laws governing the interaction between autonomous AI and weapons systems has made both the development and the implementation of autonomous weapons a controversial issue.[6] While some argue that the development of this technology is a dangerous ethical nightmare,[7] others advocate that it is the

---

[*] Duke University School of Law, J.D., expected May 2022. Thank you to Major General Charles J. Dunlap, Jr. USAF (Ret.) for his support and guidance.

[1] *See* Valerie Orleans, *Invention of Gunpowder Changed War Tactics, Political Landscapes, World History*, CAL. ST. U., FULLERTON (Nov. 1, 2006), http://calstate.fullerton.edu/news/Inside/2006/gunpower.html (explaining how the emergence of gunpowder made it's use essential in the practice of war).

[2] *Id.*

[3] *See generally* Glen Salo, Nuclear Weapons and the Revolution in Military Affairs, National Security and Strategy (May 25, 2010), https://nationalsecurityandstrategy.blogspot.com/2010/05/nuclear-weapons-and-revolution-in.html (discussing the impact that the development of nuclear weapons had on warfare).

[4] *See* Charles P. Trumbull IV, *Autonomous Weapons: How Existing Law can Regulate Future Weapons*, 34 EMORY INT'L L. REV. 533, 554 (2020) (noting that autonomous weapons systems will "fundamentally change how war is waged").

[5] *'Killer Robots:' Ban Treaty is the Only Credible Solution*, HUMAN RIGHTS WATCH (Sept. 26, 2019, 1:01 AM), http://calstate.fullerton.edu/news/Inside/2006/gunpower.html [hereinafter *Killer Robots*] ("Since 2014, more than 90 countries have met eight times at the Convention on Conventional Weapons (CCW) to discuss concerns raised by [autonomous weapons systems]. Most of the participating nations with to negotiate a new treaty with prohibitions and restrictions in order to retain meaningful human control over the use of force.")

[6] *See* discussion *infra* Section IV.A.

[7] *See Killer Robots: Survey Shows Opposition Remains Strong*, HUMAN RIGHTS WATCH (Feb. 2, 2021 12:00 AM), https://www.hrw.org/news/2021/02/02/killer-robots-survey-shows-opposition-remains-strong (explaining the reasons behind the public's opposition to

inevitable future of warfare.[8] Many have suggested that the development of such weapons ought to be prohibited, or at a minimum, regulated through an international treaty.[9] However, with countries such as China and Russia on the brink of successfully developing deadly autonomous weapons,[10] and the continued development of AI in the private sector,[11] remaining stagnant could potentially leave nations vulnerable with this new deadly technology solely in the hands of their adversaries.[12]

This paper explores the ethical arguments related to the continued development of autonomous weapons systems as well as the current laws that apply to their use. This paper ultimately aims to demonstrate that even though the continued development of autonomous weapons systems may be a necessity, the risks can be mitigated by conducting this development in a responsible fashion. Section I will focus on identifying what constitutes a fully autonomous weapons system while Section II will explain how autonomous weapons systems function differently than automatic weapons systems. Section III will outline some concerns associated with the development and implementation of these weapons systems and Section IV will then discuss the relevant law regarding autonomous weapons systems. Section V will argue that the current laws are insufficient to fully address the concerns that accompany the continued development of autonomous weapons systems and, finally, Section VI will outline a proposed method of

---

the development of autonomous weapons systems). The UN Secretary-General António Guterres has commented that "[a]utonomous machines with the power and discretion to select targets and take lives without human involvement are politically unacceptable, morally repugnant and should be prohibited by law." António Guterres (@antonioguterres), TWITTER, (Mar. 25, 2019 1:28 PM), https://twitter.com/antonioguterres?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1110232038081204224%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fthenextweb.com%2Fnews%2Fnational-security-commission-led-by-ex-google-ceo-urges-us-to-ignore-calls-to-ban-autonomous-weapons.

[8] *See* Annemarie Vazquez, *Laws and Lawyers: Lethal Autonomous Weapons Bring LOAC Issues to the Design Table, and Judge Advocates Need to be There*, 288 MIL. L. REV. 89, 104 (2020) ("So, in a word, [autonomous weapons systems] are inescapable.")

[9] *Killer Robots*, *supra* note 5.

[10] *See* Matt Bartlett, *The AI Arms Race in 2020*, TOWARDS DATA SCIENCE, (Jun. 16, 2020) https://towardsdatascience.com/the-ai-arms-race-in-2020-e7f049cb69ac (explaining how the United States, Russia and China have pulled ahead of other countries in the race to develop autonomous weapons systems).

[11] *See* Brian Seamus Haney, *Applied Artificial Intelligence in Modern Warfare and National Security Policy*, 11 HASTINGS SCI. & TECH. L.J. 61, 63 (2020) (explaining the impact that the private sectors development of AI can have on the creation of "weapons of mass destruction").

[12] *See* Trumbull *supra* note 4 at 535 ("The United States, Russia, China, and other military powers are investing heavily in AI because they believe it will provide competitive advantages.")

regulation through which autonomous weapons can be developed responsibly while ensuring that there will be agents liable for their actions.

I. WHAT ARE AUTONOMOUS WEAPONS SYSTEMS?

Defining what constitutes an autonomous weapons system poses a challenge in and of itself.[13] The Department of Defense (DOD) has defined an autonomous weapon system as "[a] weapon system that can select and engage targets without further intervention by a human operator."[14] This definition, however, encompasses weapons that require specific human inputs to create decisive action and thus has been criticized as being too broad.[15] For those that argue the absence of humans in the kill-chain is a defining feature of an autonomous weapon, this definition conflates the distinction between autonomous weapons with automatic weapons.[16] The difference between autonomous weapons and automatic weapons systems can be understood by the following.

> "Automated systems operate by clear reputable rules based on unambiguous sensed data. Autonomous systems take in data about the unstructured world around them, process that data to generate information, and generate alternatives and make decisions in the face of uncertainty."[17]

A narrower definition of autonomous weapons has been proposed by Rabecca Crootof, a former Resident Fellow at the Information Society Project at Yale University, and has been widely accepted.[18] Crootof defines an autonomous weapons system as "a weapon system that, based on conclusions derived from gathered information and preprogramed constraints, is capable of

---

[13] *See generally* Michael C. Horowitz, *Why Words Matter: The Real World Consequences of Defining Autonomous Weapon Systems*, 30 TEMPLE INT'L & COMP. L.J. 85 (2016) (arguing that it is important to establish an appropriate definition for autonomous weapons system).

[14] U.D. DEP'T OF DEFENSE, DIR. 3000.09, AUTONOMY IN WEAPONS SYSTEMS 13 (21 Nov. 2012) (C1, 8 May 2017).

[15] *See* Christopher M. Ford, *Autonomous Weapons and International Law*, 69 S. C. L. REV. 413, 421 (2017) (criticizing the DOD definition of autonomous weapons systems for being too simplistic and thus encompassing both automatic as well as autonomous weapons systems).

[16] *Id.*

[17] Cummings, Lethal Autonomous Weapons: Meaningful human control or meaningful human certification? (unpublished manuscript) (on file with Duke University Humans and Anatomy Lab) 2019-Cummings_LAW.pdf (duke.edu).

[18] Horowitz, *supra* note 13 at 89 (explaining how there is a wide consensus that Crootof's definition of autonomous weapons is appropriate despite the fact it primarily encompasses weapons systems that do not yet exist).

independently selecting and engaging targets."[19] Under this more stringent definition, to be considered an autonomous weapons system, the system must operate fully autonomously and thus, operate through algorithms that engage in machine learning.

Though Crootof's definition has been widely adopted,[20] it has also received criticism for primarily encompassing weapons systems that do not yet exist.[21] Since countries have yet to successfully develop or employ fully autonomous weapons, it is argued that for the definition of autonomous weapons to be meaningful, it must encompass semi-autonomous weapons – automatic weapons – that currently exist.[22] The issue with this criticism is that including automatic weapons in the same definition of autonomous weapons fails to acknowledge the unique concerns associated with the design, creation, and implementation of fully autonomous weapons, specifically, the absence of a human in the kill chain and the blackbox nature of AI.[23] If we include automatic weapons systems within the definition of autonomous weapons systems "then large swaths of present-day weapons systems fall into [this] categories, ranging from land mines to advanced anti-ship missiles or ballistic missile defense systems."[24] What the media has coined as "killer robots" have generated sufficient stigma and public outcry to warrant their distinction from automatic weapons that are both widely accepted and commonly implemented in military.

Regardless of the growing public outcry, there is a wide consensus that "killer robots" are inevitable.[25] Though fully autonomous weapons are not currently utilized, many countries have been devoting significant resources to their development and production. [26] With other countries developing these technologies, being able to isolate the weapons that warrant concern is imperative to generating policy surrounding thier development and use. Because this is where the controversy lies, this paper will adopt the definition of autonomous weapons put forth by Crootof.

---

[19] Rebecca Crootof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CARDOZO L. REV. 1837, 1842 (2015).

[20] *Cf.* Erica H. Ma, *Autonomous Weapons Systems Under International Law*, 95 NYU L. REV. 1435, 1441 (2020) (adopting Crootof's definition of autonomous weapons systems). *Accord* Ford, *supra* note 15 at 418 (citing Crootof as authority for defining autonomous weapons systems).

[21] Horowitz, *supra* note 13 at 89.

[22] *Id.*

[23] *See* discussion *infra* Section II.B.

[24] Heather M. Roff & David Danks *"Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems*, 17 J. MIL. ETHICS 2, 4 (2018).

[25] Vazquez, *supra* note 8.

[26] *See* Justin Haner, Denise Garcia, *The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development*, 10 GLOBAL POLICY 331 (2019) (analyzing the leaders in autonomous weapons development); Bartlett, *supra* note 10.

II. HOW AUTONOMOUS WEAPONS WORK

The development of autonomous weapons has been made possible through advances in machine learning, deep learning, and neural networks.[27] These technological advancements, however, have also brought with them problems of blackbox AI whereby the explainability and predictability of an autonomous weapons system are complex, and at times, impossible to comprehend.[28] This section will provide some background information on how machine learning and blackbox AI contribute to the creation of fully autonomous weapons.

A. Autonomous Weapons Systems and Deep Learning

Deep learning is a process that enables AI to mimic the functioning of a human brain by creating neural networks.[29] The process of deep learning happens through a machine receiving inputs form the world, and through trial and error, learning how to create an algorithm that will produce a desired result.[30] What differentiates this from a typical algorithm is that the machine itself is creating and continually altering an algorithm as it gains more data and is run through more simulations.[31]

---

[27] *Cf.* Thomas G. Warschefsk, *Alexa, Whose Fault is it? Autonomous Weapon Systems Investigations and the Importance of a Deliberate Accountability Process*, 228 MIL. L. REV. 103,108 (2020) (noting that the deep learning and neural networks that enable the development of autonomous weapons systems are the very things that lead the controversy surrounding their development); Haney, *supra* note 11 at 64.

[28] Trumbull, *supra* note 4 at 568 (discussing how the nature of a self-learning system prevents us from being able to effectively predict the actions of an autonomous weapons system as it is placed in a new and untested situation).

[29] Deep learning algorithms were designed with this mimicry in mind. "Deep learning algorithms define an artificial neural network that is designed to learn the way the human brain learns. Deep learning models require large amounts of data that pass through multiple layers of calculations, applying weights and biases in each successive layer to continually adjust and improve the outcomes." IBM Cloud Education, *Machine Learning*, IBM (July 15, 2020) https://www.ibm.com/cloud/learn/machine-learning; *see also* Vishal Maini, *Machine Learning for Humans, Part4: Neural Networks & Deep Learning*, MEDIUM.COM (Aug. 19, 2017), https://medium.com/machine-learning-for-humans/neural-networks-deep-learning-cdad8aeae49b (explaining the mechanics behind neural networks and deep learning).

[30] IBM Cloud Education, *supra* note 29 ("In data science, an algorithm is a sequence of statistical processing steps. In machine learning, algorithms are 'trained' to find patterns and features in massive amounts of data in order to make decisions and predictions based on new data. The better the algorithms, the more accurate the decision and predictions will become as it processes more data.").

[31] *Id.* ("Machine learning is a branch of artificial intelligence (AI) focused on building applications that learn from data and improve their accuracy over time without being programed to do so.")

> "[The process takes] inspiration from biology, and [learns] by observing and experiencing. . . . Instead of a programmer writing the commands to solve a problem, the problem generates its own algorithm based on example data and a desired output. The machine-learning techniques that would later evolve into today's most powerful AI systems followed the latter path: the machine essentially programs itself."[32]

The complexity of deep learning and neural networks allows AI systems to generate algorithms that are more accurate than those created by humans.[33] Because these AI systems can outstrip human's capacity to predict outcomes, the algorithms that are generated are difficult, if not at times impossible, for humans to comprehend.[34] Furthermore, these algorithms continuously alter themselves as they are presented with new data and are therefore constantly evolving and improving their capabilities.[35] Thus, the more inputs provided to the AI, the more effective it will be in producing desired outcomes.[36]

B. Autonomous Weapons Systems and Blackbox AI

One feature of deep learning AI systems that naturally follows from its ability to outstrip human capacity is the concept of blackbox AI. Blackbox AI refers to the possibility that an AI can evolve its algorithms to the point where humans are unable to understand or explain its decision-making process.[37] Even if one is able to discern the decision-making process of the

---

[32] Will Knight, *The Dark Secret at the Heart of AI*, 120 MIT Tech. Rev. (2017) https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/.

[33] *See generally* KELLEY M. SAYLER, CONGRESSIONAL RESEARCH SERVICE, REP. NO. R45178, ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY (2020), https://fas.org/sgp/crs/natsec/R45178.pdf (discussing the potential benefits and concerns that arise from the development and implementation of autonomous weapons systems).

[34] *Cf.*, Knight, *supra* note 32 ("[B]anks, the military, employers, and others are now turning their attention to more complex machine-learning approaches that could make automated decision-making altogether inscrutable.")

[35] IBM Cloud Education, *supra* note 29 ("Machine learning is a branch of artificial intelligence (AI) focused on building applications that learn from data and improve their accuracy over time without being programed to do so.")

[36] There has been much discourse surrounding the way in which autonomous vehicles (AV's) can be trained. Like autonomous weapons systems, AV's are dependent on autonomous AI systems and machine learning. This means that the more data they are presented with to learn from, the more they will be able to improve their algorithm. Paul Christianson, *Billions of Miles of Data: The Autonomous Vehicle Training Conundrum*, CLOUDFACTORY (Sep. 30, 2020), https://blog.cloudfactory.com/autonomous-vehicle-training-conundrum.

[37] Alexander Lavin, *Interpreting AI is More Than Black and White*, FORBES (June 17, 2019), https://www.forbes.com/sites/alexanderlavin/2019/06/17/beyond-black-box-

AI, by that time, the algorithm could have evolved into something else.

Because AI systems develop algorithms based on inputs and learning, there are instances when they factor considerations into an algorithm in a way that is unintended.[38] In the commercial sphere, there are documented instances where an AI programed to vet applications learned that gender was a factor to rely on when screening applicants and thus based decisions whether a candidate was qualified partly on gender.[39] Without fully understanding an AI's algorithm, it is difficult to identify these problems before the damage has already been done.

Independent of these considerations, the blackbox nature of AI is one of the features that makes it beneficial. Being able to pick up on correlations between seemingly unrelated factors and accurately predict an outcome is what these systems are prized for.[40] As such, while it is important to acknowledge the risks, it is equally important to remain mindful that the benefits that autonomous AI provides. Methods of mitigating these risks in autonomous AI systems can take place in the design process.[41] Though these steps may increase the initial cost and complexity of development,[42] balancing the need for accuracy and the assurance that certain variables are not relied on is a tradeoff that can be imposed on the implementation of blackbox AI.

III. ETHICAL ARGUMENTS AGAINST THE CURRENT LEGAL STATUS

Though utilizing fully autonomous weapons has the potential to provide significant military advantages,[43] there has been a large pushback

---

ai/?sh=4feb7ef49c47 ("[D]eveloping an AI system to be interpretable is typically challenging and ambiguous. It is often the case that a model or algorithm is too complex to understand or describe because its purpose is to model a complex hypothesis or navigate a high-dimensional space . . . [n]ot to mention what is interpretable in one application may be useless in another.")

[38] Kristian P. Humble & Dilara Altun, *Artificial Intelligence and the Threat to Human Rights*, 24 J. INTERNET L. 12, 12 (2020) ("AI systems are not seen as either 'impartial or neutral,' AI collects and learns from data that it is given, which therefore reflects the social, historical, and political conditions in which the data was created.")

[39] Jordan Weissmann, *Amazon Created a Hiring Tool Using A.I. it Immediately Started Discriminating Against Women.*, SLATE (Oct. 10,2018), https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html.

[40] SAYLER, *supra* note 33.

[41]James Manyika et al., *What Do We Do About the Biases in AI?*, HARV. BUS. REV., (Oct. 25, 2019), https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai.

[42] *Id.*

[43] Amitai Etzioni, *Oren Etzioni, Pros and Cons of Autonomous Weapons Systems*, MIL. REV., 71, 72 (May-June 2017) (arguing that autonomous weapons systems allow for a decrease in number of warfighters, an increase in capacity per each warfighter, and a reduced

against their development as well as their use. There are concerns about the impact that implementing autonomous weapons systems would have on the way in which war operates and on their ability to comply with Law of Armed Conflict (LOAC).[44] The absence of a human within the kill chain, and the autonomous decision making process of these AI systems also brings with it, concerns with regards to identifying liability.[45] Though both concerns are valid, the extent to which they ought to deter a country from continuing to develop autonomous weapons systems is questionable. Since there are currently means that mitigate the impact of these risks, the above concerns ought to serve as a guiding principle towards further regulation rather than justifications for ending development.

A. The Loss of Meaningful Human Control

The first major concern regarding autonomous weapons systems is the worry that absent a human in the kill chain, autonomous weapons systems will be unable to comply with LOAC and ultimately cause more harm than good. If autonomous weapons systems are not properly generated, it is possible that they will mistakenly engage the wrong targets more often than the right ones; ultimately increasing civilian casualties and leading to a more dangerous state of war.[46] These concerns focus on the actual ability of autonomous weapons to comply with LOAC, specifically the principles of proportionality and distinction.[47] As of now, autonomous weapons systems lack the ability to understand the nuances, such as the context of a situation, that are necessary for adequate distinction and proportionality analysis and thus some argue their use would violate LOAC.[48]

---

risk of casualties).

[44] *See* discussion *infra*, Section III.

[45] Daniel N. Hammond, *Autonomous Weapons and the Problem of State Accountability*, 15 CHI. J. OF INTL. L. 652, 662 (2015) (discussing the difficulty of assigning liability to an agent when utilizing an autonomous weapons systems).

[46] HUMAN RIGHTS WATCH, MAKING THE CASE: THE DANGERS OF KILLER ROBOTS AND THE NEED FOR A PREEMPTIVE BAN 5 (2016), https://www.hrw.org/sites/default/files/report_pdf/arms1216_web.pdf (arguing that the sensory processing capabilities of autonomous weapons systems will be insufficient in giving the weapons systems to adequately distinguish between "active combatant from a civilian or an injured or surrendering soldier").

[47] *The Threat of Fully Autonomous Weapons*, CAMPAIGN TO STOP KILLER ROBOTS https://www.stopkillerrobots.org/learn/ (last visited Apr. 14, 2021).

[48] There have been arguments that the concept of distinction and proportionality that are required in LOAC must be satisfied at the weapons level for autonomous weapons systems to be in compliance with LOAC. HUMAN RIGHTS WATCH, *supra* note 46 at 4–6. As will be discussed in Section IV, compliance at the weapons level is not only difficult to program for, but also difficult to establish. However, this does not mean that these principles of LOAC

Though there are valid concerns with respect to implementing AI systems in their current state into weapons systems, the absence of a human in the kill chain would not necessitate that utilizing these systems would destroy meaningful human control, let alone violate LOAC. Analyzing current military procedures for engaging a target reveals that there are a multitude of measures that ensure meaningful human control prior to the actual deployment of a weapon.[49] Even though a human is not at exerting control over the end of the kill chain, humans are actively involved in making decisions on when and where to deploy these autonomous weapons systems. Because of this, meaningful human control still exists, just not at the end of the kill chain.[50] Understanding that this oversight of meaningful human control still exists when using autonomous weapons systems ensures that there are checks to safeguard compliance with LOAC as will be discussed in depth in Section IV.[51] Ultimately, the absence of a human from the end of the kill chain does not mean that there are no institutional checks on or means of meaningful control over these autonomous weapons systems. The control and thus liability simply rests a level above the autonomous weapons system in the commander that decides to implement the system.[52]

B. The Liability Gap

Aside from concerns about compliance with LOAC, many have expressed concern that the implementation of autonomous weapons systems would shield governments from liability for any accidents that the autonomous weapons systems create.[53] These autonomous weapons systems

---

are automatically violated through the use of autonomous weapons systems. See discussion *supra* Section IV.

[49] Merel Ekelhof, *Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation*, 10 GLOBAL POL'Y 343, 347 ("[T]he operator's control should be considered part of a larger process where there is a division of labor. . . Considering this, meaningful human control in relation to the human-machine relationship during the deployment to weapons is not the only, nor the most appropriate, approach to comprehensively address concerns of losing human control as it fails to take into account trivial factors of military practice.");

John Lewis, Comment, *The Case for Regulating Fully Autonomous Weapons*, 124 YALE L. J. 1309, 1314 (2015) ("commanders remain responsible for the initial use of [autonomous weapons systems]. A commander must give the order to deploy a [autonomous weapons system] and set parameters for its use . . . any weapon will require for him to intervene at some point, if only to activate it. The commander is ultimately responsible for using a [autonomous weapons system] within its programing and within its legal limits. If humans must remain an integral part of the decision to take a life in order for a weapon to fulfill the condition of accountability, then [autonomous weapons systems] satisfy this requirement.")

[51] *See* discussion *infra* Section IV.A.

[52] *Id.*

[53] *See* Ilse Verrdiesen et al., *Accountability and Control Over Autonomous Weapon*

could be used as a convenient scapegoat for nations to avoid liability since it is difficult to establish the intent necessary for liability when dealing with a weapon that makes decisions and acts autonomously.[54] The nature of the autonomous AI makes it unlikely that humans could be fully liable since, the requisite level of willfulness will necessarily be absent in the case of a weapons system that operates independent of human control.[55]

Despite these practical concerns, establishing a clear, formal chain of liability is not a prerequisite for employing a certain weapons system,[56] and the absence of a legal scheme of liability, does not necessitate that countries will be able to fully escape accountability. Even if there is no legal liability imposed on a country who utilized an autonomous weapons system, it is likely that lawfare could be utilized to hold them liable in the eyes of the public.[57] The concept of lawfare is premised on the notion that the law can be utilized in leu of military force to accomplish military outcomes.[58] One means of this can be done by publicizing violations of law committed by a country which has a strong sense of respect towards legal principles.[59] Though implementing autonomous weapons is not illegal, this strategy only requires

---

Systems: A Framework for Comprehensive Human Oversight, 31 MINDS & MACHINES 137, 144 (2021) ("[T]he use of emerging technologies, including autonomous weapon systems, with weak or without norms can lead to limited or easily avoidable responsibility and accountability for states and individuals.")

[54] *See* Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 U. PENN. L. REV. 1347, 1350–51(2016) ("By definition, war crimes – serious violations of international humanitarian law that give rise to individual criminal liability – must be committed by a person acting "willfully," which is usually understood as acting intentionally or recklessly.)

[55] *Id.*

[56] Charles J. Dunlap, Jr, *Accountability and Autonomous Weapons: Much Ado About Nothing?*, 30 TEMPLE INT'L & COMP. L. J 63. 75 (2016) ("In short, the presence or absence of civil liability is not- and should not be- a necessary condition as to the legitimacy of autonomous weapons.")

[57] General Charles Dunlap, one of the foremost scholars in lawfare, explains that "Lawfare can be defined is a means of utilizing the law to a countries advantage in leu of traditional military means. One means of accomplishing this end is utilize the law and the respect that some countries have towards it as a target. This can be accomplished through publicizing what the public may consider to be a violation of legal principles such as pictures of lawful attacks, even when there were no actual laws violated. This serves as a significant deterrent for future actions, even if they are technically in compliance with LOAC. This, however, does not always necessitate that actual violations of the law occur." Charles Dunlap, *Lawfare 101: A Primer*, 97 MIL. REV. 9, 10 (2017). To parallel this point, any incident that takes place with an autonomous weapons system could be publicized in a way to garner enough public and domestic outrage that it would block a country from continuing to implement the weapons systems in the way that they were whether such implementation rose to a violation of LOAC.

[58] *Id.*

[59] *Id.*

there to be a perceived violation of law, not an actual violation.[60] With well-established opposition to the use of autonomous weapons systems, it is not inconceivable that an autonomous weapons system that killed civilians could be framed to the public as a violation of LOAC. This could generate substantial public outcry and resentment which could be detrimental to the support and effectiveness of a country's military.[61] As a result, even though there may not be legal liability, there will likely be social responsibility imposed on countries employing autonomous weapons systems.

Finally, the lack of a natural scheme of liability does not necessitate that one could not be generated.[62] Instead of pointing to this lack of liability as a justification to end the development of autonomous weapons system, a more logical approach would be a call for developing an appropriate scheme of liability for the actions of autonomous weapons systems.[63] Since countries have yet to successfully develop and employee fully autonomous weapons, the time for creating such a scheme and addressing these problems is now.

IV. CURRENT LAW GOVERNING AUTONOMOUS WEAPONS

Despite the public outcry, there are currently no laws or treaties that specifically govern the development or use of fully autonomous weapons.[64] Because of this, autonomous weapons are governed by the LOAC and therefore must comply with the principles of proportionality, distinction, unnecessary military destruction, and military necessity.[65] Because proportionality and distinction are viewed as a combat level decision, some argue that in order for an autonomous weapons system to pass muster under

---

[60] *Id.*

[61] *Id.*

[62] *See generally* Crootof *supra* note 54 at 1395 (advocating for the adoption of a new scheme of liability for autonomous weapons systems).

[63] For a proposal on a potential framework for liability in connection with autonomous weapons systems see discussion *Infra* Section VI. A.

[64] "Until the achievement of a consensus, there is not legal basis for banning or even limiting the development or employment of autonomous weapons systems. Hence, States are free to continue to research and develop such systems, with the commitment to review each system at appropriate stages in the development an employment process to ensure that the weapon can fully comply with all LOAC requirements." Eric Talbot Jensen, *The (Erroneous) Requirement for Human Judgment (and Error) in the Law of Armed Conflict*, 96 INT'L L. STUD. 26, 56 (2020).

[65] *See* Jensen, *supra* note 64 at 28–27; Allyson Hauptman, *Autonomous Weapons and the Law of Armed Conflict*, 218 MIL. L. REV. 170, 171 (2013) ("The law requires that the targeting decision maker, whether a human, machine, or human-machine team, be able to apply fully and comply with the LOAC. [Thus autonomous weapons systems] must be capable of fully applying to LOAC, including minimizing civilian death and injury and damage to civilian objects.")

LOAC, the system itself must be able to comply with these principles.[66] This section will outline the analysis of proportionality as well as distinction with respect to autonomous weapons systems at the level of the weapons system, yet ultimately demonstrate how weapons level analysis is not the only means of compliance.

A. Compliance with Proportionality

Proportionality demands that the incidental damage to property and loss of life that results from an attack must not be excessive when compared to the military advantage that the attack is expected to gain.[67] Since autonomous weapons engage with targets independent of a human command, it is not illogical to desire such systems to contain an independent means of conducting a proportionality analysis prior to engaging with a target. Unfortunately, ensuring that autonomous weapons systems are programed to conduct proper proportionality analysis can be difficult. The standard for proportionality is subjective because in order to satisfy it, the ultimate decision must be one that a "reasonably well-informed person" would have made when weighing the potential collateral damage against the potential military objective.[68] This means that the autonomous weapons system must be able to derive the value of a military objective and weigh it against the potential casualties that engaging a target may impose.[69] Though AI could enable autonomous weapons to outstrip human capacity in calculating future harms, there is an argument that without continual updates as to military objectives, autonomous weapons systems will conduct improper proportionality analyses.[70] The military advantage that would result from an autonomous weapon's various functions could drastically change between the time they are activated and the time that they engage with a target. Because of this, ensuring that autonomous weapons systems can conduct a proper proportionality analysis would require the systems to be continuously updated with new information as military objectives changed.[71]

Because this systems approach poses logistical problems, some

---

[66] *See* Amitai Etzioni & Oren Etzioni, *Pros and Cons of Autonomous Weapons Systems*, MIL. REV., May-June 2017 at 72–74. (arguing that an AI system would have to be able to distinguish between combatants and civilians as well as conduct a proper proportionality analysis for it to be implemented in compliance with existing laws).

[67] GARY D. SOLIS, LAW OF ARMED CONFLICT'S CORE PRINCIPLES, IN THE LAW OF ARMED CONFLICT: INTERNATIONAL HUMANITARIAN LAW IN WAR 293 (2d ed., 2016).

[68] Ford, *supra* note 15 at 444.

[69] *Id.*

[70] *Id.* at 445.

[71] Marco Sassoli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 INT'L L STUD. 308, 332 (2014).

contend implementing a proportionality analysis at the level of commanders rather than the system would mitigate the concerns and provide for an adequate proportionality analysis. If the proportionality analysis were to remain at the level of the commander, the commander would simply factor into their traditional proportionality analysis the extent to which an AI system could adequately conduct a proportionality analysis in the field.

> "A relevant legal question today may be whether a commander is reasonably confident that a particular target is a military objective and not a civilian object. In the future, the inquiry may be whether the commander is reasonably confident that an autonomous weapon will be capable of determining that an object is a lawful target."[72]

Ultimately, the variable nature of the weapons simply becomes another factor in a standard proportionality analysis attributable to the commander. Under this framework the absence of a measurable proportionality analysis at the systems level would not automatically create a violation of LOAC any time these weapons systems were used.[73]

Though there is no consensus on which level of proportionality analysis would be appropriate for determining whether to implement an autonomous weapons system, there is the possibility that the analysis could remain unaltered. Though proportionality is something that ought to be taken into consideration during the design of an autonomous weapons system, its overall ability to conduct a proportionality analysis is not necessarily the factor that determines it's compliance with LOAC.

B. Compliance with Distinction

The principle of distinction requires a party in an armed conflict to distinguish between civilians and military objectives, and only allows for the targeting of the latter.[74] Since autonomous weapons systems lack a human within the kill chain, there is an argument that the responsibility of distinction falls on the autonomous weapons system itself.[75] Because of this, some suggest that autonomous weapons systems must be able to distinguish as well

---

[72] Trumbull *supra* note 4 at 563.

[73] Focusing the proportionality analysis on the command rather than the weapons level opens the possibility that not utilizing autonomous weapons systems could be a violation of the proportionality analysis. One example is a fully autonomous weapon that is blowing up a building yet is programed to shut off if "senses civilians are present" Ford, *supra* note 15 at 430.

[74] SOLIS, *supra* note 67 at 269.

[75] Jensen, *supra* note 64 at 28–27; *see also* Hammond, *supra* note 45 at 674 (characterizing autonomous weapons system compliance with the principle of proportionality to take place at the weapons level).

as their human counterparts to comply with the law of distinction.[76] "If used to select and engage targets autonomously the [autonomous weapons system] must be able to distinguish combatants v. non-combatants, between military objectives and civilian objects."[77] Thus, one way of measuring whether an autonomous weapons system complies with the principles of distinction can be the extent to which the weapons system itself can adequately comply with the current LOAC principles of distinction.

As discussed above, autonomous weapons systems must be able to discern between combatants and noncombatants, and therefore, targetable, and non-targetable persons. The difficulty that accompanies this measurement, however, coincides with the fact that combatants are not required by law to distinguish themselves with specific emblems so long as they distinguish themselves from civilians in some way.[78] This means that in absence of data on what the distinguishing feature is, it would be unlikely that an autonomous weapons system would be able to adequately engage with a lawful combatant.[79] Though this inability to identify the enemy may not be the critics' greatest concern, this same inability could lead to autonomous weapons systems improperly distinguishing and engaging with civilians and non-targetable persons.[80] This potential for violations of LOAC is only exacerbated by the fact that a combatant may transition from a targetable to a non-targetable person based on the context of the situation.[81] A combatant in one situation may be targetable, but in a split second, if they appear to be surrendering, incapacitated, or are under control of an adversary, they are no longer targetable.[82] Though it is possible that an autonomous weapons system could eventually learn to differentiate context, as of now, they lack that capability and thereby pose a problem for their ability to satisfy the principle of distinction.[83]

Furthermore, being able to code for an autonomous weapons system that can distinguish at, or even better than, the level of humans is difficult to measure. Currently the capability for these systems to distinguish combatants is far from that of humans due to autonomous weapons systems inability to

---

[76] *Id.*

[77] Vazquez, *supra* note 8 at 104.

[78] Elliot Winter, *The Compatibility of Autonomous Weapons with the Principle of Distinction in the Law of Armed Conflict*, 69 INT'L & COMP. L. Q. 845 (2020).

[79] *Id.*

[80] Hauptman, *supra* note 65 at 192 ("This could potentially be solved by allowing programming systems to always err on the side of caution, but it means giving up a number of opportunities to achieve a military victory that a state may not be willing to forego. . . .")

[81] Winter, *supra* note 78.

[82] *Id.*

[83] Cummings, *supra* note 17 (explaining how the current state of autonomous AI is currently insufficient in distinction to be safely implemented).

adequately distinguish context.[84] Aside from perceiving context, since AI systems rely on patterns of pixels for recognition, slight alterations in those patterns can impede the ability of the system to adequately recognize an object making them easily deceive.[85]

As AI continues to develop, its ability to adequately distinguish targets could improve which begs the question, how do we discern when AI has reached a level of adequate distinction. There may be a time in the future where their ability to adequately distinguish targets outstrips that of humans, at which point, some have argued there would be a moral obligation to utilize autonomous weapons systems. [86] The more pressing question, however, is what level of distinction is required for it to be permissible to utilize autonomous weapons systems?[87]

Though some contend that the superior to human metric ought to apply even with regards to standards of permissible use,[88] this would require holding a weapons system to the same standard that the law imposes on agents with culpability.[89] Though autonomous weapons systems have the ability to select and engage targets, at the end of the day they are still weapons and thus their ability to distinguish can be analogized to a typical weapons accuracy. Though the accuracy of a weapon–like an autonomous weapons system ability to distinguish targets–ought to be considered when determining if using it would be permissible, it should not be dispositive and current law reflects this notion.[90] The ability of the autonomous weapons system to adequately distinguish between targetable and non-targetable person, like its ability to conduct a proportionality analysis, could simply

---

[84] Trumbull, *supra* note 4 at 576.

[85] Tobias Vestner & Altea Rossi, *Legal Reviews of War Algorithms*, 97 INT'L L. STUD. 509, 513 (2021).

[86] *See* Cummings, *supra* note 17("If autonomous targeting systems in dynamic settings could be shown to be superior to humans at some point in the future. . . not only, should we use them, but we have an obligation to do so.") *See also* 305. *The Convergence: The Policy and Law of Lethal Autonomy with Michael Meier and Shawn Steene*, MAD SCIENTIST LABORATORY (Feb. 18, 2021) https://madsciblog.tradoc.army.mil/305-the-convergence-the-policy-and-law-of-lethal-autonomy-with-michael-meier-and-shawn-steene/ (arguing that autonomous weapons will at least initially have to preform above the "better than human" standard to be implemented).

[87] Another possible factor to consider is whether or not the absence of emotions that autonomous weapons system possesses will lead them to make decisions that may overall be more effective, but may ultimately violate common consideration of what is considered right and wrong. Jesse Prinz famously argues that emotions are the means through which we learn ethics, therefore, taking emotions out of these life and death decisions may lead to consequences that are morally repugnant. *See generally* JESSE PRINZ, THE EMOTIONAL CONSTRUCTION OF MORALS (2007).

[88] Cummings, *supra* note 17.

[89] Vazquez, *supra* note 8 at 104.

[90] *Id.*

become a factor that is considered in a commander's own proportionality analysis. However, until we reach a point where we have confidence in autonomous weapons systems ability to distinguish targets, implementing them would likely violate the principle of proportionality above.[91]

V. CONCERNS UNADDRESSED BY THE CURRENT LAW

Under LOAC, as it stands, autonomous weapons are not expressly prohibited. As discussed in Section IV, the implementation of autonomous weapons does not necessarily call for a weapons level analysis to satisfy the principles of proportionality and distinction. This means that whether a weapon system is used in compliance with LOAC falls on whether a commander conducts a proper proportionality analysis. This type of compliance, however, leaves much to be desired. Implementing a command level analysis risks undermining the principles of LOAC, fails to address concerns of broad political goals and does nothing to fill the liability gap that remains when commanders conduct adequate proportionality analysis.

Looking first to concerns with respect to undermining LOAC. Though weapons systems analysis may not be necessary for compliance with LOAC, understanding and demanding that weapons systems can conduct proper proportionality analysis and distinction will be imperative for a commander to adequately conduct a proportionality analysis. The problem that will prevent ultimate compliance stems from the ever-changing nature of autonomous weapons systems. First, there is a concern that when utilizing autonomous weapons systems, commanders will give greater deference to autonomous weapons systems than their own human judgement.[92] This factor could inhibit a commander's ability to adequately weigh the potential risks of uncertainty that utilizing an autonomous weapons system imposes. Commanders would be required to base their analysis on the information they had available. The United States standard practice is to provide weapons review of weapons systems and the DOD has asserted that continual review of autonomous weapons systems will be implemented.[93] Information from

[91] *See* Lewis, *supra* note 50 at 1324 ("If a commander knowingly deploys a [autonomous weapons system] with weak targeting software in the middle of a city, and it kills dozens of civilians, most will likely agree that the commander has committed a crime, or at least should be subject to sanctions of some kind.")

[92] Professor Crootof expressed concern that humans will rely heavily on the bias that AI has superior decision-making capabilities to humans and thus defer to their judgment in cases where they ought not to. Rebecca Crootof, member, Center for New American Security's Task Force on Artificial Intelligence and National Security at the Duke Law National Security Law Conference: Artificial Intelligence and the Future of Warfighting: A Dialogue (Feb. 25, 2021).

[93] U.S. DEP'T OF DEFENSE, DIR. 3000.09, Autonomy in Weapons Systems 13 at 2 (Nov.

these reviews could be utilized by commanders, however, this information does not account for the fact that autonomous weapons systems evolve over time. Thus, along with the information given on the weapons system, when making a proportionality analysis, a commander must factor into the analysis how the weapons system could potentially evolve. The subconscious bias that tends to view AI as superior could lead to legitimate proportionality analysis based on improper widespread assumptions about the reliability of the evolution of autonomous weapons systems.

Biases within the autonomous weapons system are just as concerning as the bias commanders may harbor in favor of AI decision making. The constant evolution of AI systems that could undermine LOAC could also lead to these systems undermining broad political goals. Without knowing how an autonomous weapons system will operate in a given situation, there is a possibility that these weapons systems could act in a way that ultimately achieves it's objective, but does so at the expense of broader political goals. If a commander cannot predict exactly how an AI will operate, there is no means of ensuring that these goals, even if they are known by the commander, will be achieved correctly.

Finally, the current framework does little to alleviate the concerns surrounding that absence of liability. Though commanders under this framework could be held liable for the actions of autonomous weapons systems, they will only be held liable if they fail to adequately conduct a proportionality analysis. Thus, if an autonomous weapons system acts independently, they escape liability. Because of this, LOAC as it stands does little to fill that gap that develops with responsible implementation of autonomous weapons systems.

VI. REGULATION NOT OVERREACTION

Though the current state of fully autonomous weapons may leave much to be desired,[94] this does not mean that the future of autonomous weapons is grim. As countries continue to develop and invest in these weapons systems, and as private industries continue to develop autonomous weapons systems, their capabilities and sophistication will likely grow. With the inevitability of this development, it is important to be proactive in thinking about how these weapons systems ought to be developed and regulated.[95] Because there are valid concerns with regards to both the

21, 2017) [hereinafter DOD DIRECTIVE 3000.09].

[94] *See supra* Section IV.B. discussing the challenges AI faces with respect to distinguishing in context. *See also* Cummings, *supra* note 17 (discussing how AI systems perception is inadequate to distinguish targets independent of context).

[95] Blaine Ravert, *The Ethics of the Kill Decision: Should Humans Always be in the*

operations of these AI systems as well as possible schemes of liability, confronting these issues head on will help mitigate any negative externalities that could result from the development of this technology, and in turn, help mitigate public fear of their use. There is a myriad of ways that we can mitigate the fears associated through regulation.[96] This section will outline potential regulations regarding the development, the use, the liability and the training of autonomous weapons systems.

A. Development

Because there has been a shift in the weapons development sector from public to private companies,[97] many of these new autonomous weapon's systems are currently being developed without direct government supervision.[98] Though this may be amenable to most other weapons, requiring experts in both LOAC as well as the policy goals of the government to be involved in the design process is imperative. This requirement would alleviate the unique concerns that the autonomous nature of these weapons systems raises with regards to compliance with LOAC and those that their blackbox nature raises with regards to potential frustration of policy goals. Both these objectives can be accomplished by regulations requiring private companies developing autonomous weapons systems to consult with a new government agency charged with ensuring these ends.[99]

Loop?, THE CIPHER BRIEF (Feb. 7, 2021), https://www.thecipherbrief.com/column_article/the-ethics-of-the-kill-decision-should-humans-always-be-in-the-loop (arguing for the importance of creating a framework that will help to mitigate any negative externalities associated with the development and use of autonomous weapons systems).

[96] Andrew Figueroa, *License to Kill: An Analysis of the Legality of Fully Autonomous Drones in the Context of International Use of Force Law*, 31 PACE INT'L L. REV. 145, 157 (2018)

[97] Charles Mahoney, *Private defense companies are here to stay – what does that mean for national security?*, THE CONVERSATION (May 31, 2017 9:57 PM) https://theconversation.com/private-defense-companies-are-here-to-stay-what-does-that-mean-for-national-security-76070 ("Like it or not, government agencies responsible for national security are dependent on private defense firms.")

[98] *See generally* Kara Frederick, *The civilian private sector: part of a new arms control regime?*, OBSERVER RESEARCH FOUNDATION (Nov. 06, 2019), https://www.orfonline.org/expert-speak/the-civilian-private-sector-part-of-a-new-arms-control-regime-57345/ (discussing the interaction of the private sector in the development of autonomous weapons systems).

[99] It has been widely suggested that for the private sector, an agency be developed to oversee and ensure that the development of AI technology is done in a safe and ethical way. Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies*, and Strategies, 29 HARVARD J. OF L. TECH. 354 (2016). Though no such agency currently exists, there have been proposals that outline how that agency ought to

One potential function of the agency could be to require experts of LOAC to be integrated into the design process of autonomous weapons systems. As has been suggested throughout the paper, there are concerns that autonomous weapons systems will fall short of compliance with principles of distinction and proportionality.[100] Whether we adopt the weapons level or the command level approach to compliance with LOAC, the same framework that will ensure autonomous weapons systems comply with LOAC at a weapons level will also increase their chances of passing a proportionality analysis. Because of this, autonomous weapons ought to be programed to generate algorithms that do not blatantly violate these principles. Embedding legal reviews within the verification and validation process as opposed to conducing it after such process would serve as a necessary safeguard in ensuring the weapons systems are developed in ways that are mindful and respectful of LOAC.[101] Because of this, it is important that military Judge Advocates are directly involved in their development and design.[102] If we can develop these weapons with a focus on their compliance with LOAC at the center of their development, then not only could we ameliorate the concerns of their compliance, but also create a new type of warfare with less civilian casualties.

Along with compliance, the agency could also impose the requirement that experts in the policy goals of the country be incorporated into the design process. It is important to ensure that these programs are vetted for subconscious bias that could lead to engaging targets that would create policy problems. Therefore, it is crucial that during the design and testing process, the outcomes of the autonomous weapons systems be vetted by an entity in the pentagon. This entity would be responsible for determining whether the outcomes that the weapons systems produce would be averse to policy goals or appear to take into consideration factors that are inappropriate.

These two steps taken together at the development stage could help to mitigate the concerns and the fears about both the compliance with LOAC and fears of creating negative political externalities with little to no

function. *Id.* As will be suggested in Section, VI, creating an agency to regulate the development of autonomous systems could serve the same function of the suggested agencies in the private sector, that is ensure that the development and implementation of AI is done responsibly. Though a creation of a separate agency for autonomous weapons systems could be possible, it would also be possible to give an agency to dedicate a portion of the agency that oversees the development and certification of autonomous AI systems this task.

[100] *Supra* Section III.

[101] *See generally* Vestner, *supra* note 85 (arguing that the legal review process ought to be embedded in the development of autonomous weapons systems to combat problems related to explainability and predictability that such systems raise).

[102] *See* Lewis *supra* note 50 at 1316 ("A 'complete regulatory scheme would also tackle other thorny issues, including research, testing acquisition, development, and proliferation.")

justification. To ensure that private companies have met their design burden, there ought to be a requirement that autonomous weapons systems be certified by the agency prior to use.[103] This means that the autonomous weapons systems.

> ". . . must meet very strict certification criteria, both at the strategic layer in target selection and in the design layer with autonomous target identification. Such weapons systems should be proven through objective and rigorous testing and should demonstrate an ability to perform better than humans would in similar circumstances, with safeguards against cybersecurity attacks."[104]

Founding an agency to take regulate the design process, but also tasked with creating standards of certification for autonomous weapons systems would ultimately ensure that these weapons systems would be compliant with LOAC and the goals of the government by the time they are initially implemented.[105]

B. Periodic Audits

Because autonomous weapons are not stagnant, ensuring that they are periodically tested is crucial to ensuring that the decision of commanders to implement these weapons systems are being based on information relevant to the newly evolved AI controlling the weapons system. To combat the problem that command liability would impose with regards to improper information, after autonomous weapons systems have been developed, periodic audits on the AI systems that control the weapons should be imposed to ensure that the weapons systems both continue meeting certification requirements and to conduct an analysis on the decision-making process of the autonomous weapons system. Because the AI learns from itself, it is constantly evolving and changing, thus, a periodic recertification is necessary to ensure that the autonomous weapons system still meets the previously

---

[103] Though the DOD currently conducts weapons reviews for new weapons, the nature of autonomous weapons requires a different and more in depth analysis than is currently provided for by the weapons review process. *See generally* MICHAEL W. MEIER, *Lethal Autonomous Weapons Systems: Is it the End of the World as We Know It . . . Or Will We Be Just Fine?*, in COMPLEX BATTLESPACES 289 (Winston S. Williams & Christopher M. Ford ed. 2019). Though the DOD could amend their review process for autonomous weapons systems instead of setting up a designated agency, there would need to be a division that is devoted to the review of autonomous weapons systems because of the expertise required to evaluate the systems as well as the unique procedures and testing that would need to take place.

[104] Cummings, *supra* note 17.

[105] For a discussion on the framework for an adequate certification process see generally MEIER, *supra* note 103.

established standards of compliance.

The issue that some may raise about the audits is that AI's constant evolution harkens back to the issue of the blackbox and the inability to understand and thus defend the actions of an autonomous weapons systems. When making the decision to utilize an autonomous weapons system, the commander making the call, or the government acquiring the weapons system, must have trust in the weapons system. "This level of trust in something autonomous, whether human, animal, or machine, requires more than mere predictability and reliability; it requires the trustor to understand why the trustee does what he, she, or it does."[106] Just as one ought to be able to defend their actions in combat, one ought not utilize autonomous weapons in situations where they cannot understand why the autonomous weapon would be justified in using force, especially when targeting people.[107] Thus, understanding the way that autonomous weapons are operating and generating their decisions is critical and ought to be evaluated periodically.[108] Because autonomous AI systems continuously evolve, it would be impossible to know that every algorithm of every autonomous weapons system is both compliant and explainable. This, however, does not mean that the audits are futile. The audits serve as a safeguard and a means of ensuring the autonomous weapons systems are evolving properly.[109]

C. A Framework for Liability

Creating a clear framework for liability is another important step that should be taken prior to the implementation of fully autonomous weapons.[110] Though the United States has already demonstrated that it will hold commanders liable to some extent, holding them strictly liable for the actions of an autonomous weapons system when they had no way of knowing the potential harm is an unsavory approach.[111] This does not mean that

---

[106] Roff, *supra* note 24 at 4.

[107] Because explainability is something that may not be achievable in light of autonomous AI, it has been argued that the emphasis ought to be on imposing human responsibility in ways that do not require explainability. Warschefsk, *supra* note 27 at 112.

[108] Vazquez, *supra* note 8 at 104.

[109] Pat Huston, Assistant Judge Advocate General for Military Law and Operations at the Duke Law National Security Law Conference: Artificial Intelligence and the Future of Warfighting: A Dialogue (Feb. 25, 2021).

[110] There are suggestions that the liability can be reduced to products liability case, however such treatment would be imposing obligations on victims to bring claims in order to ensure liability on the manufacturers. Hammond, *supra* note at 45 679–80. Furthermore, this could simply result in an increase in price due to an increase in liability and will do nothing to incentivize the production of safer autonomous weapons. Vazquez, *supra* note 8 at 104.

[111] DOD DIRECTIVE 3000.09, *supra* note 93 at 2("Persons who authorize the use of,

commanders ought not bear liability in any situation, only that their liability ought only extend to the extent that they conducted an appropriate proportionality analysis. As will be discussed, all other liability ought to fall on the country that employed the autonomous weapons system.

1. Commander Liability

The fear that humans are exempt from liability for the actions of autonomous weapons systems is inconsistent with the United States current stance on commander's liability. The DOD has asserted that "Autonomous . . . weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."[112] Because of this, at least a proportionality analysis is imposed on commanders. To conduct a proper proportionality analysis each commander must have "a reasonable understanding" of the autonomous weapons system.[113]

> "This means that in order for designers, commanders, operator and others involved with autonomous weapons to avoid liability, the devices- like any weapon- must be designed and tested so that their expected actions against life and property can be reasonably anticipated. This is nothing new to the law of war practitioners."[114]

As such, commanders are obligated to be trained and educated on the capabilities and effectiveness of the autonomous weapons systems for which they are responsible, as well as the potential liabilities that utilizing such a system could impose. Aside from simply the sophistication of the weapons system,[115] other factors that a commander ought to take into consideration when conducting their proportionality analysis include whether the boundaries for which the autonomous weapons systems are employed are adequately restricted,[116] continual updates to the autonomous weapons

---

direct the use of, or operate autonomous . . . weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement.")

[112] U.D. DEP'T OF DEFENSE, DIR. 3000.09, AUTONOMY IN WEAPONS SYSTEMS 13 (21 Nov. 2012) (C1, 8 May 2017). Para. 4a.

[113] Dunlap, *supra* note 56 at 69.

[114] *Id.* at 70.

[115] *"Sophistication*: Deploying an [autonomous weapons system] that is of such advanced technological sophistication that it can identify direct participants with reasonable certainty. In the above example, such a system could be programmed to determine when--using the ICRC language--a person is physically separated from the operation." Ford, *supra* note 15 at 439–40.

[116] "*Restriction*: Limiting the [autonomous weapons systems] geographical boundaries of operation, duration of the deployment, or target set/type such that the issue of direct participation will not arise. This is most easily accomplished by deploying the system for a

system of human identified participants,[117] as well as retained control and oversight of the autonomous weapons system while it is deployed.[118] So long as an appropriate proportionality analysis was conducted at the command level, the commander ought to be discharged of liability associated with autonomous weapons systems.

Commander liability ought not hold commanders liable for the actions of autonomous weapons systems in cases where they conducted a proper proportionality analysis.

> "To hold a commander responsible for an [autonomous weapons system's] action that he could neither control nor foresee would thus go beyond the traditional scope of command responsibility. It would also cut against ethical notions about criminal liability. If commanders are not held strictly liable for their subordinates' actions, it seems 'unfair to impose liability on commanders for their fully autonomous weapons,' as these weapons will exercise a similar degree of autonomy."[119]

Some, however contend that commander's responsibility for employing autonomous weapons systems ought to be all encompassing.[120] One proponent of this view Marco Sassóli puts forth the argument that "it is a as fair to hold a commander of a robot accountable as it would be to hold accountable a commander who instructs a pilot to bomb a target he describes as a military headquarters, but which turns out to be a kindergarten."[121] In the situation, the pilot has been given orders to directly engage with a specific target and does not deviate from those orders. This, however, is a far different case than autonomous weapons systems that both engage *and* identify targets independent of a commander. A stronger analogy would be if a commander instructed a pilot to bomb a target that he describes as a military headquarters, but instead of actually bombing the target, the pilot makes the decision to bomb a

---

discrete task or for a very short period." *Id.* at 40.

[117] "*Updates*: Updating the [autonomous weapons systems] with human-identified direct participants." *Id.*

[118] "*Human Involvement*: Retaining operator control or oversight of the [autonomous weapons systems] during deployment. This would include humans on, in, or near the loop." *Id.*

[119] Hammond, *supra* note 45 at 665.

[120] James Kraska, *Command Accountability for AI Weapon Systems in the Law of Armed Conflict*, 97 INT'L L. STUD. 407, 439 (2021).

[121] Marco Sassóli, *Autonomous Weapons and International Humanitarian Law; Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 INT'L L. STUD. 308,324 (2014).

kindergarten based on the knowledge that doing so will somehow better achieve the overall military objective desired. Though this example may be extreme, it illustrates the point that once developed, autonomous weapons systems have the potential of operating independent of strict rules. In the latter example, it would be hard to say that the commander ought to be liable for the actions of the pilot because the pilot made an independent decision that the commander had no reason to suspect would be made. Because of this, holding commanders liable to the extent that they conducted a proper proportionality analysis with the information they have on the weapons system is the most appropriate standard of accountability.

2. Filling the Liability Gap

Even when a commander conducts an adequate proportionality analysis, if an autonomous weapons system acts improperly, liability ought to still be placed on the country that employed the weapon. This does not mean that liability for the actions of the autonomous weapons system fall on no agent.

> "Even where the system is acting with extreme levels of autonomy, it is--at most--an organ or agent of the State whose actions are attributable to the State. Actions will be attributable even where the system is acting in an entirely unpredictable manner and beyond the scope of the initial deployment."[122]

The actions thus fall on the state that authorized the use of the autonomous weapons system and thus the risk that utilizing such a system involved. Independent of these means however, a strict liability approach ought to be implemented.[123] Governments that knowingly employ these autonomous weapons systems are aware of the risk that they bring on and stand to benefit from their use. Thus, it makes sense to hold them strictly liable in cases where there is a liability gap for accidents involving autonomous weapons systems.

D. Bifurcated Commander Liability

Finally, there are concerns that the development and implementation of autonomous weapons system poses a threat since, to develop, these

---

[122] Ford, *supra* note 15 at 40.

[123] *See* Crootof, *supra* note 54 at 1395 (explaining how strict liability is the metric that countries ought to be accountable for when using autonomous weapons because it is difficult to determine if appropriate precautions were taken when employing the autonomous weapons systems).

systems are required to learn through actual field experience. The best way for autonomous weapons systems to get to the level of being fully autonomous is for them to learn in real life situations.[124] Questions then arise as to how and when a country ought to be permitted to implement and utilize a fully autonomous weapon, and furthermore, what regulations or limitations ought to be imposed on such utilization. While currently there is not a consensus among nations,[125] it appears that at a minimum, these weapons ought to pass a mandatory test to demonstrate their capability of complying with the LOAC as outlined above. Safety measures such as the kill switch and monitoring of the autonomous weapon system would serve as additional safeguards against any potential accidents. This measure would also serve to ensure that there was a reasonable amount of control in such situations to inhibit the algorithm from going array and thus may make a negligence approach appropriate.

Though the same level of strict liability ought to still govern nations in incidents when the autonomous weapons systems have not been certified, the level of liability for commanders who utilize uncertified systems ought to be higher. Commanders have a responsibility to ensure that their subordinates understand their obligations under LOAC and are liable if they "knew or should have known, about a situation and failed to take reasonable measures within their power to prevent, report, and punish" violations of their subordinates.[126] With this, they have an obligation to "seek information reasonably available to them" that would inform them of any potential problems.[127] If a commander were to implement an autonomous weapons system without it passing the certification process, a strong analogy could be drawn between a failure to both ensure that the subordinate (in this case, the AI system) understood its obligations under LOAC, and a failure to obtain information that was reasonable – i.e. information from mandatory audits – that would inform them about a potential violation. If an autonomous weapons system cannot pass a certification, then there is little reason to think

[124] For training Autonomous vehicles, "real world testing is the gold standard for collecting data and improving the cars' ability to operate safely." Like autonomous vehicles, autonomous weapons systems rely on the training of their AI to be effective. Thus, just as autonomous vehicles, autonomous weapons systems will function better the more they are exposed to real life training situations. Hayden Field, *Self-driving Cars are Being trained in virtual worlds while the real one is in chaos*, MIT TECH. REV. (May 22, 2020), https://www.technologyreview.com/2020/05/22/1002129/data-save-autonomous-vehicles-coronavirus/.

[125] *See* Crootof, *supra* note 92 (discussing the widespread debate among countries on the appropriate means through which they ought to be allowed to train their autonomous weapons systems).

[126] Matthew T. Miller, *Command Responsibility: A Model for Defining Meaningful Human Control*, 11 J. NAT'L SECURITY L. & POL'Y (forthcoming 2021).

[127] *Id.*

that such weapons system is safe or adequately tested. Thus, in cases where commanders choose to implement autonomous weapons systems that have either yet to pass certification or are overdue for an audit, they ought to be subject to strict liability.[128]

CONCLUSION

The future of autonomous weapons seems inevitable and therefore proper planning and preparation is crucial. Because an all-out ban does not seem plausible, developing these weapons while ensuring that they are used in compliance with the LOAC is imperative. Currently, LOAC does not give specific requirements for autonomous weapons and thus, as it stands, concerns with regard to the absence of a human in the kill chain, and the absence of a frameworks for liability, are valid. Though these concerns ought to guide future regulation, they by no means justify refusal to develop and subsequently use autonomous weapons systems. As this paper has outlined, regulation rather than overaction is the approach that ought to be taken by the United States. Regulating autonomous weapons at the development phase would help ensure that the weapon systems are not only compliant with LOAC but also with broader political goals. Ensuring that autonomous weapons systems are continuously audited will ensure continued compliance as well as provide for explainability as to the decision-making processes of the weapons system.[129] Furthermore, holding nations strictly liable for the actions of autonomous weapons while implementing a bifurcated liability approach for commanders would ensure that there is always an agent upon which liability can fall.

Though these are not the only ways that autonomous weapons systems could be regulated, they are a few that would address the predominate concerns surrounding their development. The important thing to note is that there are various options for regulation as opposed to a ban that would likely not only increase compliance but also help the autonomous weapons systems be developed in a safer and more secure way. With the

---

[128] This bifurcated liability system has been advocated for in the private sector. Balancing innovation against safety is a concern that proliferates through the discourse surround the development of AI. *See generally* Scherer, *supra* note 99. Autonomous weapons systems, like regular AI is faced with this tradeoff and thus, implementing heighted liability could serve as a saleyard to prevent the frivolous implementation of the weapons systems.

[129] Team TruEra, *Machine Learning Explainability is just the Beginning*, TRUERA (Mar. 25,2021) https://truera.com/machine-learning-explainability-is-just-the-beginning/#:~:text=Machine%20learning%20explainability%20is%20often%20viewed%20as%20a,to%20provide%20reassurances%20about%20how%20those%20models%20function.(arguing that explainability is an important insurance of employing an autonomous weapons system both at the outset and throughout it's continued evolution).

future lying in autonomous weapons systems, we have an obligation to be proactive in ensuring their design and implementation is executed responsibly in a way that avoids, when possible, imposing unnecessary negative externalities.

* * *