

Who's to Blame: Assigning Accountability to Autonomous Weapons Systems

Graham Overcash

As it becomes apparent that autonomous weapons systems (AWS) are going to play a role in future conflicts, it is important to develop a legal framework that can assign accountability when these weapon systems break international law. Who is to blame, for instance, when an AWS is unable to distinguish between a civilian truck and a military vehicle and unlawfully attacks the civilian? Who should be held accountable for this infraction of the laws of armed conflict? Should the soldier who initially launched the weapon be held accountable? Or what about the officer who ordered the strike, or even the programmer who coded the targeting system in the AWS? This blog will examine this question and attempt to ascertain how accountability should apply when AWS break the law.

The United States Department of Defense (DoD) issued a directive that defined lethal autonomous weapon systems as “weapon system[s] that, once activated, can select and engage targets without further intervention by a human operator.”¹ While nearly every country defines AWS differently, this blog will use the DoD’s definition. Without intervention by a human operator, a machine analysis, underpinned by machine learning, will make the decision about what to target.² Machine learning uses “algorithms and statistical models are used to analyse and draw inferences from patterns in data automatically.”³ These algorithms or models are “trained on a set of ‘training data’ to identify patterns or make predictions.”⁴ One issue AWS faces is the “real-world data to train AWS is limited in quantity and quality,” which may make the AWS less predictable and more prone to unintended engagements.⁵ With that in mind, it is important to be able to assign accountability if and when AWS breaches international law.

The DoD requires AWS to “allow commanders and operators to exercise appropriate levels of human judgment over the use of force.”⁶ While a human does not need to manually control the weapon, a human must ensure that the use of the AWS is “in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement.”⁷ Furthermore, there must be “broader human involvement in decisions about how, when, where, and why the weapon will be employed.”⁸ Though the DoD gives the AWS some autonomy to decide how and when the weapon is used, the DoD policy makes clear that there is a requisite level of human involvement needed to use AWS. This policy gives the law a basis to assign accountability to a human decision if the AWS breaches international law. It should be noted that although the DoD only gives a glimpse to America’s policies regulating the use of AWS, other countries have similar policies that require a certain level of human control.⁹ The requirement of

¹ Cong. Res. Serv., Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems, IF11150, at 1 (2024).

² AI in Weapon Systems Committee, House of Lords, Proceed with Caution: Artificial Intelligence in Weapon Systems, HL Paper 16, Session 2023–24, at 23 (2023).

³ *Id.*

⁴ *Id.*

⁵ *Id.* at 29.

⁶ U.S. Dep’t of Def., *DoD Directive 3000.09, Autonomy in Weapon Systems* 3 (2023).

⁷ *Id.* at 4.

⁸ Cong. Res. Serv., *supra* note 1.

⁹ AI in Weapon Systems Committee, *supra* note 2, at 32.

human control over AWS shows that these systems are “tools rather than agents,” allowing the law to assign accountability to humans when an AWS breaks the law.¹⁰

It is important to distinguish the individual who launches an AWS, the operator, from the one who orders the use of an AWS, the commander. Requiring a level of human involvement allows an investigation into a breach of international law involving an AWS “work back from the commander’s decision-making into the mechanics and design of the system.”¹¹ The individual who deploys the AWS does “not determine exactly when, where or against what force is applied,” thus making it difficult to assign accountability a decision the user may have no input over.¹² While this individual may directly responsible for the deployment of the AWS, it does not make sense for this individual to be accountable for a breach of international law committed by an AWS. The person who launches the AWS likely has no control of the system after it is launched or in the AWS targeting decisions. For this reason, the commander, as opposed to the operator, is the logical choice to be accountable for when an AWS breaches international law.

Similarly, those who programmed the AWS generally should not be held accountable for a breach of law by an AWS. This is for two main reasons. First, “[t]racing responsibility back to one of these individuals and identifying who among the various programmers, designers, data labellers and others are responsible for certain act or omission could be a significant challenge.”¹³ Second, the commander who deploys an AWS “must have a reasonable understanding of the AWS and how it will work before deploying it in a particular situation,” which transfers any issues with the programming underlying an AWS to the commander.¹⁴

The technicality of AWS systems also makes it hard to assign accountability for the actions of an AWS to an individual programmer. The programming and machine learning that underpins AWS and directs their targeting systems are incredibly complex and opaque. These machine learning models are described as “inherently opaque.”¹⁵ In fact, even the engineers that designed [an AWS] cannot reproduce the step-by-step decision as to why it chose to do A instead of B.¹⁶ This highlights the fact that assigning accountability to programmers based on the machine learning underpinning the AWS is incredibly difficult. There is a persuasive argument for why the international community needs to develop transparency standards for the targeting software used in AWS. Until those standards are developed, however, it will be prohibitively difficult to assign blame to the programmers behind an AWS.

The other compelling reason that programmers should not be accountable is because commanders should take “all necessary and reasonable measures in their power to prevent” war

¹⁰ *Id.* at 59.

¹¹ *Id.*

¹² State of Palestine, *Proposal for the Normative and Operational Framework on Autonomous Weapons Systems*, U.N. Doc. CCW/GGE.1/2023/WP.2/Rev.1, at 7 (2023).

¹³ *Id.*

¹⁴ Charles J. Dunlap, Jr., *Accountability and Autonomous Weapons: Much Ado About Nothing?*, 30 *Temp. Int'l & Comp. L.J.* 60, 69 (2016).

¹⁵ AI in Weapon Systems Committee, *supra* note 2, at 29-30.

¹⁶ AI in Weapon Systems Committee, *supra* note 2, at 29.

crimes.¹⁷ A commander who deploys an AWS without fully understanding if the AWS will act predictably and comply with the laws of armed conflict could breach international law. For example, if there was an issue with the programming that made the AWS act in an erratic way and a commander still chooses to deploy the weapon, the commander could violate this duty. Commanders “could be held accountable for knowingly deploying a potentially defective weapon.”¹⁸ If a commander deploys an AWS that was poorly coded so that it is prone to unpredictable behavior that could be a breach of international law. The commander has an obligation to ensure that weapon systems they deployed are not likely to breach the laws of war. While programmers need to be held to a high standard and cannot develop weapons that intentionally breach international law, the commander who deployed the AWS is legally responsible for the resulting actions, even if they are due to faulty programming.

Due to difficulties in assigning accountability to individual operators and programmers of AWS, the commander who orders the use of an AWS that breaks international law should be held accountable. If a commander issues orders or deploys a weapon system without “a reasonable belief that doing so would comply with the law of war,” then that commander is breaching their obligations under international humanitarian law.¹⁹ This applies to traditional weapon systems as well as new emerging systems such as AWS. Commanders must have a reasonable belief that the weapons and tactics they are using do not breach the laws of war. This is why the most appropriate individual to hold accountable when an AWS breaches international law is the commander who ordered the use of such systems. Commanders are responsible for ensuring their forces do not breach the laws of war, and this principle applies to AWS, just as it does to any other weapon system.

¹⁷ Int'l Comm. of the Red Cross, Customary IHL Database, Rule 153. Command Responsibility for Failure to Act, <https://ihl-databases.icrc.org/en/customary-ihl/v1/rule153> (last visited Sept. 29, 2024).

¹⁸ State of Palestine, Proposal for the Normative and Operational Framework on Autonomous Weapons Systems, U.N. Doc. CCW/GGE.1/2023/WP.2/Rev.1, at 7 (2023).

¹⁹ Dunlap, *supra* note 14, at 70.