

Queen Mary University of London

Autonomous Killers

The Ethical Implications of Lethal Autonomous Weapons Systems

Jaime Salmonson

Technology, Politics, and War POL303

Dr. Elke Schwarz

04.15. 2025

Introduction

Technology is constantly evolving – yesterday's invention is outdated by today's, and next week's invention is well underway. But sometimes, these changes are so significant that they completely revolutionize it's industry. The introduction of gunpowder and firearms in the 10th century was a groundbreaking technology that completely shifted the battlefield to another level – and this is what we are seeing today, another level of warfare: artificial intelligence integrated within automated weapons on the battlefields (Britannica, 2023).

Science fiction has, in a way, long served as society's mirror, reflecting our terrors about technological advancement before they materialize into reality. The Black Mirror episode “Metalhead” shows autonomous killer robots - sleek, metallic machines with advanced targeting capabilities - that murder humans. This episode's real power lies within the portrayal of the machines that operated with ruthless efficiency, having escaped meaningful human control after being engineered just a bit too well for their own creators (Netflix, 2017).

Lethal Autonomous Weapons Systems, also known as LAWS, are AI-driven military-grade technologies capable of sorting, selecting and engaging with targets *without direct human intervention* (Schwarz, 2019, p.164). These weapons range anywhere from autonomous drones to AI assisted missile defense mechanisms; LAWS represent an immense shift in what we call modern warfare, which may promise increased efficiency and less militant casualties, however, these systems raise several ethical, political, and security concerns. For this tech report in particular, I will focus on the ethical concerns of LAWS – specifically, the lack of meaningful human control and the issue of accountability. I will also analyze their impact on present day conflicts to, in turn, provide policy recommendations for regulating their use.

Lethal Autonomous Weapons Overview

Remote warfare has been used since World War I when the first introductory unmanned aerial vehicles were deployed, but it has evolved significantly since. Drones serve as a precursor to LAWS as they require human commanders to make targeting decisions, what we are witnessing today is an entirely new species of weapon: they are genuinely autonomous systems, meaning no human directly involved – a distinction that demands clarity.

Concerningly, these systems can generate "Kill Lists" based on a "triangulation of data" like lifestyle patterns, electronic footprint, and personal behavioural info (Schwarz, 2019).

Proponents like Arkin (2010) have argued that such systems could reduce civilian casualties by eliminating human weaknesses such as fatigue, emotion, and error from targeting decisions.

However, as this report will demonstrate, early deployments suggest differently. To better understand the extreme nature of these machines, we can take a look at an emerging leader in this industry: Anduril.

Anduril is a United States based defense technology startup specializing in autonomous weapons systems (AWS), including air systems, underwater vehicles, and rocket motors. These systems run on their open software platform: Lattice. What distinguishes Anduril from conventional drone manufacturers is their focus on minimizing human input while maximizing algorithmic decision-making. Their arsenal of "Air Systems" – ominously named Bolt, Barracuda, Roadrunner, Fury, Ghost, and Altius – convey a sense of aggression and dominance through their names before even considering their capabilities (Anduril, 2023a).

The Bolt system, in particular, is a perfect example of Anduril's pursuit of lethal efficiency. Marketed as a "man-packable autonomous air vehicle," Bolt operates with minimal human input in order to execute missions like intelligence, surveillance, reconnaissance (ISR), however, its true nature is revealed in the Bolt-M variant, which is explicitly designed for "simple, flexible, and lethal precision firepower." A light 12-15 pounds, and able to strike targets over 20 kilometers away, Bolt-M is essentially an aerial sniper *without the need for a human finger on the trigger*. Instead, the trigger is now a hand-held remote device miles away from the target manned by someone that, according to Anduril's website, can operate with "minimal training". It is no surprise that Bolt is partly controlled by a plastic, handheld device. Afterall, Anduril was co-founded by the CEO of Oculus (Anduril, 2023b).

Context in Present Day Battles

Anduril's mission statement asserts, verbatim, "An unfair advantage for unrivaled deterrence: an advanced software platform enabling a single human to control and coordinate a wide range of autonomous assets across ocean, land, and sky to deliver successful mission outcomes." From this mission statement, there are two key takeaways: first, understand that the use of the word "controlled" should be interpreted *very loosely*, and second, although hopefully quite readily apparent, it feels obligatory to call to your attention – “controlled” by a *single human* (Anduril, 2025). And herein is where the issue lies: When an algorithm misidentifies a target and innocent lives are lost, who bears responsibility? This "accountability gap" highlights an unresolved ethical problem – LAWS creates situations where praise and blame cannot be meaningfully assigned to any moral agent. The engineer who designed the system? The operator who merely

supervised its autonomous operation? The company that manufactured it but relinquished control to algorithms? Or political leadership that deployed such systems?

As Robert Sparrow examines in “Killer Robots”, this "responsibility gap" is not merely theoretical but represents a structural failure in our ethical frameworks for warfare. He argues that, when lethal autonomous weapons cause war crimes, it is often impossible to identify someone – man or machine – who can appropriately be held morally responsible (Sparrow 2007). The distributed nature of these decisions across human and machine nodes creates a "systematic irresponsibility" – a situation where accountability becomes so fragmented that no single person or group bears the moral burden of autonomous decisions. This lack of burden completely removes all ethos from warfare and battle. And, something even more disturbing is Asaro's (2008) findings that this ambiguity can even serve as a strategic method: "The appeal of autonomous weapons may lie precisely in their ability to obscure responsibility, allowing military and political actors to achieve objectives while maintaining plausible deniability." So this raises the question: are the weapons being purposely engineered to systematically resist accountability?

This diffusion of responsibility directly connects to a lack of meaningful control, and this is a major flaw of AI-driven warfare, as Elke Schwarz argues in "The (im)possibility of meaningful human control for lethal autonomous weapon systems." One of the three key limitations of meaningful control Schwarz looks at is cognitive limitations (Schwarz, 2018). She goes on to explain that when making decisions, people rely on two types of thinking: deliberative reasoning (slow and thoughtful), and automatic reasoning (fast and instinctive). Because automatic reasoning requires less effort, it's what we use most of the time, especially when interacting with

machines (this is called automation bias, when we put an unreasonable amount of trust in computers questioning its response).

In the context of warfare, this becomes a serious issue; if a human is overseeing an autonomous weapon, they are likely to default to automatic reasoning, meaning they would not question the machine's decision to kill or not kill. And this brings us to Schwarz's second limitation:

temporality. Autonomous weapons react in nanoseconds, much quicker than a human reaction, so the operator might not even have time to intervene. "The U.S. Department of Defense defines LAWS as systems that, 'once activated, can select and engage targets without further intervention by a human operator' (US Department of Defense, 2023)." And therefore, commensurating with the definition by the ICRC of meaningful control as "the user [having] the ability to intervene in a timely manner and that the use is linked to a certain level of accountability", meaningful human control is indeed a *false reality*. Taking these limitations and applying them to Bolt, we can see that human commanders might hold control over demands, but do not truly hold authority in a meaningful, or even sometimes literal, sense.

Impact 1 - *The Loss of Meaningful Control and Its Effects*

As Bradley Jay Strawser and Jeff McMahan explore in "Killing by Remote Control: The Ethics of Unmanned Military," autonomous weapons challenge military institutions built around human moral agency in ways that far exceed the ethical questions raised by human operated drones. For example, The U.S. Marine Corps' core values of "Honor, Courage, and Commitment" stand in stark contrast not just to drone warfare, but to the complete algorithmic decision making of

systems like those developed by Anduril, where human operators are reduced to simply supervisors rather than decision makers.

This constitutes a profound re-engineering of the warrior's relationship with killing. Shannon E. French (2004) argues in "Exploring Warrior Values Past and Present" that warrior codes across cultures distinguish warriors from murderers *by establishing ethical boundaries* that preserve moral identity while authorizing necessary violence that war calls for. While drone operators still make kill decisions and bear moral responsibility, LAWS completely disrupt this ethical equilibrium by severing the connection between the warrior's decision and lethal action.

Dr. Peter Lee discusses how autonomous systems create a "distance paradox" that exceeds even the detachment experienced with remote drone operations (Lee, 2018). Unlike drone pilots who at least engage directly in targeting decisions, LAWS users merely supervise algorithmic processes, transforming warfare into something resembling video games. However, Derek Gregory (2011) offers a counterpoint, arguing that these technologies create "a special kind of intimacy that consistently privileges the view of the hunter-killer, and whose implications are far more deadly." Katherine Hall Kindervater (2017) builds on this analysis, noting that "Through the screen, only inches away, the drone operator can see more than ever before, including the aftermath of a blast." In Anduril's autonomous lethal systems, this intimate surveillance remains, but, again, the critical human decision making element is no longer there, changing the idea of what it means to be "in control" during warfare.

The consequences extend beyond individual psychology to military institutions worldwide. While drone operators still bear direct responsibility for their actions, with autonomous weapons, the soldier's traditional moral burden doesn't vanish; it morphs into something more diffuse: a

moral blindness facilitated by technological abstraction. Anduril's vision of "a single human controlling a wide range of autonomous assets" across multiple domains exemplifies what Gregory suggests might become an immoral curiosity enabled by the technological distance of LAWS – a curiosity untempered by the weight of direct moral responsibility that even drone operators must carry.

Impact 2 - *The Lack of Accountability and Its Impacts*

Who bears responsibility when an algorithm selects the wrong target? This question has changed from theoretical debate to become an urgent real-world problem as autonomous systems increasingly determine who lives and dies through algorithmic "Kill Lists."

To understand meaningful human control, we must return to its foundational requirements: systems that are predictable, reliable, transparent, and understandable, *where accountability can be clearly established* (Roff and Moyes, 2016). Recent deployments of autonomous targeting systems demonstrate how far we have strayed from these principles.

Israel's deployment of AI-driven targeting systems in Gaza – "Gospel" and "Lavender" – is a disturbing case study. While not fully autonomous weapons in the strictest sense like Anduril's aspirational systems, these targeting algorithms demonstrate the dangers of delegating targeting decisions to algorithms with no human oversight. According to Sigal Samuel (2024), "Gospel" identifies buildings it determines Hamas is using, while "Lavender" analyzes surveillance data to rate Gaza residents on the likelihood they are militants, generating prioritized "Kill Lists." These systems have marked over 37,000 Palestinians for targeting, with an Israeli intelligence officer

acknowledging that targets are often struck in family homes because "it is easier to bomb them there than a protected military building" (Vox, 2024).

This directly contradicts Arkin's (2010) argument that AI would reduce civilian casualties. The algorithmic nature of these targeting systems creates what we might call "blameless casualties" – deaths for which no one accepts clear responsibility. Can we hold Lavender accountable? It's just code. The coder who created it? They were implementing directions from their commanders. The operator who approved the strike? They relied on algorithmic recommendations ("Kill Lists") they were trained to trust.

The opacity of these systems creates a gap between operators and the intelligence they act upon. Military personnel obey ratings these systems provide without question – a perfect example of automation bias in lethal contexts. "Within conditions of such complexity and abstraction, humans are left with little choice but to trust in the cognitive and rational superiority of this clinical authority. This relationship is often under the term 'automation bias' and is a well-documented phenomenon in the literature on human-machine interactions, whereby technological authority serves to smooth over moral tensions" (Renic and Schwarz, 2023, p.16).

Conclusion

The debut of autonomous weapons systems – which differ from human-operated drones – represents an inflection point in warfare that demands urgent regulatory attention. As demonstrated through Anduril's systems and even in the precursor to LAWS in Israel's deployment of "The Gospel," we are witnessing the rapid erosion of meaningful human control

in conflict zones, with devastating consequences for civilian populations and accountability frameworks.

With the issues identified – diffused responsibility, automation bias, and the "distance paradox" – create a perfect storm where decision-making authority is nominally human but functionally algorithmic.

The implications extend beyond individual conflicts; they challenge the very foundations of humanitarian law, which presupposes human moral agency and clear chains of accountability. As LAWS proliferate, they bring about a need for law to serve strategic interests while endangering civilian populations.

Moving forward, there needs to be an international regulatory framework for the usage of lethal autonomous weapons systems in warfare. Without such measures, we risk normalizing a form of warfare where algorithms determine who lives and dies, responsibility becomes unassignable, and the moral weight of killing is lost. The technology driving this transformation cannot be uninvented, but its implementation can — and must — be governed by principles that preserve human morals and accountability in conflict.

Policy Recommendation - *The SMART Framework*

This international regulatory framework I am recommending will be called "The SMART Framework: Supervised Military Autonomous Robotics Treaty." SMART will consist of four key principles. The first is shared accountability. Due to several different entities being involved in the creation of LAWS, there should be shared and defined responsibility for who takes

accountability. For this, we will divide responsibility by percentages in the case of a prosecution: company 40%, commanders 30%, operators 20%, engineers 10%. Each of these entities should acknowledge their involvement in a potentially rouge and dangerous machine in case of the event of prosecution. The second principle is mandatory training and certification processes for any individual involved in the creation, implementation, and operation process of LAWS. This will require vigorous training on how to use the systems to the safest of their abilities; there will be a one - three year program depending on the personnel, followed by a technical understanding assessment and ethical simulation tests with annual recertification requirements. The third principle is the requirement of transparency between nations via an international secure registry with an independent oversight committee of multinational representatives. And finally, the mandate to incorporate meaningful control to all LAWS by adding override capabilities and geographical limitations on where the weapons can be used.

The SMART Framework would be administered by a new UN-affiliated body with representatives from military powers, technology experts, ethicists, and humanitarian organizations. Of course, there are several aspects that would need to be solidified and discussed for years to come, but the process must begin somewhere; this approach understands the inevitability of autonomous weapons and does not aim for a complete ban in its use, but aims to create strict guidelines, clear accountability structures, and the assurance of qualified professionals remaining ethically involved. Even if the next realistic step is simply mandating meaningful human control, any progress toward regulation is a step toward ethical responsibility.

Reference Page

Anduril (2023a) Command and Control. Available at:

<https://www.anduril.com/command-and-control/> (Accessed: 2 April 2025).

Anduril (2023b) Profile: Palmer Luckey. Available at:

<https://www.anduril.com/profile/palmer-luckey> (Accessed: 2 April 2025).

Anduril (2025) Command and Control. Available at:

<https://www.anduril.com/command-and-control/> (Accessed: 2 April 2025).

Arkin, R. (2010) 'The Case for Ethical Autonomy in Unmanned Systems', *Journal of Military Ethics*.

Asaro, P. (2008) 'How Just Could a Robot War Be?', in Briggie, A., Waelbers, K. and Brey, P. (eds.) *Current Issues in Computing And Philosophy*.

Britannica (2023) Gunpowder. Available at: <https://www.britannica.com/technology/gunpowder> (Accessed: 2 April 2025).

Chamayou, G. (2015) *A Theory of the Drone*. New York: The New Press.

French, S. (2004) *The Code of the Warrior: Exploring Warrior Values Past and Present*. Rowman & Littlefield Publishers.

Coekelbergh, M. (2011) 'Moral Responsibility, Technology, and Experiences of the Tragic: From Kierkegaard to Offshore Engineering', *Science and Engineering Ethics*, 18(1), pp. 35-48.

Gregory, D. (2011) 'From a View to a Kill: Drones and Late Modern War', *Theory, Culture & Society*, 28(7-8)

Kindervater, K.H. (2017) 'The Technological Rationality of the Drone Strike', *Critical Studies on Security*, 5(1), pp. 28-44.

Lee, P. (2018). *Reaper Force: Inside Britain's Drone Wars*. [online] Available at: <https://raf.mod.uk/what-we-do/centre-for-air-and-space-power-studies/aspr/apr-vol21-iss3-5-pdf/> (Accessed 2 Apr. 2025).

Netflix (2017) *Black Mirror*. Season 4, Episode 5: Metalhead [TV series]. Netflix.

New York Times (2025) 'U.S. Ukraine Military War Takeaways', 30 March. Available at: <https://www.nytimes.com/2025/03/30/world/europe/us-ukraine-military-war-takeaways.html> (Accessed: 2 April 2025).

Patrick Lin; Keith Abney; George A. Bekey, "Moral Machines: Contradiction in Terms or Abdication of Human Responsibility?," in *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, 2012, pp.55-68.

Renic, N. and Schwarz, E. (2023) 'Crimes of Dispassion: Autonomous Weapons and the Moral Challenge of Systematic Killing', Queen Mary University of London Ethics & International Affairs.

Roff, H.M. and Moyes, R. (2016) 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons', Briefing Paper for Delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva.

Samuel, S. (2024) 'How AI is powering Israel's war in Gaza', Vox, 15 January. Available at: <https://www.vox.com/future-perfect/2024/1/15/24036604/israel-ai-gaza-hamas-war-lavender-gospel> (Accessed: 2 April 2025).

Schwarz, E. (2018) 'The (im)possibility of meaningful human control for lethal autonomous weapon systems', *Security Dialogue*, 54(3), pp. 216-234.

Schwarz, E. (2019). *Death Machines: The Ethics of Violent Technologies*. E-Book. Queen Mary Research Online. Available at: <<https://www.qmul.ac.uk>> (Accessed 2 Apr. 2025).

Sharkey, N. (2012) 'The evitability of autonomous robot warfare', *International Review of the Red Cross*, 94(886), pp. 787-799.

Strawser, B.J. and McMahan, J. (eds.) (2013) *Killing by Remote Control: The Ethics of an Unmanned Military*. Oxford: Oxford University Press.

US Department of Defense (2023) Directive 3000.09: Autonomy in Weapon Systems. Washington, DC: Department of Defense.

Vox (2024) 'How AI is powering Israel's war in Gaza', 15 January. Available at: <https://www.vox.com/future-perfect/2024/1/15/24036604/israel-ai-gaza-hamas-war-lavender-gospel> (Accessed: 2 April 2025).