Regulating Lethal Autonomous Systems

Introduction

In today's rapidly changing landscape of warfare, autonomous and artificial intelligence driven military systems, ranging from predictive targeting algorithms to drones that strike without human piloting, are evolving faster than legal rules can keep pace. Steven Feldstein warns that "the speed at which new, AI-driven systems are being integrated into military arsenals could outstrip any attempts at regulation" (Feldstein). His warning underlines the widening gap between rapid deployment and deliberate lawmaking. A second implication is that even well written statutes could become obsolete before they enter force. Recent conflicts in Ukraine and Gaza confirm the point by showcasing software updates that arrive on the battlefield faster than treaty negotiations progress.

The nuclear arms race offers an additional cautionary tale. Lili Xia and her colleagues calculate that "global food insecurity and famine from reduced crop, marine fishery and livestock production" would follow any large-scale conflict that disturbs the climate (Xia et al. 589). Their scenario illustrates in concrete terms how technological escalation threatens civilian life far beyond the battlefield. It also reminds policymakers that once a destructive capability spreads, its indirect effects may be both global and irreversible. These historical and contemporary signals make the search for effective regulation urgent, which sets the stage for a closer look at accountability, state behavior, and commercial incentives.

Entrusting machines with lethal decision making complicates accountability because errors might be blamed on software engineers, defense firms, or field commanders, leaving victims with no clear path to justice. Braden Allenby observes that "new technologies can undermine the laws of war" when states "are reluctant to trust international agreements that

might slow their own progress" (Allenby 24, 25). Robin Douglass adds a complementary perspective, noting that "the anxiety of being overpowered, central to Hobbes's account, remains at the heart of modern realism" (Douglass 264). His remark also underscores that weapons are never purely defensive, meaning any tool that promises protection simultaneously offers the power to strike first, which heightens suspicion among rivals. Together these insights highlight a vicious circle in which fear drives acceleration and acceleration intensifies fear. Recognizing that cycle is essential before exploring the theoretical frameworks that can interrupt it in practice.

How might legal frameworks adapt to emerging military technologies, and what evolving factors and innovations will influence this adaptation? Three theoretical lenses guide the answer. Political realism explains why states rarely slow research when rivals advance. Institutional and regulatory theory shows how bureaucracies struggle to keep up with rapid innovation. Critical political economy reveals how defense profit motives often overshadow moral or legal restraints. Each lens exposes a different barrier, and their combination reveals where meaningful regulation might emerge, where it stalls, and why. This framing will inform the structure of the argument that follows.

My thesis is that an effective legal regime for lethal autonomous systems must unite clear lines of accountability, a standing multinational inspectorate with enforcement power, and transparency rules that mandate explainability and traceability. Although realist fears, bureaucratic inertia, and powerful corporate interests all resist such a regime, layering these elements together provides the best chance to slow escalation and protect civilian lives. We will now turn to how realism shapes state behavior in this domain.

Realism and the Security Dilemma

Political realism holds that states pursue power and security above all else. Douglass explains, “far from being an archaic or purely historical curiosity, Hobbes’s model illuminates the perilous quest for security that still characterizes much of modern international relations” (Douglass 254). His claim underscores that governments perceive emerging weapons through the lens of survival rather than morality. The lesson is that no treaty will succeed unless it satisfies states that rivals face equally binding limits. That requirement raises the threshold for effective enforcement authority. This realist insight points directly toward the need for mechanisms that make cheating difficult and costly.

Douglass deepens the point when he writes, “Hobbes’s call for an absolute sovereign is not a plea for unchecked power, but a pragmatic acknowledgment that law, absent a formidable enforcer, becomes mere rhetoric when faced with competing interests” (260). This quote stresses that rules alone do not create order; a capable guardian must compel obedience. Translating that idea to autonomous weapons suggests that inspection teams must possess genuine access and sanctioning capacity. Until such an authority exists, states will continue to hedge their bets, which leads naturally to the next problem of institutional delay

Institutional and Regulatory Challenges

Large organizations rarely match the pace of technological change. Carl von Clausewitz observed that “the difficulties accumulate and end by producing a kind of friction that is inconceivable unless one has experienced war” (Clausewitz 119). His description of battlefield confusion doubles as an apt metaphor for legislatures deliberating over code they scarcely understand. This friction Clausewitz describes explains why detailed regulations often trail innovations by years, leaving an enforcement vacuum. Recognizing this lag is the first step toward devising adaptive oversight structures that can update quickly.

Andrew Collier confirms the mismatch when he writes, “the speed of innovation often far outstrips the slow, piecemeal approaches of national and international legal bodies” (Collier 45). His analysis shows that overlapping committees, security reviews, and funding cycles slow the rulemaking process. The consequence is that standards published today may govern yesterday’s algorithms but not tomorrow’s. Bridging that gap requires flexible technical annexes and rapid amendment procedures, which links the institutional problem to the economic incentives discussed next.

Critical Political Economy of the Military Industrial Complex

Critical political economy highlights how profit influences public decisions. Herbert Marcuse and Douglas Kellner argue that “large firms, profiting from war, shape public policy to reinforce those profits” (Marcuse and Kellner 37). The passage identifies lobbying and campaign donations as levers through which corporations steer legislative priorities. One result is that draft bills can be softened or delayed whenever they threaten lucrative contracts. This dynamic demonstrates why purely voluntary guidelines often prove ineffective.

Andrew Collier supplies a contemporary example by noting that some technology giants “pivot from commercial to military applications with minimal transparency” (Collier 72). The observation illustrates how dual purpose research blurs civilian and defense boundaries, complicating external review. Firms that collect vast consumer data can repackage analytic tools for targeting or surveillance without disclosing details. These incentives merge with realist fears and institutional friction to create a self reinforcing cycle of rapid innovation and weak oversight. The ongoing war in Ukraine, where local start ups and foreign contractors speed untested

systems to the front lines, vividly shows how this cycle moves from boardroom to battlefield, making it an ideal case to examine next.

Ukraine vs. Russia, the Polarizing Adaptation into Modern Warfare

The ongoing war in Ukraine demonstrates how advanced yet comparatively affordable systems can shift the balance of power. Zdzisław Śliwa notes, "the Ukrainian forces have shown that the effective synergy between modern technology and well-trained personnel can offset the numerical advantage of the Russian military" (Śliwa 3). Śliwa's observation highlights that sophisticated optics, small loitering munitions, and AI assisted reconnaissance can neutralize what once appeared to be an overwhelming troop or armor advantage. The quotation also implies that numerical strength loses decisiveness once precision and automation dominate tactical decisions. Because many of these tools arrive through foreign aid and private contracts, Ukraine's experience illustrates how global supply networks accelerate battlefield innovation even in the middle of an invasion.

Robin Douglass explains that "the anxiety of being overpowered" lies at the heart of the modern security dilemma (Douglass 264). His point helps clarify why both Kyiv and Moscow race to integrate autonomous capabilities. Ukraine relies on commercial quadcopters equipped with machine vision software to spot artillery targets, while Russia fields Lancet and KUB loitering munitions guided by neural network image classifiers. Each side also employs algorithmic systems to sift satellite and radio frequency data for troop movements, shortening the sensor to shooter cycle from hours to minutes. The result is a reciprocal escalatory spiral, where every new system adopted by one combatant compels a countervailing upgrade by the other. Defense contractors and start ups seize the opportunity to test prototypes under live conditions, compressing development timelines from years to months. This commercial military feedback

loop turns the front line into a proving ground and leaves legislators far behind, vividly confirming the institutional lag that Clausewitz describes. Because similar dynamics once propelled the spread of nuclear technology, the Ukrainian case offers a natural bridge to the historical lessons of earlier arms control efforts, which the next section explores.

Historical Analogy: Nuclear Treaties

The Non-Proliferation Treaty slowed the spread of nuclear weapons but never eliminated existential risk. Steven Feldstein warns, "AI's potential for devastation mirrors the concerns raised by early nuclear technology, yet states remain hesitant to relinquish a strategic edge" (Feldstein). His comparison reminds us that the global community required two decades, multiple crises, and extensive verification protocols before agreeing on modest limits to nuclear arsenals. Even after agreement, verification has remained a constant struggle, forcing continuing negotiations to update definitions and inspection practices.

Braden Allenby strengthens the parallel, cautioning that "the laws of war themselves can be eroded if states believe new technologies will offer a conclusive advantage" (Allenby 24). His remark underlines that weapons viewed as decisive invite preemptive use, heightening collective insecurity rather than reducing it. When read with Feldstein, the lesson is clear: a single treaty text is not enough; adaptable enforcement is required. Modern autonomous systems evolve much faster than fissile material production methods did in the 1960s, so the lag between agreement and obsolescence is far shorter. Recognizing that limitation invites a comparison with China's centralized approach to research oversight, where rapid innovation meets deliberate secrecy, as the following case study shows.

China, State Control, and Rapid AI Development

Andrew Collier reports that Beijing can “clamp down on major internet firms and encourage them to innovate, all to ensure that national strategic objectives remain paramount” (Collier 90). The quotation shows how the state uses both punishment and reward to keep companies aligned with defense goals while accelerating research. Because patent approvals, capital, and cloud resources flow quickly to favored firms, bureaucratic delay is minimal, giving China a pace advantage over pluralistic systems that require lengthy public consultations. Robin Douglass observes that “fear and distrust” sit at the center of modern realism (Douglass 254). His insight applies in two directions: Chinese planners fear falling behind foreign militaries, and foreign capitals fear that opaque Chinese advances will unsettle regional balances. This mutual suspicion drives parallel development races, illustrating how realism magnifies the strategic consequences of Collier’s observation about state guided innovation.

Institutionally, the one party system can redirect vast resources to military AI with little debate, a structural advantage when speed is the primary objective. Yet that same advantage has costs. Export control rules and state secrets laws bar outside auditors from reviewing code, limiting accountability and making it difficult to verify ethical safeguards. Collier’s description of firms pivoting from social media to weapons illustrates how state authority merges with private enterprise, producing rapid breakthroughs by granting companies access to enormous government data sets, high performance computing clusters, and joint research labs (Collier 90). Leadership rotations that move executives into party committees and back into corporate headquarters bind profit motives to national priorities rather than place them in conflict. The arrangement accelerates progress but deepens opacity, prompting foreign adversaries to assume worst case scenarios and match Chinese capabilities in kind. This escalation underscores why civilian algorithms, once repurposed without scrutiny, can spread bias and error at lethal scale.

Civilian Algorithms in Warfare

Bias uncovered in civilian software can become deadly when the same code guides weapons. Gwen van Eijk observes, “socioeconomic marginality influences how offenders are perceived and can thus systematically disadvantage the poor” (van Eijk 465). Her finding demonstrates that data reflecting historic inequality reproduces that inequality when deployed in predictive models. Transferred to a battlefield, the same skew could cause an AI sentinel to misidentify civilians from lower income regions as higher risk targets, amplifying harm rather than reducing it.

Because dual use code changes roles faster than legislators can update rules, regulators must require bias audits before any civilian model migrates into defense applications. They must also mandate that developers preserve complete data lineage to allow later investigations. Yet these safeguards hinge on technical transparency, which is difficult when firms claim proprietary rights. That difficulty leads directly to the technical requirements of explainability and traceability, topics that demonstrate why code must be designed for scrutiny from the outset.

Ethical Oversight, Explainability, and Traceability

Edward Hunter and his colleagues warn that “The inherent opacity of autonomous weapon systems significantly complicates establishing legal accountability, as these systems often operate as ‘black boxes’” (Hunter et al. 233). This quote highlights two failures; first, victims cannot receive justice when causal links remain hidden. Second, commanders cannot refine tactics when they do not know why the machine succeeded or failed. These blind spots stall both moral evaluation and strategic learning, leaving battlefield errors uncorrected and legal claims unresolved. Because opacity stands at the root of these harms, any serious regulatory project must start by making decision paths visible.

Hunter's team therefore argues that "Effective regulation must incorporate clear standards for explainability to ensure decisions taken by autonomous systems can be ethically and legally justified" (236). Explainability requires system architecture that renders each computational step interpretable to human reviewers, allowing investigators to reconstruct the chain of reasoning. Traceability extends this demand, because "detailed records of decision-making sequences" are "crucial for verifying compliance and conducting investigations post incident" (238). Together, explainability and traceability transform a sealed algorithm into an auditable record that courts, commanders, and ethicists can examine. If policymakers treat these standards as optional, they leave both soldiers and civilians vulnerable to errors that no one can later explain. Recognizing that systems must be transparent still leaves open the question of who should direct them, which leads naturally to the requirement for human oversight.

Finally, the authors insist, "Robust human oversight mechanisms are critical to bridge accountability gaps created by the delegation of decision making to autonomous weapons" (241). Human operators who can intervene slow engagements just enough to prevent runaway escalation and give political leaders time to weigh consequences. Oversight also reassures adversaries that machines will not act unpredictably, reducing incentives for preemptive strikes. Incorporating these safeguards answers many technical objections but does not remove policy debates about pace and scope of reform, so the next section evaluates incremental proposals and responds to their limitations.

Possible Objections

Critics will argue that incremental arms control is more practical than building entirely new oversight structures. Michael Horowitz and Paul Scharre state, "Incremental arms control measures have historically succeeded because they build on established norms and enforcement

mechanisms rather than introducing entirely new standards" (Horowitz and Scharre). Their point is persuasive in showing that small steps often encounter less diplomatic resistance and gain faster ratification. At the same time, incrementalism does not need to exclude comprehensive reforms; both approaches can operate in tandem if technical requirements are phased in through successive treaty protocols. Edward Hunter and his colleagues warn that opacity "significantly complicates establishing legal accountability" (Hunter et al. 233), so explainability and traceability must be included early even if other provisions are staged. Without these core safeguards, a series of small agreements risks cementing a system that remains fundamentally unaccountable.

Opponents of a robust inspectorate raise a second concern. Braden Allenby notes that "new technologies can undermine the laws of war" by outpacing legal responses, which leads states to doubt inspections (Allenby 24). Horowitz and Scharre add that "The ambiguity of dual-use capabilities complicates verification and compliance, as states may exploit civilian or commercial technological advances for military purposes" (Horowitz and Scharre). These quotations identify two related fears: first, inspectors might lag behind technical change; second, inspections might expose proprietary secrets. Conditional verification can answer both objections. Encrypted traceability records can remain sealed until a treaty body authorizes review, limiting espionage risks, while regular technical annex updates keep inspectors current with emerging designs. This pairing of flexible rules with secure data channels shows that strong verification can coexist with national security interests.

A final objection targets transparency itself. Horowitz and Scharre caution that "Effective enforcement relies heavily on transparency and mutual trust, yet autonomous weapons inherently reduce transparency due to their complexity and rapid operational tempo" (Horowitz and

Scharre). The quotation underlines a genuine tension: revealing algorithmic details might expose vulnerabilities. The response is to calibrate, not abandon, disclosure. Audit logs can employ zero knowledge proofs, allowing inspectors to confirm compliance without viewing sensitive source code. Similar cryptographic techniques already protect financial transactions and nuclear material accountancy, demonstrating feasibility. Addressing each objection as it arises clarifies that layered reforms can incorporate both incremental steps and comprehensive safeguards, a synthesis the conclusion now develops into a concrete policy blueprint.

Conclusion

My research sought to understand how legal frameworks might adapt to emerging military technologies, particularly lethal autonomous weapons systems, and identify key factors influencing this adaptation. The analysis has supported the hypothesis that an effective legal regime must include clearly defined accountability, liability, robust international oversight mechanisms, inspectorate, and stringent transparency measures.

The evidence underscores that liability must explicitly assign responsibility for errors or misuse of lethal autonomous systems. Clear accountability deters irresponsible deployment by creating legal and financial risks for states and defense contractors alike. This is crucial given the institutional tendency towards slow regulatory adaptation identified by Collier, who states, “the speed of innovation often far outstrips the slow, piecemeal approaches of national and international legal bodies” (Collier 45).

Similarly, establishing an international inspectorate, modeled after successful arms control treaties such as the Chemical Weapons Convention, is essential. An independent team authorized by treaty agreements would possess enforcement power to inspect AI weapon laboratories, verify compliance, and temporarily halt programs posing serious risks. This

enforcement is necessary to bridge the gap between rapid technological advancement and slower regulatory responses, aligning with Clausewitz's concept of friction between planning and execution (Clausewitz 119).

Transparency measures, including the publication of audit logs, enable public scrutiny by journalists and non-governmental organizations. Though transparency poses risks, its benefits outweigh the drawbacks by fostering global trust and compliance, counteracting realist fears of being strategically disadvantaged. Douglass highlights how such fears shape international relations, emphasizing "the anxiety of being overpowered" as central to nations' hesitancy in arms control agreements (Douglass 264).

Ultimately, these combined measures significantly narrow the regulatory gap, reshaping incentives to promote responsible AI development and deployment. Although incremental adjustments, as advocated by Horowitz and Scharre, may address immediate concerns, comprehensive oversight is necessary for lasting effectiveness. Looking forward, continuous evaluation and adaptation of treaties will be crucial as technological innovation evolves.

In conclusion, implementing strong accountability, oversight, and transparency measures provides meaningful guardrails against uncontrolled development of lethal autonomous weapons. This framework not only addresses immediate ethical and security concerns but also shifts power dynamics, reducing undue influence by defense contractors and reinforcing democratic oversight. Monitoring and modeling these treaty dynamics against rapid technological advances will be critical for future stability, not only in warfare but across all sectors affected by emerging technologies.

Works Cited

Allenby, Braden R. ‘Are New Technologies Undermining the Laws of War?’ The Bulletin of the Atomic Scientists, vol. 70, no. 1, Informa UK Limited, Jan. 2014, pp. 21–31, https://doi.org/10.1177/0096340213516741.

Clausewitz, Carl Von, et al. On War. Princeton University Press, 1989.

Collier, Andrew. China’s Technology War: Why Beijing Took down Its Tech Giants. 1st ed., Palgrave Macmillan, 2022, https://doi.org/10.1007/978-981-19-3042-3. Accessed 15 Feb. 2025.

Douglass, Robin. “Hobbes and Political Realism.” European Journal of Political Theory, vol. 19, no. 2, 2020, pp. 250–269. SAGE Journals, https://doi.org/10.1177/1474885116677481.

Feldstein, Steven. “Ai in War: Can Advanced Military Technologies Be Tamed before It’s Too Late?” Bulletin of the Atomic Scientists, 11 Jan. 2024, https://thebulletin.org/2024/01/ai-in-war-can-advanced-military-technologies-be-tamed-before-its-too-late/. Accessed 5 Feb. 2025.

Hunter, Edward, et al. “Regulating Lethal Autonomous Weapon Systems: Exploring the Challenges of Explainability and Traceability.” AI and Ethics, vol. 4, no. 2, 2023, pp. 229–245.

Horowitz, Michael C., and Paul Scharre. “Applying Arms-Control Frameworks to Autonomous Weapons.” Brookings Institution, 3 Dec. 2021, www.brookings.edu/articles/applying-arms-control-frameworks-to-autonomous-weapons.

Marcuse, Herbert, and Douglas Kellner. Technology, War, and Fascism. Routledge, 1998. Accessed 14 Feb. 2025.

van Eijk, Gwen. "Socioeconomic Marginality in Sentencing: The Built-in Bias in Risk

Assessment Tools and the Reproduction of Social Inequality." Punishment and Society, vol. 19, no. 4, October 2017, pp. 463-481. HeinOnline.

Xia, Lili, et al. “Global Food Insecurity and Famine from Reduced Crop, Marine Fishery and Livestock Production Due to Climate Disruption from Nuclear War Soot Injection.” Nature News, Nature Publishing Group, 15 Aug. 2022, www.nature.com/articles/s43016-022-00573-0. Accessed 5 Feb. 2025.

Zdzisław Śliwa. “THE SYNERGY BETWEEN TECHNOLOGY AND SOLDIERS IN WARFARE – THE RUSSIAN ARMED FORCES IMAGE DURING THE WAR IN UKRAINE.” Wiedza Obronna, vol. 281, no. 4, 2022, https://doi.org/10.34752/2022-d281.