

How does the delegation of lethal decision-making to autonomous weapons systems reconfigure accountability structures and alter the strategic stability of international conflict?

October 2, 2026 | SnugLab Research

Executive Summary

The delegation of lethal decision-making to autonomous weapons systems (AWS) reconfigures accountability structures by creating an "accountability gap" where existing legal frameworks struggle to attribute responsibility for harms, primarily due to the opacity of algorithmic decision-making [1, 2, 3, 4]. While existing military justice frameworks are adapting, notably through the application of command responsibility as demonstrated by International Criminal Court (ICC) arrest warrants for Russian high command [2, 3, 11], the inherent unpredictability of these systems still diffuses responsibility across states, military hierarchies, and private contractors [4, 16]. Concurrently, AWS alter strategic stability by lowering domestic political casualty thresholds, which increases the frequency and intensity of conflicts, even if collateral damage is reduced [10, 11, 12]. This is exacerbated by algorithmic brittleness and expanded cyberattack surfaces, which introduce miscalculation risks and compress diplomatic warning times, thereby undermining traditional crisis stability [1, 7, 11, 13].

Key Findings

Reconfiguration of Accountability Structures

The delegation of lethal decision-making to AWS fundamentally reconfigures accountability by creating an "accountability gap" where existing legal frameworks struggle to attribute responsibility for harms [1, 2, 3, 4]. This gap arises because there is no universally agreed-upon precise threshold for "meaningful human control" (MHC), though it generally requires humans to retain authority over critical moments of force application [2, 9, 11]. When machines select and attack targets independently, the opacity of algorithmic decision-making makes it highly problematic to prove intent (*mens rea*) or negligence among developers, commanders, or operators [1, 2, 5]. This diffusion of responsibility spreads across states, military hierarchies, and private technology

contractors [4, 16].

Despite these challenges, ultimate accountability does not transfer entirely to the machine. Command responsibility frameworks can still hold commanders liable if they knew or should have known a system under their control was likely to commit violations but failed to take reasonable preventive steps [2, 3, 11]. This adaptation is already being tested; in March-June 2024, the ICC issued arrest warrants against Russian high command for war crimes involving Shahed-136 drones autonomously destroying Ukrainian power plants, demonstrating that command responsibility remains a viable starting point for applying International Criminal Law (ICL) to AWS [3]. Scholars also propose a "distributed responsibility" model that disaggregates accountability across the entire defense ecosystem, including civilian leadership, technology contractors, and military commanders [1].

Divergent national doctrines on MHC further complicate liability attribution in multinational coalition operations. The United States formally rejects MHC as ambiguous, instead focusing on "appropriate levels of human judgment" and replacing "human operator" with "operator" in its 2023 DoD Directive 3000.09, potentially broadening the distance between human action and weapon deployment [21, 22]. In contrast, the European Union maintains a shared institutional stance requiring MHC, with Germany legally mandating it in both design and use phases, and France prohibiting fully autonomous lethal weapons [19]. Russia asserts that existing international law is sufficient and opposes new limitations [20]. These differences create legal friction and ambiguity within alliances, making it difficult to attribute *mens rea* for individual criminal liability and potentially undermining command responsibility principles in joint missions [5, 11, 12, 22].

Alteration of Strategic Stability

The delegation of lethal decision-making to AWS significantly alters strategic stability by lowering domestic political casualty thresholds, thereby increasing conflict frequency and intensity [10, 11, 12]. By removing operators from the immediate physical casualty chain and compressing decision timelines, states face fewer domestic political barriers to initiating military action [10]. For example, during the 2024 conflict in Gaza, Israeli Defense Force operators approved AI-generated targets in an average of 20 seconds with minimal scrutiny, which resulted in a 10% error rate and thousands of civilian deaths [2, 10]. This rapid approval cycle demonstrates how algorithmic delegation can accelerate targeting workflows while increasing the risk of collateral damage due to human overreliance or insufficient verification [2, 10]. This shift toward more frequent conflict

occurs even if autonomous systems successfully reduce civilian casualties through enhanced precision [10].

Furthermore, the operation of autonomous systems at machine-to-machine speeds introduces sufficient miscalculation risks to trigger inadvertent escalation, fundamentally compressing traditional diplomatic warning times [1, 11]. The chaotic nature of AI-driven technology in contested environments, combined with algorithmic brittleness, creates significant risks of miscalculation and inadvertent conflict [1, 2, 7, 8, 11, 12]. When these systems behave unexpectedly, it becomes ambiguous whether an action reflects a deliberate political choice, a tactical maneuver, or a technical failure, which undermines crisis signaling and increases the likelihood of rapid escalation [13].

The expanded cyberattack surfaces inherent in digitized military systems compound these destabilizing risks [7]. Vulnerabilities in intelligent digital devices mean that a single cyberattack could corrupt electronic decision aids or turn unmanned swarms against their users, causing simultaneous strategic losses that further confuse crisis dynamics [7].

Several incidents and exercises in the 2020s demonstrate this compression of diplomatic warning times and increased miscalculation risk:

- **Operation Epic Fury (US vs. Iran, commenced February 28, 2026):** This large-scale strike campaign utilized Project Maven, the Pentagon's AI targeting backbone, identifying and prioritizing over 1,000 targets in its first 24 hours. This compressed the intelligence-to-action cycle from days to minutes, diminishing opportunities for diplomatic intervention [18].
- **Shajareh Tayyebbeh School Strike (Minab, Iran, early 2026):** During Operation Epic Fury, this strike killed over 170 people, mostly children, due to stale human-curated data fed into Project Maven. The AI-accelerated tempo eliminated procedural friction that might have caught the mistake, triggering a global political crisis [14, 15, 18].
- **India-Pakistan Last Crisis (May 2025) and Routine Cross-Border Drone Incidents:** Rapid escalation between these nations demonstrated how quickly crises can intensify, with AI-enabled systems shortening the decision fuse. Routine cross-border drone incidents lower retaliation costs, while attribution ambiguity creates risks of deadly misunderstandings and automated air-defense escalations that outpace political leaders' ability to intervene [17].
- **Exercise Golden Eagle (Pakistan Air Force, February 2026):** This exercise focused on AI-enabled, net-centric operations, integrating cyber, space, and electromagnetic spectrum capabilities with manned-unmanned teaming to demonstrate

AI-driven ISR-to-strike loops in contested environments [17].

These cases illustrate how algorithmic acceleration shifts humans from active evaluators to procedural approvers, increasing the risk of unintended escalation and eroding international confidence in the reliability of autonomous targeting systems [6, 14, 15].

Implications

The reconfiguration of accountability structures and alteration of strategic stability due to AWS delegation carries significant implications for international relations and the future of warfare. The persistent "accountability gap" means that victims of unlawful AWS actions may face substantial barriers to justice, potentially eroding public trust in international law and military ethics [1, 2, 3]. The diffusion of responsibility across multiple actors, from states to private contractors, necessitates a re-evaluation of existing legal frameworks and potentially the development of new international norms or treaties to ensure clear attribution of liability [1, 4, 16]. The ICC's recent actions against Russian high command suggest that command responsibility will be a primary avenue for accountability, but the unique challenges of algorithmic opacity will continue to strain these frameworks [3].

For strategic stability, the lowered political threshold for conflict initiation implies a more volatile international environment where military action becomes a more readily available tool for states [10, 11, 12]. This could lead to an increase in the frequency and intensity of conflicts, even if individual engagements are more precise [10]. The compression of diplomatic warning times by machine-speed operations and the risks of algorithmic miscalculation mean that crises could escalate rapidly and inadvertently, leaving less room for human de-escalation efforts [1, 11, 13]. This shift from managed competition to algorithmic competition could fundamentally destabilize international security, requiring urgent diplomatic efforts to establish common standards and arms control measures for AWS [12].

Limitations and Caveats

The research indicates that direct operational data on the precise impact of AI-assisted targeting on casualty tolerance percentages for specific major powers (US, China, Russia) is limited [10]. While the general trend of lowered political thresholds is supported, specific quantitative shifts in national casualty tolerance since the Ukraine and Gaza conflicts are not detailed. Furthermore, while algorithmic brittleness and cyber

vulnerabilities are identified as critical destabilizing factors, the provided research does not document specific failures for named autonomous systems like the Turkish Kargu-2, Israeli Harpy, or US Switchblade due to these issues, nor their resulting escalation impacts. The precise threshold for "meaningful human control" also remains a subject of international debate, with no universally agreed-upon definition or binding domestic legislation from specific states identified in the provided findings [2, 9, 11, 12].

Sources

- [1] [peer-reviewed] Resolving responsibility gaps for lethal autonomous weapon systems - Authors: Patrick Taylor Smith - Journal: Frontiers in Big Data - <https://pmc.ncbi.nlm.nih.gov/articles/PMC9766649/>
- [2] [edu] Legal Accountability Ai Driven Autonomous Weapons - [lieber.westpoint.edu](https://lieber.westpoint.edu/legal-accountability-ai-driven-autonomous-weapons/) - <https://lieber.westpoint.edu/legal-accountability-ai-driven-autonomous-weapons/>
- [3] [edu] Lethal Autonomous Weapon Systems Laws Accountability Collate - [law.temple.edu](https://law.temple.edu/ilit/lethal-autonomous-weapon-systems-laws-accountability-collateral-damage-and-the-inadequacies-of-international-law/) - <https://law.temple.edu/ilit/lethal-autonomous-weapon-systems-laws-accountability-collateral-damage-and-the-inadequacies-of-international-law/>
- [4] [edu] Ai Defense Contracts And The Legal Limits Of Autonomous Weap - international-and-comparative-law-review.law.miami.edu - <https://international-and-comparative-law-review.law.miami.edu/ai-defense-contracts-and-the-legal-limits-of-autonomous-weapons/>
- [5] [edu] Because We Take Our Values War Analyzing Views Un Member Sta - [cjil.uchicago.edu](https://cjil.uchicago.edu/print-archive/because-we-take-our-values-war-analyzing-views-un-member-states-ai-driven-lethal) - <https://cjil.uchicago.edu/print-archive/because-we-take-our-values-war-analyzing-views-un-member-states-ai-driven-lethal>
- [6] [edu] Viewcontent.Cgi - [digital-commons.usnwc.edu](https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=3115&context=ils) - <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=3115&context=ils>
- [7] [edu] Limits To Automated Ai Controlled Military Systems Urgent Im - [dda.ndus.edu](https://dda.ndus.edu/ddreview/limits-to-automated-ai-controlled-military-systems-urgent-imperative-to-assure-human-control/) - <https://dda.ndus.edu/ddreview/limits-to-automated-ai-controlled-military-systems-urgent-imperative-to-assure-human-control/>
- [8] Autonomy And International Stability - [cnas.org](https://www.cnas.org/publications/reports/autonomy-and-international-stability) - <https://www.cnas.org/publications/reports/autonomy-and-international-stability>
- [9] [ngo] 4221 002 Autonomous Weapons Systems Full Report - [icrc.org](https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf) - <https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf>
- [10] [preprint] Html - [arxiv.org](https://arxiv.org/html/2405.01859v1) - AUTHORS UNAVAILABLE - <https://arxiv.org/html/2405.01859v1>
- [11] A 79 88 LAWS - [docs-library.unoda.org](https://docs-library.unoda.org/General_Assembly_First_Committee_Seventy-Ninth_session_(2024)/A-79-88-LAWS.pdf) - [https://docs-library.unoda.org/General_Assembly_First_Committee_Seventy-Ninth_session_\(2024\)/A-79-88-LAWS.pdf](https://docs-library.unoda.org/General_Assembly_First_Committee_Seventy-Ninth_session_(2024)/A-79-88-LAWS.pdf)
- [12] Kiseok Michael Kang 22Enemies Allies And Artificial Intellig - [georgetownsecuritystudiesreview.org](https://georgetownsecuritystudiesreview.org/wp-content/uploads/2023/10/Kiseok-Michael-Kang-22Enemies-Allies-and-Artificial-Intelligence22.pdf) - <https://georgetownsecuritystudiesreview.org/wp-content/uploads/2023/10/Kiseok-Michael-Kang-22Enemies-Allies-and-Artificial-Intelligence22.pdf>
- [13] Ai Autonomous Weapons And The Future Of Strategic Stability - [dlpforum.org](https://www.dlpforum.org/dlp-factsheets/ai-autonomous-weapons-and-the-future-of-strategic-stability-and-arms-control/) - <https://www.dlpforum.org/dlp-factsheets/ai-autonomous-weapons-and-the-future-of-strategic-stability-and-arms-control/>
- [14] The Risks - [autonomousweapons.org](https://autonomousweapons.org/the-risks/) - <https://autonomousweapons.org/the-risks/>
- [15] When Human In The Loop Isn T The Same As Human Control - [forum.effectivealtruism.org](https://forum.effectivealtruism.org/posts/wREkxujXryayX2Jov/when-human-in-the-loop-isn-t-the-same-as-human-control) - <https://forum.effectivealtruism.org/posts/wREkxujXryayX2Jov/when-human-in-the-loop-isn-t-the-same-as-human-control>
- [16] [peer-reviewed] The competence-control trade-off in military AI innovation: Autonomous weapons systems and shifting modes of state control over private experts - Authors: Andrea Johansen; Andreas Kruck - Journal: European Journal of International Security -

<https://www.cambridge.org/core/journals/european-journal-of-international-security/article/competencecontrol-tradeoff-in-military-ai-innovation-autonomous-weapons-systems-and-shifting-modes-of-state-control-over-private-experts/FE2E70CF6E618BD5B3CE35B4A51A744D>

[17] Ai Is Quietly Reshaping Nuclear Risk In South Asia - apln.network -

<https://www.apln.network/analysis/commentaries/ai-is-quietly-reshaping-nuclear-risk-in-south-asia>

[18] Algorithmic War How Mythos And Project Maven Are Rewriting T - faf.ae -

<https://www.faf.ae/home/2026/4/17/algorithmic-war-how-mythos-and-project-maven-are-rewriting-the-rules-of-combat-sovereignty-and-global-security>

[19] [gov] Crs Product - congress.gov - <https://www.congress.gov/crs-product/IF11294>

[20] [edu] Unity Principle Variation Practice European Approaches Meani - lieber.westpoint.edu -

<https://lieber.westpoint.edu/unity-principle-variation-practice-european-approaches-meaningful-human-control-laws/>

[21] [edu] GT GJIL200015 - law.georgetown.edu -

<https://www.law.georgetown.edu/international-law-journal/wp-content/uploads/sites/21/2020/03/GT-GJIL200015.pdf>

[22] [gov] Sm Netherlands - legal.un.org -

https://legal.un.org/ilc/sessions/75/pdfs/english/sm_netherlands.pdf