Autonomous Weapons Systems

Simon Rupp, Aryavrat Gupta, William Boudy

Executive Summary

The emergence of AI military technology in Autonomous Weapons Systems (AWS) paves way for significant ethical concerns. This paper examines the development and regulation of AWS, focusing particularly on the policies and capabilities of the world's two leading powers in this domain - the United States and China. Three primary concerns in this space are the potential for artificial conflict escalation, violations of the Laws of Armed Conflict (LOAC) and the loss of human accountability. We propose a technical and legal framework to address these concerns. The technical side of it is in regards to the dashboards that humans use to interact with semi-autonomous weapons systems (SAWS), and it features an example of what an interface that prioritizes human oversight and transparency looks like. The legal framework establishes an international regulatory body modeled after the Nuclear Non-Proliferation Treaty, amendments to the REAIM blueprint, and guidelines for SAWS that emphasize defensive use and clear accountability measures. While fully autonomous weapons pose significant risks, semi-autonomous systems can be effectively deployed in an ethical manner that aligns with the LOAC. The proposed solutions aim to ensure that AWS technology encourages rather than takes away human decision-making in warfare, balances the military advantages of autonomous systems with ethical considerations, and provides a path for responsible SAWS usage.

Literature Review & Thesis

Autonomous Weapons Systems (AWS) are not any specific kind of weapon but rather any system with the capability to both select and engage with targets without human interference [1]. These systems rely on the use of artificial intelligence to discern real threats and how to respond to those threats. The key reminder is that these systems do not need human supervision, thus allowing for responses that are extremely quick, but also potentially incorrect or puzzling to the humans it is designed to help. Nonetheless, the potential for AI in military applications is undeniable. There is no doubt the world recognizes this fact. Many countries and organizations have differing thoughts on how to develop and regulate AWS. To be clear, there is no unifying agreement or shared sentiment among all leaders regarding the use of AI in the military. As a result, we believe this topic to be of utmost importance and emphasize the need to adopt our proposed solutions.

As mentioned, this technology is no longer science fiction - it is here now. Countries are aggressively developing this technology, particularly the U.S. and China [2]. We'll address the current capabilities and policies surrounding AWS from the perspective of these two nations. These are the two world superpowers at the moment and as such, they are the furthest along in

their current AWS technology and in development of even more AWS. These two nations also happen to be adversaries. This serves as an interesting case study to get a sense of current AWS capabilities and policy and the tension that arises from it.

The U.S. has recently provided several documents surrounding AWS. These include the *Political Declaration on Responsible Military Use of AI and Autonomy* and DoD Directive 3000.09. The declaration (notice it's a declaration so it is non-binding and therefore not enforceable in any way) focuses on responsible use and development. Specifically, it mentions compliance with international humanitarian law and the importance of maintaining human accountability [3]. Important to note with this declaration is that there is not any mandate regarding the level of human control. In other words, the U.S. is concerned with responsible use and human accountability, but not with human autonomy. The DoD Directive goes more into the human autonomy space. It explains systems should "allow commanders and operators to exercise appropriate levels of human judgment over the use of force" [1]. Overall, the sentiment from these documents is that the U.S. is hoping to be ethical with their development and use of AWS, but it is clear they recognize the benefits of this technology and are willing to use it if they deem it to be militarily useful [4]. China takes a different, much stricter approach - at least publicly. They have called for the complete ban of fully autonomous lethal weapons systems. As seen, this is something the U.S. has clearly not done. Of course it is worth noting the People's Liberation Army (PLA) of China is known to not actually be following this stance [2]. This policy is interesting and scary because it means the two adversarial world superpowers not only do not have a unified document governing their development and use of AWS, but their individual policies are not even all that similar.

To give more specifics into the current capabilities of each nation, according to an article written in January of 2024, the Pentagon has more than 800 active military AI projects [2]. This largely falls under "Replictor", an initiative from the U.S. to make it easier for the military to uptake emerging technology. It is assumed this initiative is in response to China's aggressive pursuit of autonomous weapons with their Military-Civil Fusion strategy. This national strategy hopes to leverage civilian technology developments to improve the capabilities of the PLA [5]. As seen with these brief details, AWS use and development is occurring right now and at high speeds.

There are clear ethical issues associated with AWS. The ones we've deemed to be the main issues include the erosion of human accountability, the escalation of conflict, and the violations of Laws Of Armed Conflict (LOAC). The erosion of human accountability is a major point. In the military, decisions concerning life and death happen regularly. Traditionally, these decisions were made exclusively by humans. Humans were expected to exercise moral judgement and have some sort of ethical framework. AWS means we are now placing this extremely critical decision in the hands of an algorithm. That alone is an ethical concern. Moreover, it means we are taking the human out of the decision, and thus removing all levels of human autonomy. As a result of this lack of human autonomy, there exists this question of

responsibility. Who is accountable when an AWS makes a mistake? Is it the military officer who deployed the system or perhaps the developer who created the algorithm? Clearly, there are some unanswered questions of human accountability with AWS. The next ethical concern involves the escalation of conflict. As mentioned in the description of AWS, these systems can execute decisions faster than humans ever could. Typically, a major flaw in an AWS is a lack of transparency in the decision being made. Many people raise the concern that this is a recipe for rapid escalation of conflict. An example would be a misclassification of a commercial aircraft as hostile. With a lack of transparency and rapid decision making, this could lead to an aggressive response before humans have the chance to ensure aggression is the proper response. Finally, AWS risks violating the LOAC - specifically, distinction, proportionality, and military necessity. The potential to violate distinction is pretty clear. An algorithmic bias could easily lead to a misclassification of a civilian as a combatant, resulting in a violation of distinction. Proportionality is related largely to accessibility. AWSs requires a technologically-advanced nation, meaning only the wealthiest nations will have access to this technology. If they then engage in conflict with nations without this technology, this is potentially a violation of proportionality. Finally, military necessity is under threat of being violated by AWS based on the same example used in escalation of conflict. Misidentifying a friendly, commercial aircraft as hostile and then responding with aggression would violate this law of armed conflict. So, it is clear that while benefits of using AWS technology exist, there are glaring ethical concerns that need to be addressed.

As a result of the current context regarding AWS and the ethical concerns associated with this technology, our thesis is as follows. The development of autonomous weapon systems (AWS) must align with international ethical standards to ensure transparency, human accountability, and conflict prevention. Starting or escalating wars through autonomous technology is condemnable and undermines global stability. While fully autonomous weapons pose significant risks, the controlled use of semi-autonomous systems (SAWS) for defensive purposes can be acceptable if guided by robust legal frameworks and global collaboration.

Analysis & Design

Proposed Technical Framework

A Semi-Autonomous Weapons Systems (SAWS) dashboard is a user interface that serves as the control panel for semi autonomous weapons. The purpose of the dashboard is to fuse a variety of data streams together, like satellite imagery, infrared sensors, etc. and use AI for object recognition and to come up with strategic military responses. It would allow for a quicker response to threats, as the AI would nearly instantaneously analyze and display all available data on the situation. A SAWS dashboard is useful in that it isn't specifically a LAWS dashboard. Although it could have capacity for lethal action in high-severity scenarios, it should be a hub that allows for operators to respond to all levels of threats, from a power-grid issue, to an

intelligence risk, to environmental emergencies and everything in between, including military threats.

The US has been working on their version of this dashboard since 2017, denoted as Project Maven. Other highly developed countries are working on similar projects and these recent advancements have been sometimes referred to as an artificial intelligence arms race [6]. Although the details of these projects have been kept secret from the public, including what the dashboard looks like or what features it has to ensure safe and ethical usage of LAWS, our goal is to set a baseline for what these interfaces should contain in order to encourage ethical use of autonomous weapons.

Problems that we seek to address are dehumanization of war, insufficient operator control, and lack of escalation safeguards or ethical considerations. A simplified dashboard (one that doesn't provide the operator with all the data and omits the risk and severity of situations it is faced with) allows for the distancing of operators from the gravity of their decisions. Because operators are often physically removed from the war zone by great distances, a dashboard that hides critical details (such as how many lives are at stake) and permits major action without proper authorization further detaches them from the consequences of their decisions. Over time, this detachment can make it easier for operators to make unethical choices. To address these shortcomings, we propose a dashboard design (Figure 1) that reimagines how operators interact with semi-autonomous weapons systems. Our solution emphasizes transparency, ethical decision-making, and meaningful human control through several key innovations.

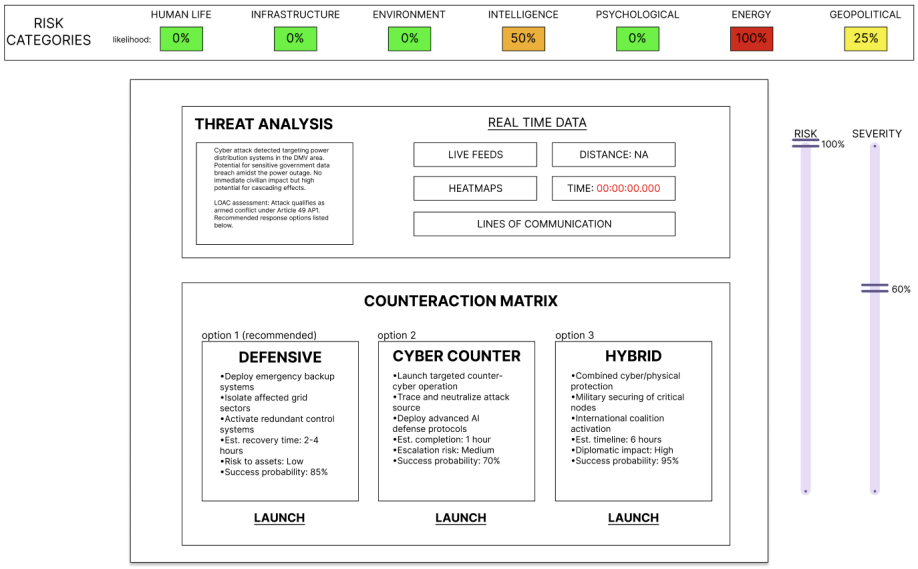

Figure 1: Proposed SAWS Dashboard Interface featuring integrated risk assessment categories, real-time threat analysis, and a multi-option counteraction matrix.

At the very top of the dashboard is a sophisticated risk assessment system that categorizes a threat based on seven domains: human life, infrastructure, environment, intelligence, psychological, energy and geopolitical. Each category visually displays a color and percentage spectrum that highlights the severity and risk of how it may be impacted. The AI estimates these categories' impacts, allowing the response to be more specialized as it can be used to quickly delegate which branches of the military or governmental departments should be involved.

Right under that is a threat analysis component. This breaks down the threat in detail and gives the operator access to all the data points that the system took into consideration. On the left it has an AI-generated threat summary that explicitly references any Laws of Armed Conflict (LOAC) in violation along with more details of the domains at risk. The real time section on the right has many tabs that the operator can navigate to in order to view the real data being used to analyze the threat. Under "Live Feeds" they might be able to view satellite imagery or camera footage. "Heatmaps" would display infrared sensors and heatmaps and "Lines of Communication" provide operators with ways to contact relevant people and groups. They would be able to easily call other countries that might be involved, including the attacker if the threat is an act of aggression or invasion, people that are on the ground where the situation is, and relevant members of the military or government. Having the opportunity to have human-to-human communication is crucial before coming to any major decision and plays a key role in preventing artificial escalation.

Most critically, the counteraction matrix presents operators with multiple response options rather than binary choices. Each option includes detailed implementation steps, estimated timelines, success probabilities, and predicted consequences. This encourages more thoughtful decision-making by arming the operator with data and making the trade-offs of each approach clear. The "Launch" button under each response opens up a deployment interface where the operator can view the logistics and approach timelines in more detail. From here, they can begin deploying the key steps and in the case of a high severity response, the system would require confirmation from a high-ranking official to begin deployment.

The interface has risk and severity sliders on the right hand side that allow for the interface to be updated dynamically. The AI initially sets the risk and severity based on its own estimates, then, if the operator sees it fit, they may move those sliders up and down manually. Changing the risk and severity levels causes the density of information and authorization requirements on the dashboard to scale. Higher severity levels trigger more details in the threat analysis component, more options in the counteraction matrix and stricter authorization protocols. Likewise, low levels of severity simplify the dashboard to allow for quicker responses to less dangerous situations. This feature ensures that the level of oversight scales appropriately with potential consequences.

By implementing these design elements, our dashboard directly addresses the ethical concerns that may arise in semi-autonomous weapons. It maintains human control in military

decision-making while leveraging AI capabilities for data analysis and strategy generation. The system's transparency and emphasis on comprehensive situation awareness help prevent the kind of rapid escalation scenarios that could result from more simplified interfaces that have too little oversight.

Proposed Legal Framework

LAWS present a profound ethical and legal challenge in modern warfare [7]. While these systems may reduce human casualties in some scenarios, they introduce significant risks, including the erosion of accountability, rapid escalation of conflict, and potential violations of the LOAC. To address these issues, this section proposes a comprehensive legal framework encompassing the establishment of an international regulatory body, amendments to the REAIM blueprint, and robust guidelines for the creation and deployment of Lethal Semi-Autonomous Weapon Systems (LSAWS) [8]. This framework ensures compliance with LOAC principles, such as distinction, proportionality, and military necessity, while emphasizing the defensive use of LSAWS under strict regulation.

Building on the technical solution outlined above, which emphasizes transparency, ethical decision-making, and maintaining meaningful human control, the legal framework institutionalizes these principles and ensures compliance on a global scale. While technical safeguards enhance operational integrity, a robust legal structure is necessary to establish accountability, regulate usage, and curb potential misuse [9].

1. Establishing an International Regulatory Body

A foundation of the proposed legal solution is the creation of an international regulatory body modeled after the Nuclear Non-Proliferation Treaty (NPT) [10]. This organization would oversee the responsible development, deployment, and non-proliferation of AWS. Its key functions would include: **(1) Non-Proliferation Oversight:** Monitoring and restricting the spread of critical AI technologies, advanced processors, and targeting algorithms to ensure that AWS capabilities are not exploited by unauthorized actors, **(2) Permissible Use Guidelines:** Defining acceptable scenarios for AWS deployment, prioritizing defensive purposes and mandating human oversight in all critical actions, **(3) Transparency Requirements:** Requiring member states to disclose AWS-related development and deployment activities, thereby fostering accountability and trust among nations. Such a body would ensure a consensus-based approach to AWS governance, addressing discrepancies in national policies that currently exacerbate global tensions.

2. Amendments to the REAIM Blueprint

Building on the Responsible AI in the Military (REAIM) framework, this proposal includes amendments to enforce stricter compliance with ethical and humanitarian principles: **(1) Mandated Human Oversight:** All AWS operations, particularly those involving life-and-death

decisions, must require human decision-makers to retain control, with systems designed to prevent fully autonomous decision-making in critical scenarios, **(2) LOAC Compliance Mechanisms:** Incorporating evaluation tools within AWS to ensure adherence to Laws of Armed Conflict (LOAC) principles, such as distinction, proportionality, and military necessity, **(3) Ethical Algorithm Safeguards:** Requiring ethical evaluations of AI algorithms to prioritize humanitarian considerations and prevent biases that could lead to unlawful or unethical actions. These amendments would strengthen existing guidelines, ensuring that AWS development aligns with international ethical standards.

3. Legal Guidelines for AWS Creation and Deployment

To ensure AWS adhere to both ethical and legal norms, the following design and deployment regulations are proposed [11]: **i) Design Regulations:** (1) *LOAC Compliance Features:* AWS must include advanced target verification systems and proportionality assessment tools to ensure compliance with LOAC, (2) *Failsafe Mechanisms:* Systems must have automatic suspension features to halt operations if LOAC violations are detected, requiring human intervention to proceed, (3) *Standardized Certification:* All AWS must undergo independent testing and certification for legal and ethical compliance before deployment, **ii) Deployment Restrictions:** (1) *Defensive Use Only:* AWS should be limited to defensive scenarios, such as protecting civilians, safeguarding critical infrastructure, or responding to imminent threats, (2) *Pre-Deployment Approvals:* In the case where there is sufficient time to evaluate decisions, all deployments must be reviewed and approved by the proposed international regulatory body, ensuring adherence to defined defensive scenarios, **iii) Accountability Mechanisms:** (1) *Clear Responsibility Chains:* Establishing an enforceable chain of accountability for developers, manufacturers, and operators to prevent ambiguity in liability, (2) *Incident Reporting and Review:* Mandating detailed reporting on AWS operations, with an international tribunal overseeing investigations into any alleged violations, (3) *Penalties for Misuse:* Imposing sanctions, including financial penalties and criminal charges, for violations of LOAC or ethical guidelines.

4. Regulation of Access and Manufacturing

To prevent the proliferation and misuse of AWS, stringent controls over their creation and distribution are essential: **(1) Export Controls:** Licensing requirements for the export of critical components, such as advanced processors and targeting technologies, with bans on sales to non-state actors or non-compliant nations, **(2) Global Registry:** Establishing an international registry to track AWS development and deployment activities, ensuring transparency and accountability, **(3) Secure Development Environments**: Mandating the use of secure facilities for AWS development to reduce risks of intellectual property theft and unauthorized modifications.

5. Integrating LOAC into the Legal Framework

The proposed legal framework embeds LOAC principles to ensure that AWS development and deployment remain aligned with international humanitarian law: **(1) Distinction:** Advanced sensors and algorithms must reliably differentiate combatants from non-combatants to minimize civilian casualties, **(2) Proportionality:** Deployments must undergo rigorous assessments to ensure military advantage justifies potential harm to civilians, **(3) Military Necessity**: AWS usage must be justified by clear military objectives, with all deployments subject to approval by the regulatory body.

Discussion & Conclusions

The deployment of AWS presents some of the most complex ethical dilemmas in modern artificial intelligence, requiring solutions that balance the advantages of automation with the need for human oversight and accountability [12]. Our proposed solutions, combining innovative technical designs with comprehensive legal frameworks, address these challenges by ensuring that human judgment remains central to critical decisions, particularly in military contexts. This research contributes significantly to the broader field of AI ethics by addressing key gaps in current frameworks.

One of the most critical aspects of our technical solution is its emphasis on preserving human agency in decision-making processes. By designing dashboards for Semi-Autonomous Weapons Systems (SAWS) that provide real-time threat analysis, detailed risk assessments, and comprehensive LOAC compliance checks, we ensure that operators retain meaningful control over life-and-death decisions. This contrasts sharply with current systems that often prioritize efficiency at the expense of ethical considerations. Our proposed design solution not only enhances human autonomy but also mitigates risks of rapid conflict escalation by incorporating safeguards such as higher authorization requirements for severe actions.

The proposed legal framework complements the technical solution by institutionalizing accountability and enforcing ethical standards through international collaboration. The creation of a regulatory body modeled after the Nuclear Non-Proliferation Treaty (NPT) provides a unified approach to AWS governance, addressing the current fragmentation in national policies. This body would enforce compliance with LOAC principles, regulate the proliferation of critical AI technologies, and require transparency in AWS development and deployment.

Emerging technologies such as self-replicating AI introduce new dimensions to the ethical debates surrounding AWS [13]. These systems amplify risks by potentially enabling uncontrollable proliferation and unintended consequences. By restricting the creation and distribution of critical AI components and emphasizing transparency through global tracking systems, our legal framework seeks to address these advanced challenges proactively. This is a significant addition to the broader field of AI ethics, which must now grapple with increasingly complex technologies.

While our solutions address many pressing issues, they also underscore the need for further research and inquiry. Algorithmic biases in risk assessments [14], the scalability of international governance frameworks, and the potential misuse of AWS by non-state actors are all areas requiring continued investigation. The dynamic and evolving nature of AI technology necessitates iterative improvements to both technical and governance strategies.

In conclusion, AWS, particularly Semi-Autonomous Weapon Systems, require careful management to ensure their deployment aligns with ethical principles. Our research provides a foundation for achieving this balance, offering technical solutions that enhance human oversight and legal frameworks that promote global stability. While significant progress can be made through these measures, ongoing collaboration, innovation, and oversight are essential to address the evolving landscape of AI technology and its implications for warfare and society.

References

[1] U.S. Department of Defense. 2017. *DoD Directive 3000.09: Autonomy in Weapon Systems.* (May 8, 2017). Retrieved December 15, 2024 from https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf

[2] Hiebert, K. (2024, January 15). The United States quietly kick-starts the Autonomous Weapons Era. Centre for International Governance Innovation. https://www.cigionline.org/articles/the-united-states-quietly-kick-starts-the-autonomous-weapons-era/

[3] U.S. Department of State. 2023. *Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.* (February 16, 2023). Retrieved December 11, 2024, from https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/

[4] Review of the 2023 US policy on autonomy in weapons systems. Human Rights Watch. (2023, February 16). https://www.hrw.org/news/2023/02/14/review-2023-us-policy-autonomy-weapons-systems#_ftnref1

[5] United States Department of State. 2020. *What is MCF One Pager.* (May 2020). Retrieved December 11, 2024 from https://www.state.gov/wp-content/uploads/2020/05/What-is-MCF-One-Pager.pdf

[6] Wikipedia. 2024. *Artificial Intelligence Arms Race.* (Accessed December 11, 2024). Retrieved from https://en.wikipedia.org/wiki/Artificial_intelligence_arms_race

[7] United Nations Office for Disarmament Affairs (UNODA). n.d. *Background on LAWS in the CCW.* (Accessed December 11, 2024). Retrieved from https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/

[8] The Readable. n.d. *REAIM Blueprint for Responsible AI Use (Military).* (Accessed December 11, 2024). Retrieved from https://thereadable.co/reaim-blueprint-for-responsible-ai-use-military/

[9] Boston University. 2023. *We Need Stronger Safeguards from Artificial Intelligence.* (2023). Retrieved December 11, 2024 from https://www.bu.edu/articles/2023/we-need-stronger-safeguards-from-artificial-intelligence/

[10] United Nations Office for Disarmament Affairs (UNODA). n.d. *The Treaty on the Non-Proliferation of Nuclear Weapons (NPT).* (Accessed December 11, 2024). Retrieved from https://disarmament.unoda.org/wmd/nuclear/npt/

[11] U.S. Department of Defense. 2023. *U.S. Endorses Responsible AI Measures for Global Militaries.* (Accessed December 11, 2024). Retrieved from https://www.defense.gov/News/News-Stories/Article/Article/3597093/us-endorses-responsible-ai-measures-for-global-militaries/

[12] Atlantic Council. n.d. *Autonomous Weapons Are the Moral Choice.* (Accessed December 11, 2024). Retrieved from https://www.atlanticcouncil.org/blogs/new-atlanticist/autonomous-weapons-are-the-moral-choice/

[13] Effective Altruism Forum. n.d. *Frontier AI Systems Have Surpassed the Self-Replicating Red Line.* (Accessed December 11, 2024). Retrieved from https://forum.effectivealtruism.org/posts/LycHN9bagozcpYTjp/frontier-ai-systems-have-surpassed-the-self-replicating-red

[14] European Labour Authority (ELA). 2023. *AI Training Handbook: Summary.* (August 2023). Retrieved December 11, 2024 from https://www.ela.europa.eu/sites/default/files/2023-08/AI-training-Handbook-summary.pdf