AI-Powered Autonomous Weapons Risk Geopolitical Instability and Threaten AI Research

Riley Simmons-Edler*
Harvard University
riley_simmons-edler@hms.harvard.edu

Ryan Badman*
Harvard University
ryan_badman@hms.harvard.edu

Shayne Longpre
MIT Media Lab

Kanaka Rajan
Harvard University

1 Extended Abstract

With the rise of powerful generative AI models such as GPT-4 and Stable Diffusion, and continued progress in fields such as robotics and reinforcement learning, concerns have grown among both experts and the public about giving AI too much power. Academic concerns have thus far centered on threats in domains such as cybersecurity, biological weapons, disinformation, fraud, and hypothetical rogue artificial general intelligence (AGI) [9, 2]. Despite these general concerns, there has been relatively little attention given to specific recent developments from military and defense-industry groups, which have already begun to deploy next-generation AI-guided Autonomous Weapon Systems (AWS). Weapons falling under the AWS label have traditionally been either remotely operated (but AI-assisted) or only autonomous within a very narrow scope. However, new fully autonomous (unmanned) AWS in development remove or reduce the human role in the control and decision making process, with the goal of removing humans from the active battlefield en mass. For example, the Pentagon's Replicator program for AI-based weapon "swarms" promises a drastic shift in warfare towards highly autonomous and cooperative AI units within the next few years [10]. AWS can serve many battlefield roles, although human-targeting lethal AWS (LAWS) have received most public attention [23]. Many of these new advanced AI and machine learning (ML) weapons systems are already seeing real-world deployment for the first time in the Ukraine War, and are being designed for every branch of the military and by many nations [21, 10, 18].

We argue that there are fundamental issues caused by removing humans from the battlefield. Human "boots-on-the-ground" can signify a commitment to following the rules of war, improve humanitarian aspects of occupation, and most importantly maintain a human cost to war for aggressor nations that prevents a state of endless war from being politically feasible [24]. We are concerned the recent embrace of AWS by global militaries is leading to a future where wars are more frequent, with such warfare having negative consequences for global stability even if AWS reduce civilian casualties relative to human soldiers. This new model of AWS-centered warfare will be supported by an increasing fusion of civilian and military AI research that will have devastating effects on research and trust in our field.

Official statements make it clear that the direction of AWS development efforts both short- and long-term is the removal of human soldiers from direct combat roles, to reduce casualties and increase combat effectiveness [10, 18, 1]. While these goals are reasonable in isolation, a lack of public attention and transparency around the rapid and increasing pace of AWS development and employment risks humanity sleepwalking into an AWS arms race between global powers. We will see these risks in the near future. China and Russia have given 2028-2030 as targets for major automatization of their militaries to begin, while the USA is set to begin deployment sooner [21, 10, 13, 26].

*equal contribution

2024 Harms and Risks of AI in the Military (HRAIM 2024).

Given these priors, we argue that because highly capable AWS lower the human costs associated with conflict initiation and escalation, they also create a large risk to global geopolitical stability. This effect worsens the more capable AWS become, even if collateral damage decreases, and cannot be solved by simply improving the ML systems involved– policy actions are needed.

In conflicts between powers with large disparities in military strength, invasion or intervention using an AWS-heavy force is politically easier than with an all-human force, since there will be fewer deaths on the aggressor side [17, 10]. However, reducing human casualties in a single conflict can be outweighed if the number of conflicts that occur increases. Indeed, the past century suggests that when dominance in a new military technology leads to regional or global hegemony, this does not always translate to greater stability, and can actually increase lower intensity conflicts and terrorism [3, 7, 11]. AWS-heavy armies with minimal human battlefield presence may lead to a rise in terrorism, attacks on civilians, and other methods extending beyond the traditional battlefield [15]. These abhorrent methods provide a way for less powerful nations who lack AWS to deter or retaliate against nations deploying AWS-heavy forces despite their inability to do so through battlefield casualties [17].

Beyond the impacts AWS have on global stability, the importance of AWS for warfare will likely lead to major negative impacts on civilian AI research. AWS will become a revolutionary military technology, as did nuclear weapons, mechanized warfare, and others historically, but with important differences in the ease of AWS proliferation and impact on civilian technology development [22, 12, 16, 27]. Recent work by Schneider argues that the prevailing military logic—that hegemonic power creates stability—can be highly erroneous when that power is based on technology that relies on a scarce or controlled resource [20]. In the case of AWS, these key resources are AI experts and knowledge, access to data, and semiconductor manufacturing—all of which have historically been dominated by a handful of countries [5, 6, 14, 8, 4]. We will see a rise in export restrictions, publication oversight and redaction, and knowledge compartmentalization in the field of AI as nations attempt to retain these resources to their military advantage [2, 19, 25].

To reduce and prevent these outcomes, action by AI researchers, policymakers, and the public will be needed. We propose several policies and actions to take in Table 1.

References

[1] S.-D. D. Bachmann and R. V. Grant. The need for an Australian regulatory code for the use of artificial intelligence (AI) in military application. *American University National Security Law Brief*, 13(2):2, 2023.

[2] Y. Bengio. My testimony in front of the U.S. Senate - the urgency to act against AI threats to democracy, society and national security. `https://yoshuabengio.org/2023/07/25/my-testimony-in-front-of-the-us-senate/`, July 2023. Accessed: 2023-12-7.

[3] E. Benvenisti. The US and the use of force: Double-edged hegemony and the management of global emergencies. *Eur. J. Int. Law*, 15(4):677–700, Sept. 2004.

[4] H. Chahal, R. Fedasiuk, and C. Flynn. Messier than oil: Assessing data advantage in military AI. *CSET Issue Brief*, 2020. Accessed: 2024-1-3.

[5] M.-C. M. Chu. China's defence semiconductor industrial base in an age of globalisation: Cross-strait dynamics and regional security implications. *Journal of Strategic Studies*, pages 1–26, 2023.

[6] M. Daniels. Controlling knowledge, controlling people: Travel restrictions of U.S. scientists and national security. *Diplomatic History*, 43(1):57–82, Jan. 2019.

[7] D. W. Drezner. Military primacy doesn't pay (nearly as much as you think). *Int. Secur.*, 38(1):52–79, July 2013.

[8] H. Farrell and A. L. Newman. Weak links in finance and supply chains are easily weaponized. `http://dx.doi.org/10.1038/d41586-022-01254-5`, May 2022. Accessed: 2023-12-16.

[9] Future of Life Institute. Pause giant AI experiments: An open letter. `https://futureoflife.org/open-letter/pause-giant-ai-experiments/`, Mar. 2023. Accessed: 2024-1-27.

Table 1: Overview of AWS issues raised and policy recommendations in this work.

TOPIC	CHALLENGES	RECOMMENDATIONS
Major power aggression	AWS replacement of human soldiers makes war more domestically palatable. Defenders turn to asymmetric warfare/terrorism for deterrence.	Require significant human battlefield presence, focus on human-AWS teaming over remote and fully autonomous AWS-centered conflict.
Escalation	Heavy use of AWS makes conflict initiation easier between major powers, risks AWS arms race.	International transparency about broad capabilities, deployment disclosures for AWS systems.
Transparency	Lack of human presence means accountability in war is harder, war crimes and battlefield under-performance less visible to leaders and the public.	Require detailed public reports on AWS capabilities, deployments, and outcomes. Embed oversight/watchdogs in AWS command centers.
Proliferation	Development and sale of AWS will be widespread, global availability of AWS is inevitable.	Avoid futile AI hardware/software restrictions.
Researcher censorship	Military AI needs lead to censorship of civilian research, reduced international collaboration, monitoring and restriction of researchers.	Universities, corporations, governments, etc. establish norms on how much military and civilian research should overlap.
Dual-use AI tech	Many AI algorithms are innately dual-use. Facial recognition, navigation, robotics, etc. Military interest in civilian research is likely to grow.	Improve university ethics oversight and transparency for military-funded AI research, and caution researchers against efforts to weaponize AI.
Over-regulation	Public backlash to AWS leads to calls for more limitations on AI research generally, hurting international research community and academic norms.	Avoid restricting basic AI research, regulate explicit AWS research and military-related datasets over general civilian hardware and AI models.
Autonomy levels	Public data on current AWS are often vague on autonomy levels, definitions of human-in-the-loop, may be more autonomous in practice.	Require governments and AWS manufacturers to clarify the degree of autonomy of AWS. Set international standards for allowed levels of autonomy.

[10] W. C. Greenwalt. DOD's Replicator program: Challenges and opportunities. *American Enterprise Institute*, Oct. 2023.

[11] H. Hegre. Democracy and armed conflict. *J. Peace Res.*, 51(2):159–172, Mar. 2014.

[12] J. Johnson. Artificial intelligence, drone swarming and escalation risks in future warfare. *The RUSI Journal*, 165(2):26–36, Feb. 2020.

[13] E. B. Kania. "AI weapons" in China's military innovation. `https://www.brookings.edu/wp-content/uploads/2020/04/FP_20200427_ai_weapons_kania_v2.pdf`, 2020. Accessed: 2024-1-3.

[14] S. M. Khan, A. Mann, and D. Peterson. The semiconductor supply chain: Assessing national competitiveness. `https://cset.georgetown.edu/wp-content/uploads/The-Semiconductor-Supply-Chain-Issue-Brief.pdf`, 2021. Accessed: 2023-12-16.

[15] J. Kwik. Mitigating the risk of autonomous weapon misuse by insurgent groups. *Laws*, 12(1):5, Dec. 2022.

[16] B. Laird. The risks of autonomous weapons systems for crisis stability and conflict escalation in future U.S.-Russia confrontations. `https://www.rand.org/pubs/commentary/2020/06/the-risks-of-autonomous-weapons-systems-for-crisis.html`, June 2020. Accessed: 2023-11-27.

[17] D. M. Moreau. Unmanned ground vehicles support of irregular warfare: A Non-Lethal approach. *United States Marine Corps University*, 2011.

[18] C. A. Pfaff, C. J. Lowrance, B. M. Washburn, and B. A. Carey. *Trusting AI: Integrating Artificial Intelligence Into the Army's Professional Expert Knowledge*. US Army War College (USAWC) Press, 2023.

[19] J.-M. Rickli and M. Ienca. The security and military implications of neurotechnology and artificial intelligence. In O. Friedrich, A. Wolkenstein, C. Bublitz, R. J. Jox, and E. Racine, editors, *Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications*, pages 197–214. Springer International Publishing, Cham, 2021.

[20] J. Schneider. The capability/vulnerability paradox and military revolutions: Implications for computing, cyber, and the onset of war. *Journal of Strategic Studies*, 42(6):841–863, Sept. 2019.

[21] S. Sharma. Unmanned ground vehicles: Global developments and future battlefield. *Manohar Parrikar Institute for Defence Studies and Analyses Issue Brief*, Apr. 2022.

[22] J. Stern. AI is like ... nuclear weapons? *The Atlantic*, Mar. 2023.

[23] M. Taddeo and A. Blanchard. A comparative analysis of the definitions of autonomous weapons systems. *Sci. Eng. Ethics*, 28(5):37, Aug. 2022.

[24] K. N. Trapp. Boots (on the ground). In J. Hohmann and D. Joyce, editors, *International Law's Objects*. Oxford University Press, Dec. 2018.

[25] A. Warren and A. Hillas. Friend or frenemy? The role of trust in human-machine teaming and lethal autonomous weapons systems. *Small Wars & Insurgencies*, 31(4):822–850, May 2020.

[26] A. Warren and A. Hillas. 'Xi Jinping Thought': Lethal autonomous weapons systems and military modernization with Chinese characteristics. *The Journal of International Relations, Peace Studies, and Development*, 7(1):6, 2022.

[27] Y. H. Wong, J. M. Yurchak, and R. W. Button. *Deterrence in the Age of Thinking Machines*. RAND Corporation, 2020.