Taming the Terminator: Pragmatic International AI Weapons Governance

Cheng-chi (Kirin) Chang

Abstract

This Article examines the legal and ethical challenges posed by artificial intelligence (AI) weapons systems in modern warfare. It analyzes the inadequacies of existing international humanitarian law in addressing autonomous weapons, explores obstacles to effective governance, and proposes a pragmatic model for regulating military AI. The Article argues that while comprehensive international regulation faces significant geopolitical hurdles, incremental progress through targeted measures is both possible and necessary.

Drawing on historical examples of arms control agreements, this Article presents a nuanced approach combining targeted prohibitions, non-binding technical standards, and domestic regulations. It contends that focusing on high-risk applications, such as autonomous control of nuclear weapons, while developing best practices for testing and human-machine interfaces, can mitigate risks without stifling innovation. The analysis emphasizes the need to balance security concerns with ethical imperatives in an era of rapid technological advancement and great power competition.

By recognizing both the limitations of current legal frameworks and opportunities for gradual progress, this Article charts a realistic path forward for the international governance of AI in warfare. It contributes to the ongoing debate on meaningful human control, accountability, and the preservation of core principles of international humanitarian law in increasingly autonomous combat systems.

About the Author

Cheng-chi "Kirin" Chang (張正麒) is the Associate Director & Academic Fellow of the AI and the Future of Work Program at Emory University School of Law. I extend my appreciation to Dr. Ifeoma Ajunwa, J.D., LL.M., Ph.D., Yinn-ching Lu, Rachel Cohen, Yilin (Jenny) Lu, Nanfeng Li, Yenpo Tseng, Zih-Ting You, Youyang Zhong, Ssu-Yuan (Iris) Yang, Ya-jou Liu, Jieyun Lin, and Yizhang (Julia) Shen, as well as the esteemed scholars who participated in *The Inaugural Emory Global AI and Law Colloquium*, including but not limited to Gabriela Arriagada-Bruneau, Anupam Chander, Colleen Chien, Ignacio Cofone, Jake Okechukwu Effoduh, Nikolas Guggenberger, Vivek Krishnamurthy, Ernest Lim, Paul Ohm, Catherine Sharkey, Katherine Strandburg, and Angela Zhang, for their valuable insights and contributions. Their expertise and engagement have significantly enriched this work and broader discussions in the field. Additionally, I am grateful to Benjamin Wu, Charlotte Pettit, Aracely Ruvalcaba, Kelsey Carpenter, and the other editors of the *UCLA Journal of International Law and Foreign Affairs* for their diligent efforts in bringing this article to publication. Any errors or oversights are my sole responsibility. The views expressed herein are my own and do not represent those of any affiliated institutions.

Keywords

International Law; Artificial Intelligence Weapons, Autonomous Weapons Systems, AI Governance in Warfare, International Humanitarian Law and AI, Military AI Regulation, AI and International Security, AI Governance

Table of Contents

Introduction

The rapid development of artificial intelligence (AI) in military applications presents unprecedented challenges to international law and global security. This Article examines the legal and ethical implications of AI weapons systems, which can select and engage targets with limited human intervention. As militaries around the world invest heavily in AI capabilities, from intelligence analysis to fully autonomous combat systems, the international community faces urgent questions about how to govern this transformative technology. While comprehensive international regulation of military AI may be politically challenging in the current geopolitical climate, this Article argues that incremental progress and targeted measures are both possible and necessary. It examines the limitations of applying traditional international humanitarian law (IHL) principles to AI weapons, analyzes the obstacles to effective governance, and proposes a realistic model for regulating military AI that balances security concerns with ethical imperatives.

AI weapons offer significant military advantages, including faster data processing, improved accuracy, and reduced risk to human soldiers. However, they also raise serious concerns about meaningful human control, compliance with IHL principles, and the potential to destabilize international security. The existing legal frameworks governing warfare, primarily based on the Geneva Conventions, were not designed to address the complexities introduced by increasingly autonomous weapons systems.

The analysis proceeds in four parts. Part I provides an overview of AI weapons systems and their implications for modern warfare. Part II examines the existing international legal framework and its inadequacies in addressing AI weapons. Part III analyzes the critical challenges in developing effective governance of military AI. Part IV proposes a pragmatic governance model that combines targeted prohibitions, non-binding standards, and domestic regulations to mitigate the risks posed by military AI. By recognizing both the limitations and opportunities in the current landscape, this Article aims to chart a pragmatic path forward for the international governance of AI in warfare.

I. Implications of AI Weapons

The integration of AI into military systems represents a fundamental shift in warfare that carries profound implications across multiple dimensions. These implications span operational capabilities, ethical considerations, and strategic dynamics. This Part examines how

AI weapons enhance military effectiveness through improved data processing and reduced human risk, while simultaneously raising serious concerns about human control, compliance with international humanitarian law, and global security stability. Understanding these multifaceted implications is crucial for developing appropriate governance frameworks.

The military deployment of AI systems has already become a reality, with global spending on military robotics reaching $7.5 billion per year in 2018, as countries worldwide expanded their arsenals of air, ground, and maritime robots.[1] Current applications primarily focus on intelligence gathering and analysis, with AI systems conducting image recognition and text analysis to enhance situational awareness and decision-making capabilities.[2] In logistics and support operations, AI technologies are extensively utilized for supply chain management, transportation optimization, and equipment maintenance.[3] Major military powers, including the United States, China, and Russia, have integrated various autonomous and semi-autonomous systems into their operations, ranging from self-driving vehicles to more sophisticated autonomous weapons platforms.[4]

Once activated, AI weapons can select and engage targets with limited human intervention.[5] These systems range from semi-autonomous weapons with human oversight, to fully autonomous systems capable of independent decision-making.[6] Military AI applications

1. Paul Scharre, Army of None: Autonomous Weapons and the Future of War 11 (1st ed. 2018) ("Global spending on military robotics is estimated to reach $7.5 billion per year in 2018, with scores of countries expanding their arsenals of air, ground, and maritime robots.").

2. Anthony King, *Digital Targeting: Artificial Intelligence, Data, and Military Intelligence*, 9 J. Glob. Sec. Stud. 1, 3 (2024), https://doi.org/10.1093/jogss/ogae009 (last visited Oct 27, 2024) ("AI has been employed for data collection, collation, and analysis. AI has been used to process data so that commanders have a better under- standing of the battlespace and have been able to target more effectively").

3. Maj. Sharlene Tilley, *Smart Logistics: Navigating the AI Frontier in Sustainment Operations*, U.S. Army (2024), https://www.army.mil/article/280377/smart_logistics_navigating_the_ai_frontier_in_sustainment_operations (last visited Oct. 27, 2024).

4. Forrest E. Morgan et al., Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World, RAND xiii (2020), https://www.rand.org/pubs/research_reports/RR3139–1.html (last visited Oct. 27, 2024) (The source discusses how the United States faces significant international competition in military artificial intelligence (AI), noting that both China and Russia are actively developing militarized AI technologies.).

5. *See generally* Anna-Katharina Ferl, *Fighting Machines: Autonomous Weapons and Human Dignity,* 36 Ethics & Int'l Affs. 539, 539–541 (2022) (This book discusses lethal autonomous weapon systems (LAWS), which are capable of selecting and engaging targets without human intervention).

6. *See generally* Aiden Warren & Alek Hillas, *Friend or Frenemy? The Role of Trust in Human-Machine Teaming and Lethal Autonomous Weapons Systems*, 31 Small Wars & Insurgencies 822 (2020), https://www.tandfonline.com/doi/full/10.1080/09592318.2020.174348

include intelligence analysis, unmanned combat systems, and simulation and training.[7]

The U.S. Department of Defense has invested in projects like the "Maven" program, using machine learning to analyze satellite imagery in real-time, which supports commanders on the battlefield.[8] AI weapons offer other significant benefits, such as quickly processing vast amounts of data, or potentially improving targeting accuracy and reducing collateral damage.[9] They may also reduce human soldier risk by undertaking dangerous tasks.[10]

However, these advantages raise serious concerns. Lowering the barriers to armed conflict can cause the numerous concerns of war and its consequences to manifest in several ways. First, the reduced risk to human personnel may make political leaders more willing to engage in military operations.[11] Second, the speed of AI-driven warfare could compress decision-making timeframes, potentially leading to conflicts escalating rapidly.[12] Third, the proliferation of relatively low-cost AI

5 (last visited Sept. 27, 2024) (This article examines the development of Human-Machine Teaming and the trust issues associated with LAWS, highlighting the varying degrees of autonomy these systems can possess).

7. *See generally* Alessandro Nalin & Paolo Tripodi, *Future Warfare and Responsibility Management in the AI-Based Military Decision-Making Process*, 14 J. Advanced Mil. Stud. 83 (2023), https://www.usmcu.edu/Outreach/Marine-Corps-University-Press/MCU-Journal/JAMS-vol-14-no-1/Future-Warfare-and-Responsibility-Management/ (last visited Sept. 27, 2024) (This source discusses the rapid growth of AI technology in military applications, including its use in intelligence analysis and unmanned systems, highlighting the integration of AI in various military operations).

8. Katrina Manson, *AI Warfare Is Already Here*, Bloomberg (Feb. 28, 2024), https://www.bloomberg.com/features/2024-ai-warfare-project-maven/ ("the ones developing and using Project Maven to identify targets on the battlefield.").

9. Robert Bruce Douglass, *2004 Command and Control Research and Technology Symposium The Power of Information Age Concepts and Technologies Topic: Effects Based Applications and Operations Military Robotics and Collateral Damage* 3 (2004), https://www.semanticscholar.org/paper/2004-Command-and-Control-Research-and-Technology-of-Douglass/02fe06f8930424127a4c6debbe7d215ba7e7647b (last visited Sept. 27, 2024) ("use of combat robots with the right rules of engagement can dramatically reduce the risk of collateral damage as compared to manned combat operations").

10. Riley Simmons-Edler et al., *AI-Powered Autonomous Weapons Risk Geopolitical Instability and Threaten AI Research*, 235 Proc. Mach. Learning Rsch. 1, 1 (2024) ("ML is already enabling the substitution of AWS for human soldiers in many battlefield roles, reducing the upfront human cost").

11. Scharre, *supra* note 1, at 11 ("They can take more risk, opening up tactical opportunities for dangerous or even suicidal missions without risking human lives").

12. *Id.* at 23 ("In 2047 technology will be able to reduce the time to complete the OODA loop to micro or nanoseconds. Much like a chess master can outperform proficient chess players . . . .While the ability to carry out attacks without communications has a mixed effect on stability, autonomous weapons' advantage in speed is decidedly negative. Autonomous weapons risk accelerating the pace of battle and shortening time for human decision-making. This heightens instability in a crisis").

weapons systems could enable smaller states and non-state actors to conduct military operations previously beyond their capabilities.[13] Additionally, AI weapons may operate beyond meaningful human control, challenging traditional notions of command responsibility and individual culpability for war crimes.[14] Their ability to comply with fundamental IHL principles, particularly distinction and proportionality rules, remains questionable.[15] The complexity of modern battlefields and the nuanced judgments required to distinguish combatants from civilians may exceed current AI capabilities.[16]

Furthermore, AI weapons' rapid development could destabilize international security.[17] Concerns about an AI arms race and lowered barriers to armed conflict have prompted calls for international regulation.[18] The proliferation of AI weapons has led to significant military investments, with China declaring in 2018 its pursuit of becoming the

13. *Id.* at 4–5 (The source notes that autonomous military technology is becoming widespread. For example, more than ninety countries already have drones patrolling the skies, and more than thirty nations have supervised autonomous weapons that can respond more quickly than humans).

14. Peter Margulies, *Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts in* Research Handbook on Remote Warfare 405, 432 (Jens David Ohlen ed., 2017) (This source discusses the challenges posed by autonomous weapons systems (AWS) to traditional notions of command responsibility and individual culpability for war crimes. It highlights that AWS can operate without specific human authorization, creating accountability gaps under international humanitarian law (IHL)).

15. *See generally* Scott Sullivan & Iben Ricket, *Targeting in the Black Box*, *in* 2024 16th International Conference on Cyber Conflict: Over the Horizon (CyCon) 207 (C. Kwan et al. eds., 2024), https://ieeexplore.ieee.org/document/10685575 (last visited Sept. 27, 2024) (This paper delves into the "black-box problem" of AI systems in military operations, particularly in targeting. It highlights the mismatch between black-box AI models and existing IHL principles, which are fundamentally rooted in nuanced context and subjective judgment. The paper argues that the unintelligibility of many AI-based systems to their developers and users poses significant challenges in ensuring compliance with IHL principles of distinction, proportionality, and precaution).

16. Ozlem Ulgen, *The Ethical Implications of Developing and Using Artificial Intelligence and Robotics in the Civilian and Military Spheres* (2017), https://www.semanticscholar.org/paper/The-ethical-implications-of-developing-and-using-in-Ulgen/b43cd247e7afc1fc9a04c170c221bc87511904af (last visited Sept. 27, 2024) (This source discusses the ethical and technical challenges associated with the use of AI and robotics in military applications, particularly in distinguishing between combatants and civilians on complex modern battlefields. It highlights the moral questions surrounding the replacement of human judgment with machine logic, which may not be capable of making nuanced decisions required in warfare.).

17. Tshilidzi Marwala, *Militarization of AI Has Severe Implications for Global Security and Warfare*, U.N. Univ. (July 24, 2023), https://unu.edu/article/militarization-ai-has-severe-implications-global-security-and-warfare (last visited Sept. 27, 2024).

18. Sam Meacham, *A Race to Extinction: How Great Power Competition Is Making Artificial Intelligence Existentially Dangerous*, Harv. Int'l Rev. (Sept. 8, 2023), https://hir.harvard.edu/a-race-to-extinction-how-great-power-competition-is-making-artificial-intelligence-existentially-dangerous/ (last visited Sept. 27, 2024).

global leader in AI technology after investing $150 billion towards this objective.[19] However, scholars have argued that, while nations are clearly competing to develop and adopt AI technology for military use, the character of that competition does not meet the traditional definition of an arms race, as military AI spending remains relatively modest—for example, U.S. defense spending on AI was only $5 billion in the fiscal year 2020, roughly 0.7 percent of the Department of Defense's budget.[20] Nevertheless, the risks of a competition over military AI remain significant, as perceptions of a "race" to field AI systems before competitors could cause nations to cut corners on testing, leading to the deployment of unsafe AI systems that could cause accidents and unintended escalation.[21] Additionally, the dynamic shift in warfare capabilities has spurred competitive development of autonomous weapons among superpowers like the United States and China, which continues to pose significant challenges to the international system.[22]

The U.S. Department of Defense has developed a crucial framework distinguishing between "human-in-the-loop," "human-on-the-loop," and "human-out-of-the-loop" systems, which directly impacts the core challenges of AI weapons governance.[23] Human-in-the-loop systems require direct human authorization for each target engagement, maintaining traditional chains of command while potentially sacrificing the speed advantages of AI.[24] Human-on-the-loop systems allow AI to operate independently while humans retain override capability, creating a precarious balance between operational efficiency and meaningful control.[25] Most controversially, human-out-of-the-loop

19. Goddy Uwa Osimen et al., *Artificial Intelligence and Arms Control in Modern Warfare*, 10.1 Cogent Soc. Scis. 1, 6 (2024), https://doi.org/10.1080/23311886.2024.2407514 (last visited Nov. 1, 2024) ("In 2018, China declared success in its pursuit of becoming the global leader in AI technology after investing $150 billion towards this objective").

20. Paul Scharre, *Debunking the AI Arms Race Theory*, 4.3 Tex. Nat'l Sec. Rev. 121, 124 (2021), https://tnsr.org/2021/06/debunking-the-ai-arms-race-theory/ (last visited Nov. 1, 2024) ("An independent estimate by Bloomberg Government of U.S. defense spending on AI identified $5 billion in AI-related research and development in fiscal year 2020, or roughly 0.7 percent of the Department of Defense's over $700 billion budget").

21. *Id.*

22. Osimen et al., *supra* note 19.

23. Amitai Etzioni & Oren Etzioni, *Pros and Cons of Autonomous Weapons Systems*, Mil. Rev., May-Jun. 2017, at 78, https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/May-June-2017/Pros-and-Cons-of-Autonomous-Weapons-Systems/ (last visited Sept. 27, 2024) ("Based on the kind of human involvement, the categories are human-in-the-loop, human-on-the-loop, and human-out-of-the-loop weapons").

24. Bonnie Docherty, *Losing Humanity: The Case Against Killer Robots*, Hum. Rts. Watch (Nov. 19, 2012), https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

25. *Id.*

systems operate with full autonomy once activated, raising fundamental questions about compliance with IHL principles and command responsibility.[26] This classification framework illustrates the central tension in AI weapons governance: as autonomy increases, military effectiveness may improve, but the ability to ensure compliance with legal and ethical principles decreases. The spectrum of human control options also demonstrates why one-size-fits-all regulations may be impractical, suggesting instead the need for context-specific governance approaches that account for different levels of autonomy and associated risks.

II. International Legal Framework

The existing international legal frameworks governing warfare, primarily embodied in the 1949 Geneva Conventions and their 1977 Additional Protocols, are insufficient to address the complexities introduced by AI weapons.[27] IHL, which encompasses both treaty law and customary international law, is built around principles designed for human-controlled systems.[28] The principle of distinction, codified in Article 48 of Additional Protocol I, requires combatants to clearly differentiate between civilians and military targets at all times.[29] This principle is reinforced by Article 51(4), which prohibits indiscriminate attacks that cannot be directed at specific military objectives.[30]

26. Neil Davison, *A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law*, 5 5–18 (2018) (states that "compliance with IHL requires limits to lawful levels of autonomy in weapon systems" and that "the greater the uncertainty and unpredictability, the greater the risk that IHL might be violated."); Yordan Gunawan et al., *Command Responsibility of Autonomous Weapons under International Humanitarian Law*, 8 Cogent Social Sciences 2139906 (2022) (states that "command responsibility is still being questioned as to whether it can be applied to AWS weapons classified as 'Human-out-of-the-Loop Weapons'").

27. Qiang Li & Dan Xie, *Legal Regulation of AI Weapons under International Humanitarian Law: A Chinese Perspective*, Humanitarian L. & Pol'y Blog (May 2, 2019), https://blogs.icrc.org/law-and-policy/2019/05/02/ai-weapon-ihl-legal-regulation-chinese-perspective/ ("AI weapons—also known as autonomous weapon systems [AWS], which have been defined by the ICRC as weapons that can independently select and attack targets, i.e., with autonomy in the 'critical functions' of acquiring, tracking, selecting and attacking targets—have raised a series of issues both legally and ethically. It is debatable whether such weapons/weapon systems with the functions of learning, reasoning, decision-making and the ability to act independent of human intervention should be employed in future battlefields").

28. *See generally* Shin-Shin Hua, *Machine Learning Weapons and International Humanitarian Law: Rethinking Meaningful Human Control*, 51 Geo. J. Int'l L. 117 (2020).

29. Protocol Additional to the Geneva Convention of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), art. 48, U.N. Doc. A/32/144, Annex I, 1125 U.N.T.S. 3, *reprinted in* 16 I.L.M. 1391 (1977) ("In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.").

30. *Rule 7: The Principle of Distinction between Civilian Objects and Military Objectives*,

AI's inability to consistently make such distinctions, especially in the ambiguous environments of modern conflict, reveals the limitations of applying traditional legal principles to new technology.[31]

Proportionality, enshrined in Articles 51(5)(b) and 57(2)(a)(iii) of Additional Protocol I, is another principle that becomes difficult to enforce in the context of AI weapons.[32] This principle prohibits attacks that cause excessive civilian harm relative to the anticipated military advantage.[33] While AI systems can process vast amounts of data faster than humans can, they do not currently possess the contextual understanding required to apply proportionality in a meaningful way.[34] A meaningful application requires nuanced judgment, weighing military necessity against civilian harm while considering cultural and social contexts—capabilities that current AI systems lack.[35] This is particularly concerning in unpredictable battlefield environments where AI, such as facial recognition algorithms used in autonomous weapons systems, might misinterpret data or fail to account for broader strategic considerations, leading to potential misidentification of combatants and civilians.[36]

International Committee of the Red Cross: Int'l Humanitarian L. Databases, https://ihl-databases.icrc.org/en/customary-ihl/v1/rule7 (last visited Sept. 27, 2024) ("The parties to the conflict must at all times distinguish between civilian objects and military objectives. Attacks may only be directed against military objectives. Attacks must not be directed against civilian objects.").

31. Cordula Droege & Eirini Giorgou, *How International Humanitarian Law Develops*, 920–921 Int'l Rev. Red Cross 1798, 1809 (2022), http://international-review.icrc.org/articles/how-international-humanitarian-law-develops-920 ("At the time of writing this article [2022], ongoing discussions on prohibiting and regulating autonomous weapon systems turn around similar issues. In both cases, the central question is whether existing IHL rules and principles are sufficient to effectively protect those affected by such weapons.").

32. Bonnie Docherty, *Mind the Gap - The Lack of Accountability for Killer Robots*, Human Rights Watch (Apr. 9, 2015), https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots.

33. *See generally* Benjamin M. Jensen et al., *Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence*, 22.3 Int'l Stud. Rev. 526 (2020) (This source discusses how LAWS capitalize on AI's ability to handle big data for rapid decision-making in combat).

34. *See generally* Edward Hunter Christie et al., *Regulating Lethal Autonomous Weapon Systems: Exploring the Challenges of Explainability and Traceability*, 4 AI Ethics 229 (2024) (This source discusses how AI struggles with the qualitative nature of legal and ethical assessments, leading to potential over-reliance on AI in military operations and risking IHL violations.).

35. Steven Umbrello, *No Machine Should Choose: Defending Human Dignity in the Age of Autonomous Weapons*, Word on Fire (Sept. 19, 2024), https://www.wordonfire.org/articles/no-machine-should-choose-defending-human-dignity-in-the-age-of-autonomous-weapons/ ("While current autonomous systems may be capable of identifying targets based on predefined criteria, they lack the nuanced understanding that human operators bring to the battlefield").

36. Jake Okechukwu Effoduh, *Weapons Powered by Artificial Intelligence Need to Be Regulated*, World Econ. Forum (Jun. 23, 2021), https://www.weforum.org/stories/2021/06/the-accelerating-development-of-weapons-powered-by-artificial-risk-is-a-risk-to-humanity/.

Article 36 of Additional Protocol I to the Geneva Conventions mandates that states review new weapons to ensure their compliance with IHL.[37] However, AI's rapid development and the opaque nature of machine learning algorithms make these reviews difficult to conduct. Current legal frameworks do not adequately account for the unpredictability of AI decision-making or the difficulty in determining how these systems will behave under combat conditions—particularly in testing AI responses to novel battlefield scenarios and measuring algorithmic reliability across varied combat situations.[38] Moreover, the evolving concept of "meaningful human control" remains undefined in legal terms.[39] As AI systems gain more autonomy, the question of what constitutes sufficient human oversight becomes increasingly complex.

Despite the growing importance of AI in military systems, international efforts to regulate its use have stalled. The Group of Governmental Experts on Lethal Autonomous Weapons Systems, operating under the framework of the Convention on Certain Conventional Weapons (CCW), has yet to produce binding regulations, reflecting the geopolitical resistance to meaningful constraints on AI weapons.[40] National initiatives, such as the U.S. Department of Defense Directive 3000.09, provide guidelines, but they fall short of establishing enforceable legal norms.[41] The European Union's Artificial Intelligence Act, while comprehensive in its regulation of civilian AI applications,

37. *Article 36 - New Weapons*, International Committee of the Red Cross: International Humanitarian Law Databases, https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-36 (last visited Sept. 27, 2024) (Article 36 of Additional Protocol I to the Geneva Conventions requires that states undertake a legal review of any new weapon, means, or method of warfare before its deployment to ensure it complies with IHL).

38. International Committee of the Red Cross, *Legal Review of New Weapons: Scope of the Obligation and Best Practices*, Humanitarian L. & Pol'y Blog (Oct. 6, 2016), https://blogs.icrc.org/law-and-policy/2016/10/06/legal-review-new-weapons/; Robert A. Coombs, *AI Integration for Scenario Development: Training the Whole-of-Force*, Mil. Rev., May 2024 (Online Exclusive), https://www.armyupress.army.mil/Journals/Military-Review/Online-Exclusive/2024-OLE/AI-Integration-for-Scenario-Development/ (last visited Nov 1, 2024).

39. *See generally* Thompson Chengeta, *Defining the Emerging Notion of "Meaningful Human Control" in Weapon Systems*, 49.3 N.Y.U. J. Int'l L. & Pol'y 833 (2017).

40. *Autonomous Weapons: ICRC Urges States to Launch Negotiations for New Legally Binding Rules*, International Committee of the Red Cross (May 6, 2023), https://www.icrc.org/en/document/statement-international-committee-red-cross-icrc-following-meeting-group-governmental (" . . . As a result, the assertion adds nothing to existing legal frameworks").

41. *See generally Review of the 2023 US Policy on Autonomy in Weapons Systems,* Human Rights Watch (Feb. 14, 2023, 8:01 AM EST), https://www.hrw.org/news/2023/02/14/review-2023-us-policy-autonomy-weapons-systems (This source discusses how Directive 3000.09 sets guidelines that require autonomous weapons to allow for "appropriate levels of human judgment over the use of force." However, it lacks clear definitions of what constitutes "appropriate" human judgment, making its practical enforcement challenging.).

explicitly excludes military applications under Article 2(3), further complicating efforts to develop comprehensive international regulations.[42]

However, it is important not to overlook the potential for legal progress, even in challenging geopolitical climates. Historical examples, such as the successful negotiation of the Nuclear Non-Proliferation Treaty (NPT) during the Cold War, show that nations can come together and establish arms control agreements even amid significant international tensions.[43] The development of AI in warfare may follow a similar trajectory, particularly given the precedent set by other weapons-specific treaties such as the Chemical Weapons Convention and the Ottawa Treaty banning anti-personnel landmines.[44] While comprehensive treaties may not be feasible in the short term, incremental progress through smaller, targeted agreements is possible. These could include restrictions on the most dangerous applications of AI, such as autonomous nuclear weapons, or agreements to maintain human oversight over critical military decisions. Legal frameworks can and have evolved, even in the face of geopolitical resistance.

III. Governance Challenges

Given these challenges, we must confront several hard realities in the development of a realistic governance framework for military AI. First, the great power competition between nations such as the United States, China, and Russia complicates any efforts to establish binding international agreements.[45] AI is seen as a strategic technol-

42. Rosanna Fanni, *Why the EU Must Now Tackle the Risks Posed by Military AI*, CEPS (Jun. 8, 2023), https://www.ceps.eu/why-the-eu-must-now-tackle-the-risks-posed-by-military-ai/; Scope of the EU Artificial Intelligence Act (AIA): Military Purposes and National Security, Eur. Ctr. for Not-for-Profit Law (2022), https://ecnl.org/sites/default/files/2022–03/ECNL%20Pagers%20on%20scope%20of%20AIA%20ECNL_FINAL.pdf.

43. Peter Bittner, *The Imperative of a US-China Accord on AI: Lessons from Arms Control*, China-US Focus (Aug. 18, 2023), https://www.chinausfocus.com/peace-security/the-imperative-of-a-us-china-*accord*-on-ai-lessons-from-arms-control (Historical examples like the successful negotiation of the Nuclear Non-Proliferation Treaty (NPT) demonstrate that nations can establish arms control agreements even during significant international tensions.).

44. *Anti-Personnel Landmines Convention*, U. N. Off. Disarmament Aff., https://disarmament.unoda.org/anti-personnel-landmines-convention/ (last visited Dec. 15, 2024) ("Since its inception more than two decades ago, the Convention has led to a virtual halt in global production of anti-personnel mines, and a drastic reduction in their deployment. More than 40 million stockpiled mines have been destroyed, and assistance has been provided to survivors and populations living in the affected areas").

45. *See generally* Michael J. Mazarr et al., *Stabilizing Great-Power Rivalries*, RAND (Nov. 29, 2021), https://www.rand.org/pubs/research_reports/RRA456–1.html (The RAND Corporation highlights that the current U.S.-China and U.S.-Russia rivalries are particularly unstable, with numerous factors pushing these relationships toward greater volatility. This instability limits the possibility of achieving long-term agreements, as mutual distrust makes meaningful cooperation

ogy, and military powers are unlikely to agree to meaningful restrictions that could undermine their competitive edge. This mistrust between military powers, coupled with the rapid pace of technological development, makes substantive cooperation difficult, if not impossible, in the short term. The situation is reminiscent of the early stages of nuclear proliferation, where national security concerns initially outweighed collaborative efforts for arms control.[46] However, it is worth noting that even during the height of the Cold War, progress was eventually made through initiatives like the Strategic Arms Limitation Talks (SALT), suggesting that patience and persistent diplomacy may yield results over time.[47]

Second, the ongoing autonomy arms race creates immense pressure on nations to deploy increasingly autonomous systems.[48] As AI capabilities improve, maintaining meaningful human control may become impossible in certain high-speed military scenarios, such as cyber warfare or missile defense. This shift threatens to erode the core principles of IHL, particularly regarding human accountability for lethal decisions. The concept of "meaningful human control" itself becomes increasingly nebulous as AI systems grow more sophisticated, raising philosophical and practical questions about the nature of human agency in warfare. This dilemma echoes past debates about the automation of nuclear launch systems, in which the desire for rapid response had to be balanced against the moral imperative of human decision-making in matters of existential importance.

Third, the dual-use nature of AI complicates efforts to regulate its military applications. Many AI technologies developed for civilian purposes, such as advanced data processing and machine learning, can be easily adapted for military use.[49] Attempts to strictly regulate

difficult.).

46. Nuno P. Monteiro & Alexandre Debs, *The Strategic Logic of Nuclear Proliferation*, 39.2 Int'l Sec. 7, 7 (2014) ("Intuitively, security is the most important factor driving nuclear acquisition").

47. *See generally Understanding SALT Treaties: Key Agreements in Military History*, Total Mil. Insight (Jun. 19, 2024), https://totalmilitaryinsight.com/salt-treaties/ (The Strategic Arms Limitation Talks (SALT) serve as a key example of how persistent diplomacy can yield meaningful results, even during periods of intense geopolitical rivalry like the Cold War.).

48. David Lague, *In U.S.-China AI Contest, the Race Is On to Deploy Killer Robots*, Reuters (Sept. 8, 2023), https://www.reuters.com/investigates/special-report/us-china-tech-drones/ (This source discusses how countries like the United States, China, and Russia are heavily investing in AI-driven military technologies, from drones to fully autonomous systems, in an effort to gain or maintain technological superiority on the battlefield.).

49. *See generally* Shama Ams, *Blurred Lines: The Convergence of Military and Civilian Uses of AI & Data Use and Its Impact on Liberal Democracy*, 60 Int'l Pols. 879 (2023) (This source discusses how AI-driven tools, initially designed for commercial purposes, are increasingly

military AI risk stifling technological innovation in other sectors, raising the question of how to balance these competing interests.[50] This challenge is not unique to AI; similar issues have arisen with other dual-use technologies like encryption and satellite imagery.[51] However, the pervasiveness and rapid evolution of AI make this problem particularly acute. Policymakers must grapple with the task of crafting regulations that are specific enough to be effective in the military domain, yet flexible enough to benefit civilian applications. Potential approaches could include establishing technical parameters for military AI capabilities and implementing separate regulatory frameworks for dual-use technologies based on their intended application and risk level. For instance, this could mean applying stricter oversight to computer vision technology in autonomous weapons compared to its use in civilian manufacturing.[52]

The software-driven nature of AI weapons also presents verification challenges. Traditional arms control measures, such as inspections and monitoring, are largely ineffective in the context of AI, in which much of the system's functionality is embedded in code.[53] Nations are unlikely to allow intrusive inspections of sensitive military AI programs, further undermining the potential for meaningful oversight. The often opaque nature of machine learning algorithms only exacerbates this problem, making it difficult for people to predict or explain system behavior. New verification methodologies, possibly involving

being integrated into military operations.).

50. *See generally* Eoin Micheál McNamara, *Nuclear Arms Control Policies and Safety in AI,* Finnish Inst. Int'l Aff. (Jan. 2024), https://www.fiia.fi/en/publication/nuclear-arms-control-policies-and-safety-in-artificial-intelligence (This source discusses the tension between AI regulation and innovation, emphasizing that overly stringent government safety regulations could inhibit technological advancement, especially when AI is treated as a dual-use technology that straddles both civilian and military domains. Such regulations can curb the flow of AI innovations from civilian industries to the military, potentially slowing progress across sectors.).

51. Guangyu Qiao-Franco & Mahmoud Javadi, *Symposium on Military AI and the Law of Armed Conflict: Navigating the Governance of Dual-Use Artificial Intelligence Technologies in Times of Geopolitical Rivalries*, Opinio Juris (Apr. 3, 2024), https://opiniojuris.org/2024/04/03/symposium-on-military-ai-and-the-law-of-armed-conflict-navigating-the-governance-of-dual-use-artificial-intelligence-technologies-in-times-of-geopolitical-rivalries/ (This source discusses how dual-use technologies often pose governance challenges because they serve both civilian and military purposes.).

52. Tobias Vestner, *Rules of Engagement as a Regulatory Framework for Military Artificial Intelligence*, Articles of War (Aug. 27, 2024), https://lieber.westpoint.edu/rules-engagement-regulatory-framework-military-artificial-intelligence/.

53. *See generally* Megan Lamberth & Paul Scharre, *Arms Control for Artificial Intelligence*, 6.2 Tex. Nat'l Sec. Rev. 96 (2023), https://repositories.lib.utexas.edu/server/api/core/bitstreams/bee19431–4d46–4fba-a5b8–4b769e019e28/content (This source discusses how traditional verification, which works well with physical weapons (e.g., nuclear), fails in the context of AI, where functionality is often intangible and hidden in software.).

secure multi-party computation or zero-knowledge proofs, may need to be developed to address these unique challenges.

Finally, as human involvement in the decision-making process decreases, traditional notions of command responsibility and individual criminal liability become increasingly problematic.[54] This accountability gap must be addressed if the deployment of AI weapons is to comply with international law. The principle of command responsibility, a cornerstone of international criminal law, may need to be re-evaluated and possibly expanded to encompass the roles of AI developers, trainers, and operators.[55] Moreover, the potential for AI systems to make decisions that were not explicitly programmed or anticipated raises complex questions about moral and legal culpability that existing legal frameworks are ill-equipped to handle.

IV. Toward a Pragmatic Governance Model

While geopolitical competition complicates international cooperation, we must not dismiss the potential for incremental progress in the global governance of AI.[56] The history of arms control provides examples of gradual progress, even in times of intense international rivalry.[57]

54. Marta Bo, *Meaningful Human Control over Autonomous Weapon Systems: An (International) Criminal Law Account*, Opinio Juris (Dec. 18, 2020), https://opiniojuris.org/2020/12/18/meaningful-human-control-over-autonomous-weapon-systems-an-international-criminal-law-account/ (This source discusses how the increasing autonomy of AWS challenges the core principle of "meaningful human control" in international criminal law, particularly in determining liability when humans no longer oversee decisions at key moments, such as targeting.).

55. Russell Buchan & Nicholas Tsagourias, *Autonomous Cyber Weapons and Command Responsibility*, 96 Int'l L. Stud. 645, 652 (2020) (The doctrine of command responsibility has a long history in international criminal law and comprises three core elements: the existence of a superior-subordinate relationship, the commander's knowledge of the crime, and the commander's failure to prevent or repress the subordinate's criminal actions.).

56. *See generally* Huw Roberts et al., *Global AI Governance: Barriers and Pathways Forward*, 100.3 Int'l Aff. 1275 (2024) (This source discusses how, despite growing geopolitical tensions, there remains space for incremental cooperation on global AI governance. For instance, frameworks like the G20 or specialized AI partnerships offer limited but meaningful opportunities for progress in setting common norms despite competition, especially between global powers like the U.S. and China.); *see also* Cheng-chi (Kirin) Chang, *The First Global AI Treaty: Analyzing the Framework Convention on Artificial Intelligence and the Eu AI Act*, 2024 U. Ill. L. Rev. Online 86, 90 (2024), https://illinoislawreview.org/wp-content/uploads/2024/12/Chang2.pdf (discussing the AI Framework Convention's creation shows broad international participation, indicating progress despite potential geopolitical hurdles); *see also* Cheng-chi (Kirin) Chang, *Global Minds, Local Governance: AI in International Law*, 2025 U. Ill. L. Rev. Online1, 1, 6 (2025), https://illinoislawreview.org/wp-content/uploads/2025/03/Chang_GlobalMinds.pdf (discussing trade agreements demonstrates that AI governance is incrementally being incorporated into existing international legal frameworks).

57. Michael Moodie & Jerry Zhang, *Bolstering Arms Control in a Contested Geopolitical Environment*, Stimson Ctr. (Oct. 31, 2022), https://www.stimson.org/2022/bolstering-arms-control-in-a-contested-geopolitical-environment/ (This source explores how arms control

Agreements on confidence-building measures, such as crisis communication hotlines and data-sharing protocols, have the potential to reduce the risks of unintended escalation driven by AI systems.[58] These agreements, while modest, could help lay the groundwork for more comprehensive future regulations. The Incidents at Sea Agreement between the United States and Soviet Union during the Cold War offers a historical parallel, demonstrating how even adversaries can find common ground in reducing the risks of unintended conflict.[59] Even during the height of the Cold War, both nations recognized the dangers posed by unpredictable military encounters, particularly at sea, and sought to reduce the possibility of accidents or miscalculations.[60] The agreement set out clear rules on how ships and aircraft from the two superpowers should interact to avoid incidents that could escalate into conflict.[61]

Non-binding technical standards and best practices for AI weapons testing, verification, and human-machine interfaces can also contribute to responsible AI development. While these standards may not have the force of law, they can provide a framework that encourages states to act responsibly, particularly if developed in cooperation with military, legal, and ethical experts. The success of initiatives like the Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies shows that voluntary frameworks

negotiations, even during periods of heightened international competition, such as during the Cold War, managed to achieve incremental successes. It emphasizes that treaties like START and the INF Treaty show that gradual steps toward disarmament were possible amidst rivalry.).

58. *Id.*

59. Alexander Klimburg, *Of Ships and Cyber: Transposing the Incidents at Sea Agreement*, Center for Strategic & International Studies (Sept. 28, 2022), https://www.csis.org/analysis/ships-and-cyber-transposing-incidents-sea-agreement (The 1972 agreement is described as "a milestone in de-escalation and confidence building" that created clear rules for naval encounters despite adversarial relations.); Sean M. Lynn-Jones, *A Quiet Success for Arms Control: Preventing Incidents at Sea*, 9 Int'l Sec. 154 (1985) (The agreement "has helped to avert potentially dangerous incidents between the U.S. and Soviet navies" and its "effectiveness appears to have survived the deterioration of U.S.-Soviet relations").

60. Klimburg, *supra* note 59 (By the time of signing, "the rapidly expanding Soviet and U.S. navies were increasingly bumping into each other—sometimes literally" and "[t]he potential for 'inadvertent escalation' [i.e., accidental war], was obvious."); Raul Pedrozo, *The U.S.-China Incidents at Sea Agreement: A Recipe for Disaster*, 6 J. of Nat'l Sec. L. and Pol'y 207, 207 (2012), https://jnslp.com/wp-content/uploads/2012/08/07_Pedrozo-Master.pdf (These incidents "culminated nearly a decade of dangerous incidents between U.S. and Soviet naval forces – close passes by low-flying aircraft, intentional shouldering [bumping] of surface ships, threatening maneuvers, and mock surface and air attacks").

61. Agreement Between the Government of The United States of America and the Government of The Union of Soviet Socialist Republics on the Prevention of Incidents On and Over the High Seas, U.S. Dep't of State (1973), https://2009–2017.state.gov/t/isn/4791.htm (last visited Sept. 27, 2024).

can have significant impact on state behavior.[62] Research on AI governance frameworks suggests that voluntary norms, particularly around transparency and accountability, are increasingly seen as essential for managing the ethical challenges posed by AI systems. These frameworks can address concerns related to fairness, bias, and the opacity of AI decision-making, while also covering key technical aspects like algorithmic transparency, safety mechanisms, and ethical considerations in data selection.[63]

Research into ethical AI development, focusing on transparency, predictability, and alignment with human values, is another area where progress can be made, even in the absence of binding treaties. Collaborative international research projects, similar to CERN in particle physics, could be established to advance our understanding of safe and ethical AI systems.[64] Such initiatives could foster a shared global understanding of AI risks and best practices, potentially paving the way for more formal agreements in the future.

Recognizing the political difficulties of establishing broad regulations, targeted prohibitions on high-risk applications of AI—such as banning autonomous control of nuclear weapons—could serve as a practical step forward. This approach mirrors successful past efforts like the Biological Weapons Convention, which focused on a specific category of weapons.[65] Similarly, doctrines and technical measures designed to maintain human judgment over critical targeting decisions are essential to preserving the integrity of IHL in an increasingly autonomous world. These measures could include mandatory override

62. *Fact Sheet: The Wassenaar Arrangement*, Ctr. for Arms Control and Non-Proliferation (Mar. 8, 2023), https://armscontrolcenter.org/fact-sheet-the-wassenaar-arrangement/ (This arrangement demonstrates that voluntary agreements, though non-binding, foster transparency and responsibility among member states by encouraging information exchange and cooperative measures to prevent the proliferation of sensitive materials.).

63. *See generally* Jonas Tallberg et al., *The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research*, 25 Int'l Stud. Rev. 1 (2023).

64. *See generally* Elliot Jones, *A 'CERN for AI' – What Might an International AI Research Organization Address? in* Artificial Intel. and the Challenge For Glob. Governance (Alex Krasodomski ed., 2024), https://www.chathamhouse.org/2024/06/artificial-intelligence-and-challenge-global-governance/02-cern-ai-what-might-international (last visited Sept. 27, 2024).

65. Laura Bruun, *Reinventing the Wheel: The AWS Debate and Existing Arms Control Agreements*, Humanitarian L. & Pol'y Blog (Jun. 13, 2024), https://blogs.icrc.org/law-and-policy/2024/06/13/reinventing-the-wheel-three-lessons-that-the-aws-debate-can-learn-from-existing-arms-control-agreements/ (This source highlights the success of the Biological Weapons Convention (BWC) as a precedent for targeted prohibitions. The BWC focuses on specific types of weapons rather than broad categories, and the lessons learned from its implementation show that limiting particular weapon systems can be a productive first step toward larger regulatory frameworks. It also notes the practical challenges of enforcing such agreements but underscores their value as confidence-building measures within international law.).

capabilities, regular human review of AI decision-making processes, and clear doctrinal guidelines on the limits of AI autonomy in different operational contexts.

Conclusions

The governance of AI weapons presents one of the most complex challenges facing international law today. While AI has the potential to revolutionize military operations, it also raises profound legal, ethical, and security concerns that existing frameworks cannot adequately address. The political realities of great power competition and the technical complexities of AI systems make a comprehensive treaty on military AI unlikely in the short term. However, history demonstrates that progress is still possible, even in difficult geopolitical environments. By embracing incremental progress, targeted prohibitions, and the development of non-binding standards, the international community can take meaningful steps towards mitigating the risks of AI-driven warfare.

The role of domestic regulation should also not be overlooked. While international consensus may be elusive, nations can still take significant steps to regulate AI within their borders. Strong national laws, ethical review boards, and democratic oversight of military AI programs can ensure that the deployment of AI weapons aligns with international law and ethical standards. As AI continues to transform warfare, the need for robust governance frameworks—both domestic and international—will only grow more urgent. By recognizing the potential for both incremental international agreements and proactive domestic regulations, the path forward for AI governance becomes clearer and more achievable.