The End of the Loop: An Argument for Meaningful Human Control of Artificial Intelligence by the Department of Defense

Major Jason R. DiNapoli*

Any ordnance or supply officer above field grade will tell you that the mathematics of modern war is far beyond the fumbling minds of mere human beings. The bigger the war, the bigger the computing machines needed.[1]

For our struggle is not with flesh and blood[.][2]

I. Introduction

The accelerated development of large-language models ("LLMs")—and, more broadly, increasingly general-purpose artificial intelligence ("AI")—raises complex and weighty questions in national security law. Is it ever acceptable, for instance, to automate lethal action?[3] How do we best regulate AI internationally?[4] Within our constitutional structure, how do we govern inherently opaque algorithms employed in an inherently secretive environment?[5]

In the answers to these questions, we find "the loop": a concept generally representative of a decision-making process.[6] It is not uncommon for legislators,[7]

---

* Judge Advocate in the United States Army currently pursuing an LL.M. at Harvard Law School. For helpful comments, I am grateful to Ashley Deeks of the University of Virginia and Fabiani Duarte at The Judge Advocate General's Legal Center and School. For feedback, insight, and support throughout the writing process, I am grateful to my wife, Kaitlin.

1. Kurt Vonnegut, *EPICAC*, Collier's Wkly., Nov. 25, 1950, at 36.

2. *Ephesians* 6:12.

3. *See, e.g.*, Rebecca Crootof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 Cardozo L. Rev. 1837 (2015) (critically examining autonomous weapon systems and arguing for intentional international regulation).

4. *See* Charlie Trumbull, *Considering a Legally Binding Instrument on Autonomous Weapons*, Lawfare (Oct. 20, 2024), https://perma.cc/7NVN-TXWT (discussing three approaches to a legally binding international instrument).

5. *See generally* Ashley Deeks, The Double Black Box: National Security, Artificial Intelligence, and the Struggle for Democratic Accountability (2025) (identifying problems created by combining government secrecy with AI opacity and advancing normative proposals to address such problems).

6. *See* Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 Vand. L. Rev. 429, 434 n.11 (2023) (using "the loop" interchangeably with an "algorithmic decisionmaking process").

7. *See* Caroline Nihill, *House AI Task Force Releases Final Report, Eyes Future Work with Trump Administration*, Fedscoop (Dec. 17, 2024), https://perma.cc/H6SX-GL83 (Rep. Bill Foster, D-Ill. "argued for making sure that AI developments and decisions are being made with 'a very strong human in the loop'").

technology companies,[8] and scholars[9] to refer to a "human in the loop" or a "human on the loop" as a framework to discuss AI governance.[10] The White House even uses the term in U.S. nuclear policy, prohibiting executive agencies from using AI intended to "[r]emove a human 'in the loop' for actions critical to informing and executing decisions by the President to initiate or terminate nuclear weapons employment."[11]

The concept undoubtedly has appeal. It expresses, in a simplified way, the relation of human activity to increasingly automated processes. "Human in the loop" generally conveys a degree of direct human involvement in a process.[12] "Human on the loop" generally indicates human oversight of an otherwise automated process.[13] And "human off the loop" (or "out of the loop") means humans are altogether absent from the process.[14] In a way, finding our place in or on the loop assures us of sufficient human involvement in AI-enabled activities and provides us some comfort that, in the end, humans are accountable.[15]

But the loop concept, helpful in discussions of simpler, linear systems, has outlived its usefulness as systems have become more complex.[16] We no longer

8. *See* Amanda McGrath & Alexandra Jonker, *What Is AI Safety?*, IBM (Nov. 15, 2024), https://perma.cc/4PS7-4KZS ("Human-in-the-loop approaches help ensure that an actual person is accountable for the actions of an AI system.").

9. *See, e.g.*, Bill Boothby, *AI Warfare and the Law*, 104 INT'L L. STUD. 1, 9 (2025) (discussing a "human operator in or on the loop"); *see also* Brian L. Cox, *Why Binding Limitations on Autonomous Weapons Will Remain Elusive*, LIEBER INST. (Jan. 27, 2022), https://perma.cc/U223-3MFK ("This system is capable of performing these functions with a human 'in' or 'on' or 'out of' the loop, depending on the preference of the operator.").

10. As an initial matter, it is necessary to recognize the distinction between AI and autonomy. The concepts are intertwined: the latter is a capability or characteristic that the former, AI, enables. *See* Boothby, *supra* note 9, at 5. Most of this article uses the terms somewhat interchangeably—for instance, an "AI-enabled" system may also be termed an autonomous or semi-autonomous system. Calling a system "AI-enabled" also avoids the challenge of assessing the degree of autonomy involved; in most cases, acknowledging the system involves some AI function serves our purposes.

11. THE WHITE HOUSE, FRAMEWORK TO ADVANCE AI GOVERNANCE AND RISK MANAGEMENT IN NATIONAL SECURITY 3 (2024) [hereinafter FRAMEWORK].

12. *See* Bonnie Docherty, Julia Fitzpatrick & Trevor Keck, *Losing Humanity: The Case Against Killer Robots*, HUM. RTS. WATCH (Nov. 29, 2012), https://perma.cc/6KFT-WEMW (defining "Human-*in*-the-Loop Weapons" as "[r]obots that can select targets and deliver force only with a human command").

13. *See id.* (defining "Human-*on*-the-Loop Weapons" as "[r]obots that can select targets and deliver force under the oversight of a human operator who can override the robots' actions").

14. *See id.* (defining "Human-*out-of*-the-Loop Weapons" as "[r]obots that are capable of selecting targets and delivering force without any human input or interaction").

15. *See generally* McGrath & Jonker, *supra* note 8 ("Human-in-the-loop approaches help ensure that an actual person is accountable for the actions of an AI system.").

16. *See generally* T. B. Sheridan, *Function Allocation: Algorithm, Alchemy or Apostasy?*, 52 INT'L J. HUM.-COMPUT. STUD. 203 (2000) (succinctly describing the evolution of automated systems: "automation has moved from open-loop mechanization of the industrial revolution, then to simple closed-loop linear control, then to non-linear and adaptive control, and recently to a mix of crisp and fuzzy rule-based decision, neural nets and genetic algorithms and other mechanisms that truly recognize patterns and learn."); see also CHARLES PERROW, NORMAL ACCIDENTS: LIVING WITH HIGH-RISK TECHNOLOGIES 75–79 (1965) (differentiating between complex and linear interactions within systems).

employ computational machines simply for discrete tasks;[17] we increasingly rely on automated systems, both at work and at home.[18] The trend toward greater automation is not new: "[t]he history of automatic control has consisted of the gradual taking-over of more and more . . . human functions."[19] When viewed in this context, the crux of the matter is not *where* a human is in relation to a decision; the question is *how* and *to what extent* humans retain control over a process or a system. Those questions figure acutely in the national security sphere.[20]

This paper advances two different, but interrelated, positions. First, the Department of Defense ("DoD") should stop evaluating AI by the human-in-the-loop (or on-the-loop) paradigm. Second, an argument concerning AI regulation: as it relates to target development, selection, and engagement, the lodestar of the DoD's AI policy should be meaningful human control, established through various means and at different stages of AI development and employment. Although the United States has not expressly embraced a universal AI standard in targeting, such a standard is consistent with current U.S. practice: the DoD effectively applies a meaningful human control standard, albeit in a disjointed form.

Part II of this paper identifies issues with the loop paradigm and argues against its use. The loop concept, despite its ubiquity, is confusing and unhelpful. Appealing to the "loop taxonomy" without deliberation "leads to ambiguity and confusion about what autonomy is and how the [DoD] intends to employ it in combat."[21]

Part III advocates for the adoption of meaningful human control as a standard to govern the DoD's use of AI in targeting. In doing so, Part III reframes the discussion about loops to a discussion about hybrid human-machine systems, explains the regulatory gap concerning AI in target development, and offers a baseline test for meaningful human control.

II. DISPOSING OF THE LOOP

A. Rise of the Loop

The loop is not a new concept. In 1965, M.J. Pedelty, on behalf of NASA, completed *A Review of the Field of Artificial Intelligence and Its Possible*

17. *See* Meg Leta Jones, *The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles*, 18 VAND. J. ENT. & TECH. L. 77, 84 (2020) ("Older automation was not mobile, held minimal purpose-specific sensors, and operated with limited processing power, but ubiquitous computing means ubiquitous automation of many functions of many tasks.").

18. *See id*. at 85 ("We have automated decisions about whether to delete emails, who to date, who to hire, what movies to watch, what search terms to enter, what to eat, where to drive, where to shop, when to sleep, when to send birthday greetings, and how to go through our days."); *see also* Crootof et al., *supra* note 6, at 432 ("Artificially intelligent algorithms are being integrated into decisionmaking processes at mind-boggling speed and scale.").

19. M. J. PEDELTY, A REVIEW OF THE FIELD OF ARTIFICIAL INTELLIGENCE AND ITS POSSIBLE APPLICATIONS TO NASA OBJECTIVES 1 (1965).

20. Boothby, *supra* note 9, at 3–4 (distinguishing between the use of AI in many civilian contexts and the use of AI "to determine who is to be killed or what is to be destroyed during an armed conflict").

21. Joseph O. Chapa, *Please Stop Saying "Human-In-The-Loop"*, INST. FOR FUTURE CONFLICT (Sep. 3, 2024), https://perma.cc/D2HP-ED5Z.

Applications to NASA Objectives.[22] He roughly divided control systems into two fields: "closed-loop systems without a human in the loop . . . and those systems containing a human in the loop." For many years, such a model satisfactorily described human relationships with machine systems.

Over time, the loop became a rhetorical device for government officials to explain national security policy.[23] The DoD, in a press release in 2021, explained that for small, unmanned aircraft systems, "[the] rules of engagement still require a human in the loop."[24] Colonel Marc Pelini, Division Chief for Capabilities and Requirements within the Joint Counter-Unmanned Aircraft Systems Office, told reporters at the time, "[W]e don't have the authority to have a human out of the loop."[25] The Air Force recently paired humans and AI systems in an "accelerated kill chain" for an exercise designed "to measure how the machine could support, not replace, the human in the decision loop."[26] Officials described it as "a clear example of what happens when we put warfighters in the loop early—at the point where concepts meet code."[27]

The DoD is not the only arm of the U.S. government focused on the loop. Lakshmi Raman, the Director of the Office of Artificial Intelligence at the Central Intelligence Agency ("CIA") recently framed the agency's approach to AI around the loop: "We understand when it's important to have a human in the loop to ensure that our AI aligns with our mission, our laws and our intelligence community standards."[28]

Nor is the loop uniquely American. Our allies, too, have used the term to describe the human-machine relationship in the national security context. British Air Chief Marshal Brian Burridge, commenting on unmanned aerial vehicles ("UAVs") in 2008, stated:

> Under the law of armed conflict, there remains the requirement to assess proportionality and within this, there is an expectation that the human at the end of the delivery chain makes the last assessment by evaluating the situation using rational judgement . . . . [W]e cannot take the human, the commander, the analyst, those who wrestle with ambiguity, out of the loop. The debate about the human-in-the-loop goes wider than that.[29]

22. *See generally* PEDELTY, *supra* note 19.

23. C. Todd Lopez, *Defense Official Discusses Unmanned Aircraft Systems, Human Decision-Making, AI*, U.S. DEP'T OF WAR (Feb. 3, 2021), https://perma.cc/7KKP-D8ST.

24. *Id*.

25. *Id*.

26. Deb Henley, *Air Force Battle Lab Advances the Kill Chain with AI, C2 Innovation*, U.S. AIR FORCE (July 11, 2025), https://perma.cc/9JZC-DYW7.

27. *Id*.

28. *AI Is a Strategic Necessity in Modern Intelligence Work, National Security*, VANDERBILT UNIV. INST. OF NAT'L SEC. (May 9, 2025), https://perma.cc/6APD-U2YR.

29. Brian Burridge, *UAVs and the Dawn of Post-Modern Warfare: A Perspective on Recent Operations*, 148 RUSI J. 18, 23 (2003).

Over time, the DoD shifted its approach to human involvement and oversight of machine functions, casting doubt on humans' permanent place in the loop. Dr. Craig Martell, the Chief Digital and Artificial Intelligence Officer for the DoD, provided insight into the altered DoD paradigm, again turning to the loop as a descriptor: "Will there always be a human in the loop for every autonomous decision? I don't think that's possible anymore."[30]

There are key problems with the use of the loop to describe AI governance. First, the loop itself is often a difficult construct to identify and delimit, and in an increasingly complex environment, there is a high degree of subjectivity in identifying the relevant loop in any particular decision. Second, the concept is meaningless insofar as it contains no qualitative threshold for human involvement in a decision. Third, the focus on "the loop" ignores upstream human involvement and encourages policymakers to "slap a human in [the process]."[31] Fourth and finally, the loop poorly frames the human-machine dynamic. While loops may have utility in describing solitary, isolated, and defined decision-making cycles involving machines, the idiom is neither an effective descriptive tool nor an employable limiting standard for the DoD's AI governance. We turn to the first of the loop's issues: the difficulty in defining it.

B. Trouble in Defining the Loop

As technology developed, M.J. Pedelty's binary became more complicated. Humans were no longer merely in or out of the loop. They were *on* the loop. They could *be* the loop.[32] In a February 2020 Armed Services Committee hearing, Senator Gary Peters posed a question about human involvement in military decision chains to General Terrence O'Shaughnessy, the Commander of North American Aerospace Defense Command ("NORAD") and U.S. Northern Command ("USNORTHCOM"). GEN O'Shaughnessy explained the U.S. approach with variations of the expression:

> Senator . . . what we have to get away from is what we now have as either the 'human in the loop' (or sometimes 'the human is the loop') in some of our systems to a 'human on the loop.' And what that will allow you to do is actually make those decisions at the speed of relevance because what can and should be done by machines and by AI and leveraging that will be done by that. But it will identify those key areas where we humans have to be the ones ethically, morally making those decisions. And so, I think 'human on the loop' is a concept we need to apply to leverage that capability while not preventing ourselves from operating at the speed of relevance.[33]

30. Ryan Lovelace, *Defense Department AI Chief: Keeping Humans in the Loop Not Always Possible Anymore*, WASH. TIMES (Feb. 23, 2024), https://perma.cc/J6DN-USJP.

31. Crootof et al., *supra* note 6, at 437 (explaining problems with a "'slap a human in it' approach").

32. *Id.* at 444 ("Put plainly, humans are everywhere, whether in the loop, on the loop, off the loop, or hidden from view."); *see also* Lovelace, *supra* note 30 (differentiating among human-is-the-loop, human-on-the-loop, and human-in-the-loop decisions).

33. *Hearing to Receive Testimony on United States Northern Command and United States Strategic Command in Review of the Defense Authorization Request for Fiscal Year 2021 and the Future Years*

Yet distinctions among the three categories (in-, on-, and off-the-loop systems) are less than clear. As one DoD official pointed out, a human's place in relation to the loop depends entirely on how one draws the loop.[34] Consequently, there is a high degree of subjectivity in the term's application.[35]

Civilian technologies provide fitting analogies. Automation in the civilian world exists all around us: in anti-lock brakes, automatic load-sensing washing machines, aircraft autopilot, smart thermostats, cruise control, automatic paper shredders. The list goes on. Each of those could be termed a human-in-the-loop or human-on-the-loop system, depending on how one characterizes the loop, as the following example illustrates.

Adaptive cruise control has been around since Mitsubishi introduced it in 1992.[36] At the time, the system provided a warning of oncoming objects.[37] Gradually, the level of automation increased: now many car manufacturers incorporate technology that detects oncoming obstructions and adjusts the vehicle's speed by downshifting, controlling the throttle, or braking without additional driver input.[38] It is easy to see how such a system's classification might differ based on the identification of the relevant system. If the decision-making process includes the decision to set and discontinue the cruise control or—even more broadly—the physical steering of the wheel and operation of the car, then the driver acts *in* the loop. But if one takes a narrower view and characterizes the relevant process instead as one that manages the speed and traveling distance of the vehicle, then the driver who monitors the adaptive cruise control is a human *on* the loop.

In the military, similar difficulties exist in defining the scope of an automated military system's "loop" and whether a human is in, on, or off it. Currently, the Air Force is developing the Collaborative Combat Aircraft ("CCA"), also known as the "loyal wingman."[39] Potential mission sets are myriad. They include air-to-air combat; air-to-ground combat; electronic warfare; intelligence, surveillance, and reconnaissance; and targeting.[40] The aim is for the unmanned aircraft's AI-driven software "[to] enable collaboration with, and take direction from, human pilots and . . . expand the fighter fleet and protect human pilots at a lower cost than current fighter jets."[41] Air Force Secretary Frank Kendall described the

---

Defense Program Before the S. Comm. on Armed Servs., 116th Cong. 58 (2020) (statement of witness Gen. Terrence O'Shaughnessy).

34. Chapa, *supra* note 21.

35. *Id.* (applying General Norman Schwarzkopf's famous "mother of all briefings" to a "swarm" of notional autonomous drones and arguing it could be considered in-the-loop, on-the-loop, or out-of-the-loop).

36. Hearst Autos Research, *What is Adaptive Cruise Control?*, CAR & DRIVER, https://perma.cc/L4SJ-UA7V.

37. *Id.*

38. *Id.*

39. *See generally* JENNIFER DIMASCIO, CONG. RSCH. SERV., IF12740, U.S. AIR FORCE COLLABORATIVE COMBAT AIRCRAFT (CCA) (2025).

40. *See id.*

41. *See id.*

concept in 2022: "The idea is the manned aircraft is essentially calling plays, and he's using these other unmanned combat aircraft basically as a formation, to do things that make sense tactically."[42]

Yet, as with adaptive cruise control, the question becomes how to define the loop. Does "play-calling" by the manned aircraft place a human in the loop?[43] What if the unmanned aircraft temporarily deviates from its mission to address an imminent threat to the manned aircraft? It depends largely on how we draw the relevant loop.[44] And it is easy to see how the loop analogy becomes muddled as we incorporate AI into complex environments involving many layers of decisions.

To further complicate the matter, the AI paradigm is changing.[45] As the October 2024 National Security Memorandum ("NSM") on AI recognized, the trend is toward general-purpose AI tools.[46] Although, by most accounts, we have not achieved artificial general intelligence ("AGI"), where computers attain a level of intelligence equal to that of a human being, current AI tools, categorized as artificial narrow intelligence ("ANI"),[47] are becoming increasingly "agentic,"

---

42. *See* Stephen Losey, *How Autonomous Wingmen Will Help Fighter Pilots in the Next War*, DEF. NEWS (Feb. 15, 2022), https://perma.cc/YL3E-5R73.

43. To use Secretary Kendall's play-calling metaphor: would we consider an offensive coordinator in football *in* a play simply because the coordinator initially calls the play? Likely not. The order from the manned aircraft appears, at the least, to be tangential to the machine's action, so it may be tempting to consider the play-caller on the loop. But disagreement even exists here. *See* Jill Aitoro, *AI Warfare Is Coming, and Some Global Leaders Say NATO Isn't Ready*, DEF. NEWS (Feb. 15, 2018), https://perma.cc/BHD3-9P9X (quoting Anders Fogh Rasmussen, former NATO secretary general, who considers a human giving an initial order as "out of the loop").

44. Hon. James Baker argues in *The Centaur's Dilemma* that a human is always in the loop: "With AI, a human is always in the loop in some manner, for a human always writes the initial code and chooses to deploy an AI-enabled machine or capacity." JAMES BAKER, THE CENTAUR'S DILEMMA 28 (2020). Despite its framing on the loop paradigm, the argument is not far removed from the one in Part III(c), *infra*, which argues that humans may exert some degree of control over AI tools in the development and testing stage of AI.

45. *See* National Security Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence, 2024 DAILY COMP. PRES. DOC. 2, § 1(d) (Oct. 24, 2024) [hereinafter NSM] ("Recent innovations have spurred not only an increase in AI use throughout society, but also a paradigm shift within the AI field . . . ."); *see also* Bernard Marr, *Agentic AI: The Next Big Breakthrough That's Transforming Business and Technology*, FORBES (Sep. 6, 2024), https://perma.cc/6E3S-8FZ2 ("agentic AI . . . is not just another industry buzzword; it's a paradigm shift that's poised to redefine the boundaries of AI capabilities.").

46. NSM, *supra* note 45, at 2, § 1(g). Those changes, in turn, spur change within the national security structure. *Id*. ("Establishing national security leadership in AI will require making deliberate and meaningful changes to aspects of the United States Government's strategies, capabilities, infrastructure, governance, and organization.").

47. *See* Alexandra Jonker & Amanda McGrath, *What Is Superalignment?*, IBM (Nov. 12, 2024), https://perma.cc/KDM2-4EY3 (explaining the differences among ANI, AGI, and artificial superintelligence ("ASI")). While there is disagreement about whether we have achieved (or will soon achieve) AGI, many experts believe that moment has not yet arrived. *See* Sam Altman, *Three Observations*, SAM ALTMAN (Feb. 9, 2025), https://perma.cc/UB8J-KHHD; Billy Perrigo, *Meta's AI Chief Yann LeCun on AGI, Open-Source, and AI Risk*, TIME (Feb. 13, 2024), https://perma.cc/ZRV5-2TDC.

going beyond traditional AI and operating with increased autonomy.[48] Agentic AI, "[i]nstead of just answering[] questions or performing simple tasks . . . can take initiative, solve complex problems, and adapt its approach based on changing circumstances."[49] Operating with greater independence, agentic AI is akin to "a digital assistant on steroids."[50]

It is easy to anticipate, as the Biden Administration forecasted in the NSM, that "many functions that to date have been served by individual bespoke tools may, going forward, be better fulfilled by systems that, at least in part, rely on a shared, multi-purpose AI capability."[51] As AI becomes more general and adaptive, it will become more difficult to adequately identify the loop of import. If we augment adaptive cruise control with other automated tasks through more generalized AI—say, automated parking or steering—what system do we evaluate in the event of an accident? Similarly, it may be easy to conceptualize the loop if our loyal wingman merely flies from Point A to Point B, but if the aircraft functions with greater agency, if its multi-purpose AI is designed to address dynamic environments, how do we view the CCA's use of AI? Does it further the discussion to say we are on or off the loop? Fundamentally, what *is* the loop?

The authors of a recent law review article, *Humans in the Loop*, attempted to nail down the loop construct, defining it in a way "[that] regulators often implicitly employ."[52] They defined "a 'human in the loop' as an individual who is involved in a single, particular decision made in conjunction with an algorithm."[53] Yet, as the authors point out, the definition often employed by regulators is both "narrow" and "misleading."[54] It conflates the in- and on-the-loop concepts, redefining "in the loop" beyond its putative meaning.[55] Moreover, because the definition focuses on only one particular decision, it "obscures the fact that humans are everywhere in automated decision-making systems."[56]

The struggle to clearly define the loop exhibits the problem with using it as a categorical term: the loop confuses more than it clarifies.[57] Although its application to automated systems predates the Observe-Orient-Decide-Act ("OODA")

48. Marr, *supra* note 45.
49. *Id.*
50. *Id.*
51. NSM, *supra* note 45, at 2, § 1(g).
52. Crootof et al., *supra* note 6, at 440.
53. *Id.*
54. *Id.* at 441–43.
55. *Id.* at 441 n.37 ("As we define it . . . the 'human in the loop' includes folks that might otherwise be considered a 'human on the loop.'").
56. *Id.* at 443.
57. Over a decade ago, two law clerks observed in a law review article, "Today's debate about humans, autonomy, and 'the loop' relies on language too imprecise to draw out and analyze the relevant differences between drones and predecessor technologies. What does it mean for a human to be 'inside,' 'outside,' or 'on' the loop?" William C. Marra & Sonia K. McNeil, *Understanding "The Loop": Regulating the Next Generation of War Machines*, 36 HARV. J. L. & PUB. POL'Y 1139, 1142 (2013). Despite the time that has passed, the language has not gotten clearer.

loop (a concept ubiquitous in the military), many now equate the two.[58] And as an inherently muddled concept, the loop neither describes military processes nor provides a helpful framework through which to establish norms and standards. It exists nowhere in DoD policy.[59] Yet, in continuing to use the term, "we obscure the importance of human judgment and its connection to the laws of war and just war principles."[60]

C. When Humans are Rubber Stamps

The second major issue with the loop is that the concept contains no inherent, qualitative threshold for human involvement. It fails to distinguish a human capable of understanding inputs and shaping outputs from a human automaton who presses the "fire" button every time a red light comes on.[61] For a human to be in or on the loop, it matters not whether that human is awake or asleep at the switch; it only matters that a human is present in a decision cycle.

In their review of "the law of the loop," the authors of *Humans in the Loop* identify nine roles humans play in the loop, "highlight[ing] the variety of (admirable and distasteful) reasons policymakers might wish to encourage human involvement in algorithmic decisionmaking processes":

> Humans may play (1) corrective roles to improve system performance, including error, situational, and bias correction; (2) resilience roles to act as a failure mode or alternatively stop the whole system from working under an emergency; (3) justificatory roles to increase the system's legitimacy by providing reasoning for decisions; (4) dignitary roles to protect the dignity of the humans affected by the decision; (5) accountability roles to allocate liability or censure; (6) "stand-in" roles to act as proof that something has been done or stand in for other humans and human values; (7) friction roles to slow the pace of automated decisionmaking; (8) "warm body" roles to preserve human jobs; and (9) interface roles to link the systems to human users.[62]

---

58. *Id.* at 1144 ("When commentators debate whether drone warfare leaves humans 'in the loop,' 'out of the loop,' or simply 'on the loop,' it is the OODA Loop they are talking about."); *see also* Ameya Kelkar, *Artificial Intelligence in Military Decision Making—A Case for the Human to Stay in the OODA Loop*, CENT. FOR LAND WARFARE STUD. (Jan. 6, 2024), https://perma.cc/V846-ULB5 (stressing the need "to keep the human element in all essential parts of the OODA (Observe, Orient, Decide and Act) Loop"); *see also* Gary Olson, *Keeping the Human in the OODA Loop*, FED. TIMES (Oct. 31, 2022), https://perma.cc/YA2G-6Y6P (arguing that, "[d]espite the information overload, keeping the human in the loop is an essential element in analysis and decision making, which is where the . . . OODA Loop comes in"). The Observe-Orient-Decide-Act ("OODA") loop was the vision of John Boyd, an F-86 Sabre pilot. *See* Dan Grazier, *Why the OODA Loop is Forever*, TASK & PURPOSE (Sep. 13, 2019), https://perma.cc/7WR9-D8NX (arguing the OODA loop "is not synonymous with decision-making"; rather "OODA is an action cycle").

59. Chapa, *supra* note 21.

60. *Id.*

61. ARTICLE 36, KEY AREAS FOR DEBATE ON AUTONOMOUS WEAPONS SYSTEMS 2 (May 2014), https://perma.cc/C3GS-D8LN (arguing that a human who "simply pressed a 'fire button' every time a light came on without having any other information" fails to exercise meaningful control over that system).

62. Crootof et al., *supra* note 6, at 473–74.

Those varied and disparate roles—some active, others passive—illustrate the ineffectiveness of the concept as a governing tool. If a human acting in any of those varied and distinct roles qualifies as a "human in the loop," then the loop has little import as a governing principle. A human acting as a "warm body" is a human in the loop to the same extent as a human acting in a corrective role. There must be some qualitative component to the human's role; it is not enough to say a human is present.

Take, for example, the Army's Counter Rocket, Artillery, and Mortar ("C-RAM") system. The C-RAM system automatically detects, tracks, and engages incoming rockets, artillery, and mortar rounds.[63] Although a human operator certifies the target,[64] the window is typically small before the weapon automatically shoots hundreds of 20-millimeter rounds at the incoming targets. With such a small window—and the system's ability to operate independently of human input—the question becomes whether a human truly is "in the loop." A former Centurion C-RAM's operator elucidated: "A [h]uman operator just presses the button that gives the authorization to the weapon to track, target, and destroy the incoming projectile . . . . [T]his kind of limited human involvement in the process starts to reach the limits of what humans can react to in a short time span."[65] In those circumstances, some have argued humans are not actually in the loop; they are out of the loop.[66]

Peter Asaro, co-founder of the International Committee for Robot Arms Control ("ICRAC") and a professor at the New School, contends that for any lethal decision process to be legitimate, it must hold a trained, well-informed human accountable and responsible, and it must afford him sufficient time to deliberate and decide to act.[67] If humans have inadequate (or at least limited) opportunity to exercise judgment over an AI-enabled decision, their involvement in the process is ineffective and merely perfunctory.[68] In those cases, terming humans' involvement as in-the-loop or on-the-loop is not just confusing; to the extent the concept implies meaningful involvement in the process, it is misleading. The absence of a qualitative threshold makes the loop a deceptive metaphor.

---

63. *See* Caitlin Kenney, *C-RAM Transforms Defense Tactics*, U.S. ARMY (Apr. 26, 2012), https://perma.cc/J9RJ-9M8U.

64. *See id.*

65. *See* Ray Kurzweil, *The Proposed Ban on Offensive Autonomous Weapons Is Unrealistic and Dangerous*, KURZWEIL LIBR. (Aug. 5, 2015), https://perma.cc/XCT5-Z3JU.

66. *See* Docherty et al., *supra* note 12 (arguing that human-on-the-loop processes with limited human supervision are more properly characterized as out-of-the-loop); *see also* Bradley Peniston, *Not By Widgets Alone*, ARMED FORCES J. (Feb. 1, 2011), https://perma.cc/LTE2-FEJU (arguing that automation bias in U.S. Patriot air and missile defense units effective cedes control responsibility from the human operator to the machine).

67. Peter Asaro, *On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making*, 94 INT'L REV. RED CROSS 687, 695 (2012).

68. Jeffrey S. Thurnher, *No One at the Controls: Legal Implications of Fully Autonomous Targeting*, 67 JOINT FORCES Q. 77, 83 (2012) ("oversight would not be effective if the human operator were merely a rubber stamp to approve an engagement").

D. Ignoring Upstream Human Involvement

Another issue with the loop is that in focusing on the employment of AI, it deemphasizes upstream human involvement during the design and testing phases of the technology. There is a tendency to blame the human who operates or employs a human-machine system, a bias Madeleine Elish terms a "moral crumple zone."[69] Yet, "[f]ocusing on human-in-the-loop error can obscure bigger system design problems, such as how the system's interface fosters confusion and how discrete errors can cascade."[70] The USS *John S. McCain* accident is a case in point.[71]

On August 21, 2017, the U.S. Navy destroyer *John S. McCain* crashed into the Liberian-flagged oil tanker *Alnic MC* while both vessels transited the Singapore Strait, one of the busiest channels in the world.[72] Ten U.S. Sailors died in the accident, forty-eight others were injured, and the vessel sustained more than $100 million in damage.[73]

The Navy investigation conducted in the aftermath pointed to the humans in the loop—*i.e.*, the Sailors involved in the operation of the ship—as the cause of the accident.[74] The collision "was caused by a loss of situational awareness due to mistakes in operating the ship's steering and propulsion system while in a high density of maritime traffic, [the ship's operators] failed to follow the International Nautical Rules of the Road, and watchstanders had insufficient proficiency and knowledge of their systems."[75] The report concluded the collision "resulted primarily from complacency, over-confidence and lack of procedural compliance."[76] In other words, the cause was human error on the part of the crew. The Navy later charged the ship's captain with homicide and a chief petty officer, responsible for training on the ship's navigation system, with dereliction of duty.[77]

But further investigation revealed another cause of the collision: flaws in the ship's Integrated Bridge and Navigation System ("IBNS"), the touch-screen system that controlled the USS *McCain*'s steering and thrust.[78] An investigation conducted by the National Transportation Safety Board ("NTSB") found the design of the IBNS, "a relatively new system" on the USS *McCain*, "increased the likelihood of the operator errors that led to the collision." [79] As the USS *McCain*

69. Madeleine Clare Elish, *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction*, 5 ENGAGING SCI., TECH. & SOC'Y 40, 42 (2019).

70. Crootof et al., *supra* note 6, at 469–71.

71. *Id.* at 471–72 (arguing the accident was indicative of the "tendency to blame the human in the loop").

72. NAT'L TRANS. SAFETY BD., MARINE ACCIDENT REPORT: COLLISION BETWEEN US NAVY DESTROYER *JOHN S MCCAIN* AND TANKER *ALNIC MC* SINGAPORE STRAIT, 5 MILES NORTHEAST OF HORSBURGH LIGHTHOUSE 1 (2017), https://perma.cc/8Y6N-5U3J.

73. *Id.* at viii.

74. *See* Rich Abott, *Navy Releases First Report on DDG Collisions, Finds Crew Failures*, DEF. DAILY (Nov. 1, 2017), https://perma.cc/CK4S-2GRE.

75. *Id.*

76. *Id.*

77. T. Christian Miller, Megan Rose, Robert Faturechi & Agnes Chang, *Collision Course*, PROPUBLICA (Dec. 20, 2019), https://perma.cc/E58Y-5HM2.

78. *See id.*

79. NAT'L TRANS. SAFETY BD., *supra* note 72, at 3, 33.

transited the strait, it initially operated in backup manual mode, a mode which "allow[ed] other stations to unilaterally take steering control."[80] And when operating in backup manual mode, the IBNS software was designed to make "'automatic offers' to other steering stations on the ship allowing them to take unilateral control."[81] After steering control inadvertently shifted from the helm to the lee helm of the ship in the early hours of August 21, 2017, the crew became confused.[82] And, in crucial minutes leading up to the crash, "[t]he steering log showed that steering control remained at the lee helm station in computer-assisted mode."[83] Although "the crew attempted to regain control of the vessel . . . the [USS *McCain*] unintentionally turned to port into the path of the *Alnic MC*."[84]

A subsequent Navy investigation acknowledged "[the] tendency of designers to add automation based on economic benefits (e.g., reducing manning, consolidating discrete controls, using networked systems to manage obsolescence), without considering the effect to operators who are trained and proficient in operating legacy equipment."[85] Yet those involved in "pre-loop" activity (development, training, implementation, etc.) have largely gone unscathed: as one outlet reported, "[N]o one responsible for the development or deployment of the technology has faced any known consequences[.]"[86]

When we consider AI in terms of "the loop," we orient ourselves to the operation and employment of AI. Our discussion of AI would do better to avoid the loop and engage in more comprehensive analysis.

E. An Inverted Relationship

The last problem with the DoD's use of the loop is perhaps the most fundamental: it poorly frames the human-machine dynamic and misrepresents the character of war.[87] It orients us to a machine-focused paradigm in which we, humans, position ourselves relative to the machine upon which we rely.

But war is inherently human. *Humanity* is central. As Peter Asaro identifies, the critical moral question is "whether a computer, machine or automated process ought to make these decisions of life and death at all."[88] Kathleen Lawand, the head of the arms unit for the International Committee of the Red Cross ("ICRC"), is unequivocal: "Humans cannot delegate the decision to use force and violence to machines. Decisions to kill, injure and destroy must remain with humans. It is humans who apply the law and are obliged to respect it."[89]

---

80. *See id*. at 26.
81. Miller et al., *supra* note 77.
82. NAT'L TRANS. SAFETY BD., *supra* note 72, at 25.
83. *Id*. at 13.
84. *Id*. at ii.
85. Miller et al., *supra* note 77.
86. *Id*.
87. Chapa, *supra* note 21.
88. Asaro, *supra* note 67, at 699.
89. News Release, Int'l Comm. of the Red Cross, Autonomous Weapons: States Must Agree on What Human Control Means in Practice (Nov. 20, 2018), https://perma.cc/8X9A-TFYR [hereinafter ICRC News Release].

In warfare, AI must forever be a national-security tool; it must not become the carpenter.[90] We need a standard that adequately assures subordination of the machines to human command. That standard is meaningful human control.

III. IN SUPPORT OF MEANINGFUL HUMAN CONTROL

A. Framing the Discussion

Part II's argument against the loop presents a problem at first blush. If not oriented on the loop, how should we think of AI? After all, the loop is more than a governing principle; it is—to its supporters, at least—a tool that enables regulation by framing governance of autonomous and AI-enabled systems.[91] If we dispose of the loop taxonomy altogether, how do we view the human-machine relationship and its governance in national security?[92]

Part III answers that question by, at the outset, reframing the discussion of AI around hybrid, human-machine systems rather than loops. The paper then examines meaningful human control in the context of the targeting process.

Accordingly, the scope of the paper shifts from, in Part II, the general application of AI in national security to, in Part III, a crucial subset of the DoD's AI use: AI in targeting. There are two principal reasons for the focus on targeting. First, it is in targeting that the DoD's employment of AI is truly unique—and ethically and legally complicated. The DoD has ample opportunity to apply AI to various, more mundane aspects of its operations, among them, human resources, legal, finance, and public affairs. In these, the DoD's use of AI may hardly differ from, say, the Department of the Treasury's use of AI or any other executive agency's use of the technology. AI in targeting presents different considerations.

Second, regulation of AI in targeting acts as a microcosm—or even an upper bound—for the regulation of AI in military activities across the DoD.[93] Targeting, where consequences may be lethal, demands the greatest attention and restraint.

We begin with a discussion of human-machine systems.

1. Viewing AI Technology as a Component of Hybrid Systems

Rather than the loop, a more apt construct is to view human-machine interactions within the DoD as part of hybrid, human-machine systems—webs of

90. *See* Victoria Cardiel, *Pope Leo XIV on AI: 'All of Us Are Concerned for Children and Young People'*, CATH. NEWS AGENCY (Jun. 20, 2025), https://perma.cc/4FNL-WPXF (quoting Pope Leo XIV as framing AI as "an exceptional product of human genius, [but it is] above all else a tool").

91. *See* discussion *supra* note 57.

92. This is not to say it has no utility in other contexts, whether in or out of the military. *See* Nicholas Freund, *Using Data to Leverage the OODA Loop*, FORBES (Jan. 10, 2023), https://perma.cc/NZ3X-4XXR.

93. The arguments here apply equally to missile defense, for instance. On the other hand, the paper makes no argument about whether the DoD should require a certain degree of control over other potential applications of AI, including in logistics, training, administrative matters, etc. Those uses, on their face, appear less problematic, and in the parlance of the FRAMEWORK, *supra* note 11, are not "high impact" uses.

factors, inputs, outputs, decisions, weights, and relationships. The loop evokes the image of a single, isolated decision-making cycle.[94] Such a framing may well represent a fighter pilot's decision cycle,[95] but it misrepresents the complexity of the national security ecosystem and the dispersed nature of decision-making within it, where human-machine systems intermesh and where decision-making is broadly distributed.[96] Anna Greipl, a researcher at the Geneva Academy, points out, "[H]umans are *already dependent on and intertwined with* AI systems across a wide range of military decision-making processes."[97] Professors Rebecca Crootof, Margot Kaminski, and W. Nicholson Price assert, "[G]overnance should regulate the [hybrid] system as a whole"; focusing narrowly on the loop "frequently lead[s] to failure."[98]

The key, as Part III addresses, is *how* and *to what extent* humans retain control over those systems. The loop tells us a human's figurative position relative to a single decision cycle. Meaningful human control over hybrid, human-machine systems, on the other hand, provides an adaptable, normative framework capable of governing AI within the DoD.

This framing, however, invites more questions—one door leads to another—namely, why the focus on autonomy in target development, selection, and engagement? Is that approach the equivalent of regulating autonomous weapon systems?

2. Governing Autonomous Functions in the Targeting Process

Regulation of AI in targeting differs from regulation of autonomous weapon systems. Principally, regulation of AI in targeting encompasses key functions in the targeting cycle: development, selection, and engagement of targets. Conversely, regulation of autonomous weapon systems governs, as one might

94. Advocates of using John Boyd's OODA loop as a regulatory principle acknowledge that terming humans as "'in the loop' or 'out of the loop,' full stop, would be too simplistic. Machine systems do not operate in a single loop; rather, they cycle through the loop hundreds or thousands of times in a single mission." William Marra & Sonia McNeil, *Understanding "The Loop": Humans and the Next Drone Generations*, ISSUES IN GOVERNANCE STUD., Aug. 2012, at 1, 6 (2012), https://perma.cc/BA4C-DMP6. *See also* Sheridan, *supra* note 16, at 65 (describing systems in four layers of complexity); Crootof et al., *supra* note 6, at 439 (arguing for regulation of hybrid systems using lessons drawn from engineering).

95. *See* discussion *supra* note 58 (discussing John Boyd's OODA loop).

96. Merel Ekelhof, *Responsible AI Symposium—Translating AI Ethical Principles into Practice: The U.S. DOD Approach to Responsible AI*, LIEBER INST. (Nov. 23, 2022), https://perma.cc/ECP7-AM6X. It also oversimplifies AI, which is itself a "not a single piece of hardware or software, but rather a constellation of technologies that depend on interrelated elements." NAT'L SEC. COMM'N ON ARTIFICIAL INTEL., FINAL REPORT 31 (2021), https://perma.cc/868N-HVT2.

97. *See* Anna Rosalie Greipl, *Artificial Intelligence Systems and Humans in Military Decision-making: Not Better or Worse but Better Together*, LIEBER INST. (Jun. 14, 2024), https://perma.cc/6LVM-666D.

98. Crootof et al., *supra* note 6, at 488. The authors adhere to the loop construct, however, terming these human-machine systems "complex human-in-the-loop systems." *Id.* at 497. *See also* Marra & McNeil, *supra* note 94, at 1179 ("If we want to regulate the design and operation of increasingly autonomous technologies, we should adopt [a] more nuanced view of system operations.").

expect, autonomous (and semiautonomous) weapon systems. The DoD governs those in DoD Directive ("DoDD") 3000.09.

In DoDD 3000.09, the DoD regulates the design, testing, and use of autonomous weapon systems. An autonomous weapon system, as defined by the policy, is "[a] weapon system that, once activated, can select and engage targets without further intervention by an operator."[99] And a "semiautonomous weapon system" is "[a] weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by an operator."[100] Yet the directive expressly carves out, among others, "[a]utonomous or semi-autonomous systems that are not weapon systems" and "[u]narmed platforms."[101] In this, the policy leaves regulatory gaps, chiefly among them: the use of AI in target development.

Existing policy addressing AI in target development is scant. The DoD, like most of the international community, focuses its attention on autonomous weapon systems, provocatively dubbed "killer robots," rather than broadly considering the use of AI in targeting.[102] Nevertheless, the notion of AI in targeting captures more fully the breadth of autonomy present in the military today. As Michael Horowitz and Paul Scharre of the Center for a New American Security ("CNAS") allude to, "autonomy is already used for a wide variety of military tasks, including many related to the use of force. These include: identifying, tracking, prioritizing and cueing targets; deciding the timing of when to fire a weapon; maneuvering and homing in on targets; and detonation timing."[103] All of those tasks fall within the ambit of targeting, and many arise beyond the scope of current governance focused on autonomous weapon systems themselves.[104] Open questions remain, for instance, about whether it is permissible or advisable for AI to nominate targets.[105]

---

99. DoD Directive 3000.09, Autonomy in Weapon Systems 21 (2023) [hereinafter DoDD 3000.09].

100. *Id.* at 23. For brevity, and because the distinction between autonomous and semiautonomous weapon systems matters little for these purposes, we collectively refer to both in this paper as autonomous weapon systems.

101. *Id.* at 4.

102. Anthony King notes that "[up] to now, the literature has tended to concentrate on AI-enabled lethal autonomous weapons." But he argues "the attention to lethal autonomous weapons is exaggerated," the greater application of AI is instead in targeting, as militaries "have principally employed AI, not to automate weapons but to help process data." Anthony King, *Digital Targeting: Artificial Intelligence, Data, and Military Intelligence*, 9 J. GLOB. SEC. STUD. 1, 1 (2024).

103. *See* Paul Scharre & Michael C. Horowitz, An Introduction to Autonomy in Weapon Systems 3 (Feb. 2015) (working paper) (on file with Ctr. for a New Am. Sec.), https://perma.cc/2FR9-XPAG.

104. *See* DoDD 3000.09, *supra* note 99, at 23.

105. *Compare* Brad Boyd, *There Are Some Things We Shouldn't Do. . .*, KILLER ROBOT COCKTAIL PARTY (Feb. 19, 2024), https://perma.cc/4PCM-LW7Y (arguing "there are some things that AI just shouldn't do, and nominating targets is one of those things"), *with* John Tramazzo, *Guest Post: Is AI Nominating Targets . . . Bad?*, KILLER ROBOT COCKTAIL PARTY (Mar. 22, 2024), https://perma.cc/99WD-GXLD ("[a]llowing AI to nominate targets is not inherently illegal, unethical, or immoral.").

3. A Brief Overview of the Targeting Process

Under U.S. joint doctrine, targeting is "the process of selecting and prioritizing targets and matching the appropriate response to them, considering operational requirements and capabilities."[106] It "requires a continuous, analytic process to identify, develop, and affect targets to meet commander objectives."[107] That iterative process in joint operations is called the joint targeting cycle, which comprises six steps, beginning with receipt of the commander's objectives and guidance and ending with a combat assessment.[108] Through this process, the joint force "systematically analyzes and prioritizes entities (persons, places, or things) considered for possible engagement or action and matches appropriate capabilities to those entities to create specific effects, accounting for operational constraints and restraints and the results of previous assessments."[109] Steps two through five (target development, capabilities analysis, commander's decision, and force execution, respectively) are the heart of the targeting cycle.[110]

Of course, not every target is planned.[111] Like the joint targeting process used for planned (or deliberate) targets, a six-step process exists for unplanned (or dynamic) targets.[112] But for both deliberate and dynamic targets, an iterative process exists; and underpinning each is intelligence collection and target identification, data-heavy functions that invite the use of AI to improve speed and processing. Given that, it is perhaps easy to see the potential for states to incorporate AI into targeting on the battlefield. In fact, they already have.

4. Automation in the Targeting Cycle

Israel's platform, Habsora (or "the Gospel"), provides a current example of a state enhancing its targeting with AI.[113] Per press reports, Israel employed the Gospel, an AI-enabled platform that makes targeting recommendations to human analysts, in its operations in Gaza.[114] The Israeli Defense Forces ("IDF") briefly described the platform in a posting on its website, shortly after beginning operations in Gaza in response to the massacre on October 7, 2023:

---

106. JOINT CHIEFS OF STAFF, JOINT PUBLICATION 3-60: JOINT TARGETING vii (2013).

107. *Id.* at viii.

108. *See id.* at x–xi.

109. *Id.* at I-6.

110. *See id.* at x–xi.

111. Planned targets are, in military parlance, "deliberate" targets. Many targets are instead targets-of-opportunity or "dynamic" targets. *Id.* at ix–x. Timing primarily distinguishes deliberate targets from dynamic targets. *Id.*

112. This is the find, fix, track, target, engage, assess ("F2TEA") process. *Id.* at I-9–I-10, II-23. Because this process is used in a compressed period of time, the F2TEA process is abbreviated compared to the process used for deliberate targets.

113. *See* Geoff Brumfiel, *Israel Is Using an AI System to Find Targets in Gaza. Experts Say It's Just the Start*, NAT'L PUB. RADIO (Dec. 14, 2023), https://perma.cc/U8U3-JMBM; *see also* Harry Davies, Bethan McKernan & Dan Sabbagh, *'The Gospel': How Israel Uses AI to Select Bombing Targets in Gaza*, GUARDIAN (Dec. 1, 2023), https://perma.cc/R9PS-Q8XE.

114. Brumfiel, *supra* note 113.

> [The Gospel] is a system that enables the use of automated tools to produce targets at a rapid pace, and works by improving accurate and high-quality intelligence material according to demand. Using artificial intelligence, and through the rapid and automatic extraction of up-to-date intelligence, it creates a recommendation for the researcher, with the goal of having a full match between the machine's recommendation and the identification carried out by humans.[115]

Although there is no open-source reporting on what data, exactly, the Gospel ingests, an AI-enabled target-nomination system such as the Gospel may analyze "large sets of information from a range of sources, such as drone footage, intercepted communications, surveillance data and information drawn from monitoring the movements and [behavior] patterns of individuals and large groups."[116] And though Israel has faced some criticism for its use of AI in target nomination, the use of AI "has significantly accelerated" the targeting process.[117] According to one source, the Gospel "processes enormous amounts of data that 'tens of thousands of intelligence officers could not process,' and recommends bombing sites in real time."[118] Whereas in prior conflicts Israel's target engagements outpaced its target nominations, meaning "the IDF often ran out of quality targets," incorporation of AI into its existing targeting capabilities "helped the IDF cross the point where they can assemble new targets even faster than the rate of attacks."[119]

Three U.S. technologies tested at Project Convergence 2020, an annual demonstration of emerging DoD technology, demonstrate various applications of AI in the targeting cycle.[120] The first, TITAN is a ground station that provides targeting information by compiling and enriching data from various sources using AI and machine learning.[121] The second capability, Prometheus, ingests data from TITAN, recognizes threats on the battlefield, and generates targeting coordinates.[122] Prometheus then sends those coordinates to FIRESTORM, the third technology, which uses AI to pair targets with weapons.[123]

---

115. *See* Israel Defense Force, *A Glimpse into the IDF's Round-the-Clock Target Enterprise*, WAR DIARY (Nov. 2, 2023), https://perma.cc/BJ3D-TXLR [translated using Microsoft Edge].

116. Davies et al., *supra* note 113.

117. *Id.* In addition, critics fail to acknowledge that little attention has been given to target nomination in international discussions; the focus instead is on autonomous weapons systems.

118. *See* Yuval Abraham, *'A Mass Assassination Factory': Inside Israel's Calculated Bombing of Gaza*, +972 MAG. (Nov. 30, 2023), https://perma.cc/356J-8MDN. Although the system has improved the quantity of targets, some reports have questioned the quality of those targets and predict increased risk to civilians. *Id.*; *see also* Davies et al., *supra* note 113.

119. *See* Yonah Jeremy Bob, *IDF Bombs Whole Gaza Neighborhoods To Hit Hamas Targets-Official*, JERUSALEM POST (Oct. 11, 2023), https://perma.cc/335M-NQCZ.

120. Michael W. Meier, *Responsible AI and Legal Review of Weapons*, LIEBER INST. (Dec. 27, 2022), https://perma.cc/P3ZK-DR8J.

121. *See Titan*, PALANTIR, https://perma.cc/V9CJ-N7FU. Palantir characterizes the technology as "the Army's first AI-defined vehicle." *Id.*

122. *See* Devon Suits, *AI Software Prototype Expedites Army Sensor, Shooter Capabilities*, U.S. ARMY (Oct. 27, 2020), https://perma.cc/9WGW-5JPS.

123. Meier, *supra* note 120. The United States has used autonomous systems such as the AN/TPQ-53 Counterfire Radar System to direct artillery counterfires since at least 2010. *See* U.S. Mission Geneva,

Michael W. Meier, the Special Assistant to the Army Judge Advocate General for Law of War Matters, points out: none of the three technologies is a weapon system; thus, DoDD 3000.09 governs none of them.[124] Unless embedded in an autonomous or semiautonomous weapon, a Gospel-like target-nomination system within the DoD has no public-facing governance. And many AI capabilities in target development fall outside formal weapons reviews, such as those required by Article 36 of the 1977 Additional Protocol I ("AP I") to the 1949 Geneva Conventions.[125]

The United States, though not a party to AP I, still adheres to the protocol by conducting weapons reviews by policy.[126] It does not, however, find any obligation under international law to conduct reviews of novel "means and methods" of warfare.[127] For that reason, among others, some have argued that the weapons review process is an ineffective mechanism of ensuring ethical and responsible AI use by the DoD.[128]

The DoD should employ a standard to govern AI use across the spectrum of warfighting functions—not merely when embedded in weapons systems, but also when leveraged in target development and nomination. To be clear, autonomy in targeting functions is not necessarily a bad thing. On the contrary, the United States should and must leverage AI in these areas. But the DoD should do so responsibly and surely by emplacing clear standards that govern those functions. That is the place for meaningful human control.

B. Characterizing Meaningful Human Control

The term "meaningful human control" has its roots, as an international governing principle, in a briefing paper to the 2014 Experts Meeting on Lethal Autonomous Weapon Systems.[129] The paper posited the existence, implicit in

---

U.S. Statement at the GGE on Laws During the Discussion of Agenda Item 5(D), U.S. MISSION TO INT'L. ORGS. IN GENEVA (Aug. 5, 2021), https://perma.cc/YK4A-3ARG.

124. Meier, *supra* note 120.

125. *See* Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, art. 36 (available at https://perma.cc/93GR-WR5R). Article 36 requires: "In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party." *See* Meier, *supra* note 120 (saying "[i]f the weapon review was the mechanism [by which to review AI systems] it is easy to see that systems such as these would never get that review"). Meier raises the question as to whether "the United States should expand legal reviews to such systems that incorporate AI that inform humans' decision-making related to targeting." *Id.*

126. Although not a party to AP I, the United States is one of a small handful of countries that formally conducts legal reviews of weapons, as Article 36 of AP I requires. INT'L COMM. OF THE RED CROSS, A GUIDE TO THE LEGAL REVIEW OF NEW WEAPONS, MEANS AND METHODS OF WARFARE: MEASURES TO IMPLEMENT ARTICLE 36 OF ADDITIONAL PROTOCOL I OF 1977 5 (2006), https://perma.cc/LNT3-8YSH.

127. Meier, *supra* note 120.

128. *Id.*

129. ARTICLE 36, *supra* note 61; *see also* Rebecca Crootof, *A Meaningful Floor for "Meaningful Human Control"*, 30 TEMPLE INT'L & COMP. L.J. 53, 58 (2016).

international law and existing in current state practice, of "an expectation that human control is exercised over when, where and how weapons are used, and over their effects."[130] That expectation may erode, however, with the dawn of autonomous weapon systems, which are "best understood as being composed of disparate soft- and hardware elements that work together—including sensors, algorithmic targeting and decision-making mechanisms, and the weapon itself."[131]

States, to date, have not agreed on a definition of "meaningful human control."[132] But in the years since, various organizations have proffered opinions on its essential components, in particular the ICRAC[133] in 2014, the CNAS[134] in 2015, Article 36[135] in 2016, and the ICRC[136] in 2017. The following table identifies the elements of meaningful human control, as defined by each organization:

Proposed Definitions of Meaningful Human Control	
ICRAC	1. [A] human commander (or operator) must have full contextual and situational awareness of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack. 2. [T]here must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of attack, and likely incidental and possible accidental effects of the attack. 3. [T]here must be a means for the rapid suspension or abortion of the attack.
CNAS	1. Human operators are making informed, conscious decisions about the use of weapons. 2. Human operators have sufficient information to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action. 3. The weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon.

130. ARTICLE 36, *supra* note 61, at 1.

131. *Id.* at 2. The paper broadly defines autonomous weapon systems and its like terms. *Id.* at 1 ("sometimes referred to as 'lethal autonomous robots' (LARs), 'lethal autonomous weapons systems' (LAWS) or 'killer robots', concerns around AWS are not limited to the killing of human beings, but extend to any infliction of harm by means of an AWS, including incapacitation or injury of human beings, and material damage to the human and natural environment").

132. Crootof, *supra* note 129, at 58.

133. *See* Denise Garcia, *ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting*, INT'L. COMM. FOR ROBOT ARMS CONTROL (May 14, 2014), https://perma.cc/TV26-BMJQ.

134. Paul Scharre & Michael C. Horowitz, Meaningful Human Control in Weapon Systems: A Primer 4 (Mar. 2015) (working paper) (on file with Ctr. for a New Am. Sec.), https://perma.cc/YN8W-RYG7.

135. *See generally* ARTICLE 36, *supra* note 61.

136. *Expert Meeting on Lethal Autonomous Weapons Systems*, INT'L COMM. OF THE RED CROSS (Nov. 15, 2017), https://perma.cc/S4RA-3EH6 [hereinafter *Expert Meeting*].

Continued

Proposed Definitions of Meaningful Human Control	
Article 36	1. Predictable, reliable and transparent technology. 2. Accurate information for the user on the outcome sought, the technology, and the context of use. 3. Timely human judgement and action, and a potential for timely intervention. 4. Accountability to a certain standard.
ICRC	1. Predictability. 2. [H]uman supervision and ability to intervene. 3. [V]arious operational restrictions, including on tasks, types of targets, the operating environment, time frame of operation, and scope of movement. 4. [O]perator [has] sufficient information on and understanding of the weapon system and operating environment, and the interaction between them.

Except for the definition proposed by CNAS, each definition requires the ability to intervene.[137] But that requirement unreasonably narrows the scope of meaningful human control, and in any event, such a requirement is at odds with state practice.[138] The United States has repeatedly rejected international proposals for meaningful human control, and with it, its presumed requirement for human intervention.[139] In its rejection of "control" as a governing feature, the United States appears, at least on some occasions, to have taken too narrow a view of the term.[140]

This article proposes the following baseline definition of meaningful human control of an AI-enabled system: human control of a system is generally meaningful if a human maintains the ability to understand and shape its outcomes. Such a definition, intended as a starting point for any assessment of meaningful

137. ICRAC requires "a means for the rapid suspension or abortion of the attack." Garcia, *supra* note 133. Article 36 requires "a potential for timely intervention"; and the ICRC requires "human supervision and ability to intervene." *Expert Meeting*, *supra* note 136.

138. *See* INT'L COMM. OF THE RED CROSS, ICRC POSITION ON AUTONOMOUS WEAPON SYSTEMS 5 (2021), https://perma.cc/A8XS-VVM2 (acknowledging "[s]ome AWS are already in use for specific tasks in narrowly defined circumstances . . . ."). *See also* Scharre & Horowitz, *supra* note 134, at 12 (identifying "[a]t least 30 nations" that use autonomous or semi-autonomous defensive systems).

139. U.N. Secretary-General, *Lethal Autonomous Weapons Systems*, 115, U.N. Doc. A/79/88 (July 1, 2024) (alluding to "practical steps" a state may take at "different stages of the weapon design, development, and deployment process" to ensure a machine effectuates the intent of commanders and operators).

140. *See generally* United States, *Working Paper: Human–Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, U.N. Doc. CCW/GGE.2/2018/WP.4 (Aug. 28, 2018) (repeatedly referring to control as "manual control," which seems to imply some degree of physical control).

human control, is little more than the sum of its parts. Meaningful, according to Merriam-Webster, means "having meaning or purpose."[141] And control means "to exercise restraining or directing influence over" or "to incorporate suitable controls in" (as in, for instance, "a controlled test").[142] As compared to the proposed definitions above, the working definition proposed here is simpler and more adaptable. We diminish the concept's utility by either proposing too high or too defined a standard for meaningful human control.[143]

One may argue—rightfully—that the definition introduces the notion of understanding, a criterion which is not immediately evident from the piecemeal addition of Merriam-Webster's definitions. But understanding is critical to meaningful human control and implicit in the very notion of control itself; without it, the definition encompasses situations where control is illusory.[144] Perhaps recognizing this, the DoD has already embraced "understanding" both as an ethical tenet in the DoD AI Ethical Principles[145] and as a requirement throughout its policy on autonomous weapon systems.[146]

Requisite understanding, of course, varies by position. The operator of an AI-enabled system need not understand the computer code comprising the system, just as a mortarman need not understand Newtonian physics. In both circumstances, the operator of the AI system and the operator of the mortar need only to understand how to operate the system or weapon, which often comes through training and repetition.[147] The DoD would hesitate to place the mortar in the

141. *Meaningful*, MERRIAM-WEBSTER, https://perma.cc/BV56-UDHX.

142. *Control*, MERRIAM-WEBSTER, https://perma.cc/5MNQ-KCCU.

143. Crootof, *supra* note 129, at 58 (arguing the concept's imprecision is a strength).

144. Let us use a hypothetical to demonstrate: An intelligence analyst receives orders to provide a list of ten priority targets in a particular area. She relies on an AI platform, such as Palantir's AIP, to help generate those in the same way a person might rely on an agentic AI platform to form a list of potential restaurants in a city. As the system's operator, the analyst exercises control of its output insofar as she controls her own input: she crafts the parameters for the ultimate list through a carefully articulated prompt for the AI platform. But if the developer of the platform lacked understanding of the way in which the platform generates its output, whether because of opaque design or inadequate testing, there is no meaningful human control of the system. This lack of understanding, called the "black box problem," is of real concern. *See* Lou Blouin, *AI's Mysterious 'Black Box' Problem, Explained*, U. MICH. DEARBORN NEWS (Mar. 6, 2023), https://perma.cc/Z4S4-BDGV (describing the "inability for us to see how deep learning systems make their decisions"). The analyst's limited ability to shape the outcome creates the illusion of control in the absence of human understanding of the system's functionality.

145. As stated in the DoD AI Ethical Principles (and restated in DoDD 3000.09), "The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation." DoDD 3000.09, *supra* note 99, § 1.2(f)(3).

146. *Id.* § 1.2(a)(3) (requiring "the human-machine interface for autonomous and semi-autonomous weapon systems" to be "readily understandable to trained operators" in order "[f]or operators to make informed and appropriate decisions regarding the engagement of targets"); *id.* § 4.1(d)(4) (requiring, before fielding new autonomous weapons, verification that "[a]dequate training, TTPs, and doctrine are available, periodically reviewed, and used by system operators and commanders to understand the functioning, capabilities, and limitations of the system's autonomy in realistic operational conditions").

147. *Id.* § 4.1(d)(4) (mandating "[s]ystem design and human-machine interfaces" of autonomous weapons to be "readily understandable to trained operators").

mortarman's hands, however, if human developers of the weapon did not understand how the mortar functioned.

The proffered definition of meaningful human control also reflects two aspects of control: first, human agents may exercise control through various means; second, control is inherently context dependent. To illustrate the interplay between means and context, let us take an example, one pulled from everyday life. A father, walking through the park with his four-year-old child may exert control of his child verbally, by providing guidance "not to walk too far ahead and not to stray from the path." The ambulatory, sentient child acts with a degree of autonomy, of course, and although bends in the path occasionally obscure the father's view, he is generally able to monitor the child's behavior to ensure she conforms to his intent. That is one form of control. The father may, however, find the need to exercise control of the child physically. As the father and child near the end of the path, they approach a busy road. Here, the father tells the child to stop at the end of the path. He takes the child's hand, and together they cross the street, hand-in-hand. The father has retained control of the child, but the mode of control varies based on context.

Now consider the same principle in the national security sphere. In some cases, an operator may have reasonable certainty that an AI-enabled drone will act consistent with the operator's intent and directs the drone to take certain action. In others—perhaps in situations of heightened risk—the operator may monitor the AI-enabled drone and directly intervene. In both circumstances, the operator exerts meaningful control over the system, just as the father retains control over the child, which our definition of meaningful human control reflects.

C. Three Phases of Meaningful Human Control

Meaningful human control should not be equated with direct control of a system in the final stages; individuals throughout the national security structure may ensure control of the system through various means and at different stages of the AI platform's employment. Rigorous human testing, for instance, may establish meaningful human control in the design phase. So, too, would human-machine integration (roughly equivalent to the notion of a "human in the loop") and system oversight ("human on the loop"). The United States generally adopted this three-phase approach in its May 23, 2024 response to the U.N. Secretary-General concerning lethal autonomous weapons systems.[148]

Doubts exist as to whether upstream human involvement sufficiently establishes human control of an autonomous weapon system. The ICRC, for one, finds "broad agreement on the need for human control over weapons and the use of force" but remains uncertain "whether human control at the stages of the development and the deployment of an autonomous weapon system is sufficient to

148. U.N. Secretary-General, *supra* note 139 (alluding to "practical steps" a state may take at "different stages of the weapon design, development, and deployment process" to ensure a machine effectuates the intent of commanders and operators).

overcome minimal or no human control at the stage of the weapon system's operation—that is, when it independently selects and attacks targets."[149]

This paper argues that humans may exert control in three codependent modes: (1) as designer and tester, before deployment of an autonomous system; (2) as operator, integrated into the system; and (3) as supervisor, overseeing operation of the system. Less human involvement in one requires greater human involvement in another. It is through this lens that meaningful human control comes into focus and may be applied generally to state practice.[150] We turn to the first phase or mode of control.

1. Human Control During Design and Testing

The first point at which humans may begin to exert meaningful human control is during the design and testing phase of AI algorithms and of the systems that contain them—for example, fine-tuning an LLM.[151] The DoD already requires rigorous testing of acquired capabilities, which, in turn, supplies greater understanding and refinement of those capabilities.[152]

Human control permeates the existing design and testing process. The DoD requires integrated testing "at the earliest opportunity," which enables "early identification of concerns to improve the system design, and guides the system development during the engineering and manufacturing development phase."[153] The testing itself can be extensive.[154] And by statute, the product of DoD testing is a report from the Director of Operational Test and Evaluation ("DOT&E") to Congress about the adequacy of testing and the effectiveness and suitability of the items and components tested.[155]

The DoD also regulates human involvement in the development and testing of autonomous weapon systems.[156] DoDD 3000.09 requires "[s]ystems [to] go

149. INT'L COMM. OF THE RED CROSS, *supra* note 126.

150. Some scholars have equated the terms "human in the loop" and "meaningful human control." *See* Jovana Davidovic, *What's Wrong With Wanting a "Human in the Loop"?*, WAR ON THE ROCKS (June 23, 2022), https://perma.cc/52L4-BXNT ("Systems that are meant to have meaningful human control or a human-in-the-loop are tested and evaluated (and rightly so) as socio-technical systems (with operators in place obviously).").

151. Google, *LLMs: Fine-tuning, Distillation, and Prompt Engineering*, MACH. LEARNING, https://perma.cc/FN54-EG9L.

152. *See* DoD Instruction 5000.89, Test and Evaluation 1 (2020) (establishing policy, responsibility, and procedures "for test and evaluation (T&E) programs across five of the six pathways of the adaptive acquisition framework") [hereinafter DoD Instruct. 5000.89].

153. *Id.* § 3.1(c).

154. *See* David Jordan, *Army Moving Forward with Next Generation Squad Weapon Program*, U.S. ARMY (Mar. 15, 2023), https://perma.cc/85VV-99FY (describing testing by Sig Sauer of the Next Generation Squad Weapon ("NGSW"), the XM7 Rifle, XM250 Automatic Rifle, and 6.8 mm ammunition). In connection with Sig Sauer's new weapon's capability, the testing team "conducted over 100 technical tests, fired over 1.5 million rounds of 6.8 mm ammunition and executed over 20,000 hours of Soldier testing." *Id.*

155. 10 U.S.C. § 2399(b).

156. DoDD 3000.09, *supra* note 99, § 3 (Verification and Validation and Testing and Evaluation of Autonomous and Semi-Autonomous Weapon Systems). Testing of these capabilities also falls under DoDI 5000.89, which applies more generally. *See generally* DoD Instruct. 5000.89, *supra* note 152, § 3.

through rigorous hardware and software verification and validation (V&V) and realistic system developmental and operational test and evaluation (T&E)."[157]

While it may seem counterintuitive, significant control in building and testing algorithms may ensure meaningful human control over processes from which humans are ostensibly absent. The C-RAM provides an example. The system operates with a human directly involved in the process, overseeing the process, or outside the process.[158] But even when the system operates in a fully autonomous mode, humans may still exert meaningful control:

> [T]he [C-RAM] weapon processes data received by sensors and applies pre-programmed software code to determine whether to engage an incoming target. In doing so, it autonomously applies the judgment of human software engineers who have analyzed data to develop computer programs that can determine whether the incoming object constitutes a threat. So, it is humans, not algorithms, who truly make the targeting decision—even if it is a machine that later autonomously carries out the function it has been programmed to perform.[159]

Although this type of programmed control may strike some as illusory, we can see evidence of it in other contexts. Let us revisit the example of the father and child walking through the park. In the example above, the father told his child to stop at the end of the path before crossing the street. There, he took the child's hand. But if we change the visual somewhat—perhaps the father walked through the park with not one, but two children: his four-year-old child and his neighbor's four-year-old child. As he approaches the end of the path, he hurries ahead to ensure both children stop at the end of the path before the road. Although he has confidence that his own child will comply with his verbal instruction to stop before entering the road, he is less certain of his neighbor's child, whose hand he may need to grab if the child shows any indication of darting into the road.

The difference is the measure of meaningful control the parent employed over his own child's behavior through countless prior interactions, appraising, teaching, reprimanding—in a word, shaping—before they ever stepped foot outside. Because the parent had less involvement with his neighbor's child, he exerts more control over that child during their walk. The analogue in AI development is the designing and testing of algorithms, the stage at which humans, first, may define *ex ante* bounds that govern an AI system's functions and, second, may ensure, through rigorous testing, that the system will function in accordance with an operator's intent.[160]

---

157. DoDD 3000.09, *supra* note 99, § 1.2(a)(1).

158. *See* Cox, *supra* note 9.

159. *Id.*

160. In this way, human understanding of an AI system may crystalize through empirical knowledge derived from extensive testing, even when humans' ability to map the innerworkings of an AI system is limited. By way of comparison, we can say with confidence that Johannes Kepler, through empirical research, understood the motion of planets, though he died over 50 years before Sir Isaac Newton published the *Principia*, which explained Kepler's laws of planetary motion. *See generally* JOHANNES KEPLER, EPITOME OF COPERNICAN ASTRONOMY: IV AND V THE HARMONIES OF THE WORLD (1619),

2. Human-Machine Integration

The second point at which humans may exert meaningful human control is during employment of the AI tool through human-machine integration (or, if one remains unconvinced by the argument in Part II: the human in the loop). This is perhaps the easiest place to see how effectively a human may exercise meaningful control, since the human remains an integral part of the system's operation. The key in determining whether control is meaningful at this stage, and as addressed in Part II, is whether the human operator has real understanding and decision-making ability—and is not a reflexive red-button-pusher.

3. System Oversight

The third and final means by which humans may exercise meaningful control over an AI system is through system oversight. This requires a human to oversee the actions of a particular system, not necessarily each individual targeting decision. The ability to shape outcomes of (or to control) a system does not imply control over individual outcomes of a particular system.

To better draw the distinction, let us picture a human overseeing a swarm of drones.[161] The human may be unable to oversee the actions of each particular drone in real time; however, if she receives immediate feedback from the system's outputs, she is able to correct the actions of the drone swarm. If she receives information that some of the drones took actions inconsistent with the mission and her intent, she can adjust.[162] In this sense, she exerts control over the AI-enabled system of drones through oversight.[163]

---

reprinted in 16 THE GREAT BOOKS OF THE WESTERN WORLD (1939); ISAAC NEWTON, MATHEMATICAL PRINCIPLES OF NATURAL PHILOSOPHY (1686), *reprinted in* 34 THE GREAT BOOKS OF THE WESTERN WORLD (1934).

161. This is but one method of "command and control" used with drone swarms. *See* U.S. GOV'T ACCOUNTABILITY OFF., SCI. & TECH SPOTLIGHT: DRONE SWARM TECH. 1 (Sep. 14, 2023), https://perma.cc/8WL8-8C44 (dividing drone swarms into centralized, decentralized (semi-autonomous), and decentralized (autonomous) categories of operation).

> Drone swarms can use various methods of command and control, including preprogrammed missions with specific predefined flight paths, centralized control by a ground station or a single control drone, or distributed control where the drones communicate and collaborate based on shared information. More advanced methods of control include swarm intelligence, inspired by the collective behaviors of insect colonies and flocks of birds, as well as artificial intelligence techniques to teach drone swarms to respond to new or unexpected situations.

Id. Drones have played a massive role in the current war between Ukraine and Russia. *See The War in Ukraine Has Changed—and It's Deadlier Than Ever*, N.Y. TIMES (March 3, 2024), https://perma.cc/UR4F-KTR9 (reporting on the deaths due to drones during the war in Ukraine and Russia). Ukraine officials report that drones now kill more individuals and destroy more armored vehicles than all other weapons combined. *Id*.

162. Such oversight bears resemblance to human oversight of existing defensive autonomous systems, where "[g]iven the short time required for engagements, human controllers may not be able to intervene before any inappropriate engagements occur, but they can terminate further functioning of the system." *See* Scharre & Horowitz, *supra* note 103, at 13.

163. By way of analogy, an individual who decides to microwave popcorn may not have control over the popping of individual kernels; but if the popcorn starts to burn, he can stop the microwave.

Human oversight, too, occurs through *ex post* checks and reviews. Currently the DoD conducts these with autonomous weapons.[164]

D. Why Human Control is Possible

Scholars have argued against a standard involving human control for a more practical reason: they argue human control of AI is not, or soon will not be, possible.[165] The most common contention is that human control is impossible because machines can ingest more data, more quickly, than their human counterparts. Machine capacities outpace human capacities to such a degree, the argument goes, that any pretension of human control is folly.[166] A recent article in *Foreign Affairs* observes: "Future AI-enabled warfare will be even faster and more data-dependent, as AI-enabled weapons systems (e.g., swarms of drones) can be deployed swiftly and at scale. Humans will have neither the time nor the cognitive ability to assess that data independently of the machine."[167]

It is a logical fallacy, however, to insist that human control is illusory simply because humans are unable to independently evaluate the machines' decisions. Human control does not require a human to review everything the AI system reviewed or even require the human to have the cognitive capacity of the machine. That is a bit like arguing that a music conductor must play every instrument in the orchestra to ensure the musicians play harmoniously and in time. The controlling human is not a redundancy; the human and machine play separate roles, operating as a hybrid human-machine system. The musician's ability and the conductor's control are separate concepts.

Some make the slightly different contention that human control must cede to machine control for the United States to maintain a strategic advantage.[168] But such logic could argue against *any* rule or restraint in national security.[169] We should instead aim to outpace our adversaries by leveraging technology within clear ethical and legal bounds.

E. An Adaptable Concept

As an adaptable legal construct, meaningful human control's "indefiniteness is its strength."[170] Some scholars have argued meaningful human control's flexibility

164. DoDD 3000.09, *supra* note 99, § 4.1.

165. Elke Schwarz, *The (Im)possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems*, HUMANITARIAN L. & POL'Y (Aug. 29, 2018), https://perma.cc/W4RR-8SBK.

166. Sebastian Elbaum & Jonathan Panter, *AI Weapons and the Dangerous Illusion of Human Control*, FOREIGN AFFS. (Dec. 6, 2024), https://perma.cc/33T9-NEBY.

167. *Id.*

168. *Id.* ("The pressures of security competition point in only one direction: ever-accelerating automation in warfare."). The authors allow that "tactics and operations are not the only considerations in war." *Id.*

169. The rule of proportionality, bans on chemical weapons, requirements for the human treatment of *hors de combat*—all of those impose some cost of compliance.

170. Crootof, *supra* note 129, at 58.

is imprecision—a flaw that inhibits any utility.[171] Yet as Professor Charlie Trumbull observes:

> The ambiguity of "meaningful human control," or some similar standard, is not necessarily contrary to the interests of states with advanced militaries. Many IHL rules incorporate ambiguous standards, such as the requirement to take "all feasible precautions" or the prohibition on weapons of a nature to cause "superfluous injury." The ambiguity in these rules affords states some margin of discretion in implementing their obligations while also allowing the meaning of these ambiguous terms to be clarified through subsequent state practice.[172]

In its elasticity, the meaningful human control standard resembles the reasonable-person standard in tort law, a legal regime that similarly developed as a result of technological development.[173] As with the reasonable-person standard, the meaningful human control standard offers "the attraction of a simple, general standard, as opposed to more detailed . . . rules or . . . particularized criteria[.]"[174]

This adaptable framework also ensures adequate responsibility for state actions. In the employment of any AI-enabled system, the threshold question must be whether the state ensured meaningful human control over that system. Assuming a state has ensured meaningful human control, accountability naturally flows: responsibility lies with those actors who controlled the autonomous system. Accordingly, humans at various stages may be accountable for state actions—behind every kinetic strike, for instance, there will be humans who bear responsibility due to actions they took (or failed to take) during design and testing, human-machine interaction, or oversight of an AI system.

In a simplified example, a commander directs an AI-enabled drone to target a thirty- to forty-year-old male wearing a blue hat within a ten-digit grid. If the drone erroneously strikes a civilian, who bears responsibility? Assuming the state ensured meaningful human control over the system, the answer is: it depends. The issue could be one of faulty inputs: the parameters were too broad (perhaps there were many blue-hat-wearing thirty- to forty-year-old males in the grid) or flat-out wrong (the real target was not wearing a hat at all). In those cases, responsibility primarily lies with the humans directly involved in the decision-making process: the commander, the intelligence agents, etc. On the other hand, if the

171. Cox, *supra* note 9 ("[T]he 'meaningful human control' description that has emerged among humanitarian advocates as the standard from which autonomous weapons must not deviate is not adequately precise.").

172. Trumbull, *supra* note 4 (discussing three approaches to a legally binding instrument).

173. *See* Rebecca Crootof & BJ Ard, *Structuring TechLaw*, 34 HARV. J. L. & TECH. 347, 373 (2021) https://perma.cc/55JF-Y86U.

174. Kenneth W. Simons, *The Hegemony of the Reasonable Person in Anglo-American Tort Law*, OXFORD STUD. IN PRIV. LAW THEORY 65 (Paul B. Miller & John Oberdiek eds., 2020), https://perma.cc/KDD2-C45D (considering the positive and negative aspects of the hegemony of the reasonable person standard).

drone mistook blue for yellow, targeting a civilian in a yellow hat rather than the actual target in a blue hat, responsibility may lie with the individuals who exerted meaningful control further upstream in the process: those involved in the designing and testing of the AI system. And if the drone was altogether incapable of distinguishing blue from yellow—a limitation on the AI system's capability unknown to the commander directing the system—perhaps responsibility lies with the humans who trained the system's operators. Regardless of where responsibility lies, the crucial point is that meaningful human control ensures accountability.[175]

F. International Support

There are three main approaches to international restrictions on AI.[176] First, states could impose positive obligations—such as meaningful human control—on the use of AI.[177] Second, states could create categorical bans, prohibiting certain autonomous weapons systems. For instance, states may ban autonomous weapons that are not under meaningful human control. Austria has proposed such a ban.[178] Third and finally, states could require a certain amount of due diligence or impose procedural requirements on the design and employment of autonomous weapons—to include through an Article 36 weapons review.[179] A fourth option also exists: an admixture of some or all of those.

The United States' adoption of a meaningful human control standard, a positive obligation on its use of AI, would surely receive support internationally.[180] Among the Group of Governmental Experts ("GGE"), "[t]here is widespread recognition that *human control is essential* to ensure responsibility and accountability, compliance with international law and ethical decision-making[.]"[181] By embracing meaningful human control and shaping its meaning, the United States could garner the support of the international community and place its own practices at the center of an emerging international standard.

G. The Importance of the Human

Inevitably, we must address the purpose behind human involvement in AI systems. The preference for human involvement is evident in current DoD policy,

175. In practice, this may be hard to determine. The starting point for any thorough examination and assessment of responsibility requires traceability and record-keeping.

176. *See* Trumbull, *supra* note 4 (discussing not two, but three approaches to a legally binding instrument).

177. This tack still requires states to identify the scope of those obligations, whether the standard applies to autonomous weapon systems or, as this paper argues, more broadly to targeting functions.

178. *See* Thomas Hajnoczi, Austrian Ambassador and Permanent Representative to the U.N. in Geneva, Statement of Austria to the Group of Governmental Experts on Lethal Autonomous Weapon Systems in Geneva on the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (Apr. 9–13, 2018), https://perma.cc/8XJN-866W.

179. *See* Trumbull, *supra* note 4.

180. *See generally* U.N. Secretary-General, *supra* note 139.

181. *Id.* ¶ 89 (emphasis added).

where "appropriate levels of human judgment" are required[182]; it is evident in discussions about banning autonomous weapons.[183] And, of course, it is evident in the argument for human control. It would be easy to gloss over the human element, which is integral to almost any proposed AI regulation scheme.[184] It helps, however, to evaluate the human operator's *raison d'être*.

Some have argued that human involvement is necessary because AI will be unable to adequately comply with international law's requirements of proportionality and distinction.[185] But to the extent these require evaluating data (and predictive modeling) within the bounds of human-created parameters, this seems a specious prediction, particularly given AI's demonstrated capabilities. AI has shown the ability to diagnose medical conditions,[186] provide instant intelligence of CCTV cameras,[187] and predict human emotions.[188] The machine never gets tired, hungry, or bored. It is not hard to envision circumstances in which an AI system outperforms a human system in evaluating key aspects of proportionality and distinction, considering more factors, more rapidly, and with more accuracy.

Humans' role in the process is not, as some have argued, merely to ensure compliance with existing international law. If that were so, there would be no need to ban autonomous weapons in the first place or create new bounds on their use: use of autonomous weapon systems without human control agents would be *de facto* inconsistent with existing law.

Rather, human involvement is necessary on one side in war in order to recognize humanity on the other. Such recognition has been a fundamental component of war throughout history.[189] Yet the further removed the human, the easier it

182. DoDD 3000.09, *supra* note 99, § 1.2(a).

183. INT'L COMM. OF THE RED CROSS, *supra* note 126, at 8 (addressing "fundamental ethical concerns for humanity" raised by autonomous weapon systems).

184. *See, e.g.*, CCW/MSP/2019/9, annex III, Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2019), https://perma.cc/4WE8-KYJ7 (reaffirming human responsibility and human-machine interaction as guiding principles for autonomous weapon systems).

185. Bonnie Docherty, *Mind the Gap: The Lack of Accountability for Killer Robots*, HUM. RTS. WATCH (Apr. 9, 2015), https://perma.cc/64YL-5BC5.

186. *See* GOOGLE CLOUD, *New Way Now: Hackensack Meridian Health's Epic Cloud Migration Paves the Way for AI Innovation* (YouTube, June 24, 2024), https://perma.cc/23P6-TJY8 (discussing an AI model that provides initial medical diagnoses to physicians for review).

187. Anisa Menur A. Maulani, *Lytehouse Improves Workplace Safety With AI-based Video Intelligence, Teams Up with Google Cloud*, E27 (Sep. 27, 2024), https://perma.cc/3C3U-YVMK ("Lytehouse offers real-time video intelligence that automates risk management by transforming any CCTV camera feed into virtual co-workers, enhancing efficiency and reducing costs.").

188. Ala N. Tak & Jonathan Gratch, *GPT-4 Emulates Average-Human Emotional Cognition from a Third-Person Perspective*, ARXIV (Aug. 11, 2024), https://perma.cc/32U9-G3DE. According to Ala N. Tak and Jonathan Gratch, GPT-4 accurately predicts emotions from a third-party viewpoint. In this, it aligns with how humans perceive the feelings of others.

189. One prominent example is the famous exchange that took place in the first World War, during the Christmas Truce of 1914, when British and German soldiers paused fighting and exchanged Christmas greetings. *See* CAPTAIN SIR EDWARD HULSE, LETTERS WRITTEN FROM THE ENGLISH FRONT IN FRANCE BETWEEN SEPTEMBER 1915 AND MARCH 1915 56–62 (1916), https://perma.cc/947L-62P7. *See also* Crootof, *supra* note 3, at 1845 (noting "autonomous weapon systems threaten one of the law of

becomes to lose that recognition.[190] It is not hard to see how autonomous systems create more distance between the decision to use force and the results of that force in the context of armed conflict.

Further, meaningful human control is consistent with the nature of warfare, an inherently human struggle. As one scholar articulates, our entire governing system is anthropocentric:

> The very nature of IHL [international humanitarian law], which was designed to govern the conduct of humans and human organizations in armed conflict, presupposes that combatants will be human agents. It is in this sense anthropocentric. Despite the best efforts of its authors to be clear and precise, applying IHL requires multiple levels of interpretation in order to be effective in a given situation. IHL supplements its rules with heuristic guidelines for human agents to follow, explicitly requires combatants to reflexively consider the implications of their actions, and to apply compassion and judgement in an explicit appeal to their humanity. In doing this, the law does not impose a specific calculation, but rather, it imposes a duty on combatants to make a deliberate consideration as to the potential cost in human lives and property of their available courses of action.[191]

An autonomous system is ill-equipped to incorporate characteristically human concepts such as honor, courage, respect, or mercy. Yet here is where the machine's lack of humanity may alter outcomes for the worse, even while it complies with international humanitarian law.

Consider the following hypothetical: *An armed, autonomous MQ-9 Reaper tracks a group of enemy combatants walking through a crowded urban street. The Reaper's objective is to destroy all enemy combatants in the area. As the group walks among civilians, the path the enemy combatants take is clear: they are heading to a known location, identified as an enemy command post within the city. As anticipated, the group turns down a less populated city street. There are civilians nearby, but not nearly as many as before. In another block, the group will be in a clearing, where there are no civilians. Still, the AI-enabled MQ-9,*

---

armed conflict's most fundamental assumptions: that, ultimately, a human being decides whether another human being lives or dies").

190. ROGER BROWN, SOCIAL PSYCHOLOGY 22–23 (2d ed. 1986) (describing Milgram's shock experiments and the "proximity condition"). Brown applies the lessons drawn from Milgram's experiments to acts of war: "This is an important series of experiments because it captures the essence of the difference between dropping bombs from 10,000 feet and attacking someone close up. It is an experimental series that accounts for the horror we feel at the close-up atrocities of My Lai and Auschwitz and the probably lesser horror we feel at the remote leadership that was primarily responsible. Adolf Eichmann took care to see as little as he possibly could of the sufferings his office helped to bring about." *Id*. *See also* Crootof, *supra* note 3, at 1845 (pointing out that "[w]eapons systems with increasingly autonomous capabilities may allow for an increasingly attenuated temporal and geographic link between a human being's decision to deploy a weapon and the use of force"); DEEKS, *supra* note 5, at 112 (noting that "the military use of predictive algorithms and machine learning tools" will likely increase the "perceived dehumanization of lethal action").

191. Asaro, *supra* note 67, at 695.

which has been coded to comply with IHL, fires. The lethal strike takes out the group and kills a number of nearby civilians.

Under international law, the deaths may not have been excessive in relation to the concrete and direct military advantage anticipated.[192] And by waiting until the targets transited to a less populated area, the autonomous Reaper took feasible precautions to minimize civilian harm—even if it did not wait until totally clear of civilian bystanders.[193] In other words, the drone complied with IHL. It balanced the anticipated military advantage with the rule of proportionality. But if it had waited for the group to travel another block, the drone might have avoided all civilian deaths. Would a human decide to delay fire, despite being legally entitled to fire sooner? In this, there is perhaps tension between preprogrammed rules—based on international law and DoD policy—and humanity's desired outcome. The drone achieved the objective in compliance with international law. But the hypothetical illustrates problems that arise when we remove humanity from warfare. Meaningful human control is necessary. The U.N. Secretary-General recently declared, "Machines that have the power and discretion to take human lives are politically unacceptable and morally repugnant, and should be banned by international law."[194]

Increasingly, scholars have considered whether it is possible to adequately program AI to comply with the law of armed conflict, essentially encoding within the algorithm the myriad rules and policies that govern decisions in war.[195] But even within a single unit within a particular country at a defined point in time, morality varies. Person A's moral compass is not Person B's moral compass. Will the AI system have the moral code of Courtney Massengale or Sam Damon?[196]

H. The United States' Position, as Articulated and in Practice

On the global stage, the United States has emphatically and consistently eschewed a "human control" standard.[197] Rather, the United States points to DoDD 3000.09 and refocuses the discussion on ensuring "machines help effectuate

192. U.S. DEP'T OF DEF., LAW OF WAR MANUAL § 5.12 (Proportionality) (rev. ed. July 2023) ("Military objectives may not be attacked when the expected incidental loss of civilian life, injury to civilians, and damage to civilian objects would be excessive in relation to the concrete and direct military advantage expected to be gained.").

193. *Id.* § 5.11.

194. U.N. Secretary-General, *supra* note 139, ¶ 89.

195. Ashley Deeks, *Coding the Law of Armed Conflict: First Steps*, LIEBER INST. (July 14, 2022), https://perma.cc/X33U-LEGP.

196. *See generally* ANTON MYRER, ONCE AN EAGLE (1968). In Myrer's classic, Courtney Massengale is an ambitious staff officer driven primarily by career advancement. Sam Damon is the quintessential "Soldier's Soldier," who puts duty and his soldiers' well-being above his own interests.

197. *See, e.g.*, U.N. Secretary-General, *supra* note 139, at 115–17; Josh Dorosin, Deputy Legal Adviser, U.S. Dep't of State, Statement of the U.S. Delegation on Possibly Policy Options to the Meeting of the Group of Governmental Experts on Lethal Autonomous Weapons Systems (Aug. 29, 2018) (available at https://perma.cc/C7KK-M33U).

the intention of commanders and the operators of weapon systems."[198] The United States' May 2024 submission to the U.N. Secretary-General on lethal autonomous weapons systems explained:

> [T]he key issue, as reflected in Department of Defense Directive 3000.09 and in United States working papers to the Group of Governmental Experts, is ensuring that machines help effectuate the intention of commanders and the operators of weapons systems. This is done by, *inter alia*, taking practical steps—at different stages of the weapon design, development, and deployment process—to reduce the risk of unintended engagements and to enable personnel to exercise appropriate levels of human judgment over the use of force.[199]

Such statements beg the question: how else—*but through control*—might a state ensure machines "effectuate the intention" of their human commanders and operators? The answer, of course, is that it is only through human control at various stages that a state can ensure effectuation of human intent: to do so, humans must necessarily "exercise restraining or directing influence over" or "incorporate suitable controls in" (*i.e.*, control) the AI system.[200] Despite its continued resistance to the meaningful human control standard, in practice, the United States' position on autonomous weapons very much resembles such a standard: the United States employs rigorous testing and evaluation, human-machine integration, and human oversight at various levels for various weapon systems. In taking a narrow view of the term "control," the United States misses an opportunity to embrace the majority's standard[201] and shape what the international community means by control.

IV. CONCLUSION

As we debate the regulatory guardrails that should bind the United States in its use of AI, we should be intentional and precise in our language. Implementing AI governance is a monumental undertaking; we compound the challenge by using the loop—a confusing and muddled concept. Rather, we should refocus on governing human-machine systems.

Moreover, we cannot totally cede control to machines for targeting. As it relates to weapons systems and intelligence, DoD's AI policy should explicitly require meaningful human control, established at various stages of AI use. The DoD, in practice, already adheres to the standard; it has the opportunity to expressly embrace meaningful human control in targeting and lead the international community in the next era of warfare.

---

198. U.N. Secretary-General, *supra* note 139, at 115.

199. *Id.*

200. *See Control*, *supra* note 142.

201. *See* ICRC News Release, *supra* note 89. The ICRC President, Peter Maurer, observed, "It is now widely accepted that human control must be maintained over weapon systems and the use of force, which means we need limits on autonomy[.]" *Id.* Accordingly, "[n]ow is the moment for States to determine the level of human control that is needed to satisfy ethical and legal considerations." *Id.*