



Background Briefing

Review of the 2023 US Policy on Autonomy in Weapons Systems

February 2023

Introduction

A new directive on autonomy in weapons systems issued on January 25, 2023 shows the United States Department of Defense (DoD) is serious about ensuring it has policies and processes in place to guide its development, acquisition, testing, fielding, and use of autonomous weapons systems as well as semi-autonomous weapons systems, such as remotely operated armed drones. The directive, however, constitutes an inadequate response to the serious ethical, legal, accountability, and security concerns and risks raised by autonomous weapons systems.

The 2023 DoD Directive 3000.09 on Autonomy in Weapons Systems revises, but does not radically change, the department's original policy on the topic, released on November 21, 2012.¹ The 2023 directive is valid for a limited period of up to 10 years, as was the case for the 2012 directive.²

The 2023 directive maintains the basic structure and substance of its predecessor, and as a result, misses an opportunity to address its shortcomings.³ For example, it continues to allow certain waivers to senior reviews before the development and fielding of autonomous weapons systems, and it does not remove ambiguity surrounding key terms.

The directive contains some improvements such as the inclusion of additional design constraints and review requirements, but it also adds some problematic amendments, particularly with regards to definitions and terms.

¹ US Department of Defense, "Directive 3000.09: Autonomy in Weapon Systems," Office of the Under Secretary of Defense for Policy, January 25, 2023, <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.PDF> (accessed January 26, 2023).

² This was also the case for the 2012 policy, which was reissued in 2017 without substantive change for another five years. DoD Directive 5025.01 specifies that all department directives should be updated or cancelled within 10 years of their publication date. Department of Defense Instruction 5025.01, "DoD Issuances Program," Office of the Chief Management Officer of the Department of Defense, May 22, 2019, <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/502501p.pdf> (accessed February 1, 2023).

³ See Human Rights Watch and Harvard Law School's International Human Rights Clinic, *Review of the 2012 US Policy on Autonomy in Weapons Systems*, April 15, 2013, <https://www.hrw.org/news/2013/04/15/review-2012-us-policy-autonomy-weapons-systems> (accessed February 1, 2023).

When the original DoD directive on autonomous weapons was issued in 2012, the United States was the first country to put such a detailed policy in the public domain. Since then, many countries have developed their own positions, and dozens have expressed interest in adopting a new treaty that prohibits autonomous weapons systems that operate without meaningful human control or that target people, and that regulates all other autonomous weapons systems to ensure they operate only with meaningful human control.⁴ These proposed elements are equally applicable to national policy, but they are not fully reflected in the 2023 directive. The directive thus diverges from the current position of a majority of states, the International Committee of the Red Cross (ICRC), and civil society organizations engaged on this issue, which see an urgent need for a legally binding instrument that contains prohibitions on and regulations of autonomous weapons systems.

As with the original policy, the 2023 directive applies only to the Department of Defense. A decade on, there is still no US government-wide policy governing autonomous weapons systems and their use in law enforcement, border control, and other circumstances outside of armed conflict. For example, the 2023 directive also does not apply to the Central Intelligence Agency (CIA), which has played an active role in the use of armed drones.

Recommendations

Human Rights Watch and Harvard Law School's International Human Rights Clinic (IHRC) urge the United States to:

- Ensure that DoD reviews of autonomous weapons systems under the 2023 DoD Directive 3000.09 are transparent and that use of the waivers is limited;
- Adopt national law and policy with prohibitions and regulations on autonomous weapons systems; and
- Work for a new international treaty that prohibits and regulates autonomous weapons systems.

Comparison to Previous Policy

Similar Shortcomings

The new directive, as with its predecessor, has a broad scope and covers autonomous and semi-autonomous weapons. Its definition preserves the 2012 directive's definition of an autonomous

⁴ See, for example, Human Rights Watch and Harvard Law School's International Human Rights Clinic, *New Weapons, Proven Precedent: Elements of and Models for a Treaty on Killer Robots*, October 20, 2020, <https://www.hrw.org/report/2020/10/20/new-weapons-proven-precedent/elements-and-models-treaty-killer-robots> (accessed February 1, 2023).

weapon system as “a weapon system that, once activated, can select and engage targets without further intervention by a human operator.”⁵ The 2023 directive removes the word “human” before “operator,” but defines an “operator” as “a person who operates a platform or weapon system.”⁶

The 2023 directive also continues to recognize numerous weaknesses and risks involved in autonomous weapons systems. In its glossary’s definition of “failure,” the 2023 directive repeats a list of possible causes of failures raised by autonomy in weapons systems including “human error, faulty human-machine interaction failures, malfunctions, communications degradation, software coding errors, enemy cyber attacks or infiltration into the industrial supply chain, jamming, spoofing, decoys, other enemy countermeasures or actions, or unanticipated situations on the battlefield.”⁷

The 2023 directive does not radically alter the 2012 policy, and as a result, many of the earlier policy’s shortcomings and weaknesses remain. The 2023 directive contains some of the same significant loopholes as the original directive. Its requirements for senior review before formal development or fielding can be waived by specific high-level department officials under certain circumstances, such as “in cases of urgent military need.”⁸

In 2019, a DoD spokesperson reportedly stated that, “to date, no weapon has been required to undergo the Senior Review in accordance with DOD Directive 3000.09.”⁹ A 2019 Congressional Research Service report on the US policy on autonomous weapons systems found that “no weapon system is known to have gone through the senior level review process” provided by the 2012 directive.¹⁰ When asked to confirm if any weapon systems has undergone review in accordance with the 2012 directive, an unnamed DoD civilian official told a January 30 briefing for civil society that he could not comment “on individual weapons systems.”¹¹ Transparency is needed to build trust, facilitate monitoring, and promote accountability with the policy. Clarity is

⁵ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 21, glossary.

⁶ Ibid, p. 22.

⁷ Ibid.

⁸ Ibid, p. 17, section 4.2.

⁹ Sydney J. Freedberg Jr., “Fear & Loathing In AI: How The Army Triggered Fears Of Killer Robots,” Breaking Defense, Breaking Defense, March 6, 2019, <https://breakingdefense.com/2019/03/fear-loathing-in-ai-how-the-army-triggered-fears-of-killer-robots/> (accessed February 9, 2023).

¹⁰ The report found that “as the United States is not currently known to be developing LAWS, no weapon system is known to have gone through the senior level review process to date.” Congressional Research Service, “Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems,” first issued March 27, 2019 and updated November 14, 2022, <https://www.everycrsreport.com/reports/IF11150.html> (accessed January 30, 2022).

¹¹ In response, the official said: “We are constantly having conversations about weapon systems to see how they fall under the directive, to see whether they maybe should or should not be reviewed. I’m not able to comment on individual weapons systems.” Department of Defense briefing for civil society on the 2023 DoD Directive 3000.09 on Autonomy in Weapon Systems held by conference call, January 30, 2023.

also needed to better understand the person or agency that is responsible for determining when and if a review process is required.¹²

As with its predecessor, the 2023 directive does nothing to curb the proliferation of autonomous weapons systems. The US, as well as Australia, China, India, Iran, Israel, South Korea, Russia, Turkey, and the United Kingdom are investing heavily in the military applications of artificial intelligence and related technologies to develop air, land, and sea-based autonomous weapons systems. Section 1.2(e) of the directive allows for “international sales and transfers” that are “approved in accordance with existing technology security and foreign disclosure requirements and processes.”¹³ Once these weapons leave the country, the United States would lose exclusive control over them. The directive thereby facilitates continued US development and acquisition of autonomous weapons systems as long as it is done in accordance with existing law, DoD processes and ethical principles.

The 2023 directive misses an opportunity to add clarity to areas of ambiguity in the 2012 directive. Under Section 1.2(a), the 2023 directive borrows its predecessor’s provision that “autonomous and semi-autonomous weapon systems will be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.”¹⁴ While the language recognizes the value of human judgment, which is essential for ensuring compliance with international humanitarian law, neither the previous nor current directive is clear on what constitutes an “appropriate level” of human judgment and who will determine it.

The new policy also replaces the term “shall,” which indicates a legal obligation, with “will,” which suggests the events will be certain to occur. An unnamed Pentagon official informed a civil society briefing that this was not “an intentional change,” but was done to be consistent with the DoD’s style guide.¹⁵

Problematic Revisions

While the fundamental policy has not dramatically changed, the 2023 directive contains some problematic revisions compared to its predecessor.

¹² Email from Michael Klare, Senior Visiting Fellow, Arms Control Association, February 9, 2023.

¹³ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 5, section. 1.2(e).

¹⁴ *Ibid*, p. 5, section. 1.2(a).

¹⁵ Department of Defense briefing for civil society on the 2023 DoD Directive 3000.09 on Autonomy in Weapon Systems, January 30, 2023. Held under the Chatham House rule. Notes taken by Human Rights Watch. Under “Helping Verbs” the DoD style guide says: “Do not use ‘shall.’” See “DoD Issuance Style Guide,” section 1.24, January 26, 2023, https://www.esd.whs.mil/Portals/54/Documents/DD/iss_process/standards/DoD_Issuance_Style_Guide.pdf?ver=NOeX5yOYKGTsuig-h-WYj-g%3D%3D (accessed February 1, 2023).

The directive amends the definition of “unintended engagement” in a potentially troublesome way. The purpose of the directive is to establish guidelines designed to minimize “the probability and consequences of failures ... that could lead to *unintended engagements*.”¹⁶ It requires that autonomous weapons be tested to ensure they avoid such engagements. The term “unintended engagement” was previously defined as “*the use of force resulting in damage* to persons or objects that operators did not intend to be the targets of U.S. military operations....”¹⁷ The 2023 directive now says: “*the use of force against persons or objects* that commanders or operators did not intend to be the targets of U.S. military operations.”¹⁸

The difference is that the policy now seems to limit system failures to those that involve the use of force against unintended targets, rather than those that cause any unintended harm. It would therefore, for example, seek to avoid using an autonomous weapon system that targeted a civilian, but it would not take steps to avoid use of a system that damaged civilian infrastructure that in turn caused civilian harm. The DoD official stated that the reason for this change was not to narrow the definition of “unintended engagement” but to broaden it by designating “anything that is targeted, that wasn’t designed to be targeted, [as] unintended in some ways.”¹⁹

The 2023 policy also deletes several references to the word “control” – a word that the US has objected to repeatedly in multilateral meetings on autonomous weapons systems.²⁰ Section 1.2(a)(1)(c), for example, calls for measures to provide “sufficient confidence” that autonomous weapon systems “are sufficiently robust to minimize the probability and consequences of failures.”²¹ It deleted the rest of the 2012 policy’s phrase, which read: “... that could lead to unintended engagements or to loss of control of the system to unauthorized parties.”

“Control” is an appropriate word to use in the context of autonomous weapons systems because it encompasses both the mental judgment and the physical act needed to prevent the systems from posing moral, ethical, legal, and other threats. When asked about the word’s removal from the new directive—as described above—the unnamed DoD official said it was removed “for technical accuracy” and because the term “control” had generated questions “about what exactly ‘control’

¹⁶ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 1. Emphasis added.

¹⁷ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” November 21, 2012, p. 15, glossary. Emphasis added.

¹⁸ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 23, glossary. Emphasis added.

¹⁹ Department of Defense briefing for civil society on the 2023 DoD Directive 3000.09 on Autonomy in Weapon Systems, January 30, 2023.

²⁰ See, for example, Statement of the US, Convention on Conventional Weapons Group of Governmental Experts meeting on lethal autonomous weapons systems, Geneva, 25 September 2020, <https://documents.unoda.org/wp-content/uploads/2020/09/Intervention-by-the-United-States.pdf> (accessed February 1, 2023).

²¹ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p.4, section 1.2(a)(1)(c).

means and how it's similar or different than 'judgment' in this case.”²² According to the DoD official, “it was cleaner, essentially to remove ‘control’ there and focus on the continuing requirement ... for there to be appropriate level of human judgment over the use of force.” Yet, as mentioned earlier, the new directive still does not clarify what amounts to an appropriate level of human judgment.

The 2023 directive, in addition, revises the previous directive’s rules for reviewing and approving autonomous and semi-autonomous weapons systems before developing and approving them. For example, it requires reviews of existing technology that have had “changes to the system algorithms, intended mission set, intended operational environments, intended target sets, or expected adversarial countermeasures [that] substantially differ from those applicable to the previously approved weapon system.”²³ The directive contains a “flow chart to help determine if senior review and approval is required.”²⁴

These revisions could be said to address criticisms that early armed drones escaped appropriate review and approval as they were first deployed for surveillance purposes and did not carry weapons. The greater specificity of development procedures and other elements of the directive, however, may also help accelerate the development of autonomous weapons systems.²⁵ While the procedures in the earlier directive sought to “ensure” that the systems were “sufficiently robust to minimize failures,” the 2023 directive states that testing and evaluation measures “*will provide sufficient confidence* that autonomous and semi-autonomous weapon systems” are “sufficiently robust to minimize the probability and consequences of failures.”²⁶ The 2023 directive thus suggests that if those measures are conducted, the systems will be “sufficiently robust.” The phrase “sufficient confidence,” which is weaker than “ensure,” does not appear in the original directive, but appears six times in the 2023 directive.²⁷

Other Observations

In one positive development, the directive imposes additional constraints for the design of autonomous weapons systems. It requires that “before a decision to enter formal development,” a system must be designed to complete engagements not only within a timeframe consistent with

²² Department of Defense briefing for civil society on the 2023 DoD Directive 3000.09 on Autonomy in Weapon Systems, January 30, 2023.

²³ *Ibid*, p. 15, section 4.1(a).

²⁴ *Ibid*, p. 18, figure 1.

²⁵ Email from Michael Klare, Senior Visiting Fellow, Arms Control Association, February 9, 2023.

²⁶ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 3, section 1.2(a)(1).

²⁷ Email from Michael Klare, Senior Visiting Fellow, Arms Control Association, February 9, 2023.

commander and operator intentions, as per the previous policy, but also within the “geographic area, as well as other applicable environmental and operational parameters.”²⁸

The 2023 directive brings policy on autonomy in weapons into line with other DoD policies put in place since 2012. It reiterates the department’s set of five ethical principles on artificial intelligence adopted on February 21, 2020.²⁹ Thus, the directive reflects the commitment of the DoD “to take deliberate steps to minimize unintended bias in AI capabilities.” It affirms the need to “detect and avoid unintended consequences” and “to disengage or deactivate deployed systems that demonstrate unintended behavior.”³⁰

The 2023 directive also removes gendered language such as the terms “manned” and “unmanned.” According to the unnamed DoD official, these revisions were “purposeful and deliberate.”³¹

The 2023 directive reflects organizational changes made at the Department of Defense since 2012 with respect to the names, roles, and responsibilities of certain offices and officials.

It also establishes an “Autonomous Weapons Systems Working Group” comprised of Department of Defense employees and active service members to provide advice during the process of reviewing autonomous weapons systems.

Falling Short of Emerging Policy Convergence

The 2023 directive should be compared not only to its 2012 predecessor but also to other policy proposals for addressing the ethical, legal, accountability, and security threats posed by autonomous weapons systems. Diplomatic talks on lethal autonomous weapons systems were not initiated until November 2013, so the 2012 DoD policy was an important, if imperfect, step toward recognizing and addressing the problems raised. A decade later, however, the 2023 policy is out of step with widely supported international proposals for a new treaty to prohibit and regulate autonomous weapons systems.

Over the past few years, a majority of states, although not the US, have expressed support for governing autonomous weapons through a legally binding instrument that would contain a

²⁸ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 15, section 4.1(c)(2).

²⁹ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, pp. 5-6, section 1.2(f). See also “DOD Adopts Ethical Principles for Artificial Intelligence,” DoD press statement, February 24, 2020, <https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/> (accessed February 6, 2023).

³⁰ Department of Defense Directive 3000.09, p.6, section 1.2(f)(5).

³¹ Department of Defense briefing for civil society on the 2023 DoD Directive 3000.09 on Autonomy in Weapon Systems held by conference call, January 30, 2023.

combination of prohibitions and regulations.³² More than 70 countries as well as nongovernmental organizations and the ICRC regard a new treaty with prohibitions and restrictions as necessary, urgent, and achievable. Since 2018, United Nations Secretary-General António Guterres has called for “internationally agreed limits” on weapons systems that could, by themselves, target and attack human beings, describing such weapons as “morally repugnant and politically unacceptable.”³³

The US has opposed proposals to negotiate new international law on autonomous weapons systems and instead argued for voluntary commitments. Its 2021 proposal to use multilateral talks to produce a common “code of conduct” to guide the development and use of autonomous weapons systems, however, won little support.³⁴

As articulated by Human Rights Watch and Harvard Law School’s International Human Rights Clinic, to address the problems of autonomous weapons systems adequately, states should:

- Prohibit the development, production, and use of weapons systems that by their nature select and engage targets without meaningful human control;
- Prohibit the development, production, and use of autonomous weapons systems that target people; and
- Adopt positive obligations (or regulations) to ensure other autonomous weapons systems cannot be used without meaningful human control.³⁵

These basic elements of a new legal instrument have been reiterated in whole or in part by many states, the ICRC, and the Stop Killer Robots campaign, which Human Rights Watch co-founded in 2012. While raised in the context of calls for an international treaty, the elements are equally applicable to national laws and policies. The US DoD directive addresses some of the same areas but fails to reach the bar set by the proposed elements.

The directive fails to incorporate either of the proposed prohibitions because it does not absolutely prohibit the development, production, and use of any autonomous weapons systems.

³² See, for example, International Committee of the Red Cross, “ICRC position on autonomous weapon systems,” May 12, 2021, <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems> (accessed February 2, 2023).

³³ Statement of António Guterres, UN Secretary-General, UN General Assembly, February 6, 2023, <https://www.un.org/sg/en/content/sg/statement/2023-02-06/secretary-generals-briefing-the-general-assembly-priorities-for-2023-scroll-down-for-bilingual-delivered-all-english-and-all-french-versions> (accessed February 6, 2023); and statement of António Guterres, UN Secretary General, Paris Peace Forum, November 11, 2018, <https://www.un.org/sg/en/content/sg/speeches/2018-11-11/address-paris-peace-forum> (accessed February 6, 2023).

³⁴ “International Code of Conduct on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems,” US concept note distributed to the Convention on Conventional Weapons, December 5, 2021.

³⁵ Human Rights Watch and Harvard Law School International Human Rights Clinic, *Crunch Time on Killer Robots: Why New Law Is Needed and How It Can Be Achieved*, December 1, 2021, <https://www.hrw.org/news/2021/12/01/crunch-time-killer-robots> (accessed February 2, 2023).

As explained above, it imposes some restrictions and those activities generally require senior review. It still allows the activities if they are approved, however, and “in cases of urgent military need,” the review process may be waived. The goal of the directive is to “minimize the probability and consequences in autonomous and semi-autonomous weapon systems that could lead to unintended engagements.” It makes clear in a new addition to its definition of failure that “minimizing” failure “does not mean achieving the lowest possible level of risk by never engaging targets.”³⁶ A prohibition, by contrast, might find that “never engaging targets” is appropriate in cases where the weapon system operates without meaningful human control or when it targets people.

While the directive recognizes that humans should have a role in the use of force, it takes a different approach to defining that role. It relies on the term “appropriate levels of human judgment” rather than “human control” or “meaningful human control,” which are more commonly used at the international level.

The 2023 directive, like the proposed treaty elements, acknowledges the difference between systems that target people (anti-personnel systems) and ones that do not (anti-materiel systems), but it does so primarily by loosening rules for the latter. The directive states that senior reviews of autonomous weapon systems are not required for semi-autonomous systems “used to select and engage materiel targets for local defense to intercept attempted time-critical or saturation attacks.”³⁷ While such an exception may be legitimate, the use of autonomous weapons systems to target people is still allowed with senior review, which can be waived in certain circumstances.

The directive does include several requirements that, if properly applied, could help ensure autonomous weapons systems are not used without some level of human involvement. To be truly effective, however, the directive would have to accompany those regulations with prohibitions of autonomous weapons systems that cross moral and legal red lines, and transparency about how the process is implemented.

Ultimately, the United States should adopt a national law or policy, not limited to the Department of Defense, that takes these elements into account and extends them across the government.

³⁶ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 22, glossary.

³⁷ Department of Defense, “Directive 3000.09: Autonomy in Weapon Systems,” January 25, 2023, p. 5, section 1.2(d)(2).