Center for a New American Security I April 2021

Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities

Robert O. Work

Center for a New American Security

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

About the Author

Robert Work was the 32nd Deputy Secretary of Defense, serving alongside three Secretaries of Defense from May 2014 to July 2017. In 2001, he retired as a colonel in the United States Marine Corps after 27 years on active duty. He subsequently was a Senior Fellow and Vice President and Director of Studies at the Center for Strategic and Budgetary Assessments. In May 2009, he was confirmed as the 31st Under Secretary of the Navy in the first Obama administration. Mr. Work stepped down as the Under Secretary in March 2013 to become the chief executive officer for the Center for a New American Security (CNAS). He remained in that position until he assumed the role of Deputy Secretary of Defense in May 2014. He currently is the president and owner of TeamWork, LLC, which specializes in defense strategy and policy, programming and budgeting, military-technical competitions, revolutions in war, and the future of war.

Acknowledgments

I'm grateful to Paul Scharre, Shawn Steene, Jason Stack, and Michael Horowitz for their valuable feedback and suggestions on the report draft. Thank you to Maura McCarthy, Emma Swislow, Melody Cook, Chris Estep, and Megan Lamberth for their role in the review, production, and design of the report. A special thanks to those who participated in the series of CNAS workshops on developing principles for lethal autonomous weapons. Their insights and expertise helped shape this report. Any errors that remain are the responsibility of the author alone.

About CNAS

The mission of CNAS is to develop strong, pragmatic and principled national security and defense policies. Building on the expertise and experience of its staff and advisors, CNAS engages policymakers, experts and the public with innovative, fact-based research, ideas and analysis to shape and elevate the national security debate. A key part of our mission is to inform and prepare the national security leaders of today and tomorrow.

Located in Washington, CNAS was established in February 2007 by cofounders Kurt M. Campbell and Michèle A. Flournoy. CNAS is a 501(c)3 tax-exempt nonprofit organization. Its research is independent and nonpartisan.

As a research and policy institution committed to the highest standards of organizational, intellectual, and personal integrity, CNAS maintains strict intellectual independence and sole editorial direction and control over its ideas, projects, publications, events, and other research activities. CNAS does not take institutional positions on policy issues, and the content of CNAS publications reflects the views of their authors alone. In keeping with its mission and values, CNAS does not engage in lobbying activity and complies fully with all applicable federal, state, and local laws. CNAS will not engage in any representational activities or advocacy on behalf of any entities or interests and, to the extent that the Center accepts funding from non-U.S. sources, its activities will be limited to bona fide scholastic, academic, and research-related activities, consistent with applicable federal law. The Center publicly acknowledges on its website annually all donors who contribute.

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Table of Contents

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Introduction

An international debate over lethal autonomous weapon systems (LAWS) has been under way for nearly a decade.[1] In 2012, the Department of Defense (DoD) issued formal policy guidance on weapon systems with autonomous functionalities,[2] and nations have come together since 2014 to discuss LAWS through the United Nations Convention on Certain Conventional Weapons (CCW). The discussions at the CCW have been hampered by the lack of an agreed-upon definition for LAWS.[3] However, states party to the CCW agreed in 2019 that "human responsibility" for the decisions over the use of weapon systems and the use of force "must be retained."[4] Accordingly, discussions now tend to focus on the type and degree of human involvement required to ensure compliance with international humanitarian law and satisfy ethical concerns.[5]

Several scholars argue these discussions should focus on "developing objective, commonly held, and *function-based understandings of autonomy in the military context*" (emphasis added).[6] The premise of this paper is that the best way to achieve such an understanding is to develop, debate, and agree upon some commonly accepted principles for the employment of weapon systems with autonomous functionalities in armed conflict.[7] This is where the legal, ethical, and moral questions about autonomy in warfare are most acute and deserve the most attention.

This paper offers a starting point for these discussions. The seven principles proposed in this paper are intended to complement and build on existing DoD guidance, including DoD Directive (DoDD) 3000.09, *Autonomy in Weapon Systems*, and DoD's *Artificial Intelligence (AI) Principles*.[8] They are also consistent with the 11 guiding principles adopted in 2019 by the CCW in its "Meeting of the High Contracting Parties to 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."[9] These seven new principles concentrate on the responsible use of autonomous functionalities in armed conflict in ways that preserve human judgment and responsibility over the use of force and help minimize the probability of loss of control of the system or unintended engagements, especially against noncombatants.

This paper is organized into four sections. The first details the history of U.S. weapon systems with autonomous functionalities. It is intended to give the reader a notion of how these weapons have historically been used, why autonomous functionalities are so useful, and why the DoD retains the right to use them. The second section explains why weapons with autonomous functionalities are now being improved through the addition of AI, an important development that aims to make the weapons more discriminate in the application of force. The third section explains why the DoD should consider publishing a new set of principles for the combat employment of weapon systems with autonomous functionalities. The final section outlines seven proposed principles for consideration.

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

A Short History of Weapon Systems with Autonomous Functionalities

The first mass-produced U.S. weapon system with autonomous functionalities in its engagement-related functions was an air-dropped, passive acoustic homing torpedo developed during World War II. The Mk-24 "Fido" made its combat debut in May 1943, using hydrophones arrayed around the midsection of the torpedo to listen for, locate, track, and home in on German U-boats attacking allied transatlantic shipping.[10]

Soon after the war, the U.S. military began to introduce autonomous functionalities into larger weapon systems, particularly air defense combat systems. This move was prompted first by the intense kamikaze raids off Okinawa in 1945 and then accelerated by the threat of atomic air attack on the American homeland. The semi-round environment (SAGE) was designed to direct and control U.S. continental air defense starting in the late 1950s. It could take inputs from a variety of radar sensors dotted around the periphery of the continental United States, autonomously generate "tracks" of reported targets, and highlight to human operators any air defenses within range that were capable of conducting an intercept.[11] The operators then would order the appropriate defenses to engage the targets. Later, SAGE could provide updates directly to "shooters" without intermediate human intervention.[12] The Navy began development of the Naval Tactical Data System (NTDS)—a smaller version of the SAGE built to control the air defense of naval task forces—in 1956.[13]

As computers became smaller, and especially after digital microprocessors appeared, combat control systems of all types—in aircraft, ships, ground combat vehicles, and artillery and missile fire control systems—proliferated across the force. Over time, as technology improved, the military added greater autonomy into engagement-related functions of both munitions and weapon systems, including, but not limited to: acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing when to fire; or providing terminal guidance to home in on selected targets.[14]

As these functions suggest, an engagement is a sequence of actions that ends with an attack on the intended target. Such a sequence is often referred to in military parlance as a "detect-to-engage" sequence, or "kill chain." Eventually, the U.S. military developed, tested, and deployed weapon systems that combined autonomous operations across all engagement-related functions. For munitions, these activities resulted in weapons that, once fired by a human operator, had a degree of self-governance over their behavior that allowed them to complete an attack sequence entirely on their own. These include fire-and-forget guided munitions and two-state fire-and-forget guided munitions.

"*Fire-and-forget" guided munitions* can independently home in on specific targets or aimpoints selected by human operators. Examples include the aforementioned Fido and the Navy's SWOD-9 BAT, which in 1945 became the first autonomous, radar-guided antiship glide bomb used in combat.[15] After the war, fire-and-forget weapons proliferated. The AIM-9 heat-seeking infrared guided air-to-air missile debuted in combat in 1958,[16] laser-guided weapons were first used operationally in the Vietnam War, and GPS-guided missiles and bombs were used during Desert Storm and since.[17]

Two-stage fire-and-forget guided munitions are designed to engage specific groups of concentrated targets selected by human operators. The first stage consists of a guided payload bus that releases guided submunitions over the target group. Each submunition then selects and engages a specific target in the group without human intervention. For example, the Army Tactical Missile System (ATACMS) was designed to deliver six Brilliant Anti-Armor Technology (BAT) submunitions to ranges of 190 miles; each BAT was capable of searching for and attacking enemy armored vehicles. While the first ATACMS entered service in 1991, the variant designed to carry the BAT never was fielded.[18] One two-stage fire-

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

and-forget weapon that was fielded and employed was the air-dropped, CBU-105 Wind-Corrected Munition Dispenser (WCMD) that guides over the target group and releases 40 small Sensor Fuzed Weapons, or "skeets." Each skeet is capable of independently selecting and engaging an armored vehicle, using a combination of laser and infrared sensors. Optimally, a single CBU-105 can attack target groups in an area of 1,500 by 500 feet. However, by releasing the skeets at higher altitude, a WCMD can engage target groups spread over an area of 15 acres. This munition was deployed in 1999 and used during the 2003 invasion of Iraq with devastating effect.[19]

As these examples attest, the U.S. military has incorporated weapon systems with autonomous functionalities for eight decades. They have proven effective, reliable, and safe in combat as part of human-activated kill chains. Consequently, U.S. warfighters long ago gave weapons their proxy to select and engage targets at the end of an engagement sequence, especially when those targets are beyond their line-of-sight. Once activated, the weapon system navigates to the vicinity of the target or specific group of targets, detects them with onboard sensors, classifies and selects a particular target in its field of view, and completes the attack—all on its own. However, because a human selects the target or specific group of targets to be attacked, the DoD considers these weapons to be semi-autonomous.[20] As stated in DoDD 3000.09, semi-autonomous weapon systems "must be designed such that . . . the system does not autonomously select and engage individual targets or specific target groups *that have not been previously selected by an authorized human operator*" (emphasis added).[21]

In contrast, once activated, autonomous weapon systems can select and engage targets that have not been previously designated for attack by a human operator.[22] Such weapons were developed to operate in environments where humans cannot or to search actively for targets over wider areas. Because autonomous weapon systems select and engage targets completely on their own, the risks that such weapons carry out an unintended engagement against friendly or allied forces or noncombatants are higher than for semi-autonomous weapons.

Accordingly, the U.S. military has been cautious about developing and employing such weapon systems, and their operations have been purposely restricted in two ways. First, they are designed to engage only specific classes of targets (e.g., ships or guided missile launchers) coded into sophisticated automated target recognition algorithms. These are pattern-matching algorithms that compare potential target characteristics with a library of approved targets. If a potential target is not in the library, the weapon will not initiate an attack.[23] Also, their search parameters are restricted by the size of their assigned search areas and the duration of an authorized search.

Autonomous weapon systems have come in three distinct types: static search weapons, bounded search weapons, and human-supervised autonomous weapon systems.[24]

Static search weapons include the CAPTOR (encapsulated torpedo), a deep water mine fielded in 1979 during the height of the Cold War.[25] As designed, this weapon system was to be emplaced in deep water, anchored to the ocean floor, and activated. It had its own upward-looking sonar system that ignored surface ships and listened only for submarines. In the event of war, when detecting a hostile (Soviet) submarine, CAPTOR would release its torpedo, which then would home in on the sub and sink it. In other words, once the mine was emplaced and activated, the weapon system could detect, classify, and attack its own target without any further human oversight or intervention.[26] However, the risk of any unintended engagement with CAPTOR was extremely low: There were no civilian objects in the undersea operating domain; the mine's engagement logic ignored surface ships, looking only for a particular type of acoustic signature; and friendly and allied submarines would know the locations of CAPTOR minefields and avoid them.

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Bounded search weapons can surveil a prescribed search area (often called a "kill box") to hunt down and attack imprecisely located groups or classes of targets. These often are referred to as "loitering weapons." Examples include the Tomahawk Anti-Ship Missile (TASM) and Low-Cost Autonomous Attack System (LOCAAS). The radar-guided TASM, fielded in the early 1980s, was fired at an enemy ship on a generated target bearing with an estimated range to target. At the end of its fly out, if it did not detect a target, the TASM would begin a radar search pattern to cover the area of uncertainty resulting from how far the target ship could have moved at maximum speed since weapon launch. The TASM never was used in combat and is no longer in service.[27]

The LOCAAS was developed after Operation Desert Storm to find ballistic missile launchers that were hiding and practicing "shoot and scoot" tactics. It could fly out as far as 70 miles, search a kill box of 62 square miles, and destroy any target found whose signatures matched those in its approved target library. Against closer targets, the LOCAAS could search a larger area since it would have more residual fuel for the search portion of its mission. Although the LOCAAS was developed and successfully tested, it never was fielded because DoD leadership worried the risks of unintended engagements were too high, especially as the search area expanded or the duration of the mission was extended. As explained in DoDD 3000.09, weapon systems with autonomous functionalities needed to be designed "to complete engagements in a time frame consistent with commander and operator intentions and, if unable to do so, to terminate engagements or seek additional human operator input before continuing the engagement."[28] While a data link could solve this problem, it would add additional costs to the system and introduce new operational vulnerabilities. Consequently, the weapon never was fielded.[29]

Human-supervised autonomous weapon systems are systems that, once activated, can select and engage targets on their own but are designed to allow human operators to override their operation if the risk of unintended engagements becomes too high.[30] These "human-on-the-loop" systems include air defense systems that include an automated or automatic mode designed to cope with large air or missile raids that would overwhelm human operators.[31] These types of systems have been around since the 1980s, when the Army introduced the Patriot air and missile defense system and the Navy its Aegis combat system for ship air and missile defense. Although capable of supervised autonomous operations once activated, both systems can revert quickly to human control, if necessary. This is especially important when friendly aircraft are operating in the defended airspace.

One evident difference between these autonomous weapons and the aforementioned semi-autonomous weapons is when the weapon is "activated." For a semi-autonomous weapon, the human chooses the target or specific target group and then activates the weapon. For an autonomous weapon, the human activates the weapon, and the weapon selects and engages its target.

These examples attest that the U.S. military has pursued autonomous weapon systems only for rigidly prescribed situations. By 2009, however, at the start of the Obama administration, autonomous technologies had advanced to the point that weapon designers sought guidance from the Office of the Secretary of Defense on the allowable limits for autonomous functionalities in weapon systems. Such guidance came in the form of the aforementioned DoDD 3000.09, *Autonomy in Weapon Systems*, published in November 2012.

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

The Next Step: Exploiting Improved AI

DoDD 3000.09 established official Department of Defense policy and assigned responsibilities for the development and use of autonomous and semi-autonomous functions in weapon systems, including manned and unmanned systems. The directive requires that commanders and operators always must exercise appropriate levels of human judgment over the use of force.[32] The directive's primary aim was to "minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements."[33]

As required by the law of war, avoiding unintended engagements has long been a high priority for U.S. combat commanders and operators. To date, the primary way that autonomous functionalities in weapon systems have contributed to this goal is by improving the accuracy of both sensors and weapons. The key characteristic of unguided weapons warfare was that most projectiles, bombs, torpedoes, and rockets missed their intended targets, and the miss distance increased rapidly over range. Weapon accuracy was measured by circular error probable (CEP), the radius of a circle, centered on the intended target, in which 50 percent of all shots fired fall. For example, the CEP of U.S. bombs dropped over Germany in World War II was 3,300 feet.[34] As a result, the U.S. Army Air Corps concentrated formations of up to 1,000 bombers over a target to increase the statistical probability that the intended target actually would be hit. And, as half of all bombs dropped exploded more than 3,300 feet away from their targets, collateral damage to civilians and civilian infrastructure was an expected and accepted fact of warfare.

Now, however, improved autonomous functionalities in navigation, target identification, and mid-course and terminal guidance have led to a wholesale shift to guided weapons that are far more accurate than previous generations of unguided weapons, with average miss distances of tens of feet or less regardless of the range to target. Guided munitions therefore allow for smaller but more accurate salvos, cutting collateral damage substantially. Moreover, increased accuracy allows for smaller warheads to achieve the same desired effect on target, which reduces collateral damage even more.[35]

The next advancement in weapon development will be the introduction of improved AI-enabled autonomous functionalities. One expectation is that "intelligent weapons" will allow for new collaborative weapons that can share target information and autonomously coordinate their strikes after launch. Such collaborative weapon salvos will help confuse, overwhelm, or evade enemy defenses, and compensate for weapons lost to enemy defenses. This will allow attack planners to further reduce the size of a salvo necessary to achieve effects on a target.[36] AI-enabled autonomous functionalities also will allow a special type of collaborative attack using swarms of small, low-cost munitions, which also will present defenses with difficult problems.[37] These new AI-enabled functionalities are expected to help conserve U.S. joint force "magazine depth," which is critical for overall force effectiveness and staying power in expeditionary operations.[38]

AI-enabled functionalities also are likely to help mitigate the biggest cause of unintended combat engagements: target misidentification. In an analysis of combat operations in Afghanistan, target misidentifications were the cause of about half of all U.S.-caused civilian casualties.[39] The majority of these misidentifications were made by human operators. Target misidentification also is a leading cause for fratricide (i.e., friendly units firing on friendly or allied units). AI-enabled control systems can improve target discrimination in certain domains, such as air defense and air combat, reducing both civilian casualties and friendly fire. For example, the USS *Vincennes* shootdown of Iran Air Flight 655 in 1988, which killed all 290 civilians on board, was due to cognitive overload of human commanders on board the *Vincennes*, who were dealing with simultaneous threats from enemy aircraft and gunboats near a commercial airway.[40] Improved autonomous functionality to help fuse and process data might have prevented the incident.[41]

AI-enabled autonomous identification and terminal guidance functions thus have the potential to dramatically improve target identification and discrimination, resulting in:

- fewer "blue-on-blue" incidents (unintended attacks on friendly U.S. units);
- fewer "blue-on-green" incidents (unintended attacks on friendly allied and partner forces);
- fewer unintended engagements of noncombatants, with a reduction in civilian casualties; and
- less damage to civilian infrastructure.

For these reasons, the DoD continues to pursue the promise of weapon systems with improved AI-enabled autonomous functionalities. Eight decades of combat experience demonstrate that, if used appropriately, autonomous functionalities combined with human-machine teaming can continue to improve the discriminate use of force on the battlefield. Moreover, the DoD's cautious deployment to date of fully autonomous functionalities in weapons demonstrates its ability to employ such weapons in ways consistent with the laws of war and moral and ethical obligations.

Nevertheless, the U.S. military is ever mindful of the need to verify the combat reliability and safety of weapon systems with autonomous functionalities. It is working to improve its test, evaluation, validation, and verification (TEVV) procedures to protect against security and safety vulnerabilities. Commanders and operators also must guard against expecting too much from AI given its current brittleness when confronted by unexpected circumstances or changing context.[42] Thus, improved training and understanding of the capabilities and limits of AI-enabled weapon systems are necessary going forward. But the historical record clearly shows that the U.S. military has demonstrated its willingness to scrap or forgo deployment of promising new weapon systems that cannot confidently be deemed capable of being used in compliance with the law of war or are judged to be too risky for operational use (i.e., LOCAAS).

One clarifying point in this regard: Some who read DoDD 3000.09 conclude that DoD policy is that all weapon systems with autonomous functionalities must be controlled by either a human-in-the-loop or human-on-the-loop during the entire engagement sequence.[43] In the former case, the weapon system would perform a task in the engagement sequence and await the human user to take an action before continuing.[44] And as previously discussed, while a human-on-the-loop weapon system can sense, decide, and act on its own, a human supervises its operation and can intervene and abort its operation, if desired.[45] In fact, DoDD 3000.09 does *not* mandate human-in-the-loop or on-the-loop control schemes. Instead, it establishes broad policies and an internal bureaucratic process for senior leaders to approve or reject novel uses of autonomy in weapons, including fully autonomous weapons. Nevertheless, some have insisted "meaningful" human control should require the ability to intervene and deactivate the weapon at any step in the engagement sequence, mandating a human-in- or on-the-loop.[46]

However, human accountability for the results of engagements of weapons does not and should not necessarily mandate human oversight over every step of the kill chain. Once an operator initiates an engagement against a target or group of targets expected to end in the application of lethal force, then subsequent steps in the attack sequence may be completed autonomously without further human oversight. If there is significant uncertainty in the behavior and outcomes of one or more steps of an engagement plan, humans must take responsibility for the uncertainty and associated variance of outcomes. When feasible and valuable, system design can include points of human observation and guidance at intermediate steps in a sequence of automated actions.[47] At such points, a human controller would review the system's status and decide whether to move forward (e.g., stop, continue execution, or modify a plan). But a blanket policy requiring real-time human supervision with the ability to deactivate systems in all instances is neither realistic nor desirable. Indeed, such a policy instead could spur commanders to use less precise, unguided weapon systems that might result in greater levels of collateral damage.

For example, imagine if a wind-corrected munition dispenser navigated over a group of targets and released 40 skeets. The time between the release of the skeets and their attacks is measured in seconds. Requiring a human-in-the-loop would therefore require 40 human operators to monitor the action of one skeet and permit or abort its attack—a prohibitive personnel requirement. As this example suggests, requiring human-in- or on-the-loop control schemes for every single step of a weapon system with autonomous selection and engagement functions would be impractical and extraordinarily burdensome in combat operations—establishing a standard that has not been required even for unguided weapons. For this reason, these control schemes are discretionary, not mandatory, in DoD policy. They are implemented when a weapon's expected tactics, techniques, and procedures call for heightened human supervision.

Toward Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities

The law of war does not specifically prohibit or restrict the use of autonomy to aid in the operation of weapons.[48] Neither does it expressly approve of its use. DoD policy is that any and all weapons, including weapon systems with autonomous functionalities, must be developed and used in compliance with the law of war, policy, applicable treaties, weapon system safety rules, ethical guidance, and rules of engagement. Weapon systems with autonomous engagement functionalities have met this standard for eight decades.

The DoD's current policy guidance on autonomy in weapons, DoDD 3000.09, gives the DoD the freedom to pursue and employ new, more advanced munitions and weapon systems with AI-enabled autonomous functionalities, including fully autonomous weapons. It also outlines the internal departmental process to ensure their responsible design, test, evaluation, approval, and use—a process that remains in place and is useful to this day.

As outlined in DoDD 3000.09, *Autonomy in Weapon Systems*, the development and TEVV of any munition or weapon system with autonomous functionalities must demonstrate that it can reliably and repeatedly meet mission objectives in realistic operating environments while conforming to the law of war, policy, applicable treaties, weapon system safety rules, ethical guidance, and rules of engagement. In addition to TEVV, a separate legal review of the weapon and its intended use also is required to ensure compliance with the law of war and DoD policy, as is the case for all weapons developed by the DoD.[49] A specific goal of these activities is to minimize the probability and consequences of failures that could lead to unintended engagements, especially against civilians, civilian objects and infrastructure, and other protected entities.

Given these circumstances, it is reasonable to ask why additional principles for weapons with autonomous functionalities are needed. There are two interrelated reasons. First, since DoDD 3000.09's adoption in 2012, the understanding of autonomous functionalities and the use of AI in weapons have matured considerably, and the debate over LAWS has become sharper and broader. Consequently, the time is ripe for DoD to demonstrate leadership on weapons with autonomous functionalities by working to establish norms for their employment.

Second, additional guidance is needed because existing policies are not specific enough. Beyond very broad guidance such as ensuring that weapons shall be designed to allow "appropriate levels of human judgment over the use of force," DoDD 3000.09 does not delve deeply into the connection between a human decision to employ a weapon with autonomous functionalities and its subsequent actions.

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Similarly, the recently published *DoD AI Principles* provide high-level guidance for how the Department should approach AI, but not on how to use AI-enabled autonomous functionalities in armed conflict.

Accordingly, the principles proposed below are intended to build on both DoDD 3000.09 and the *DoD AI Principles* by giving additional guidance for the battlefield employment of semi-autonomous and autonomous munitions and weapon systems. Consistent with DoD policy, a key focus of these principles is to preserve human judgment over the use of force in armed conflict and to minimize the probability and consequences of failures that could lead to unintended engagements, especially against noncombatants.

The DoD, working with the White House, Department of State, and other relevant federal agencies, should consider adopting these principles to help guide the combat employment of weapon systems with autonomous functionalities and to shape U.S. positions in international discussions on these types of weapons.

Proposed DoD Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities

While TEVV and legal reviews ensure baseline compliance with the law of war, policy, applicable treaties, weapon system safety rules, ethical guidance, and rules of engagement, weapon systems with autonomous functionalities raise additional questions regarding the appropriate scope of human judgment over the use of force and how to further minimize unintended engagements. The following principles are intended to provide guidance on these questions.

Nothing in these principles is intended to contradict existing laws or policies.

1. **Any use of weapon systems with autonomous functionalities must be guided and overseen by a responsible chain of human command and control.** This chain must lay out objectives, methods, rules of engagement, special instructions, and expressed limitations to ensure all weapons use, including any with autonomous behavior, meets mission objectives while conforming to law of war, policy, applicable treaties, weapon system safety rules, ethical guidance, and rules of engagement.

2. **Decisions to initiate a sequence of actions, including autonomous actions, that may result in the loss of human life through the use of force (i.e., a kill chain) are the sole province of human intent and judgment.** Whether mediated by humans or machines, all acts, but especially acts related to the use of force, always must be governed by the chain of responsible human command and control.[50] This includes decisions to activate autonomous weapon systems that can select and engage targets without further human intervention.

3. **Human responsibility for decisions over the use of force cannot be transferred to machines under any circumstances.** Human beings are responsible for law of war obligations such as distinction, proportionality, and precautions in attack. The law of armed conflict does not allow weapons to make legal determinations. Rather, it is persons who must comply with the law of war; only they are accountable for their determinations and decisions.[51]

4. **To make a valid determination about the lawfulness of an attack on a specific target, any person who authorizes the use of, directs the use of, or operates weapon systems with autonomous functionalities must have sufficient information about the system's expected performance and capabilities, doctrine for use, the intended target, the environment, and the context for use (e.g., the presence of noncombatants in the engagement area).**[52] Clear doctrine, tactics, techniques, and procedures and adequate training are necessary for commanders and operators to understand the functions, capabilities, and limitations of a weapon system's autonomy in realistic operational conditions.[53]

5. **Once a human being initiates a sequence of actions that is intended to end with the application of lethal force, weapon systems with autonomous functionalities may complete the sequence on their own without further human oversight.** This includes autonomously detecting, classifying, and engaging targets or specific groups of targets designated for attack by human operators, in a manner consistent with weapon system performance and within authorized sets of legal, ethical, operational, spatial, and temporal bounds.

6. **As long as a weapon system's selection and engagement of a target occurs as part of a sequence of actions tied directly to a deliberate human decision to carry out a lawful attack, the standard of appropriate human judgment over the use of lethal force is met.** Once such a decision is made, as with the use of weapon systems with autonomous functionalities today, direct control of every single step in the subsequent engagement sequence would be impractical and would impose undue burdens on operators engaged in combat. As such, human on-the-loop or in-the-loop control schemes are discretionary, not mandatory; they are contextually determined by temporal and spatial parameters and are implemented consistent with expected weapon use and as necessary to ensure compliance with these principles.

7. **Commanders must take appropriate action if they obtain evidence that weapon systems with autonomous functionalities may be operating in a manner contrary to expected performance, the law of war, policy, applicable treaties, ethical guidance, and rules of engagement.** Any unintended engagement against noncombatants must be investigated to determine its causes—which might include, but are not limited to, faulty weapon design, inadequate testing of possible failure modes, operator error/improper weapon employment, poor operator training, faulty intelligence, target misidentification, weapon malfunction, or adversary action (i.e., hacking, spoofing).

Conclusion

Weapon systems with autonomous functionalities have been used safely and reliably in combat for eight decades. They will continue to be used in the future. Indeed, the addition of AI-enabled applications into these weapon systems is expected to make them even more discriminate in the application of force and lead to a reduction in unintended engagements—an aim entirely consistent with international humanitarian law.

Nevertheless, opponents of these weapons are concerned that their use will lead to problematic ethical, moral, and legal outcomes in armed conflict. The United States should be at the forefront of advanced TEVV protocols and legal reviews to demonstrate that weapons with autonomous functionalities will perform as they are intended. The United States also should strive to demonstrate that it is committed to employing weapons in ways that can meet mission objectives while conforming to the law of war, policy, applicable treaties, weapon system safety rules, ethical guidance, and rules of engagement. One way to do this is to adopt principles for the combat employment of weapon systems with autonomous functionalities and institutionalize these principles through acquisition processes, training, education, and field exercises. The seven proposed principles are meant to jump-start such an effort and provide the foundation for the adoption of international norms.

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

1. In 2012, the DoD defined an autonomous weapon system as one that, "once activated, can *select* and *engage* targets without further intervention by a human operator" (emphasis added). As autonomous weapon systems could be designed to apply lethal or nonlethal, kinetic or nonkinetic force, a *lethal* autonomous weapon system refers only to one that can take human life. Department of Defense Directive 3000.09 (hereafter DoDD 3000.09), *Autonomy in Weapon Systems*, November 21, 2012, https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf, 13.
2. A weapon system is, "A combination of one or more weapons with all related equipment, material, services, personnel and means of delivery and deployment (if applicable) required for self-sufficiency." Weapon systems include simple devices operated manually by a single man or woman (e.g., a rifle), and munitions. A munition is defined as a "complete device charged with explosives, propellants, pyrotechnics, initiating composition; or chemical, biological, radiological, or nuclear material for use in operations including demolitions." Whereas weapon systems generally are designed for sustained use, a munition generally is a single-use, expendable weapon designed to impart kinetic effect on a target (e.g., shell, bomb, missile, torpedo, etc.). Newer single-use, expendable weapons include those designed to impart nonkinetic effects, such as jamming. Both definitions can be found in: Department of Defense, *DOD Dictionary of Military and Associated Terms,* (January 2021), https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf.
3. See Austin Wyatt, PhD, *So Just What is a Killer Robot? Detailing the Ongoing Debate around Defining Lethal Autonomous Weapon Systems*, Washington Headquarters Services, Department of Defense, June 8, 2020, https://www.whs.mil/News/News-Display/Article/2210967/so-just-what-is-a-killer-robot-detailing-the-ongoing-debate-around-defining-let/.
4. "Autonomy, artificial intelligence and robotics: Technical aspects of human control," (International Committee of the Red Cross, August 2019), 4.
5. See for example, Daniele Amoroso and Guglielmo Tamburrini, "What Makes Human Control over Weapons Systems 'Meaningful'?" (International Committee for Robot Arms Control, August 2019), https://www.icrac.net/wp-content/uploads/2019/08/Amoroso-Tamburrini_Human-Control_ICRAC-WP4.pdf.
6. Wyatt, *So Just What is a Killer Robot?*
7. This paper adopts the same position as DoDD 3000.09; as used herein, "weapon systems" does not apply to autonomous or semi-autonomous cyberspace systems for cyberspace operations; unarmed, unmanned platforms; or unguided munitions; DoDD 3000.09, 2.
8. Department of Defense, *DoD Adopts Ethical Principles for Artificial Intelligence*, (February 24, 2020), https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.
9. Annex III, "Meeting of the High Contracting Parties to 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 Effect," November 13–15, 2019, 10.
10. Barry D. Watts, *Six Decades of Guided Weapons and Battle Networks, Progress and Prospects* (Washington: Center for Strategic and Budgetary Assessments, March 2007), 3, 103–4.
11. A track is, "1. A series of related contacts displayed on a data display console or other display device. 2. To display or record the successive positions of a moving object."; Department of Defense, *DOD Dictionary of Military and Associated Terms*.
12. Watts, *Six Decades of Guided Weapons and Battle Networks, Progress and Prospects,* 123–26.
13. Norman Friedman, *U.S. Destroyers*, revised edition (Annapolis, MD: Naval Institute Press, 2004), 207.
14. DoDD 3000.09, 14.
15. Michael Peck, "How the ASM-N-2 Bat Gave Birth to a New Class of Anti-Ship Missiles," *The National Interest*, May 10, 2020, https://nationalinterest.org/blog/buzz/how-asm-n-2-bat-gave-birth-new-class-anti-ship-missiles-152576.
16. The AIM-9 Sidewinder was first used in air combat between the Republic of China (Taiwan) and the People's Republic of China; Watts, *Six Decades of Guided Weapons and Battle Networks, Progress and Prospects*, 127.
17. For a thorough discussion of laser and GPS guided bombs, see Watts, *Six Decades of Guided Weapons and Battle Networks, Progress and Prospects*.
18. "ATACMS Block II/Brilliant Anti-armor Technology (BAT), Federation of American Scientists, https://fas.org/man/dod-101/sys/land/atacms-bat.htm.
19. N.R. Jenzen-Jones, "CBU-97/CBU-105 'Sensor Fuzed Weapon' Cluster Munition," Armament Research Services, https://armamentresearch.com/us-cbu-97cbu-105-sensor-fuzed-weapon-cluster-munition/; Area coverage figures from "CBU-97 Sensor Fuzed Weapon," Wikipedia, https://en.wikipedia.org/wiki/CBU-97_Sensor_Fuzed_Weapon.
20. A semi-autonomous weapon is "a weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator."; DoDD 3000.09, 14.
21. DoDD 3000.09, 3; Some interpret the wording as saying that since the selection of specific targets within a group were not previously selected by the operator, this type of weapon is not consistent with DoD policy. However, the DoD directive explicitly states that semi-autonomous weapon systems include those in which human operators select "specific target groups." See DoDD 3000.09, 3, 14. Once a human approves an attack on a specific group of targets, the destruction of any target within the group can be done autonomously, and the weapon is still considered a semi-autonomous weapon system.
22. "A weapon system that, once activated, can select and engage targets without further intervention by a human operator." DoDD 3000.09, 13.
23. For an explanation of automatic target recognition systems, see James A. Ratches, "Review of current aided/automatic target acquisition technology for military target acquisition tasks," *Optical Engineering,* 50 no. 7 (March 2011): https://www.spiedigitallibrary.org/journals/optical-engineering/volume-50/issue-07/072001/Review-of-current-aided-automatic-target-acquisition-technology-for-military/10.1117/1.3601879.full?SSO=1.
24. The terms "static search weapons" and "bounded search weapons" are not drawn from DoDD 3000.09. They are adopted here to help readers understand the different types of weapons that have been fielded by DoD.
25. DoDD 3000.09 excludes sea and land mines from its purview. Therefore, although it otherwise possesses the attributes of an autonomous weapon, the CAPTOR is not officially considered to be one; DODD 3000.09, 2.
26. See "Mark 60 CAPTOR," Weapon Systems.net, https://weaponsystems.net/system/449-Mark+60+CAPTOR.
27. See Carlo Kopp, "Tomahawk Cruise Missile Variants, BGM/RGM-109B Tomahawk Anti-Ship Missile (TASM)," Air Power Australia, http://www.ausairpower.net/Tomahawk-Subtypes.html.
28. DoDD 3000.09, 7.
29. Watts, *Six Decades of Guided Weapons and Battle Networks, Progress and Prospects*, 281–83.
30. DoDD 3000.09, 14.

[31] In human-on-the-loop weapon systems, the system can sense, decide, and act on its own, but a human supervises its operation and can intervene and abort its operation, if desired; Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W.W. Norton & Company, 2008), 29.
[32] DoDD 3000.09, 2.
[33] DoDD 3000.09, 1.
[34] U.S. Air Force, "Historical PGM Analysis—Increasing Precision of Munitions," PowerPoint presentation.
[35] U.S. Air Force, "Historical PGM Analysis."
[36] Theresa Hitchens, "AFRL's Golden Horde 'Swarms" Would Increase Fighter Lethality," Breaking Defense, January 8, 2021, https://breakingdefense.com/2021/01/afrls-golden-horde-swarms-would-increase-fighter-lethality/; See also Defense Advanced Research Project Agency, "CODE Demonstrates Autonomy and Collaboration with Minimal Human Commands," https://www.darpa.mil/news-events/2018-11-19.
[37] John Arquilla and David Ronfeldt, *Swarming and the Future of Conflict*, (RAND https://www.rand.org/content/dam/rand/pubs/documented_briefings/2005/RAND_DB311.pdf.
[38] The U.S. Joint Force generally will fight an "away game." It brings its fighting networks with it, to include all its ammunition and munitions. Conserving their expenditures is important to keep the force in the fight.
[39] Lawrence "Larry" Lewis, an analyst at the Center for Naval Analyses, has analyzed over 1,000 incidents in Afghanistan resulting in civilian casualties. He found that in 50 percent of the incidents, the person making the engagement decision accidently misidentified civilians as a lawful target; Larry Lewis, *Redefining Human Control: Lessons from the Battlefield for Autonomous Weapons* (Arlington, VA: Center for Autonomy and AI, March 2018); and Larry Lewis email to the author dated May 21, 2020.
[40] "Shooting Down of Iran Air 655," Four Corners on Australia Broadcasting Corporation, 2000, https://www.youtube.com/watch?v=Onk_wI3ZVME.
[41] Scharre, *Army of None*, 169–70.
[42] This is especially true for AI-enabled systems that rely on machine learning.
[43] See, for example, Sydney J. Freedberg Jr., "Fear and Loathing in AI: How the Army Triggered Fears of Killer Robots," Breaking Defense, https://breakingdefense.com/2019/03/fear-loathing-in-ai-how-the-army-triggered-fears-of-killer-robots/.
[44] Scharre, *Army of None*, 29.
[45] Scharre, *Army of None*, 29.
[46] "Human-on-the-loop supervision, intervention and the ability to deactivate are absolute minimum requirements for countering this risk, but the system must be designed to allow for meaningful, timely, human intervention – and even that is no panacea."; "Autonomy, artificial intelligence and robotics: Technical aspects of human control," 3.
[47] Eric Horvitz, Commissioner on the National Security Commission on Artificial Intelligence, in emails to the author, January 3, 2021.
[48] Department of Defense, *Law of War Manual* (Washington: Office of the General Counsel, Department of Defense, June 2015; updated December 2016), 353.
[49] DoDD 3000.09, 7.
[50] Defense Science Board, Summer Study on Autonomy (Washington: Department of Defense, June 2016), 16.
[51] Department of Defense, *Law of War Manual*, 354.
[52] DoDD 3000.09, 7–8.
[53] DoDD 3000.09, 7–8.

Center for a New American Security
1152 15th Street NW, Suite 950, Washington, DC 20005
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc