June 11, 2021

Documents Reflecting U.S. Practice Related to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems.

The United States appreciates the Chair's invitation to share relevant national policies and practices.

The United States has long advocated for the sharing of national practices and policies related to the implementation of international humanitarian law (IHL). IHL establishes rules governing the use of weapon systems in armed conflict, no matter the type of technology incorporated in the weapon systems. Through robust national implementation measures, States can ensure the effective implementation of IHL, and through the sharing of practices, practitioners in one State can benefit from lessons learned in another.

In the context of the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts (GGE) on emerging technologies in the area of lethal autonomous weapons systems, the United States has appreciated the efforts of other States to share their practice, including a growing number of States that are developing and publicly articulating their national policies on ensuring the responsible use of emerging technologies. The United States has also sought to share U.S. practice. For example, the United States has shared in past GGE discussions U.S. practice:

- on the *Counter-Rocket, Artillery, and Mortar System* (April 2018);
- on a system to counter naval mines, the *Single Sortie Detect to Engage* (Sept. 2018); and
- on the *AN/TPQ-53 Counterfire Radar System* (March 2019).

More generally, U.S. Department of Defense (DoD) Directive 3000.09, *Autonomy in Weapon Systems*, provides policy guidance on the development and use of autonomous and semi-autonomous functions in weapon systems. (Attachment 1). For example, this policy establishes guidelines designed to minimize failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements.

DoD also requires the legal review of weapon systems. DoD Directive 2311.01, *DoD Law of War Program*, provides that "[t]he intended acquisition, procurement, or modification of weapons or weapon systems is reviewed for consistency with the law of war." ¶1.2.d. The Directive also establishes the DoD Law of War Working Group, which among other functions, develops and coordinates law of war "analysis regarding the legality of new means or methods of warfare under consideration by DoD components." ¶3.1. Under DoD Directive 2311.01, the *DoD Law of War Manual* serves as the authoritative statement on the law of war within DoD and includes extensive guidance in Chapter VI regarding the legal review of new weapons and legal rules specific to certain types of weapons. In addition, DoD Directive 5000.01, *The Defense Acquisition System*, provides that:

> The acquisition and procurement of DoD weapons and information systems must be consistent with all applicable domestic law, and the resulting systems must comply with applicable treaties and international agreements (for arms control agreements, see DoD Directive (DoDD) 2060.01), customary international law, and the law of armed conflict (also known as the laws and customs of war). An attorney

authorized to conduct such legal reviews in the DoD must conduct the legal review of the intended acquisition of weapons or weapons systems.

¶1.2v. The Military Departments within DoD have implemented this requirement and provided further guidance on the legal review of weapons in issuances for their personnel:

- Department of the Army: Army Regulation 27-53, *Legal Review of Weapons and Weapon Systems*, Sept. 23, 2019.
- Department of the Navy: Paragraph 10 of Enclosure 3 of SECNAV Instruction 5000.2F, *Defense Acquisition System and Joint Capabilities Integration and Development System Implementation*, March 26, 2019, addresses "Mandatory Legal Review of Potential Weapons & Weapon Systems."
- Department of the Air Force: Part 2 of Air Force Instruction 51-401, *The Law of War*, Aug. 3, 2018, addresses "Legal Reviews of Weapons and Cyber Capabilities." *See also* Air Force Policy Directive 51-4, *Operations and International Law*, July 24, 2018.

U.S. military practice in conducting the legal review of weapons was described extensively in a *2017 DoD submission* to a study by the Stockholm International Peace Research Institute (SIPRI).

Artificial Intelligence (AI) has received attention both within DoD and in our GGE discussions. AI applications across all sectors of life, such as transportation and health care, present the possibility of great benefits to society, especially in light of the rapid pace of ongoing developments in this field. However, there are also concerns that AI could be misused or misapplied. Organizations across various sectors are seeking proactively to develop principles to guide the responsible development and use of AI. Similarly, a key focus area of *DoD's Strategy on Artificial Intelligence* is "Leading in military ethics and AI safety." In February 2020, the Secretary of Defense reaffirmed "that the Department will use AI consistent with applicable domestic and international law, in particular the law of war and adopted Artificial Intelligence Ethical Principles for the Department of Defense. (Attachment 2). On May 26, 2021, the Deputy Secretary of Defense reaffirmed these principles and established DoD's holistic, integrated, and disciplined approach for implementing them, including articulating six foundational tenets. (Attachment 3). As AI has the potential to transform positively a whole spectrum of DoD activities, these memos are not limited to weapon systems. The DoD AI Ethical Principles apply to all DoD AI capabilities, of any scale, including AI-enabled autonomous systems, for warfighting and business applications.

We have shared these references and documents in order: (1) to continue to provide transparency on U.S. practice; (2) to encourage others to share their practice and to consider U.S. practice; and (3) to facilitate further discussion about the development and use of emerging technologies in the area of lethal autonomous weapons systems. We are happy to discuss U.S. practice with other delegations to the GGE.

Attachments:

1. U.S. Department of Defense Directive 3000.09, *Autonomy in Weapons Systems*, Nov. 21, 2012, incorporating Change 1, May 8, 2017

2. Secretary of Defense, *Artificial Intelligence Ethical Principles for the Department of Defense*, Feb. 21, 2020
3. Deputy Secretary of Defense, *Implementing Responsible Artificial Intelligence in the Department of Defense*, May 26, 2021

Tab 1

(This page intentionally left blank)

Department of Defense

DIRECTIVE

NUMBER 3000.09
November 21, 2012
Incorporating Change 1, May 8, 2017

USD(P)

SUBJECT: Autonomy in Weapon Systems

References: See Enclosure 1

1. PURPOSE. This Directive:

a. Establishes DoD policy and assigns responsibilities for the development and use of autonomous and semi-autonomous functions in weapon systems, including manned and unmanned platforms.

b. Establishes guidelines designed to minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements.

2. APPLICABILITY. This Directive:

a. Applies to:

(1) OSD, the Military Departments, the Office of the Chairman of the Joint Chiefs of Staff and the Joint Staff (CJCS), the Combatant Commands, the Office of the Inspector General of the Department of Defense, the Defense Agencies, the DoD Field Activities, and all other organizational entities within the DoD (hereinafter referred to collectively as the "DoD Components").

(2) The design, development, acquisition, testing, fielding, and employment of autonomous and semi-autonomous weapon systems, including guided munitions that can independently select and discriminate targets.

(3) The application of lethal or non-lethal, kinetic or non-kinetic, force by autonomous or semi-autonomous weapon systems.

b. Does not apply to autonomous or semi-autonomous cyberspace systems for cyberspace operations; unarmed, unmanned platforms; unguided munitions; munitions manually guided by the operator (e.g., laser- or wire-guided munitions); mines; or unexploded explosive ordnance.

3. DEFINITIONS. See Glossary.

4. POLICY. It is DoD policy that:

a. Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.

(1) Systems will go through rigorous hardware and software verification and validation (V&V) and realistic system developmental and operational test and evaluation (T&E) in accordance with the guidelines in Enclosure 2. Training, doctrine, and tactics, techniques, and procedures (TTPs) will be established. These measures will ensure that autonomous and semi-autonomous weapon systems:

(a) Function as anticipated in realistic operational environments against adaptive adversaries.

(b) Complete engagements in a timeframe consistent with commander and operator intentions and, if unable to do so, terminate engagements or seek additional human operator input before continuing the engagement.

(c) Are sufficiently robust to minimize failures that could lead to unintended engagements or to loss of control of the system to unauthorized parties.

(2) Consistent with the potential consequences of an unintended engagement or loss of control of the system to unauthorized parties, physical hardware and software will be designed with appropriate:

(a) Safeties, anti-tamper mechanisms, and information assurance in accordance with DoD Instruction 8500.01 (Reference (a)).

(b) Human-machine interfaces and controls.

(3) In order for operators to make informed and appropriate decisions in engaging targets, the interface between people and machines for autonomous and semi-autonomous weapon systems shall:

(a) Be readily understandable to trained operators.

(b) Provide traceable feedback on system status.

(c) Provide clear procedures for trained operators to activate and deactivate system functions.

b. Persons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement (ROE).

c. Autonomous and semi-autonomous weapon systems intended to be used in a manner that falls within the policies in subparagraphs 4.c.(1) through 4.c.(3) will be considered for approval in accordance with the approval procedures in DoD Directive 5000.01 (Reference (b)), DoD Instruction 5000.02 (Reference (c)), and other applicable policies and issuances.

(1) Semi-autonomous weapon systems (including manned or unmanned platforms, munitions, or sub-munitions that function as semi-autonomous weapon systems or as subcomponents of semi-autonomous weapon systems) may be used to apply lethal or non-lethal, kinetic or non-kinetic force. Semi-autonomous weapon systems that are onboard or integrated with unmanned platforms must be designed such that, in the event of degraded or lost communications, 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.

(2) Human-supervised autonomous weapon systems may be used to select and engage targets, with the exception of selecting humans as targets, for local defense to intercept attempted time-critical or saturation attacks for:

(a) Static defense of manned installations.

(b) Onboard defense of manned platforms.

(3) Autonomous weapon systems may be used to apply non-lethal, non-kinetic force, such as some forms of electronic attack, against materiel targets in accordance with DoD Directive 3000.03E (Reference (d)).

d. Autonomous or semi-autonomous weapon systems intended to be used in a manner that falls outside the policies in subparagraphs 4.c.(1) through 4.c.(3) must be approved by the Under Secretary of Defense for Policy (USD(P)); the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)); and the CJCS before formal development and again before fielding in accordance with the guidelines in Enclosure 3, References (b) and (c), and other applicable policies and issuances.

e. International sales or transfers of autonomous and semi-autonomous weapon systems will be approved in accordance with existing technology security and foreign disclosure requirements and processes, in accordance with DoD Directive 5111.21 (Reference (e)).

5. RESPONSIBILITIES. See Enclosure 4.

6. RELEASABILITY. Cleared for public release. This Directive is available on the DoD Issuances Website at http://www.dtic.mil/whs/directives.

7. SUMMARY OF CHANGE 1. The changes to this issuance are administrative and update organizational titles and references for accuracy.

8. EFFECTIVE DATE. This Directive is effective November 21, 2012.

Ashton B. Carter
Deputy Secretary of Defense

Enclosures
1. References
2. V&V and T&E of Autonomous and Semi-Autonomous Weapon Systems
3. Guidelines for Review of Certain Autonomous or Semi-Autonomous Weapon Systems
4. Responsibilities

Glossary

ENCLOSURE 1

REFERENCES

(a) DoD Instruction 8500.01, "Cybersecurity," March 14, 2014
(b) DoD Directive 5000.01, "The Defense Acquisition System," May 12, 2003, as amneded
(c) DoD Instruction 5000.02, "Operation of the Defense Acquisition System," January 7, 2015, as amended
(d) DoD Directive 3000.03E, "DoD Executive Agent for Non-Lethal Weapons (NLW), and NLW Policy," April 25, 2013, as amended
(e) DoD Directive 5111.21, "Arms Transfer and Technology Release Senior Steering Group and Technology Security and Foreign Disclosure Office," October 14, 2014
(f) DoD Directive 2311.01E, "DoD Law of War Program," May 9, 2006, as amended
(g) DoD Directive 1322.18, "Military Training," January 13, 2009, as amended

ENCLOSURE 2

V&V AND T&E OF AUTONOMOUS AND SEMI-AUTONOMOUS WEAPON SYSTEMS

To ensure autonomous and semi-autonomous weapon systems function as anticipated in realistic operational environments against adaptive adversaries and are sufficiently robust to minimize failures that could lead to unintended engagements or to loss of control of the system, in accordance with subparagraph 4.a.(1) above the signature of this Directive:

a. Systems will go through rigorous hardware and software V&V and realistic system developmental and operational T&E, including analysis of unanticipated emergent behavior resulting from the effects of complex operational environments on autonomous or semi-autonomous systems.

b. After initial operational test and evaluation (IOT&E), any further changes to the system will undergo V&V and T&E in order to ensure that critical safety features have not been degraded.

(1) A regression test of the software shall be applied to validate critical safety features have not been degraded. Automated regression testing tools will be used whenever feasible. The regression testing shall identify any new operating states and changes in the state transition matrix of the autonomous or semi-autonomous weapon system.

(2) Each new or revised operating state shall undergo integrated T&E to characterize the system behavior in that new operating state. Changes to the state transition matrix may require whole system follow-on operational T&E, as directed by the Director of Operational Test and Evaluation (DOT&E).

ENCLOSURE 3

GUIDELINES FOR REVIEW OF CERTAIN AUTONOMOUS OR SEMI-AUTONOMOUS WEAPON SYSTEMS

1. Autonomous or semi-autonomous weapon systems intended to be used in a manner that falls outside the policies in subparagraphs 4.c.(1) through 4.c.(3) above the signature of this Directive must be approved by the USD(P), USD(AT&L), and CJCS before formal development and again before fielding.

a. Before a decision to enter into formal development, the USD(P), USD(AT&L), and CJCS shall ensure:

(1) The system design incorporates the necessary capabilities to allow commanders and operators to exercise appropriate levels of human judgment in the use of force.

(2) The system is designed to complete engagements in a timeframe 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.

(3) The system design, including safeties, anti-tamper mechanisms, and information assurance in accordance with Reference (a), addresses and minimizes the probability or consequences of failures that could lead to unintended engagements or to loss of control of the system.

(4) Plans are in place for V&V and T&E to establish system reliability, effectiveness, and suitability under realistic conditions, including possible adversary actions, to a sufficient standard consistent with the potential consequences of an unintended engagement or loss of control of the system.

(5) A preliminary legal review of the weapon system has been completed, in coordination with the General Counsel of the Department of Defense (GC, DoD) and in accordance with References (b) and (c), DoD Directive 2311.01E (Reference (f)), and, where applicable, Reference (d).

b. Before fielding, the USD(P), USD(AT&L), and CJCS shall ensure:

(1) System capabilities, human-machine interfaces, doctrine, TTPs, and training have demonstrated the capability to allow commanders and operators to exercise appropriate levels of human judgment in the use of force and to employ systems with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable ROE.

(2) Sufficient safeties, anti-tamper mechanisms, and information assurance in accordance with Reference (a) have been implemented to minimize the probability or consequences of failures that could lead to unintended engagements or to loss of control of the system.

(3) V&V and T&E assess system performance, capability, reliability, effectiveness, and suitability under realistic conditions, including possible adversary actions, consistent with the potential consequences of an unintended engagement or loss of control of the system.

(4) Adequate 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.

(5) System design and human-machine interfaces are readily understandable to trained operators, provide traceable feedback on system status, and provide clear procedures for trained operators to activate and deactivate system functions.

(6) A legal review of the weapon system has been completed, in coordination with the GC, DoD, and in accordance with References (b), (c), (f), and, where applicable, Reference (d).

2. The USD(P), USD(AT&L), and CJCS may request a Deputy Secretary of Defense waiver for the requirements outlined in section 1 of this enclosure, with the exception of the requirement for a legal review, in cases of urgent military operational need.

ENCLOSURE 4

RESPONSIBILITIES

1. USD(P). The USD(P) shall:

a. Provide policy oversight for the development and employment of autonomous and semi-autonomous weapon systems.

b. In coordination with the USD(AT&L) and CJCS, review and consider for approval weapon systems submitted in accordance with paragraph 4.d. above the signature of this Directive.

c. Review, as necessary, the appropriateness of guidance established in accordance with this Directive given the continual advancement of new technologies and changing warfighter needs.

d. Approve the DoD position on international sales or transfers of autonomous and semi-autonomous weapon systems in accordance with existing technology security and foreign disclosure requirements and processes.

2. USD(AT&L). The USD(AT&L) shall:

a. Provide principal oversight responsibility for the establishment and enforcement of standards for testing, safety and reliability, hardware and software V&V, anti-tamper mechanisms, and information assurance in accordance with Reference (a), for autonomous and semi-autonomous weapon systems in order to minimize the probability and consequences of failures that could lead to unintended engagements or to loss of control of the system.

b. Provide principal oversight responsibility for the establishment of science and technology and research and development priorities for autonomy in weapon systems, including the development of new methods of V&V and T&E.

c. Oversee adequate developmental testing of autonomous and semi-autonomous weapon systems to assess the risk of failures that could lead to unintended engagements or to loss of control of the system.

d. In coordination with the USD(P) and CJCS, review and consider for approval weapon systems submitted in accordance with paragraph 4.d. above the signature of this Directive.

3. UNDER SECRETARY OF DEFENSE FOR PERSONNEL AND READINESS (USD(P&R)). The USD(P&R) shall, consistent with DoD Directive 1322.18 (Reference (g)), oversee and provide policy for:

a. Individual military training programs for the Total Force relating to autonomous and semi-autonomous weapon systems.

b. Individual and functional training programs for military personnel and the collective training programs of military units and staffs relating to autonomous and semi-autonomous weapon systems.

4. DOT&E. The DOT&E shall:

a. Provide principal oversight responsibility for the development of realistic operational T&E standards for semi-autonomous and autonomous weapon systems, including standards for T&E of any changes to the system following IOT&E, in accordance with subparagraph 4.a.(1) above the signature of this Directive and Enclosure 2.

b. Evaluate whether semi-autonomous and autonomous weapon systems under DOT&E oversight have met sufficient V&V and T&E in realistic operational conditions, including potential adversary action, in order to minimize the probability and consequences of failures that could lead to unintended engagements or to loss of control of the system to unauthorized parties.

5. GC, DoD. The GC, DoD, shall, in accordance with References (b), (c), (f), and, where applicable, Reference (d), provide for guidance in and coordination of legal reviews of weapon systems submitted in accordance with paragraph 4.d. above the signature of this Directive.

6. DEPARTMENT OF DEFENSE CHIEF INFORMATION OFFICER (DoD CIO). The DoD CIO, shall monitor, evaluate, and provide advice to the Secretary of Defense regarding information assurance for autonomous and semi-autonomous weapon systems, in accordance with subparagraph 4.a.(2)(a) above the signature of this Directive and Reference (a).

7. ASSISTANT TO THE SECRETARY OF DEFENSE FOR PUBLIC AFFAIRS (ATSD(PA)). The ATSD(PA) shall coordinate and approve guidance on public affairs matters concerning autonomous and semi-autonomous weapon systems and their use.

8. SECRETARIES OF THE MILITARY DEPARTMENTS; COMMANDER, U.S. SPECIAL OPERATIONS COMMAND (USSOCOM); AND THE HEADS OF THE DEFENSE AGENCIES AND DoD FIELD ACTIVITIES. The Secretaries of the Military Departments; the Commander, USSOCOM; and the Heads of the Defense Agencies and DoD Field Activities shall:

a. Develop and implement employment concepts, doctrine, experimentation strategies, TTPs, training, logistics support, V&V, anti-tamper mechanisms, physical hardware and software-level safeties, information assurance in accordance with Reference (a), and

developmental and operational T&E appropriate for autonomous and semi-autonomous weapon systems.

(1) Design autonomous and semi-autonomous weapon systems in such a manner as to minimize the probability and consequences of failures that could lead to unintended engagements or to loss of control of the system.

(2) Perform rigorous and realistic developmental and operational T&E and V&V, including T&E of any changes to the system following IOT&E, in accordance with subparagraph 4.a.(1) above the signature of this Directive and Enclosure 2.

(3) Design autonomous and semi-autonomous weapon systems with sufficient safeties, anti-tamper mechanisms, and information assurance in accordance with subparagraph 4.a.(2) above the signature of this Directive and Reference (a).

(4) Design human-machine interfaces for autonomous and semi-autonomous weapon systems to be readily understandable to trained operators, provide traceable feedback on system status, and provide clear procedures for trained operators to activate and deactivate system functions, in accordance with subparagraph 4.a.(3) above the signature of this Directive.

(5) Certify that operators of autonomous and semi-autonomous weapon systems have been trained in system capabilities, doctrine, and TTPs in order to exercise appropriate levels of human judgment in the use of force and employ systems with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable ROE.

(6) Establish and periodically review training, TTPs, and doctrine for autonomous and semi-autonomous weapon systems to ensure operators and commanders understand the functioning, capabilities, and limitations of a system's autonomy in realistic operational conditions, including as a result of possible adversary actions.

b. Ensure that legal reviews of autonomous and semi-autonomous weapon systems are conducted in accordance with References (b), (c), (f) and, where applicable, Reference (d). Legal reviews should ensure consistency with all applicable domestic and international law and, in particular, the law of war.

c. Consider for support only those autonomous and semi-autonomous weapon systems that are technically feasible and that conform to this Directive. Submit to the USD(P), USD(AT&L), and CJCS for review, in accordance with paragraph 4.d. above the signature of this Directive, any autonomous or semi-autonomous weapon system intended to be used in a manner that falls outside the policies in subparagraphs 4.c.(1) through 4.c.(3) above the signature of this Directive before a decision to enter into formal development and again before fielding of any such system.

9. CJCS. The CJCS shall:

a. Advise the Secretary of Defense on the capability needs and employment of autonomous and semi-autonomous weapon systems.

b. Assess military requirements for autonomous and semi-autonomous weapon systems, including applicable key performance parameters and key system attributes.

c. Develop and publish joint doctrine, as appropriate, to incorporate emerging capabilities of autonomous and semi-autonomous weapon systems.

d. In coordination with the USD(P) and USD(AT&L), review and consider for approval autonomous weapon systems submitted in accordance with paragraph 4.d. above the signature of this Directive.

10. COMMANDERS OF THE COMBATANT COMMANDS. The Commanders of the Combatant Commands shall:

a. Use autonomous and semi-autonomous weapon systems in accordance with this Directive and in a manner consistent with their design, testing, certification, operator training, doctrine, TTPs, and approval as autonomous or semi-autonomous systems.

b. Employ autonomous and semi-autonomous weapon systems with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable ROE, in accordance with paragraph 4.b. above the signature of this Directive.

c. Ensure that weapon systems are not employed or modified to operate in a manner that falls outside the policies in subparagraphs 4.c.(1) through 4.c.(3) above the signature of this Directive without specific approval in accordance with paragraph 4.d. above the signature of this Directive.

d. Integrate autonomous and semi-autonomous weapon systems into operational mission planning.

e. Through the CJCS, identify warfighter priorities and operational needs that may be met by autonomous and semi-autonomous weapon systems.

GLOSSARY

PART I. ABBREVIATIONS AND ACRONYMS

ATSD(PA)	Assistant to the Secretary of Defense for Public Affairs
CJCS	Chairman of the Joint Chiefs of Staff
DoD CIO	Department of Defense Chief Information Officer
DOT&E	Director of Operational Test and Evaluation
GC, DoD	General Counsel of the Department of Defense
IOT&E	initial operational test and evaluation
ROE	rules of engagement
T&E	test and evaluation
TTP	tactics, techniques, and procedures
USD(AT&L)	Under Secretary of Defense for Acquisition, Technology, and Logistics
USD(P)	Under Secretary of Defense for Policy
USD(P&R)	Under Secretary of Defense for Personnel and Readiness
USSOCOM	U.S. Special Operations Command
V&V	verification and validation

PART II. DEFINITIONS

These terms and their definitions are for the purpose of this Directive.

automated regression testing. A type of regression testing that uses testing tools and repeatable test scripts.

autonomous weapon system. A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of

the weapon system, but can select and engage targets without further human input after activation.

electronic attack. Division of electronic warfare involving the use of electromagnetic energy, directed energy, or antiradiation weapons to attack personnel, facilities, or equipment with the intent of degrading, neutralizing, or destroying enemy combat capability and is considered a form of fires.

failures. An actual or perceived degradation or loss of intended functionality or inability of the system to perform as intended or designed. Failures can result from a number of causes, including, but not limited to, human error, 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.

human-supervised autonomous weapon system. An autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur.

operating state. A variable or vector reflecting the status of the system.

operator. A person who operates a weapon system.

regression testing. A type of software testing that seeks to uncover new deficiencies (i.e., regressions) in the existing functional and non-functional areas of a system created by changes to the software, including enhancements, patches, emergency transports, or configuration changes.

semi-autonomous weapon system. 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. This includes:

Semi-autonomous weapon systems that employ autonomy for engagement-related functions including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement.

"Fire and forget" or lock-on-after-launch homing munitions that rely on TTPs to maximize the probability that the only targets within the seeker's acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by a human operator.

state transition matrix. A matrix that characterizes the ability of a system to transition from one operating state to another.

target selection. The determination that an individual target or a specific group of targets is to be engaged.

unintended engagement. The use of force resulting in damage to persons or objects that human operators did not intend to be the targets of U.S. military operations, including unacceptable levels of collateral damage beyond those consistent with the law of war, ROE, and commander's intent.

unmanned platform. An air, land, surface, subsurface, or space platform that does not have the human operator physically onboard the platform.

Tab 2

(This page intentionally left blank)

SECRETARY OF DEFENSE
1000 DEFENSE PENTAGON
WASHINGTON, DC 20301-1000

FEB 2 1 2020

MEMORANDUM FOR CHIEF MANAGEMENT OFFICER OF THE DEPARTMENT OF DEFENSE
SECRETARIES OF THE MILITARY DEPARTMENTS
CHAIRMAN OF THE JOINT CHIEFS OF STAFF
UNDER SECRETARIES OF DEFENSE
CHIEF OF THE NATIONAL GUARD BUREAU
GENERAL COUNSEL OF THE DEPARTMENT OF DEFENSE
DIRECTOR OF COST ASSESSMENT AND PROGRAM EVALUATION
INSPECTOR GENERAL OF THE DEPARTMENT OF DEFENSE
DIRECTOR OF OPERATIONAL TEST AND EVALUATION
CHIEF INFORMATION OFFICER OF THE DEPARTMENT OF DEFENSE
ASSISTANT SECRETARY OF DEFENSE FOR LEGISLATIVE AFFAIRS
ASSISTANT TO THE SECRETARY OF DEFENSE FOR PUBLIC AFFAIRS
DIRECTOR OF NET ASSESSMENT
DIRECTORS OF DEFENSE AGENCIES
DIRECTORS OF DOD FIELD ACTIVITIES

SUBJECT: Artificial Intelligence Ethical Principles for the Department of Defense

Artificial intelligence (AI) is beginning to change the global security environment. I expect this technology to affect the full spectrum of DoD activities. Our 2018 AI Strategy directs the Department to accelerate the adoption of AI and the creation of a Joint Force fit for our time.

Although technology changes, the Department's commitment to the Constitution, the Law of War, and the highest standards of ethical behavior does not. I reaffirm that the Department will use AI consistent with applicable domestic and international law, in particular the law of war. In addition, drawing upon this Nation's proud history and culture of technological excellence from both Government and industry, the Department adopts the following DoD AI Ethical Principles for the design, development, deployment, and use of AI capabilities:

1. Responsible: DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.

2. Equitable: The Department will take deliberate steps to minimize unintended bias in AI capabilities.

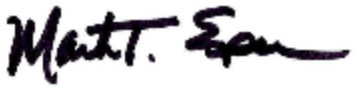
3. Traceable: The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.

4. Reliable: The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.

5. Governable: The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

These principles, informed by the rigorous work of the Defense Innovation Board, and the Department's work in implementing them, will stand as a model for other countries to follow. This adoption of principles stands in contrast to the efforts of governments that do not share similar values, including the importance of and attention to ethics, in their pursuit of AI capabilities. This Department is committed to leading in the technologies that will continue to help secure our national interests and those of U.S. allies and partners.

The Chief Information Officer, through the Joint Artificial Intelligence Center, will serve as the Department's lead for coordination of oversight and implementation of these principles.



Tab 3

(This page intentionally left blank)

DEPUTY SECRETARY OF DEFENSE
1010 DEFENSE PENTAGON
WASHINGTON, DC 20301-1010

MAY 2 6 2021

MEMORANDUM FOR SENIOR PENTAGON LEADERSHIP
COMMANDERS OF THE COMBATANT COMMANDS
DEFENSE AGENCY AND DOD FIELD ACTIVITY DIRECTORS

SUBJECT: Implementing Responsible Artificial Intelligence in the Department of Defense

As the DoD embraces artificial intelligence (AI), it is imperative that we adopt responsible behavior, processes, and outcomes in a manner that reflects the Department's commitment to its ethical principles, including the protection of privacy and civil liberties. A trusted ecosystem not only enhances our military capabilities, but also builds confidence with end-users, warfighters, and the American public. By leading in military ethics and AI safety, we reflect our Nation's values, encourage Responsible AI (RAI) development globally, and strengthen partnerships around the world. To that end, I reaffirm the DoD AI Ethical Principles adopted by the Department on February 21, 2020, for the design, development, deployment, and use of AI capabilities.

The DoD AI Ethical Principles are:

1. **Responsible**: DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.

2. **Equitable:** The Department will take deliberate steps to minimize unintended bias in AI capabilities.

3. **Traceable:** The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of technology, development processes, and operational methods applicable to AI capabilities, including transparent and auditable methodologies, data sources, and design procedure and documentation.

4. **Reliable:** The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across AI capabilities' entire life-cycle.

5. **Governable:** The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

The DoD AI Ethical Principles build on and complement the existing ethical, legal, safety, and policy frameworks that are the hallmark of our Department. They apply to all DoD AI capabilities, of any scale, including AI-enabled autonomous systems, for warfighting and business applications. To ensure our Nation's values are embedded in the AI capabilities, as the Department develops, procures, and deploys AI, these principles will be implemented not only in

technology, but also in enterprise operating structures and organizational culture. This memorandum establishes and directs the Department's holistic, integrated, and disciplined approach for RAI.

The Department will implement RAI in accordance with the following foundational tenets:

1. ***RAI Governance***: Ensure disciplined governance structure and processes at the Component and DoD-wide levels for oversight and accountability and clearly articulate DoD guidelines and policies on RAI and associated incentives to accelerate adoption of RAI within the DoD.

2. ***Warfighter Trust***: Ensure warfighter trust by providing education and training, establishing a test and evaluation and verification and validation framework that integrates real-time monitoring, algorithm confidence metrics, and user feedback to ensure trusted and trustworthy AI capabilities.

3. ***AI Product and Acquisition Lifecycle***: Develop tools, policies, processes, systems, and guidance to synchronize enterprise RAI implementation for the AI product throughout the acquisition lifecycle through a systems engineering and risk management approach.

4. ***Requirements Validation***: Incorporate RAI into all applicable AI requirements, including joint performance requirements established and approved by the Joint Requirements Oversight Council, to ensure RAI inclusion in appropriate DoD AI capabilities.

5. ***Responsible AI Ecosystem***: Build a robust national and global RAI ecosystem to improve intergovernmental, academic, industry, and stakeholder collaboration, including cooperation with allies and coalition partners, and to advance global norms grounded in shared values.

6. ***AI Workforce***: Build, train, equip, and retain an RAI-ready workforce to ensure robust talent planning, recruitment, and capacity-building measures, including workforce education and training on RAI.

The Joint Artificial Intelligence Center (JAIC) serves as the Department's coordinator for development and implementation of RAI strategy, guidance, and policy. The Director of the JAIC will develop, assess, and report on the implementation of a DoD RAI ecosystem, with support from the Office of the Secretary of Defense Components, the DoD Privacy and Civil Liberties Office, the Joint Staff, and the Military Departments and Services. The JAIC will also work in close coordination with stakeholders across DoD, as appropriate, including the Joint Staff, the Joint All-Domain Command Control Cross-Functional Team, Directorates, and Programs to ensure alignment and deconfliction of RAI ecosystem developmental activities.

To accelerate the adoption and implementation of RAI across the Department at scale, the

JAIC Director will, through the RAI Working Council, coordinate the following actions:

- ***RAI Working Council & Training***: Provide O-6/civilian equivalent representatives to an RAI Working Council: Military Departments, Joint Staff, U.S. Special Operations Command, Director of Cost Assessment and Program Evaluation, General Counsel of the Department of Defense, Inspector General of the Department of Defense, Office of the Under Secretary of Defense (OUSD) Acquisition and Sustainment, OUSD Comptroller, OUSD Intelligence and Security, OUSD Policy, OUSD Personnel and Readiness, OUSD Research and Engineering, Director, Operational Test and Evaluation, Defense Privacy, Civil Liberties, and Transparency Division (Privacy), Office of the Chief Data Officer, and Office of the Chief Information Officer. The RAI Working Council may include representatives from other DoD Components as approved by JAIC. The Working Council will be an initial RAI working body to ensure input and coordination across the Department. The JAIC will provide RAI training to the Working Council (based on its Responsible AI Champions pilot). No later than fourteen (14) days from the date of the signature of this memorandum, representatives will be designated in writing to the JAIC Director. RAI training will be provided by the JAIC no later than thirty (30) days from the identification of Working Council designees.

- ***RAI Strategy & Implementation Pathway***: The RAI Working Council will develop a DoD RAI Strategy & Implementation Pathway with executable and practical actions based on the RAI foundational tenets listed above. The Pathway will include proposed actions, with corresponding metrics (as applicable), and timelines, as well as the future role of the Working Council, while leveraging existing efforts, processes, policies, and structures for RAI integration across the Department. The Initial Pathway is due no later than ninety (90) days from the date of the signature of this memorandum. The Final Pathway is due no later than one hundred fifty (150) days from the date of the signature of this memorandum.

- ***RAI Workforce Talent Management:*** The RAI Working Council will develop a talent management framework and identify required skills to build a cadre of RAI experts and an RAI-literate workforce. A presentation to the Deputy's Workforce Council is due no later than October 1, 2021.

- ***RAI Acquisition:*** The RAI Working Council will provide recommendations on the integration of RAI into the AI acquisition requirements, on process, and on any policy modifications to enable RAI considerations within existing supply chain risk management practices. A report is due no later than one hundred twenty days (120) days from the date of the signature of this memorandum.

Ensuring a culture of ethical and responsible AI across the Department is a collective effort that requires strong leadership, robust governance, oversight, and sustained engagement at all levels of our organization. Applying RAI across a wide range of warfighting, enterprise support, and business practices is essential for ensuring military advantage, supporting our people, and safeguarding the Nation.