



*I would gladly agree with all the world to lay aside the use of arms, and settle matters by negotiations; but, unless the whole world wills, the matter ends, and I take up my musket, and thank heaven he has put it in my power. . . . We live not in a world of angles. The reign of Satan has not ended, neither can we expect to be defended by miracles.*

-Thomas Paine  
July, 1775

## Dimensions of Autonomous Decision-making

A First Step in Transforming the Policies and Ethics Principles Regarding Autonomous Systems into Practical System Engineering Requirements

Michael F. Stumborg, Becky Roh, Mark Rosen

With contributions from Sam Bendett, Kevin Pollpeter

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

## Abstract

This study identifies the dimensions of autonomous decision-making (DADs)—the categories and causes of potential risk that one should consider before transferring decision-making capabilities to an intelligent autonomous system (IAS). The objective of this study was to provide some of the tools needed to implement existing policies with respect to the legal, ethical, and militarily effective use of IAS. These tools help to identify and either mitigate or accept the risks associated with the use of IAS that might result in a negative outcome. The 13 identified DADs were developed from a comprehensive list of 565 “risk elements” drawn from hundreds of documents authored by parties in favor of, and opposed to, the use of autonomy technology in weapons systems. We record these elements in the form of a question so that they can be used by the acquisition community to develop requirements documents that ensure the ethical use of autonomous systems, and be used by military commanders as a risk assessment checklist to ensure that autonomous systems are not used in an unethical manner. In this way, the Department of Defense can make fully-informed risk assessment decisions before developing or deploying autonomous systems.

---

This document contains the best opinion of CNA at the time of issue.

It does not necessarily represent the opinion of the sponsor

## Distribution

**DISTRIBUTION STATEMENT A.** Approved for public release: distribution unlimited.

Public Release.

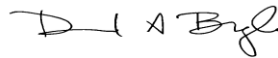
12/30/2021

This work was performed under Federal Government Contract No. N00014-16-D-5003.

**Cover image credit:** Thomas Paine c. 1792 by Auguste Millière, after an engraving by William Sharp, after George Romney. (National Portrait Gallery; public domain)

**Approved by:**

**December 2021**



Dr. David Broyles, Director  
Special Activities and Intelligence Program  
Operational Warfighting Division

| REPORT DOCUMENTATION PAGE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |             |                         |                            | Form Approved<br>OMB No. 0704-0188                                     |                                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|-------------------------|----------------------------|------------------------------------------------------------------------|-----------------------------------------------------------|
| Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>                                                                                                                                                                                                                                                                                                                                          |             |                         |                            |                                                                        |                                                           |
| 1. REPORT DATE (DD-MM-YYYY)<br>December 2021                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |             | 2. REPORT TYPE<br>Final |                            | 3. DATES COVERED (From - To)                                           |                                                           |
| 4. TITLE AND SUBTITLE<br>(U) Dimensions of Autonomous Decision-making: A First Step in Transforming the Policies and Ethics Principles Regarding Autonomous Systems into Practical System Engineering Requirements                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |             |                         |                            | 5a. CONTRACT NUMBER<br>N00014-16-D-5003                                |                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         |                            | 5b. GRANT NUMBER                                                       |                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         |                            | 5c. PROGRAM ELEMENT NUMBER<br>0605154N                                 |                                                           |
| 6. AUTHOR(S)<br>Michael F. Stumborg, Becky Roh, Mark Rosen, Sam Bendett, Kevin Pollpeter                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |             |                         |                            | 5d. PROJECT NUMBER<br>R0148                                            |                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         |                            | 5e. TASK NUMBER<br>E317.00                                             |                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         |                            | 5f. WORK UNIT NUMBER                                                   |                                                           |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br><br>Center for Naval Analyses<br>3003 Washington Blvd<br>Arlington, VA 22201                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |             |                         |                            | 8. PERFORMING ORGANIZATION REPORT NUMBER<br><br>DRM-2021-U-030642-1Rev |                                                           |
| 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>Office of Naval Research<br>Navy Department<br>Washington, DC 20350                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |             |                         |                            | 10. SPONSOR/MONITOR'S ACRONYM(S)                                       |                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         |                            | 11. SPONSOR/MONITOR'S REPORT NUMBER(S)                                 |                                                           |
| 12. DISTRIBUTION / AVAILABILITY STATEMENT<br><br>DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         |                            |                                                                        |                                                           |
| 13. SUPPLEMENTARY NOTES                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |             |                         |                            |                                                                        |                                                           |
| 14. ABSTRACT<br>This study identifies the dimensions of autonomous decision-making (DADs)—the categories and causes of potential risk that one should consider before transferring decision-making capabilities to an intelligent autonomous system (IAS). The objective of this study was to provide some of the tools needed to implement existing policies with respect to the legal, ethical, and militarily effective use of IAS. These tools help to identify and either mitigate or accept the risks associated with the use of IAS that might result in a negative outcome. The 13 identified DADs were developed from a comprehensive list of 559 "risk elements" drawn from hundreds of documents authored by parties in favor of, and opposed to, the use of autonomy technology in weapons systems. We record these elements in the form of a question so that they can be used by the acquisition community to develop requirements documents that ensure the ethical use of autonomous systems, and be used by military commanders as a risk assessment checklist to ensure that autonomous systems are not used in an unethical manner. In this way, the Department of Defense can make fully-informed risk assessment decisions before developing or deploying autonomous systems. |             |                         |                            |                                                                        |                                                           |
| 15. SUBJECT TERMS<br>artificial intelligence, autonomy, ethics.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         |                            |                                                                        |                                                           |
| 16. SECURITY CLASSIFICATION OF:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |             |                         | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES                                                    | 19a. NAME OF RESPONSIBLE PERSON                           |
| a. REPORT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | b. ABSTRACT | c. THIS PAGE            |                            |                                                                        | Knowledge Center/Dana Smith                               |
| U                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | U           | U                       | SAR                        | 114                                                                    | 19b. TELEPHONE NUMBER (include area code)<br>703-824-2123 |

# Executive Summary

---

This study identifies the dimensions of autonomous decision-making (DADs)—the categories of potential risk that one should consider before transferring decision-making capabilities to an intelligent autonomous system (IAS). The objective of this study was to provide some of the tools needed to implement existing policies with respect to the legal, ethical, and militarily effective use of IAS. These tools help to identify and either mitigate or accept the risks associated with the use of IAS that might result in a negative outcome.

The 13 DADs identified by this study were developed from a comprehensive list of 565 “risk elements” drawn from hundreds of documents authored by a global cadre of individuals both in favor of, and opposed to, the use of autonomy technology in weapons systems. Additionally, these risk items go beyond current DOD policies and procedures because we expect those to change and IAS technologies to evolve. We captured each risk element in the form of a question. Each can then be easily modified to become a “shall statement” for use by the acquisition community in developing functional requirements that ensure the legal and ethical use of autonomous systems.<sup>1</sup> This approach can elevate artificial intelligence (AI) ethics from a set of subjectively defined and thus unactionable policies and principles, to a set of measurable and testable contractual obligations.

The risk elements can also be used by military commanders as a (measurable and testable) pre-operational risk assessment “checklist” to ensure that autonomous systems are not used in an unethical manner. In this way, the Department of Defense (DOD) can make fully informed risk assessment decisions before developing or deploying autonomous systems. Because our study results were specifically designed for use within the defense acquisition system and the military planning process, they provide a first step in transforming the policies and ethics principles regarding autonomous systems into practical system engineering requirements.

The 13 DADs we identified are as follows:

- Standard semantics and concepts: ensures the use of common terminology and concepts throughout the life cycle of an IAS and amongst the different user communities to prevent risks that would occur due to miscommunication.

---

<sup>1</sup> For the purposes of this study, when we say “ethical use” it encompasses the range of external influences on the development and deployment of IAS including domestic and international law, including the laws of war and laws of peace, regulatory requirements, civil legal requirements, and finally, ethical considerations.

- Continuity of legal accountability: ensures that a human is legally accountable for the IAS at all times, with no gaps in accountability during fast-paced and dynamic military operations.
- Degree of autonomy: ensures that adjustments can be made to the degree of system autonomy to accommodate dynamic operational conditions and match changing risk tolerance levels.
- Necessity of autonomy: ensures that use of an IAS provides a military advantage (to include reducing the probability of collateral damage) commensurate with any additional risk introduced by its use.
- Command and control: ensures that all practicable measures are taken to prevent the loss of command and control over an IAS and ensures that the IAS can detect and prevent unintended consequences and deactivate systems that may engage in unintended behaviors.
- Presence of persons and objects protected from the use of force: ensures that the IAS can identify and not intentionally harm persons or objects in a manner that would violate laws, policies, or the rules of engagement.
- Pre-operational audit logs: ensures positive control over all aspects of an IAS during its acquisition by documenting the provenance of data, software, hardware, personnel interactions and processes executed from pre-acquisition inception to delivery to the fleet.
- Operational audit logs: ensures that inputs, actions, interactions, and outcomes are recorded for post-operational analysis, supporting legal accountability, sharing of lessons learned, and making improvements to future tactics, techniques, procedures and technologies.
- Human-machine teaming: ensures that human judgement is exercised (particularly when the use of force is involved).
- Test and evaluation adequacy: ensures that the depth, breadth, and complexity of the contemplated operational environment are represented to the greatest extent practicable during test and evaluation.
- Autonomy training and education: ensures everyone associated with the development and use of the IAS understands its attributes well enough to execute their responsibilities to act to avoid illegal and unethical use.
- Mission duration and geographic extent: ensures that a mission's time length and spatial extent do not invalidate pre-mission risk assessments and planning factors.

- Civil and natural rights: ensures that the IAS, when used in other than a lethal autonomous weapons application, is engineered to safeguard both civil and natural rights and to identify and mitigate the bias sometimes present in autonomous systems.

This study makes six recommendations on how best to use the 13 DADs and their 565 risk elements to aggressively move the DOD AI ethical principles from the articulation phase to the implementation phase:

- Make the presence of ethical use enablers a mandatory key performance parameter for IAS: turns ethics principles into measurable and testable contractual obligations.
- Incorporate IAS risk mitigation checklists into doctrine and planning: provides the doctrinal foundation needed to make IAS-related risk assessment a mandatory component of long-term strategic, and shorter-term operational planning.
- Maintain an authoritative and standardized Joint Autonomy Risk Elements List (JAREL): transforms the list of 565 risk elements into the primary tool for implementing IAS-related ethics principles in a repeatable and tailorable way.
- Make the JAREL publicly available to the greatest extent possible: promotes public trust in DOD use of IAS, improves DOD's ability to leverage and attract an IAS development workforce and improves the US' ability to attract allies and partners.
- Reimagine the approach to "defining" standard terminology: removes the barrier to implementation created by the use of poorly defined or undefined subjective terminology in ethics-related policy that are prone to misinterpretation or differing interpretations.
- Create a research and development portfolio: provides the technologies that enable ethically conforming IAS.

Finally, our findings support the DOD's commitment to the ethical use of IAS by taking a transparent approach to implementing the DOD AI ethical principles. To demonstrate this transparency, the sponsors of this study agreed to make this report publicly available. Doing so can reduce the misinformation, miscommunication, and misinterpretation of statements and intentions made by the many organizations and communities involved in the debate over the development and use of AI in warfare systems.

This page intentionally left blank.

# Contents

---

|                                                                                         |           |
|-----------------------------------------------------------------------------------------|-----------|
| <b>Motivation .....</b>                                                                 | <b>1</b>  |
| <b>What This Report Is (and Is Not).....</b>                                            | <b>4</b>  |
| <b>Study Methodology.....</b>                                                           | <b>5</b>  |
| The study questions .....                                                               | 5         |
| The overarching study approach.....                                                     | 6         |
| Task 1: Identify and gather risk elements.....                                          | 6         |
| Task 2: Assemble DADs from the risk elements.....                                       | 7         |
| Task 3: Validate the DADs.....                                                          | 9         |
| <b>The 13 DADs .....</b>                                                                | <b>11</b> |
| Putting the 13 DADs in context.....                                                     | 11        |
| DAD#1: Standard semantics and concepts.....                                             | 16        |
| DAD#2: Continuity of legal accountability.....                                          | 16        |
| DAD#3: Degree of autonomy.....                                                          | 17        |
| DAD#4: Necessity of autonomy.....                                                       | 17        |
| DAD#5: Command and control.....                                                         | 18        |
| DAD#6: Presence of persons and objects protected from the use of force.....             | 18        |
| DAD#7: Pre-operational audit logs.....                                                  | 19        |
| DAD#8: Operational audit logs.....                                                      | 19        |
| DAD#9: Human-machine teaming .....                                                      | 19        |
| DAD#10: Test and evaluation adequacy.....                                               | 20        |
| DAD#11: Autonomy training and education .....                                           | 21        |
| DAD#12: Mission duration and geographic extent.....                                     | 21        |
| DAD#13: Civil and natural rights.....                                                   | 21        |
| <b>How to Use the DADs and Their Elements .....</b>                                     | <b>23</b> |
| Defining a key performance parameter for the presence of ethical use enablers.....      | 23        |
| Conducting operational risk assessments.....                                            | 24        |
| Other potential uses .....                                                              | 25        |
| <b>Conclusion .....</b>                                                                 | <b>28</b> |
| <b>Recommendations.....</b>                                                             | <b>29</b> |
| Mandate a KPP for the presence of ethical use enablers for IAS .....                    | 29        |
| Incorporate risk mitigation checklists into doctrine and planning.....                  | 30        |
| Maintain an authoritative and standardized joint autonomy risk elements list (JAREL) .. | 30        |
| Make the JAREL publicly available to the greatest extent possible.....                  | 31        |
| Reimagine the approach to “defining” standard terminology .....                         | 32        |

|                                                                                |            |
|--------------------------------------------------------------------------------|------------|
| Create a research and development portfolio for ethically conforming IAS ..... | 34         |
| <b>Appendix A: Bibliography of documents consulted .....</b>                   | <b>37</b>  |
| Academia.....                                                                  | 37         |
| Books .....                                                                    | 39         |
| Public policy advocacy groups.....                                             | 39         |
| International governmental organizations .....                                 | 42         |
| National governments.....                                                      | 45         |
| Think tanks and research centers .....                                         | 57         |
| <b>Appendix B: IAS risk elements .....</b>                                     | <b>59</b>  |
| DAD#1: Standard semantics and concepts.....                                    | 59         |
| DAD#2: Continuity of legal accountability.....                                 | 60         |
| DAD#3: Degree of autonomy.....                                                 | 62         |
| DAD#4: Necessity of autonomy.....                                              | 64         |
| DAD#5: Command and control.....                                                | 67         |
| DAD#6: Presence of persons and objects protected from the use of force.....    | 71         |
| DAD#7: Pre-operational audit logs.....                                         | 78         |
| DAD#8: Operational audit logs.....                                             | 82         |
| DAD#9: Human-machine teaming .....                                             | 85         |
| DAD#10: Test and evaluation adequacy.....                                      | 88         |
| DAD#11: Autonomy training and education .....                                  | 90         |
| DAD#12: Mission duration and geographic extent.....                            | 93         |
| DAD#13: Civil and natural rights.....                                          | 94         |
| <b>Abbreviations.....</b>                                                      | <b>100</b> |
| <b>References.....</b>                                                         | <b>101</b> |

# Motivation

---

Artificial intelligence (AI) and robotics, and the autonomous functionality enabled by them, hold great promise to improve the effectiveness and the efficiency of many human endeavors. This promise, coupled with the known weaknesses and pitfalls of the underlying technology, generates considerable angst, debate and analyses. The lengthy bibliography in this report (Appendix A: Bibliography of documents consulted) attests to that and it is admittedly just a miniscule fraction of the total corpus of available work on this subject.

Perhaps the greatest concern—because their potential negative effects are irreversible—is the use of AI in lethal autonomous weapon systems (LAWS). The Department of Defense (DOD) created a review process in 2012 to “minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements” [1]. This policy applies to “autonomous and semi-autonomous weapon systems...that can independently select and discriminate targets,” including weapons that deliver nonkinetic effects. Overall, the goal of these reviews is to ensure that commanders and operators can “exercise appropriate levels of human judgment over the use of force.” To date, no weapon system that would require such a review has been put forward.<sup>2</sup>

This and other initiatives [5-6] make it clear that DOD is committed to the ethical use of AI in weapon systems. Other organizations disagree, or fear that the technology will fall into less-responsible hands [7], so they seek a preemptive ban on its development. The stated objective of most of these groups aligns with that of the Campaign to Stop Killer Robots: a preemptive ban on the development of “fully autonomous weapons [that] would decide who lives and dies, without further human intervention” [8]. We are not aware of any DOD system, actual or contemplated, that fits this description. Additionally, Robert O. Work, a former Deputy Secretary of Defense and a central figure in the use of AI in warfare, flatly stated in a recent presentation that “no sane commander” would use such a system [9].

After an extensive review (see Appendix A: Bibliography of documents consulted) of the literature on this topic, we conclude that there is a fair amount of misinformation, miscommunication, and/or misinterpretation regarding the statements and intentions of the parties involved in the development of AI for warfare systems and those who oppose this development. This report seeks, in part, to reduce this miscommunication by increasing

---

<sup>2</sup> While every weapon is reviewed for DOD’s Law of War compliance [2-3], special and additional considerations apply to nonlethal weapons [4] and to autonomous weapons [1].

transparency with respect to the considerable lengths to which DOD is prepared to go to ensure that the use of autonomous systems is legal, ethical, and militarily effective. While the primary concern is with LAWS, DOD must also contend with autonomous systems that have potentially negative, but reversible nonlethal effects.<sup>3</sup>

To that end, the Department of the Navy (DON) sponsors of this study agreed that this report should be releasable to the public. This study identifies the **Dimensions of Autonomous Decision-making (DADs)**—the things that must be considered before transferring decision-making capabilities to a system possessing autonomous functionality to ensure that its use is legal, ethical, and militarily effective.

This begs a second question: How does one know, *and how can one verify*, that a particular autonomous system can be used ethically?<sup>4</sup> Shifting to the vernacular of the Defense Acquisition System [2], ethical conformity measures could be embedded in functional requirements documents for autonomous systems and even be designated as a key performance parameter (KPP) for these systems.

The Defense Acquisition System defines a KPP as follows:

An attribute of a system considered critical or essential to the development of an effective military capability. KPPs are contained in the Capability Development Document (CDD) and the updated CDD and are included verbatim in the Acquisition Program Baseline (APB). KPPs are expressed in terms of parameters which reflect measures of performance (MOPs) using a threshold/objective format. KPPs must be measurable, testable, and support efficient and effective Test and Evaluation (T&E) [10].<sup>5</sup>

Given its importance in DOD policy [1, 5], we assert that the presence of technologies that enable ethical use is a “critical or essential” attribute of DOD autonomous systems possessing decision-making capabilities, and DOD should therefore consider elevating the presence of these ethical use enabling technologies to KPP status.<sup>6</sup> The present report explores this possibility in depth.

The critical part of the KPP definition for the purposes of this study is the requirement that KPPs be “measurable and testable.” Our intention in developing the DADs is to provide DOD

---

<sup>3</sup> A nonlethal autonomous system could result in a nonlethal but still negative outcome. For example, an unmanned vehicle designed to evacuate casualties could wound an evacuee. An AI-enabled decision aid could deny a Sailor’s rightfully earned pay and benefits.

<sup>4</sup> The DOD already has well-established processes to ensure military effectiveness [2] and legal compliance [3].

<sup>5</sup> The threshold/objective format is described in the “Defining a key performance parameter for the presence of ethical use enablers” section of this report.

<sup>6</sup> A key system attribute (KSA) is a performance attribute of a system considered important to achieving a balanced solution/approach to a system, but not critical enough to be designated as a KPP [6].

with an initial draft of measurable and testable risk “elements” that can be used to construct functional requirements—which may also rise to the level of a KPP—that enable the ethical use of autonomous systems.

We provide a “first draft” of these risk elements in recognition of the fact that our list is based on current DOD policies, ethics standards from multiple organizations, legal analyses of the system requirements, and the current state of the art in autonomy technologies. We expect all of these to evolve over time, and the DADs must evolve with them.

We use the term risk “elements” to reflect the fact that the DADs are just the overarching categories of the many things that must be considered before developing or using a system possessing autonomous functionality. As such, the DADs are not measurable and testable on their own. The risk elements that make up the DADs (see this report’s “Study Methodology” section and Appendix B: IAS risk elements) *are* measurable and testable. Some elements are quantifiable measures. Most elements are a simple yes/no test—a checklist. We designed the elements for use by acquisition professionals to construct functional requirements documents for autonomous systems, and for use by military commanders to conduct pre-deployment operational risk assessments.

Providing risk elements for the construction of functional requirements documents and for conducting operational risk assessments was the initial motivation for this study. It became clear during the early phase of the study that these risk elements could have the fortuitous side effect of demonstrating to the public that the DOD is going to great lengths to implement the policies that ensure the legal and ethical use of AI and robotics[1, 5]. This became an additional and equally important motivation. The subtitle to this report reflects this motivation:

*A First Step in Transforming the Policies and Ethics Principles Regarding  
Autonomous Systems into Practical System Engineering Requirements*

We note that DOD system engineers and program managers have the duty of carrying these principles into practice during system development and acquisition. Military commanders carry them into practice during system use. The output of this study is intended to assist these individuals in their implementation of DOD AI ethics policies and principles.

# What This Report Is (and Is Not)

---

While this report was developed in consultation with multiple US government individuals and organizations, it was authored by employees of the Center for Naval Analyses (CNA), who are solely responsible for its content. As the Navy's federally funded research and development center (FFRDC), CNA is required under the Federal Acquisition Regulations and various DOD and Navy Directives to operate in the public interest with objectivity and independence [11-12].

As such, this report contains the best opinion of CNA at the time of issue and does not necessarily represent the opinion of the US government, the DOD, or the DON. This report need not reflect, nor is it bound by, any policy or position of any US government individual or organization. Of course, to the extent that this report identifies "black letter" requirements which are grounded in law or policy, then the US Government will presumably incorporate those requirements into its overall planning.

This report contains a comprehensive collection of concerns that many people have expressed regarding the use of autonomous systems—groups from within and from outside the DOD and the United States, and groups in favor of and opposed to the use of autonomy in weapons systems (see Appendix A: Bibliography of documents consulted). No DOD (or CNA) endorsement of any of these concerns is to be implied by their inclusion in this report. For this reason alone, nothing in this document should be construed as the position of, or to be binding upon, the DOD or the DON.

CNA provides information and analysis and makes recommendations to its study sponsors. The sponsors are free to accept, modify, or reject our recommendations. The output of this study is also intended to inform the debate that must occur among DOD stakeholders who have the authority to articulate and implement official DOD policy positions. While much of this debate will by necessity be conducted internally, this report was made public now to demonstrate that this debate is occurring and that organizations within DOD are taking the steps necessary to implement their commitment to the legal, ethical [6] and responsible use of AI [5].<sup>7</sup>

---

<sup>7</sup> Responsible AI is one of the five DOD AI ethical principles. The other four are equitable, traceable, reliable, and governable AI.

# Study Methodology

---

The autonomy portfolio manager at the Office of Naval Research and the warfare integration branch (N9IX) at the Deputy Chief of Naval Operations for Warfighting Requirements and Capabilities co-sponsored this study. These offices published the Science and Technology Strategy for Intelligent Autonomous Systems [13] and the DON Unmanned Campaign Framework [14], respectively. The strategy considers intelligent autonomous systems (IAS) to be “autonomy plus its intersections with unmanned systems (UxS) and AI” [13].<sup>8</sup> This study provides some of the tools required to implement the objectives identified in each of these DON guidance documents, and those articulated in DOD policy that require the ethical use of AI [5].

## The study questions

Our primary study question was as follows:

- What are the dimensions of autonomous decision-making (DADs) that must be considered in delineating what decision-making capability<sup>9</sup> should/must remain with humans versus what can be transferred to an IAS?

Rephrasing this question in the context of the military planning processes [16] yields the following:

- Transferring the ability to make a decision to an IAS incurs a risk that the IAS may make a decision resulting in an unintended negative consequence. What attributes of the IAS, the operational environment, friendly forces, and enemy forces must commanders consider and evaluate during the military planning process to prevent or minimize the risks created by transferring decision-making capabilities to an IAS?

A second study question facilitates implementing the findings of the primary study question:

---

<sup>8</sup> We use this definition for IAS for the purposes of this study because it comes from the study’s sponsors. We recognize there are many definitions for AI, autonomy, and other associated terms, but to date, the absence of definitions in Joint Publication 1-02, *Department of Defense Dictionary of Military and Associated Terms* [15] indicates that there are no official and accepted definitions.

<sup>9</sup> These questions have been reworded slightly from the original in an attempt to avoid anthropomorphizing the IAS. For example, “ceding decision-making authority” became “transferring decision-making capabilities.” The legal consensus is that machines can never be given authority or responsibility, or be held accountable. The humans that design, build, or operate them do have ultimate legal responsibility for the IAS’s actions.

- How can DOD incorporate the DADs into a framework to guide the development, fielding, integration, and employment of militarily effective, legally compliant, and ethically conforming IAS?

The rephrased first question explicitly states the risks of concern to the study sponsor: the risk that the IAS will not be militarily effective, legally compliant, or ethically conforming. Since this is a DON-sponsored study, the framework we propose leverages risk-reduction opportunities within existing DON processes: the DON implementation of the Joint Capabilities Integration and Development System [17] and the Joint Planning Process [16].<sup>10</sup>

## The overarching study approach

Rather than take a “top-down” approach to identifying the DADs, we took a “bottom-up” approach. The CNA study team has access to numerous subject matter experts (SMEs) who could have provided their perspective on what the DADs are or should be. In fact, the study sponsor provided us with an initial list of DADs compiled during ongoing discussions with just such a group of SMEs—members of the Office of the Secretary of Defense Autonomy Community of Interest. This would have been a top-down approach to our first study question.

Instead, we hypothesized that we could synthesize the written record on the use of IAS to collect a list of concerns that knowledgeable people have regarding the legal and ethical use of IAS. These are the risk “elements” referred to earlier. We further hypothesized that these risk elements would fall into a mutually exclusive and collectively exhaustive list of “affinity groups” and that each of these groups would then constitute one of the DADs. With an eye toward addressing our second study question, we posed each of the risk elements in the form of a question. This list of questions, now categorized under their respective DADs, can then be used to assemble “checklists” to serve as the measurable and testable parameters of acquisition requirements documents and to help military commanders create pre-deployment risk assessment checklists similar to a pilot’s pre-flight checklist. This was our top-down approach. We describe it in detail below.

## Task 1: Identify and gather risk elements

The documents we analyzed to extract DAD risk elements are listed in Appendix A: Bibliography of documents consulted. We sought to maximize the intellectual diversity of the

---

<sup>10</sup> While we focus on DOD processes for our DON study sponsors, our proposals are readily adaptable to nonmilitary and nongovernmental applications.

expert thinking on the legal, ethical, and militarily effective use of IAS. To that end, we purposefully included the following:

- Documents from advocacy groups adamantly opposed to LAWS
- National-level AI strategies from a wide variety of countries with differing perspectives on technology, human oversight, ethics, privacy, and civil rights
- Military targeting doctrine even though it is mostly limited to non-autonomous systems
- AI, autonomy and robotics roadmaps, strategies, directives, etc. from multiple and varied departments and agencies of the US federal government
- Commercial IAS applications such as self-driving cars
- Academic journal articles, books, and media reports about IAS
- Documents concerning biometric algorithms—particularly facial recognition—since these algorithms are for all intents and purposes autonomous “targeting” algorithms,<sup>11</sup> are already fielded, and have already created negative effects and risks in need of mitigation
- Documents concerning automated decision-making algorithms—such as for loan applications or parole recommendations (for the same reasons listed for biometrics).

This analysis resulted in 4,641 individual risk elements. Many duplicate elements exist in this list because various analysts worked independently and in parallel, and because we erred on the side of inclusion, not knowing until after we read all of these documents what our final criteria for inclusion would be. We refined this initial list in Task 2.

## **Task 2: Assemble DADs from the risk elements**

To analyze the 4,641 risk elements identified and gathered in Task 1, we entered them into a Microsoft Excel spreadsheet, sortable by columns.

### **Task 2a: Reduction and refinement**

The first step in the analysis was to purge duplicate risk elements from our list. This was a largely manual process, consisting of key word searching, grouping elements with matching key words, and manually adjudicating these to identify the duplicates.

---

<sup>11</sup> A “targeting” algorithm picks one object out of the environment or out of a group of objects for some further action. For military algorithms, that action is often lethal; for other algorithms, it is not. Both lead to potential negative implications for the targeted object, so the lessons learned from one application domain are largely applicable to the other and vice versa.

As noted in Task 1, we erred on the side of inclusion. Taking into account what we learned about the issues as we analyzed additional documents, the next step in the analysis was to delete entries that no longer fit our emerging criteria for what constituted a risk element of a DAD. Specifically, we deleted the following:

- Concerns that are important, but are not introduced by the use of autonomy technology
- Concerns that are present because of the introduction of autonomous functionality, but that are not of an ethical or legal nature
- Risk mitigation capabilities that do not address risks introduced by the use of autonomy technology
- Risk mitigation capabilities that do address risks introduced by the use of autonomy technology, but those risks are not of a legal or ethical nature
- Concerns and risks that are not measurable, testable, or otherwise actionable.

After closer examination, some risk elements were found to contain more than one thought and were broken out into two or more distinct elements.

We chose not to analyze the remaining risk elements for rank or importance. This would have produced an unduly subjective result for two reasons:

- The importance of each risk element is highly dependent on the operational scenario in which the IAS is to be used. The DADs and their risk elements are intended to be used by a wide variety of IAS end users who must select the risk elements important to them based on their current operational environment.
- The number of times a risk element appears in the reviewed literature is not a good proxy for its importance. For example, the ability for an IAS to recognize an act of surrender is clearly important in many scenarios, but appeared in only two of the documents we reviewed [7, 18]. Similarly, the four principles of the Law of Armed Conflict (military necessity, distinction, proportionality, and prevention of unnecessary suffering [19]) appear repeatedly, but just restate the original four principles and all cite the same Law of Armed Conflict.

## Task 2b: “Affinitization” and categorization

Carrying out the reduction and refinement step above created additional familiarization with the risk elements that allowed us to start to recognize the naturally occurring affinity groups in which many but not all of them belonged. We postulated candidate groups—candidate DADs—and began assigning risk elements to the appropriate DAD. We then created additional DADs to accommodate any remaining risk elements. Subsequent iterations of categorization (and further refinement) were informed by the ultimate intended use of these DADs as risk

elements to build acquisition functional requirements documents and pre-deployment risk mitigation checklists. Task 2 resulted in 565 risk elements assigned to 13 DADs.

## Task 3: Validate the DADs

Recognizing that our very small study team—even after analyzing thousands of pages from almost 200 documents to identify the risk elements of the 13 DADs—could not possibly capture all risk elements of concern to all stakeholders, we reached out to SMEs to validate our findings.

We forwarded the Task 2 results to DOD SMEs at various organizations that build, review, test and evaluate (T&E), experiment, exercise, or demonstrate IAS capabilities at one of four “waypoints” before the system is finally delivered to military commanders at scale:

1. **DOD senior level review:** As previously mentioned, DOD has a senior-level review process for autonomous weapon systems [1] (which includes a legal review). Certain IAS must be approved by the Under Secretary of Defense for Policy; the Under Secretary of Defense for Acquisition, Technology, and Logistics; and the Chairman of the Joint Chiefs of Staff at two points: before formal development and before fielding.
2. **Developmental T&E:** This is a complicated and multifaceted process, but for the purposes of this study we can limit our description to its stated objective: Developmental T&E enables DOD to acquire systems that work [20]. This includes systems and subsystems with autonomous functionality.
3. **Operational T&E:** Also a complicated and multifaceted process, Operational T&E can likewise be described here by its stated objective: to assess systems for effectiveness, suitability, survivability, and lethality *in near-real-world combat conditions* to determine if the system does what it’s supposed to, the warfighter can use it safely, and can depend on it in combat [21]. The requirements tested in the developmental phase are tested again here, but in a realistic operational environment (which is specifically required for autonomous weapons [1]).
4. **Fleet introduction:** IAS in the DON are by definition [13] strongly associated with UxS [14]. These platforms are novel enough that the Navy set up several developmental squadrons to take possession of the first instantiations of them and figure out the best way to leverage their capabilities before large-scale delivery to the fleet. Developmental squadrons exist for unmanned surface ships [22], unmanned underwater vehicles [23], and unmanned aerial systems [24]. While waypoints 1, 2, and 3 above will use the DADs for requirements development, these squadrons must (among their many other roles and responsibilities) understand the operational risks created by the introduction of new systems—including IAS—and determine how to

mitigate them. We see them as the initial users of the DADs for developing risk mitigation checklists.

We asked individuals to review our Task 2 findings and provide comments on both the structure and completeness of the DADs and their underlying risk elements. We incorporated these (sometimes-conflicting) comments to the greatest extent practicable. Some of the recommended changes were not incorporated because they fell outside the intended scope of our study (see the criteria in Task 2a above).

Other commenters “non-concurred” with the inclusion of some DAD risk elements or their wording because they ran counter to the position of the respondent’s home organization. We did not address these concerns here because our DAD risk elements list is meant to be a comprehensive collection of risk concerns from many organizations—many of which clearly disagree with one another. CNA does not have the inherently governmental authority (nor did this study team have the resources) to adjudicate and resolve conflicting comments. We do recommend later in this report that such an adjudication occur (see the recommendation to maintain an authoritative and standardized joint autonomy risk elements list).

The final list of DADs is included in its entirety in the next section and the comprehensive list of risk elements that comprise them is included in Appendix B: IAS risk elements.

Given resource restrictions and the breadth of DOD interest in AI and autonomy, it was impossible to consult every stakeholder organization or every SME within those organizations we did consult. The organizations that provided comments were a wide though necessarily incomplete representation of all DOD interests.

It is also important to note that we consulted *individuals* within these organizations who were not necessarily serving as representatives of their organizations. As such, nothing in this report represents an official policy position of any organization consulted. The individuals consulted acted strictly as SMEs.

# The 13 DADs

---

We describe below each of the DADs resulting from our analysis. Appendix B: IAS risk elements provides the full list of risk elements that fall under each of these 13 DADs. As noted previously, each risk element is presented in the form of a question to better serve as a risk mitigation “checklist” for IAS developers, operators, and commanders. Before we describe each DAD, several observations help put them in the proper context.

## Putting the 13 DADs in context

### Requirements versus considerations

These 13 DADs and their associated risk elements are things that should be considered and *might* result in a formal requirement. These are not requirements that *must* be fulfilled for every IAS.<sup>12</sup> This is a subtle but important distinction within the context of risk mitigation. All risks should be identified, but not all risks can be mitigated.

For example, the ability to answer “yes” to several of the risk elements under each DAD requires that the IAS be able to communicate with a human operator. For example:

- *Can the IAS communicate system malfunctions to the human operator?*

*This is not meant to be a requirement.* It is a question that the military commander must ask as a way of assessing risk. For the example above, if the commander determines that adversary capabilities (or the weather, operator fatigue, system malfunction, etc.) cannot possibly cause an interruption of communications with the IAS, then the answer is “yes” and there is no risk of unknown system malfunction that requires mitigation. However, if the commander determines there is a possibility that the answer may be “no,” then a risk of IAS malfunction may be present. The risk must be either mitigated or accepted.

All 13 DADs (and their associated risk elements) should be considered, but given resource limitations (funding constraints for developers, and time constraints for military commanders and operators), not all risks can be mitigated. Some will have to be accepted. A similar example could also be constructed for IAS developers: Does the acquisition program manager have the funding to incorporate assured communications, or will the developers pass the risk of

---

<sup>12</sup> In fact, it may be impossible to impose these requirements on some IAS. For example, operational audit logs would be difficult, if not impossible to require of disposable unmanned systems.

unknown system malfunctions on to the fleet? The DADs and their risk elements help to identify risk, allowing developers and commanders to make fully informed risk assessments.

## Legal versus ethical versus operationally effective

Our second study question seeks an actionable framework to “guide the development, fielding, integration, and employment of militarily effective, legally compliant, and ethically conforming IAS.” We provide that framework in the “How to Use the DADs and Their Elements” section of this report. DOD already has processes in place to address military effectiveness and legal compliance. We saw no indication that IAS characteristics create a need to change the processes that ensure military effectiveness. The only significant impact of IAS characteristics on legal processes is the need to ensure continuity of legal accountability.

We therefore focused our study on IAS ethical conformity. We must note here, however, that among the documents we researched and the individuals we consulted, no clear consensus exists that unequivocally distinguishes ethical concerns from legal concerns. The DAD descriptions in this section, and Appendix B: IAS risk elements, should be read with this in mind.

## Human judgement versus human control

Autonomous weapons systems are a topic of considerable, longstanding, and ongoing debate amongst the nations (High Contracting Parties) represented by a Group of Government Experts at the United Nations Convention on Certain Conventional Weapons. Some nations contend that “appropriate levels of human judgement” are sufficient to mitigate the risks associated with the use of LAWS. Other nations (and advocacy groups) contend that “meaningful human control” is required, the assertion being that human control provides a higher level of risk mitigation than does human judgement. This assertion remains unproven.

The “bottom up” approach of this study described earlier analyzed documents from both sides in this debate, so elements of judgement and elements of control appear throughout our list of risk elements. While these risk elements appear in several of the 13 DADs, human judgement features most prominently in “DAD#9: Human-machine teaming,” and human control features most prominently in “DAD#5: Command and control.” The fact that we have 13 DADs, and not just two, indicates that neither human judgement, nor human control alone (or even together) are sufficient to the task of mitigating the risks associated with IAS<sup>13</sup> use.

---

<sup>13</sup> The UN debate is restricted to LAWS. Our study addresses IAS, which includes non-lethal applications of autonomy technology. With the possible exception of “DAD#13: Civil and natural rights,” our IAS risk elements are equally applicable to LAWS.

It should go without saying, but given the heated nature of the ongoing debate we are obliged to say that our analysis is agnostic to the “judgement versus control” debate. In fact, one could argue from the extensiveness of our risk elements list, and the fact that many elements have little to do with human judgement or control, that to claim that either of these approaches alone (or again, even together) is capable of fully mitigating the risks associated with the use of LAWS or IAS is incomplete. Judgement and control are just two components of a larger solution space.

Our study approach purposely gathered and cataloged as many legal and ethical risk elements as we could find; recognizing that only a subset of the entire list would be applicable in any given IAS application, with the operational context driving the composition of that subset. From this, one could conclude that relying solely on either human judgement or human control to mitigate risk is to ignore the operational context in an attempt to provide a “one size fits all” solution to risk mitigation. Context matters, and context can be complex.

Finally, we remain agnostic to the “judgement versus control” debate because neither term is authoritatively defined or agreed upon such that they can serve as a standard for analytical assessment. Both terms are preceded by a subjective adjective—“appropriate” in the case of judgement, and “meaningful” in the case of control. We cannot measure appropriateness or meaningfulness.

Our risk elements list can be used to mitigate the risks associated with the use of LAWS or IAS right now, even as this debate continues. The elements on this list may be changed by the eventual outcome of this debate (if ever it concludes), but the approach of using a risk elements list is designed to survive intact, no matter the outcome. The DAD descriptions in this section, and Appendix B: IAS risk elements, should be read with this in mind.

## Command and control

To make it even more abundantly clear that our analysis is agnostic to the “judgement versus control” debate, we chose the (military) term “command and control” to describe our control-centric DAD. We also chose this term because unlike “appropriate levels of human judgement,” and “meaningful human control,” this term has an accepted definition (at least in the DOD). The DOD defines “command and control” as “the exercise of authority and direction by a properly designated commander over assigned and attached forces in the accomplishment of the mission [15].” Depending on one’s interpretation of the term “attached forces,” this might appear to limit command and control to directing the activities of the subordinate human beings that constitute those forces. There is, however, an extensive body of literature<sup>14</sup> that makes it clear that military command and control also applies to the “autonomous entities”

---

<sup>14</sup> One need only enter the term “command and control of autonomous systems” (or “vehicles”) into any search engine to get a sampling of this body of work.

employed by those forces. We note also that this literature is not limited to military applications. In fact, one could argue that “bosses who direct employees to do a job” is analogous to the military definition of command and control, so our use of this military term need not limit the application of our DADs and risk elements to military operations. The DAD descriptions in this section, and Appendix B: IAS risk elements, should be read with this in mind.

## Anthropomorphization

Anthropomorphization is the practice of imparting human characteristics to animals or other non-human objects—like robots. We have attempted to expunge anthropomorphization from the DAD descriptions and the risk elements because legal reviewers of this document correctly note that machines<sup>15</sup> cannot be held accountable for the results of military actions under the Law of Armed Conflict (LOAC). Expunging this anthropomorphization proved to be a difficult task because scientists and engineers are busily developing machines that have attributes and capabilities (physical and cognitive) that previously were the exclusive purview of humans.

Many scientists and engineers we interacted with during this study naturally and comfortably slipped into anthropomorphic dialog. Less so, the academic, legal and policy professionals. Our attempts to expunge anthropomorphization sometimes led to what might appear to be verbose or cumbersome language. The DAD descriptions in this section, and Appendix B: IAS risk elements, should be read with this in mind.

## Ethical machines versus ethical use of machines

One very good reason to avoid anthropomorphization is to avoid ascribing human agency to a machine—and then by extension assigning accountability to it in violation of the LOAC as noted above. In discussions regarding ethics and IAS, we rightly note that a machine cannot “be” ethical, but a machine can be “used in an ethical manner.” This distinction reserves and reinforces the assignment of ethical accountability to the human operator (or commander in military applications).

We must note, however, that machines can increasingly do many things that were once the exclusive purview of humans, and the portfolio of human actions (not traits, but actions) that machines can do continues to grow. This portfolio includes many of the activities that are required to “act” ethically. Many of these actions cannot be done by machines, or at least not done well, and at least not yet.

---

<sup>15</sup> We consider an IAS, even one with only software components, to be a machine.

The DAD elements in this report contain many such technologically immature capabilities. As machines continue to amass these capabilities they may appear to “be” ethical, but in fact they are only “acting (more) ethically.” The IAS may not ever be ethical from a philosophical or legal perspective, but it will for all practical purposes “appear to be” ethical from an engineering perspective because it will be able to execute the functions that an ethical entity must be able to execute.

We make this point here to reinforce that we concur with the notion that machines cannot “be” ethical, even if they “act” more and more ethically as autonomy technologies advance. We do not intend to blur the legal distinction between humans and machines, even if some of the capabilities inherent in the autonomy risk elements may appear to do so. The DAD descriptions in this section, and Appendix B: IAS risk elements, should be read with this in mind.

## Decision making versus decision aids

The most difficult steps taken by the authors to avoid anthropomorphization had to do with decision “making” given that our primary objective was to identify the concerns that must be addressed before transferring decision-making capabilities to an IAS. While machines are legally barred from making certain decisions, they are capable of selecting among courses of action based on predetermined selection parameters. Transfer of decision-making capabilities does not imply the authority to make decisions that are reserved by law for humans, nor does it imply the transfer of accountability for any negative consequences of “decisions” that machines are legally permitted to make.

IAS can serve as powerful decision *aids* [25] in support of military commanders. We assert that, an IAS can make a decision—even a decision affecting the four LOAC principles, without violating the principle that only humans are accountable for those decisions. So long as the IAS does not *act on* that decision without human approval the principle of human accountability is not violated. That human approval process means that the human has made the decision and the machine has been relegated to the role of a decision aid. The DAD descriptions in this section, and Appendix B: IAS risk elements, should be read with this in mind. The IAS is a decision aid, and not necessarily the entity that takes an action based on that decision—at least not without human approval.

## The assumption of increased risk

All of the DADs address the need to mitigate the risk associated with the use of IAS. Some of them appear to make the assumption that with increased use of IAS comes increased risk. This is generally the case but may not be true in all situations. We do recognize that increased IAS use does not always increase, or even introduce risk, but this report addresses risk mitigation. The situations where risk is not introduced or increased by IAS use require no new

corrective measures and therefore, are not considered here. The DAD descriptions in this section, and Appendix B: IAS risk elements, should be read with this in mind.

## **DAD#1: Standard semantics and concepts**

- *Are all parties (developers, decision-makers, commanders, and operators) using a common and agreed-upon IAS lexicon so as not to inadvertently introduce risk via miscommunication?*

As the number of organizations developing IAS grows, the need for common definitions and consistency in the semantics that characterize IAS concepts also grows. The main objective of this DAD is to identify terms that, if misunderstood or miscommunicated, could lead to increased risk in the use of an IAS. It is especially important for all persons involved in the life cycle of an IAS (e.g., designers, developers, operators) to take a consistent approach in defining the terms to prevent risks that can inadvertently result in mishaps. Terms that require particular attention are those that are highly subjective and strongly depend on user interpretation and situational context. Consistency is not only important among the different people involved but also across the different phases of military operations. Having standard semantics and concepts may not completely eliminate risk, but reduces the risk of the operator misusing an IAS because of a miscommunication of the designer's original intention.

## **DAD#2: Continuity of legal accountability**

- *For every point in time during an operation, and for every component that may be used<sup>16</sup> during that operation, is there a human who understands how the component functions, is aware of any autonomous functionality within the component, understands the risks associated with its use, and is legally accountable for authorizing its use?*

Autonomous functionality can exist in every "link" of the "kill chain," not just at the pinnacle of the decision to apply lethal force. Similarly, autonomous functionality can exist in every precursor activity of a nonlethal autonomous decision aid. Perceived gaps in accountability magnify legal and ethical concerns in the debate on IAS. Ensuring the continuity of legal

---

<sup>16</sup> DOD seeks to move from static, predefined "kill chains" to adaptive "kill webs" [26-27], where capabilities residing within disaggregated force components can dynamically formulate adaptive webs based on all of the options available [28]. Under this construct, operators and commanders will not know what components will constitute the optimal solution until perhaps the last moment, when the "chain" is assembled from the best options available within the "web."

accountability through every phase of an operation and amongst every interaction between the subcomponents of the overall IAS can help alleviate these concerns.

### DAD#3: Degree of autonomy

- *Can the degree<sup>17</sup> of autonomous functionality be adjusted to accommodate different or changing degrees of risk tolerance?*

The degree of autonomy exists on a spectrum from zero autonomy, in which a human executes all functions, up to full autonomy, in which an IAS executes all functions. Increasing the degree of a system's autonomy can increase the risk of an unintended negative consequence. Additionally, changes in risk tolerance come with changing situations, conditions, and commanders in any dynamic operational environment. Thus, it is advantageous to have the ability to adjust the degree of autonomous functionality to match different risk tolerances—reducing it when risk tolerance is low and increasing it when risk tolerance is high.

### DAD#4: Necessity of autonomy

- *Does use of the IAS impart a military advantage over a non-autonomous alternative system, to include reducing the probability of collateral damage, that justifies accepting the potential risks associated with the use of the IAS?*

The role of autonomy is to extend and complement human capabilities, and it is especially valuable in situations that are “dull, dirty, or dangerous.” Autonomous functionalities also become necessary when an IAS can perform tasks more effectively, such as decision-making at speeds beyond human capability, and when an IAS can help prevent cognitive overload. Ultimately, the potential risks that come with the use of an IAS are justified if an IAS provides a clear advantage over a non-autonomous alternative system. However, if both perform similarly, then the user may opt to use the non-autonomous system to reduce the risks regarding legal and ethical concerns associated with the use of an IAS (when and if these risks are not also present in the non-autonomous alternative).

---

<sup>17</sup> We use the term “degree” rather than “level,” noting only that autonomy exists on a spectrum. Levels imply rigidly defined positions or bands on that spectrum, and attempting to define these levels impedes progress [29].

## DAD#5: Command and control

- *Has every available and practicable measure been taken by the commander to; prevent loss of command and control, relinquish command and control (but not accountability) only in the most dire of circumstances where no other option is available, ensure transfer of command and control without its loss, detect and prevent unintended behaviors, deactivate systems that demonstrate unintended behavior, i.e. to fail safe when command and control is lost, and re-establish command and control as quickly and safely as possible when it is lost due to error, miscommunication, malfunction or enemy action?*

One major concern with IAS is that in some critical circumstances command and control may be difficult or impossible. This DAD seeks to maintain command and control of the IAS to the greatest extent possible and practicable, mitigate against negative consequences when command and control is lost, and minimize *but accept* the situations where the commander must transfer decision-making capabilities to an IAS. The risk elements under this DAD allow the commander to make a risk-informed decision to transfer decision-making capabilities to an IAS. The commander thus remains accountable for any negative consequences that may result.

## DAD#6: Presence of persons and objects protected from the use of force

- *Will there be persons or objects present that cannot be subjected to the effects of the IAS, or that require justification in accordance with applicable laws, policies, and rules of engagement before being subjected to the effects of the IAS?*

IAS use must comply with the LOAC, which requires distinction, proportionality, military necessity, and reducing unnecessary suffering in military operations [30]. LOAC applies to all forms and all instruments of warfare, including IAS. It obliges commanders to consider the presence of certain persons and objects in the operational environment. This DAD addresses specific classes of persons and objects protected from the use of force that use of an IAS must be able to identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement). This DAD specifically does not address the use of IAS that do not use force but can still cause harm, such as an automated or autonomous decision aids. We address those systems in DAD#13: Civil and natural rights.

## DAD#7: Pre-operational audit logs

- *Has positive control been retained over all components, and all processes used to develop the IAS, from the time of its pre-acquisition inception to its delivery to the fleet?*

The purpose of pre-operational audit logs is provenance in the data, software (algorithms), hardware, personnel activities, and processes that are involved in creating an IAS. In order to take action or make a decision, an IAS depends on all of these, but most heavily on the data. Thus, it is critical to ensure that the data are not manipulated, or improperly deleted, and pre-operational audit logs do so by keeping track of authorized users who accesses and processes the data in much the same way as chain of custody is established and maintained in legal proceedings. It is also important to ensure that pre-operational audit logs have detailed descriptions of the origin and trace the development of the algorithms. Pre-operational audit logs also have information on inspecting computer hardware for repairs and modifications. Proper and complete documentation has an important role in establishing a reliable IAS by helping to retain positive control over all aspects of an IAS *during its acquisition*.

## DAD#8: Operational audit logs

- *Are all environmental conditions, sensor inputs, operator interactions, internal processes, IAS actions, and operational outcomes recorded for post-operation reconstruction and analyses to support accountability and improvements to tactics, techniques, and procedures (TTP) and technology?*

An important consideration in establishing confidence in, and acceptance of an IAS is transparency. Documenting inputs, actions, interactions, and outcomes will help create a transparent IAS. The documentation must be detailed and clear enough to promote confidence and acceptance and to enable post-operation reconstruction and analyses. Proper documentation enables users to share suggestions on improvements and lessons learned from previous systems. It also promotes ethical use of an IAS because the traceability created by these logs create the records required to enforce accountability.

## DAD#9: Human-machine teaming

- *Does design of the human-machine team (particularly when the use of force is involved) leverage the human's superior ability to exercise judgment in the use of the IAS in order to reduce the risks associated with its use?*

A common theme running through the literature regarding IAS is that the risks of its use come from its inability to adapt, generalize, or put information into a wider context—thus the concepts of “narrow artificial intelligence” [31] and “brittleness” [32]. It is unfortunate then that warfare (and many other application domains where IAS are used) occurs in an unforgiving environment replete with thinking adversaries that actively seek to push enemy commanders out of their tested and preferred concepts of operation and seek to push the technologies they depend on beyond their original design parameters.

One solution to this problem is to put a human operator “back” into or on the loop. This creates a conundrum in that doing so must not reintroduce the negative characteristics of human performance that the IAS was intended to remove in the first place. Human-machine teaming attempts to design system interactions whereby the respective strengths of the human and the machine are retained, and the weaknesses of each are not. IAS weaknesses can introduce additional legal and ethical risks. Human weaknesses can introduce additional operational risk.<sup>18</sup> Therefore, human-machine team designs provide an opportunity to balance legal and ethical risks against operational risks. The objective of this study was to identify and mitigate the legal and ethical risks, but this must be balanced against operational risk. Hence the inclusion of human-machine teaming as a DAD.

## DAD#10: Test and evaluation adequacy

- *Will/did the IAS test and evaluation procedure reflect the breadth, depth, and complexity of the contemplated operational environment to test and evaluate the system attributes unique to the use of autonomy technologies to the greatest extent practicable?*

The potential negative results of systems used by the DOD include unintended deaths. Such high stakes demand an exhaustive approach for test and evaluation to reduce the risk of these negative consequences. Nondeterministic IAS present a situation where the potential configurations are practically infinite in number, negating any attempt to test each and every possible IAS configuration exhaustively in each and every scenario.

Since there is little hope of eliminating all risk, ways must be found to minimize it to the greatest extent possible within the confines of time and funding constraints. Test and evaluation of IAS is further complicated by subjective terminology in the policy documents that mandate it [1, 5].

This DAD aids the development community in the maximization of test and evaluation, and the minimization of risk, to provide a balance between the depth, breadth, and complexity of

---

<sup>18</sup> We do recognize that humans are not legally or ethically infallible.

testing conditions, recognizing that exhaustively testing along all three axes is certainly not practicable and may not be possible.

## **DAD#11: Autonomy training and education**

- *Does everyone associated with the use of IAS—developers, civilian and military leaders, and operators—understand the attributes of IAS enough to avoid illegal or unethical use?*

Initial and continuing training and education prepares developers, operators, and commanders to establish competence in recognizing the characteristics of an IAS that can lead to illegal or unethical use. Also, through training, human operators prepare for situations that require them to be more active and reestablish command and control for safety or security reasons. Training and education also illuminate the limitations of an IAS, the central role of its data and algorithms, and AI-specific failure modes so that the user can clearly understand the IAS technologies and avoid their misuse.

## **DAD#12: Mission duration and geographic extent**

- *Will the time required to conduct the contemplated mission be so long, and/or the geographic extent of the IAS' operating envelope so extensive, that pre-mission risk mitigation conditions and planning factors will change to an extent that might increase risk?*

The conditions at the beginning of a long-term mission may change with time. They may also change as the geographic range increases and possibly becomes different from what the IAS design calls for. Thus, initial risk assessments and preprogrammed plans may become invalid by the time an IAS needs to execute its task. Considering an IAS' performance with respect to time and space is crucial in preventing misconstrued confidence that an IAS will execute its task with appropriate risk. In response to changing risk mitigation conditions and planning factors, it is important that an IAS be able to operate effectively and be equipped to disable autonomous functionalities.

## **DAD#13: Civil and natural rights**

- *Have all available measures been taken to ensure that the IAS algorithms and/or machine learning (ML) training data employed in automated and autonomous decision aids do not*

*contain any biases that might incorrectly, disproportionately, or otherwise unfairly affect an individual or a group of persons?*

An IAS must be engineered to safeguard human rights—*civil rights*, which are explicitly codified in law, and *natural rights*, which are implicitly accepted to be self-evident by most freedom-loving persons and nations. This DAD contains natural rights that some public policy advocacy groups assert should be codified in law to become civil rights.

DOD intends to develop IAS as LAWS for use against enemy combatants, but also as non-lethal autonomous decision aids that can affect enemy combatants, US servicemembers and their dependents, DOD civilian employees, and third parties (including civilian populations). Clearly, each group has different expectations and protections with respect to civil and natural rights. Enemy combatants, for example, are not protected by many of the civil rights listed in this DAD.

A primary (but by no means only) source of risk to human rights infringement is bias. Therefore, IAS developers must be able to recognize algorithms and data that contain biases that will unfairly affect an individual or a group. Initial mitigation of bias in the software itself involves the developers being aware of a large variety of potential bias sources that can bias an IAS' decisions against a particular individual or group of persons. This includes sources of "analytical" bias that are not inherently biased against any group or person but could become so if not identified and addressed.

Notifications of when an IAS is in use further mitigate the effects of bias, as they bring awareness to the persons affected by the IAS' actions and decisions. Not all persons enjoy this right to notification. It is clearly not practical to extend this right to suspected criminals in law enforcement or some homeland security applications. This DAD also addresses the bias that occurs when operating in coalition and allied environments, such as bias that stems from different sets of treaties, ROE, or cultural norms. Overall, it is important for all persons involved—the developer, operator, commander, affected person, and oversight specialists—to be vigilant in detecting bias that may result in human rights violations.

# How to Use the DADs and Their Elements

---

The common theme across the entire list of 565 risk elements that constitute the 13 DADs is the ability to identify and mitigate the risks associated with the use of IAS—not the technical or operational risks, but the legal and ethical risks. Our list of risk elements is extensive, so it is clearly impractical to use every element to mitigate every risk in every situation.

Risk mitigation requires the expenditure of scarce resources; program managers developing the IAS must balance the competing forces of cost, schedule, and performance. For the military commanders using the IAS, time is a scarce resource. Risk mitigation procedures often have an opportunity cost of some sort. Risk mitigation resources are finite, so risk management entails mitigating risk when resources are available to do so, accepting risk when they are not, or simply not using the system that generates the risk (while acknowledging that nonuse may create its own risks).

In this resource-constrained environment, some risk cannot be mitigated. Risk mitigation must be prioritized, with threats to human health or safety clearly being the highest priorities. Our goal is to help IAS developers and users identify risks and risk mitigations so they can make *fully informed* risk mitigation or acceptance decisions based on the constraints of their operational environment, be it a budget environment or a combat environment.

## Defining a key performance parameter for the presence of ethical use enablers

A program manager seeking to comply with a KPP is, as discussed above, operating in a resource-constrained (budget) environment. Requirements intended to meet a KPP are thus written in a threshold/objective format:

- Threshold requirements are “must have” requirements. If they cannot be met with available resources, then more resources must be obtained, or the requirement must be relaxed, and additional risk must be accepted—else the program must be halted.
- Objective requirements are “nice to have” requirements. They are desired and, in our case, mitigate additional risk. But they are not critical to system operation; they can be addressed as resource availability permits.

In consultation with the various oversight boards that are part of the defense acquisition system’s processes [1, 33], the program manager determines which requirements are

objective, and which are threshold. In doing so, they tacitly accept some risks while mitigating others.

To use the DAD risk elements to define a KPP for the presence of ethical use enablers for a contemplated IAS, program managers would survey the entire list, discard any risk elements not applicable to the system based on its envisioned use, and then separate the remaining risk elements into threshold and objective requirements. In doing so, they define the resources required to meet the KPP via the threshold requirements and also define the programmatic risk tolerance via the objective requirements. It should be noted that any risk not mitigated during acquisition will be passed along to military commanders who take delivery of the system and are expected to use it; they must conduct their own operational risk assessment before using the system.

## Conducting operational risk assessments

As previously noted, an IAS will not be delivered to the fleet if it is inherently illegal, but a legal weapon can still be used in an illegal manner. Similarly, with the adoption of the DOD AI ethical principles, an IAS will not be delivered to the fleet if its use would be inherently unethical, but it can still be used in an unethical manner. Military commanders must therefore make their own risk assessment prior to IAS use to ensure legal compliance and ethical conformity.

Ideally, many of these risk management decisions are made long before an operation is about to commence. Risks are mitigated or accepted during acquisition, as described above, but risk mitigation also occurs during mission planning [16]. Military commands make plans for potential operations in their areas of responsibility so that they are prepared if and when that operation must be conducted.<sup>19</sup> Part of that planning entails identifying and understanding potential risks and weighing those against the military necessities required to accomplish the military objectives.

For legal and ethical issues, military commanders receive advice and support from lawyers—principally, the Staff Judge Advocate (SJA) [30]. This support occurs during long-range planning for potential future operations, and during short-range planning immediately preceding an operation that is about to commence.

During both the long- and short-term planning exercises, the supporting SJA could use the DAD risk elements as a ready reference for the many IAS-associated risks that must be mitigated or accepted. Ideally, the risk elements will be incorporated into standing, theatre, and mission specific ROE long before operations commence. As was done by the acquisition program

---

<sup>19</sup> For example, War Plan Orange, the plan to conquer the Japanese empire in the Pacific, was a decades-long planning process prior to World War II [34].

manager, the SJA, in consultation with technical and operational SMEs on the commander's staff [35], can determine which risk elements are relevant to the contemplated operation and which risk elements are relevant to the IAS available to conduct it. Of those risk elements that are relevant, the SJA can make an assessment as to the feasibility of implementing available risk mitigation measures and then advise the commander accordingly. Using the DAD risk elements as a pre-operational "checklist" ensures that all IAS-associated risks that must be considered *are* considered. The risk assessment is thus fully informed.

## Other potential uses

We crafted the risk elements that constitute the DADs to be customizable to any IAS application domain simply by choosing the risk elements appropriate to the desired application and disregarding those that are not. For example, risk elements that address the application of lethal force are not required when mitigating the risk that an IAS at a bank might deny a mortgage application. Risk elements that address privacy concerns are not required for battlefield applications. The customizable use of generic risk elements means that the DAD risk elements have the potential to mitigate the legal and ethical risks of IAS use far beyond the military centric KPP and operational checklist use cases above that were the primary objective of this study. We detail below other potential use cases.

### DOD contracting officers

The DOD AI ethical principles require that all AI-enabled systems be responsible, equitable, traceable, reliable, and governable [5-6]. Presumably, contractors that provide AI to DOD must deliver solutions that adhere to these five principles. Unfortunately, these terms are open to subjective interpretation, making their enforcement in DOD acquisition problematic.

DOD could improve enforcement and adherence if contracting officers provide potential vendors an explicit checklist in any request for proposals that, if met, would constitute contractual adherence to the five principles. We assert that the list of DAD risk elements developed in this study could serve as just such a measurable and testable checklist. Depending on the application, contracting officers could draw from the list to develop *explicit* contract requirements that enable adherence to the DOD AI ethical principles. Contract requirements are often written in the form of "shall statements." For example, the following risk element, in the form of a question:

- Can the IAS recognize symbols that designate persons and objects protected from the use of force, such as a Red Cross or Red Crescent?

Is easily convertible into a contractual requirement in the form of a declarative shall statement:

- The IAS shall recognize symbols that designate persons and objects protected from the use of force, such as a Red Cross or Red Crescent.

The DOD AI ethical principles are already a requirement for at least one DOD request for proposals. The Joint AI Center's Data Readiness for AI Data (DRAID) request states that "orders executed with the DRAID will explicitly include a task requiring the contractors to demonstrate how their products and solutions address or instantiate the DOD AI Ethical Principles" [36]. Providing prospective contractors with a checklist that would explicitly state how their products and solutions can conform to DOD's AI ethical principles would ease their task and lower the bar to entry for the small and innovative companies that this contracting vehicle explicitly seeks to encourage. This would also provide government contract officers with objective and measurable requirements upon which to base contract adherence. As an illustrative example, the ethical use principle of *traceability* could be met by choosing from the risk elements listed under the two "audit log" DADs above (see the recommendation below on reimagining the approach to "defining" standard terminology).

## Law enforcement

We noted the many similarities between DOD and law enforcement use of AI when we noted that facial recognition algorithms are essentially "targeting" algorithms. Indeed, just as the Campaign to Stop Killer Robots is a coalition of groups that seeks a ban on AI in LAWS, the American Civil Liberties Union heads a coalition of groups seeking a presidential moratorium on the use of facial recognition technology in the United States [37]. The group European Digital Rights calls for a similar ban in Europe [38]. Nineteen of the documents in Appendix A deal directly with concerns regarding facial recognition. Thus, it should not be surprising that our list of DAD risk elements could be useful for law enforcement applications too. These organizations also acquire and field IAS that have the potential to result in negative consequences for individual persons and groups.

## Automated decision systems

Automated decision systems are ubiquitous, and while the stakes are lower than for DOD use of LAWS, these systems still have the potential to negatively impact individual persons or groups [39]. Several documents in Appendix A deal with these systems; hence, as with the law-enforcement use case, our list of DAD risk elements could be used to develop customizable acquisition requirements and risk mitigation frameworks for these systems too.

## SMEs

Responses from our SME request for comment included specific potential uses:

- Concept and doctrine development for overall system and subsystems

- Decision-making criteria for all stages in the systems engineering process
- Generation and decomposition of the system design process requirements
- Interrogation of uses for, and impacts of, the system to drive design decisions
- Risk assessment for design evaluation and changes
- Criteria for automated test case development in DevSecOps<sup>20</sup> pipelines.

Maintaining an interdisciplinary autonomy risk element list across scientific, engineering, acquisition, doctrinal, policy, and warfighting user communities will enable a broader, more transparent understanding of the development and use of IAS. It will help clear up miscommunication and establish coherence between different user communities.

---

<sup>20</sup> Development, security, and operations.

# Conclusion

---

The DOD is committed to the ethical use of the AI and robotics technologies that enable IAS, including LAWS. This commitment is articulated in policy [1, 5], which is a necessary, but not sufficient first step toward achieving ethical use. To carry these policies into practice, the scientists, engineers and acquisition professionals that develop and acquire autonomous systems, and the strategic-, operational-, and tactical-level military commanders that order their use or use them require additional tools.

This CNA analysis sought to provide at least some of the required tools in the form of the DADs and the list of risk elements that constitute them. These tools facilitate the requirements definition needed to acquire autonomous systems. We purposefully developed them with an eye toward use within current Defense Acquisition System processes [2]. These tools also facilitate the risk assessment that military commanders conduct when deciding whether a contemplated use of an IAS adheres to the LOAC, DOD policies, DOD AI ethical principles, and ROE. Thus we also purposefully developed them with an eye toward use within the Joint Planning Process [16]. They are also adaptable to the planning processes of the other services, and to those on non-military organizations as well.

This CNA analysis also sought to demonstrate the depth of DOD's commitment to the ethical use of IAS by publishing these "implementation enablers" in a publicly available document, subject to continued public scrutiny, debate, and collaboration. We expect the transparency of this approach to lead to improved public trust in DOD's use of IAS and a reduction in the misinformation, miscommunication, and misinterpretation of intent present in the current debate over the use of AI and robotics technologies.

# Recommendations

---

## Mandate a KPP for the presence of ethical use enablers for IAS

This report provides a tool that acquisition professionals can use to develop measurable, testable, and thus *enforceable* KPPs. Policy is a necessary but not sufficient first step toward the ethical use of IAS. There is plenty of evidence from the DOD's long and unsuccessful attempts to "require" data sharing suggesting that policy without a credible enforcement mechanism<sup>21</sup> seldom achieves the policy's objectives [40].

When a requirement is written down in an acquisition document, it becomes part of a legally enforceable contract between the government and the vendor. Generally, if the requirement is not met, the vendor is not paid.<sup>22</sup> KPPs and their measurable and testable parameters are one vehicle by which requirements become contractual obligations. A policy that makes a KPP mandatory is thus an enforceable policy.

We recommend that the presence of ethical use enablers be made a mandatory KPP for all IAS. Legal use is already mandatory for *all* weapons systems [2-3], so a mandatory KPP to accomplish legal use is not required for IAS. Presumably, legal use would encompass ethical use, but perhaps not.<sup>23</sup> We also note that DOD policy places additional review requirements for both nonlethal weapons [4] and autonomous weapons [1]. The additional requirement that the presence of ethical use enablers become a mandatory KPP for IAS is therefore deemed necessary and is consistent with exiting DOD policy.

Using the term "presence of ethical use enablers" instead of just "ethical use" is cumbersome, but necessary. Recall from an earlier section of this report "Putting the 13 DADs in context" that the IAS cannot "be" ethical, so policy requires "ethical use" by the human operator or commander. The defense acquisition system levies requirements on acquired machines, not

---

<sup>21</sup> "One more unenforceable policy, law, framework, or strategy document will not help. If you are drafting such a document, and are not including enforcement mechanisms, put down your pen" [40]. It should also be noted that even if enforcement mechanisms exist, they are equally ineffective without funds for implementation.

<sup>22</sup> We say "generally" because in the acquisition of complex military systems, vendors are sometimes granted relief from the government if a requirement cannot be met due to circumstances beyond their control.

<sup>23</sup> In his book, *Army of None: Autonomous Weapons and the Future of War* [7], Paul Scharre recounts an incident in Afghanistan where a young girl was scouting his position in support of Taliban fighters. While it would have been legal to fire on the girl, Scharre and his fellow US Army squad members concluded that it would not be ethical and did not fire.

human operators or commanders. An “ethical use” KPP is therefore nonsensical and not executable. The “requirements” placed on the humans that DOD “acquires,” (to become operators and commanders) are levied by the service-level personnel systems, not the defense acquisition system, and come in the form of knowledge, skills, and abilities, not requirements.

We note without further comment (because it is beyond the scope of the present study) that the merging of human and machine capabilities in human-machine teaming applications may well blur the lines between the human-centric and machine-centric DOD systems and processes if DOD is to fully leverage the promised human-machine synergies.

Making the presence of ethical use enablers one of just five mandatory KPPs for IAS (four already exist for all DOD acquisitions) would send a clear and unequivocal signal to all concerned parties that the DOD is committed to the ethical use of autonomy technologies. Failure to do so sends the opposite signal and opens up the DOD to continued criticism and opposition from advocacy groups opposed to LAWS.

## **Incorporate risk mitigation checklists into doctrine and planning**

This report also provides a tool that commanders can use to mitigate the risks associated with IAS use. Since those risks are of a legal and/or ethical nature, the person on the commander’s staff primarily responsible for this effort would be the Legal Counsel, General Counsel, or Staff Judge Advocate, depending on the level of command. Joint Publication 1-04 *Legal Support to Military Operations* [30] describes how these legal advisors support the commander.

We recommend revising Joint Publication 1-04 (and its subordinate service-level publications) to incorporate the use of checklists that mitigate the legal and ethical risks created by the use of IAS. The checklists should be derived from an *authoritative and standardized* list of risk mitigation elements (see next recommendation).

## **Maintain an authoritative and standardized joint autonomy risk elements list (JAREL)**

The execution of this study made it clear that developing the elements of a KPP or a risk mitigation checklist “from scratch” is a time-consuming and labor-intensive endeavor. We cannot reasonably expect the developers and users of IAS to do so for every new IAS or for every substantive modification of an existing IAS. KPP and checklist development requires *a menu of elements in a common language, which serves as the foundation for KPP and risk mitigation-checklist development.*

This approach is not without precedent: It is analogous to the use of the universal joint task list (UJTL) to develop mission essential task lists (METLs) tailored to specific missions in specific operational environments under specific ROE. The UJTL is *“a menu of tasks in a common language, which serves as the foundation for joint operations planning across the range of military and interagency operations”* [41]. The UJTL is maintained, updated quarterly, and made available to mission planners via the Joint Electronic Library website.

We recommend the creation and continued maintenance of a “joint autonomy risk elements list” (JAREL) similar in form and function to the UJTL. We further recommend that the risk elements developed in this study (listed in Appendix B: IAS risk elements) serve as the initial JAREL.

## **Make the JAREL publicly available to the greatest extent possible**

There may be JAREL elements whose inclusion in a publicly available list would disclose potential vulnerabilities of US military forces. Clearly, risk elements of this nature should not be publicly disclosed and may even have to be classified. We anticipate that the vast majority of JAREL elements would be releasable to the public because they would not become controlled or classified until tied to a particular region, operation, or joint warfighting concept. Harkening back to the analogy between the UJTL and the JAREL, the “disassociated” UJTL is unclassified, but the METLs derived from them and associated with a specific mission may then become classified.

Making all JAREL elements that do not disclose potential vulnerabilities publicly available has several significant advantages:

- The transparency created by publishing the JAREL reduces the occurrence of misinformation, miscommunication, and/or misinterpretation regarding the statements and intentions of the parties involved in the development of AI for warfare systems and those who oppose this development.
- Public trust of the military in its use of autonomy technology would likely be elevated from its current low levels. [42]
- Private sector vendors (and current and prospective DOD employees) that might hesitate to participate in DOD IAS development [43] may now be willing to do so, given their better understanding of DOD’s legal and ethical safeguards. DOD can ill afford to forego their contributions.

- Public participation and input into the quarterly updates to the JAREL become an opportunity to engage and collaborate with industry, academia, and the public to further build trust in DOD’s use of IAS.
- Public availability allows non-DOD developers and users of nonmilitary IAS—in government and in the private sector—to build measurable and testable functional requirements and risk mitigation checklists for their own products. In this way, a significant and wide-ranging social good can be realized from DOD investment in the JAREL.
- A wider pool of JAREL users will increase the lessons learned from its use, and those lessons can flow to, and from, the DOD—again realizing a significant and wide-ranging social good from DOD’s investment.
- As more IAS practitioners from government, industry, and academia begin to use the same standard JAREL for requirements definition and risk reduction, collaboration among them becomes more seamless. Administrative, organizational, and contractual “friction” decreases.
- DOD will become the intellectual leader in this area, rather than the follower or adopter of private industry achievements. This can make DOD more competitive with private industry employers in the battle to attract an IAS development workforce.
- DOD leadership in the implementation of legal and ethical IAS applications can also provide a comparative military advantage vis-à-vis other nations in the competition to attract allies and partners.

## Reimagine the approach to “defining” standard terminology

One of our DADs (*standard semantics and concepts*) addresses the need for all parties to use consistent terminology to prevent the operational risk that can result from miscommunication. IAS development spans a diverse set of stakeholder groups with different perspectives, interpretations, and even objectives. Even when objectives are not in conflict, arriving at a common lexicon can be an exceedingly difficult thing to do<sup>24</sup> and can cause working groups to get bogged down in definitional and terminological debates and squabbles that impede progress toward their ultimate objective. The years-long inability to arrive at consensus definitions with respect to LAWS has been pointed out as one of the factors impeding progress

---

<sup>24</sup> One member of the CNA study team (Stumborg) participated in the development of the Weapons Technical Intelligence Improvised Explosive Device Lexicon [44], which was eventually adopted and published by the United Nations Mine Action Service [45]. This observation stems from that experience.

among delegates to the United Nations Convention on Certain Conventional Weapons as they grapple with the appropriate use of autonomy in weapons systems [46-48].

Again, this report notes that policy is a necessary but not sufficient first step toward the ethical use of IAS—one of the previously stated reasons for this is that policy in and of itself may not provide the tools required to implement it. Another reason that policy is an insufficient first step is the use of subjective terminology that is prone to misinterpretation. For example, DOD Directive 3000.09 uses the term “appropriate levels of human judgment,” but the Campaign to Stop Killer Robots prefers “meaningful human control.” At least one DOD legal scholar considers this to be a distinction without a difference [49]; other individuals that the study team consulted disagree. This ongoing “definitional conflict” creates much of the miscommunication and misunderstanding in the debates over the ethical use of autonomy in warfare.

Similarly, the five DOD AI ethical principles [5] are not yet actionable or implementable as written because they contain words that are subject to misinterpretation.

We propose a different approach. Rather than attempt to define and then attempt to adhere to these terms using still more terms that may in turn also be subject to misinterpretation, we propose that DOD define what it means to adhere to each term by choosing from the list of measurable and testable JAREL entries that, if met, would constitute adherence. The JAREL represents explicit conditions that an engineer can work to meet. As an example, one of the five DOD AI ethical principles is “traceable,” defined as follows:

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. [5]

Troublesome subjective terms in this definition include “relevant,” “appropriate,” “transparent,” and “auditable.”

We contend that the 53 risk elements that constitute the “pre-operational audit logs” DAD are an actionable “yes/no checklist” that could be adopted to “define” an IAS as “traceable” when the engineer is able to answer “yes” to all 53 conditions.<sup>25</sup> We do acknowledge that while most of the 53 risk elements are extremely difficult to misinterpret (i.e., “Is the IAS training data retention policy documented?”), a small number could be subject to misinterpretation (i.e.,

---

<sup>25</sup> We previously discussed the difference between *threshold* and *objective* requirements. In this particular example, all 53 requirements are treated as threshold requirements. This is only an example. In practice, some subset of the 53 requirements would likely be threshold, with the remainder being objective requirements or not adopted as requirements at all.

“Does the IAS use any open-source training data with absent or suspect provenance documentation?” This depends on one’s definition of the subjective term “suspect.”).

In a similar fashion, adherence to the “equitable” principle could be defined using the risk elements from the “civil and natural rights” DAD. Adherence to the DOD Directive 3000.09 requirement that a LAWS be tested in a “realistic operational environment” (“realistic” being the troublesome subjective term here) could be defined using the risk elements under the “test and evaluation adequacy” DAD.

## Create a research and development portfolio for ethically conforming IAS

Many of the 565 IAS risk elements listed in this report also comprise a list of capabilities that, if realized, would enable ethically conforming IAS. Some of these capabilities can be realized by new or modified TTPs, and others are amenable to a technology solution. The latter should therefore be considered for investment by the DOD research and development (R&D) enterprise.

IAS development requires many other technologies not listed here (navigation, communications, etc.), so the recommended R&D effort would be just a portion of the overall IAS R&D portfolio—the portion that enables IAS to be used in an ethically conforming manner.

The listed capabilities currently exist at various levels of technical maturity, from those that would likely require extensive basic scientific research, such as:

- Can the IAS distinguish between benign and hostile intentions?

To those that require little more than an integration of existing technologies, such as:

- Can the IAS detect both sharp and gradual changes in its own performance and provide alerts to the human operator and other systems based on criticality?

To accelerate the delivery of ethically conforming IAS, we recommend that DOD research program directors with responsibilities for autonomy technologies use the results of this study to identify and prioritize their R&D program’s investments using a four-step process:

1. Identify which of the 565 autonomy risk elements in this report describe capabilities that could be achieved by developing a technical solution.
2. Evaluate the technology readiness level of each and assign responsibility to either basic research program managers, or engineering development program managers, as appropriate.
3. Prioritize the list of potential investments based on:

- a. the impact its development would have across contemplated DOD IAS investments,
  - b. the impact its development would have on IAS-dependent concepts of operation,
  - c. the time and funding required to develop and deliver the technology.
- 4. Monitor JAREL development for new candidate technology investments.

This page intentionally left blank.

# Appendix A: Bibliography of documents consulted

---

The documents listed below were the source of the 565 risk elements we used to identify the 13 DADs. They are categorized for convenience only, noting that several documents could be listed in more than one of our chosen categories. For example, at least one listed nongovernmental organization (Campaign to Stop Killer Robots) is an umbrella organization made up of multiple organizations, some of which appear elsewhere in this list. In addition, the documents listed under national governments<sup>26</sup> and international governmental organizations are written *by* them or *about* their approach to AI and autonomy. All documents are available online or are cleared for public release.

## Academia

- *A Roadmap for US Robotics from Internet to Robotics 2013 Edition*. Georgia Institute of Technology, Carnegie Mellon University, University of Pennsylvania, University of Southern California, Stanford University, University of California–Berkeley, University of Washington, Massachusetts Institute of Technology. Mar. 2013.  
<http://archive2.cra.org/ccc/files/docs/2013-Robotics-Roadmap>.
- Challen, Robert, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, and Krasimira Tsaneva-Atanasova. *Artificial Intelligence, Bias and Clinical Safety*. 2019. *BMJ Quality and Safety* 28 (3): 231-237.  
<https://qualitysafety.bmj.com/content/28/3/231>.
- Dankar, F. K., and M. Ibrahim. *Fake It till You Make It: Guidelines for Effective Synthetic Data Generation*. *Journal of Applied Science*. 2021. 11(5), 2158.  
<https://www.mdpi.com/2076-3417/11/5/2158>.
- Ekelhof, Merel A.C. *The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting*. 2019. PhD, Vrije Universiteit Amsterdam.

---

<sup>26</sup> The astute reader of this bibliography will note that the People's Republic of China and the Russian Federation are heavily over-represented by document count. This was intentional. These countries have expressed a desire to lead the world in AI, have large militaries, and have governance and civil rights structures very different from most of the other nations sampled in this bibliography. Additionally, CNA has a China studies group and a Russia studies group with analysts that read and speak the Mandarin Chinese and Russian languages respectively.

<https://research.vu.nl/ws/portalfiles/portal/90547655/cover.pdf>.

- Ekelhof, Merel A.C. *Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation*. Sep. 2019. *Global Policy* 10 (3): 343-348. doi: 10.1111/1758-5899.12665.  
<https://onlinelibrary.wiley.com/doi/pdf/10.1111/1758-5899.12665>.
- Hobbs, Alan, and Beth Lyall. *Human Factors Guidelines for Unmanned Aircraft Systems*. Apr. 2016. *Ergonomics in Design* 24 (3): 23-28.  
[https://www.researchgate.net/publication/301716592\\_Human\\_Factors\\_Guidelines\\_for\\_Unmanned\\_Aircraft\\_Systems](https://www.researchgate.net/publication/301716592_Human_Factors_Guidelines_for_Unmanned_Aircraft_Systems).
- Johnson, Matthew. *Coactive Design: Designing Support for Interdependence in Human-Robot Teamwork*. Sep. 2014. PhD, Delft University of Technology.  
<https://repository.tudelft.nl/islandora/object/uuid:35cfe91a-bd59-427d-89a9-e2019f9b0c28?collection=research>.
- Johnson, Matthew, et. al., *Team IHMC's Lessons Learned from the DARPA Robotics Challenge: Finding Data in the Rubble*. Sep. 2016. *Journal of Field Robotics* 00 (1): 1-21. doi: 10.1002/rob.21674.  
[https://www.researchgate.net/publication/308340634\\_Team\\_IHMC%27s\\_Lessons\\_Learned\\_from\\_the\\_DARPA\\_Robotics\\_Challenge\\_Finding\\_Data\\_in\\_the\\_Rubble](https://www.researchgate.net/publication/308340634_Team_IHMC%27s_Lessons_Learned_from_the_DARPA_Robotics_Challenge_Finding_Data_in_the_Rubble).
- MacLachlan, Robert A., and Christoph Mertz. *Tracking of Moving Objects from a Moving Vehicle Using a Scanning Laser Rangefinder*. Sep. 2006. Proceedings of the IEEE Intelligent Transportation Systems Conference, Toronto, Canada.  
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1706758>.
- Marge, Matthew, Carol Espy-Wilson, and Nigel G. Ward. *Spoken Language Interaction with Robots: Research Issues and Recommendations*. ArXiv:2011.05533v1. Nov. 11, 2020. <https://arxiv.org/abs/2011.05533>.
- Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. *A Survey on Bias and Fairness in Machine Learning*. Sep. 17, 2019. <https://arxiv.org/abs/1908.09635>.
- Phillips, P. J., A. N. Yates, Y. Hu, C. A. Hahn, E. Noyes, K. Jackson, J. G. Cavazos, G. Jeckeln, R. Ranjan, S. Sankaranarayanan, J. C. Chen, C. D. Castillo, R. Chellappa, D. White, and A. J. O'Toole. *Face Recognition Accuracy of Forensic Examiners, Superrecognizers, and Face Recognition Algorithms*. Proceedings of the National Academy of Sciences. 2018. 115 (24): 6171-6176. <https://www.ncbi.nlm.nih.gov/pubmed/29844174>.
- Righetti, Ludovic, Raj Madhavan, and Raja Chatila. *Unintended Consequences of Biased Robotic and Artificial Intelligence Systems (Ethical, Legal, and Societal Issues)*. 2019. *IEEE Robotics & Automation Magazine* 6 (3): 11-13.

<https://ieeexplore.ieee.org/document/8825881>.

## Books

- Brynjolfsson, Eric, and Andrew McAfee. *The Second Machine Age*. 2014. New York: WW Norton & Company.
- Daugherty, Paul R., and James Wilson. *Human + Machine: Reimagining Work in the Age of AI*. 2018. Boston, MA: Harvard Business Review Press.
- Hughes, CAPT Wayne P. Jr. USN (Ret.) and RADM Robert P. Girrier USN (Ret.). 2018. *Fleet Tactics and Naval Operations*. 3<sup>rd</sup> ed. Annapolis, MD: Naval Institute Press.
- O'Neil, Cathy. *Weapons of Math Destruction*. 2016. New York: Crown.
- Polany, Michael. *The Tacit Dimension*. 1966. Garden City, New York: Doubleday and Company.
- Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. 2018. New York: W. W. Norton & Company.

## Public policy advocacy groups

### Ada Lovelace Institute

- *Beyond Face Value: Public Attitudes to Facial Recognition Technology*. Sep. 2019. <https://www.adalovelaceinstitute.org/wp-content/uploads/2019/09/Public-attitudes-to-facial-recognition-technology-v.FINAL.pdf>.
- *Examining the Black Box: Tools for Assessing Algorithmic Systems*. Apr. 23, 2020. <https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/>.
- *Exploring Legal Mechanisms for Data Stewardship*. Mar. 2021. <https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/>.
- *Inspecting Algorithms in Social Media Platforms*. Aug. 6, 2020. <https://www.adalovelaceinstitute.org/wp-content/uploads/2020/11/Inspecting-algorithms-in-social-media-platforms.pdf>.
- *Transparency Mechanisms for UK Public-sector Algorithmic Decision-making Systems*. <https://www.adalovelaceinstitute.org/wp-content/uploads/2020/10/Transparency-mechanisms-explainer-1.pdf>.

## AI Now Institute

- Kak, Amba, ed. *Regulating Biometrics: Global Approaches and Urgent Questions*. Sep. 1, 2020. <https://ainowinstitute.org/regulatingbiometrics.pdf>.
- Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker. *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*. Apr. 2018. <https://ainowinstitute.org/aiareport2018.pdf>.
- Richardson, Rashida. *Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*. Dec. 2019. <https://ainowinstitute.org/ads-shadowreport-2019.pdf>.

## American Civil Liberties Union

- Ruane, Kate. *Face Recognition Technology Moratorium Coalition Letter*. Feb. 16, 2021. [https://www.aclu.org/sites/default/files/field\\_document/02.16.2021\\_coalition\\_letter\\_requesting\\_federal\\_moratorium\\_on\\_facial\\_recognition.pdf](https://www.aclu.org/sites/default/files/field_document/02.16.2021_coalition_letter_requesting_federal_moratorium_on_facial_recognition.pdf).

## Article 36

- *Regulating Autonomy in Weapon systems*.  
<https://article36.org/wp-content/uploads/2020/10/Regulating-autonomy-leaflet.pdf>.

## Campaign to Stop Killer Robots

- *Key Elements of a Treaty on Fully Autonomous Weapons*. Nov. 2019.  
<https://www.stopkillerrobots.org/wp-content/uploads/2020/04/Key-Elements-of-a-Treaty-on-Fully-Autonomous-WeaponsvAccessible.pdf>.

## Computing Community Consortium & Association for the Advancement of Artificial Intelligence

- Gil, Yolanda, and Bart Selman. *A 20-Year Community Roadmap for Artificial Intelligence Research in the US*. Aug. 2019. Computing Community Consortium & Association for the Advancement of Artificial Intelligence. <https://cra.org/ccc/wp-content/uploads/sites/2/2019/08/Community-Roadmap-for-AI-Research.pdf>.

## Human Rights Watch

- *Losing Humanity: The Case Against Killer Robots*. Nov. 19, 2012.  
<https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>.

- *Review of the 2012 US Policy on the Autonomy in Weapon systems.* Apr. 15, 2013. <https://www.hrw.org/news/2013/04/15/review-2012-us-policy-autonomy-weapons-systems>.
- *Shaking the Foundations: The Human Rights Implications of Killer Robots.* May 12, 2014. <https://www.hrw.org/report/2014/05/12/shaking-foundations/human-rights-implications-killer-robots>.

## International Committee of the Red Cross

- *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons.* Mar. 2016. [https://icrcndresourcecentre.org/wp-content/uploads/2017/11/4283\\_002\\_Autonomus-Weapon-Systems\\_WEB.pdf](https://icrcndresourcecentre.org/wp-content/uploads/2017/11/4283_002_Autonomus-Weapon-Systems_WEB.pdf).

## International Committee for Robot Arms Control

- Amoroso, Daniele, and Guglielmo Tamburrini. *What Makes Human Control over Weapon systems "Meaningful"?* Aug. 2019. [https://www.academia.edu/40112926/WHAT\\_MAKES\\_HUMAN\\_CONTROL\\_OVER\\_WEAPON\\_SYSTEMS\\_MEANINGFUL](https://www.academia.edu/40112926/WHAT_MAKES_HUMAN_CONTROL_OVER_WEAPON_SYSTEMS_MEANINGFUL).
- Gubrud, Mark, and Jürgen Altmann. *Compliance Measures for an Autonomous Weapons Convention. Working Paper #2.* May 2013. [https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Altman-Compliance-Measures-AWC\\_ICRAC-WP2.pdf](https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Altman-Compliance-Measures-AWC_ICRAC-WP2.pdf).
- Sharkey, Noel. *Guideline for Human Control of Weapon systems. International Committee for Robot Arms Control. Working Paper for CCW GGE.* Apr. 2018. [https://www.icrac.net/wp-content/uploads/2018/04/Sharkey\\_Guideline-for-the-human-control-of-weapons-systems\\_ICRAC-WP3\\_GGE-April-2018.pdf](https://www.icrac.net/wp-content/uploads/2018/04/Sharkey_Guideline-for-the-human-control-of-weapons-systems_ICRAC-WP3_GGE-April-2018.pdf).

## International Panel on the Regulation of Autonomous Weapons

- *Focus on Human Control. Report No. 5.* Aug. 2019. [https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-09\\_iPRAW\\_HumanControl.pdf](https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-09_iPRAW_HumanControl.pdf).
- *A Path Toward the Regulation of LAWS.* May 2020. [https://www.ipraw.org/wp-content/uploads/2020/05/iPRAW-Briefing\\_Path-to-Regulation\\_May2020.pdf](https://www.ipraw.org/wp-content/uploads/2020/05/iPRAW-Briefing_Path-to-Regulation_May2020.pdf).
- *Verifying LAWS Regulation - Opportunities and Challenges. iPRAW Working Paper.* Aug. 2019. [https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-16\\_iPRAW\\_Verification.pdf](https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-16_iPRAW_Verification.pdf).

## The Open Data Institute

- *Getting Data Right: Perspectives on the UK National Data Strategy 2020*. Nov. 24, 2020. <https://theodi.org/article/getting-data-right-perspectives-on-the-uk-national-data-strategy-2020/>.

## PAX for Peace

- *Killer Robots: What are they and What Are the Concerns?* <https://paxforpeace.nl/media/download/pax-booklet-killer-robots-what-are-they-and-what-are-the-concerns.pdf>.
- Slijper, Frank. *Slippery Slope: The Arms Industry and Increasingly Autonomous Weapons*. Nov. 2019. <https://paxforpeace.nl/what-we-do/publications/slippy-slope>.

## Stockholm International Peace Research Institute

- Boulanin, Vincent, Neil Davison, Netta Goussac, and Moa Peldan Carlsson. *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*. Jun. 2020. <https://www.sipri.org/publications/2020/other-publications/limits-autonomy-weapon-systems-identifying-practical-elements-human-control-0>.
- Boulanin, Vincent, and Maaïke Verbruggen. *Mapping the Development of Autonomy in Weapon Systems*. Nov. 2017. [https://www.sipri.org/sites/default/files/2017-11/siprireport\\_mapping\\_the\\_development\\_of\\_autonomy\\_in\\_weapon\\_systems\\_1117\\_1.pdf](https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf).

## International governmental organizations

### Europe/multinational

- *Capstone Report: Robotic and Autonomous Systems in a Military Context*. Jan. 2021. The Hague Centre for Strategic Studies. <https://hcss.nl/report/capstone-report-robotic-and-autonomous-systems-in-a-military-context/>.
- *Communication from the Commission to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe*. European Commission. Apr. 25, 2018. SWD(2018) 137 final. <https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF>.

- *Regulation on a European Approach for Artificial Intelligence*. 2020. The European Parliament and The Council of the European Union. <https://fpf.org/wp-content/uploads/2021/04/4858d4e3-fece-4e56-8a57-66a5d440c361-AI-Regulation-draft.pdf>.
- *White Paper on Artificial Intelligence - A European Approach to Excellence and Trust*. Feb. 19, 2020. European Commission. [https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf).
- Jakubowska, Ella, and Diego Narajo. *Ban Biometric Mass Surveillance: A Set of Fundamental Rights Demands for the European Commission and EU Member States*. May 13, 2020. European Digital Rights. <https://edri.org/our-work/blog-ban-biometric-mass-surveillance/>.
- Marischka, Christoph. *Artificial Intelligence in European Defence: Autonomous Armament?* Nov. 2020. The Left in the European Parliament. <https://documentcloud.adobe.com/link/track?uri=urn:aaid:scds:US:1884c966-f618-4110-a5f3-678899e4c8ee>.
- Turek, Helen. *Open Algorithms: Experiences from France, the Netherlands and New Zealand*. Jun. 22, 2020. Open Government Partnership. <https://www.opengovpartnership.org/stories/open-algorithms-experiences-from-france-the-netherlands-and-new-zealand/>.

## North Atlantic Treaty Organization

- Ekelhof, Merel A.C. *Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting*. Naval War College Review. 2018. 71 (3): 62-94. <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=5125&context=nwc-review>.
- Williams, Andrew P. and Paul D. Scharre. ed. *Autonomous Systems: Issues for Defence Policymakers*. Norfolk, VA: NATO Supreme Allied Command Transformation. [https://www.researchgate.net/publication/282338125\\_Autonomous\\_Systems\\_Issues\\_for\\_Defence\\_Policymakers](https://www.researchgate.net/publication/282338125_Autonomous_Systems_Issues_for_Defence_Policymakers).

## Organisation for Economic Co-operation and Development

- *Recommendation of the Council on Artificial Intelligence*. 2021. <https://legalinstruments.oecd.org/api/print?ids=648&lang=en>.

## United Nations

### United Nations Economic Commission for Europe

- May 8, 2020. *Common Functional Performance Requirements for Automated and Autonomous Vehicles*. FRAV-03-05.  
<https://wiki.unece.org/display/trans/FRAV+3rd+Session>.

### United Nations Institute for Disarmament Research

- *Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies*. 2018. <https://unidir.org/files/publications/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf>.
- *The Human Element in Decisions About the Use of Force*. 2020.  
[https://www.unidir.org/sites/default/files/2020-03/UNIDIR\\_Iceberg\\_SinglePages\\_web.pdf](https://www.unidir.org/sites/default/files/2020-03/UNIDIR_Iceberg_SinglePages_web.pdf).
- *Increasing Transparency, Oversight and Accountability of Armed Unmanned Aerial Vehicles*. 2017. <https://unidir.org/publication/increasing-transparency-oversight-and-accountability-armed-unmanned-aerial-vehicles>.
- *Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies*. 2016. <https://unidir.org/files/publications/pdfs/safety-unintentional-risk-and-accidents-en-668.pdf>.
- *The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches*. 2017.  
<https://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf>.
- Kostopoulos, Lydia. *The Role of Data in Algorithmic Decision-Making*. 2019.  
<https://unidir.org/sites/default/files/publication/pdfs/the-role-of-data-in-algorithmic-decision-making-en-815.pdf>.

### United Nations Security Council

- Choudhury, Lipika Majumdar Roy. *Final Report of the Panel of Experts on Libya Established Pursuant to Security Council Resolution 1973 (2011)*. Mar. 8, 2021. S/2021/229. <https://undocs.org/S/2021/229>.

### Other United Nations groups

- *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapon systems*. U.S. Mission. Aug. 28,

2018. <https://geneva.usmission.gov/2020/09/30/group-of-governmental-experts-on-lethal-autonomous-weapons-systems-laws-agenda-item-5c/>.

- *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*. Dec. 13, 2019. GE.19-21466(E). <https://undocs.org/CCW/MSP/2019/9>.

## National governments

### Commonwealth of Australia

- Devitt, Kate, Michael Gan, Jason Scholz, and Robert Bolia. *A Method for Ethical AI in Defence*. 2020. Australian Government Department of Defence, Defence Science and Technology Group. DSTG-TR-3786. <https://www.dst.defence.gov.au/sites/default/files/publications/documents/A%20Method%20for%20Ethical%20AI%20in%20Defence.pdf>.

### Dominion of Canada

- *Pan-Canadian Artificial Intelligence Strategy*. Sep. 8, 2018. Canadian Institute for Advanced Research. <https://cifar.ca/wp-content/uploads/2020/11/AICan-2020-CIFAR-Pan-Canadian-AI-Strategy-Impact-Report.pdf>.

### People's Republic of China

- *China Questions Safety Issues in Military Inteligentization*. Dec. 2020. Foreign Military Studies Office: OE Watch. 10 (12): 32. <https://community.apan.org/wg/tradoc-g2/fmso/p/oe-watch-issues>.
- Bo Wanjuan, Yang Wenzhe, and Xu Chunlei. *Intelligent Warfare, What Stays the Same?* (智能化战争，不变在哪里?). China Military Online (中国军网). Jan. 14, 2020. [http://www.81.cn/jfjbmap/content/2020-01/14/content\\_252163.htm](http://www.81.cn/jfjbmap/content/2020-01/14/content_252163.htm).
- Chai Shan. *The Essence of Winning an Intelligent War* (智能化战争的制胜精髓). China Military Online (中国军网). Jun. 4, 2019. [http://www.81.cn/jfjbmap/content/2019-06/04/content\\_235225.htm](http://www.81.cn/jfjbmap/content/2019-06/04/content_235225.htm).
- Chen Xiaonan and Cong Hanwen. *Talking About the Wisdom of Intelligent Warfare*, (话说智能化战争之智). China Military Online (中国军网). Dec. 27, 2019. [http://www.81.cn/jfjbmap/content/2019-12/27/content\\_250879.htm](http://www.81.cn/jfjbmap/content/2019-12/27/content_250879.htm).

- Creemers, Rogier, Graham Webster, Paul Tsai, Paul Triolo, and Elsa Kania. *State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan* (新一代人工智能发展规划). China's State Council. (translated by the New America Cybersecurity Initiative). Jul. 20, 2017.  
<https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>.
- Dong Jianmin. *Are You Ready for Intelligent Warfare?* (智能化战争，你准备好了吗?). Seeking Truth (求是). Jun. 12, 2019. [http://www.qstheory.cn/defense/2019-06/12/c\\_1124611640.htm](http://www.qstheory.cn/defense/2019-06/12/c_1124611640.htm).
- Li Dapeng. *How to Take Advantage of Intelligent Warfare* (如何夺取智能化战争优势). China Youth Daily (中国青年报). May 23, 2019. [http://www.xinhuanet.com/mil/2019-05/23/c\\_1210141455.htm](http://www.xinhuanet.com/mil/2019-05/23/c_1210141455.htm).
- Li Minghai. *Where is the Winning Mechanism of Intelligent Warfare?* (智能化战争的制胜机理变在哪里?). Xinhua. Jan. 15, 2019. [http://www.xinhuanet.com/mil/2019-01/15/c\\_1210038327.htm](http://www.xinhuanet.com/mil/2019-01/15/c_1210038327.htm).
- Lu Zhisheng. *Draw a Picture of the Future Intelligent Warfare* (为未来智能化战争画个像). People's Daily. Oct. 18, 2018.  
<http://military.people.com.cn/n1/2018/1018/c1011-30348113.html>.
- Ma Rongsheng. *The Study of Intelligent Warfare is Inseparable from Dialectical Thinking* (智能化战争研究离不开辩证思维). Seeking Truth (求是). Jul. 4, 2019.  
[http://www.qstheory.cn/defense/2019-07/04/c\\_1124710009.htm](http://www.qstheory.cn/defense/2019-07/04/c_1124710009.htm).
- Shi Xiaogang. *Intelligent Warfare Forms and Countermeasures* (智能化战争形态及应对策略). China Social Sciences Today (中国社会科学报). Jul. 5, 2018.  
[http://news.cssn.cn/zx/bwyc/201807/t20180705\\_4496198.shtml](http://news.cssn.cn/zx/bwyc/201807/t20180705_4496198.shtml).
- Wang Chunfu. *Let Military Intelligence Enter the Track of Scientific Development* (让军事智能化步入科学发展轨道). China Military Online (中国军网). Mar. 26, 2019.  
[http://www.chinamil.com.cn/jwgd/2019-03/26/content\\_9459996.htm](http://www.chinamil.com.cn/jwgd/2019-03/26/content_9459996.htm).
- Wang Ronghui. *Insight into The Future of Intelligent Warfare* (透视未来智能化战争的样子). China Military Online (中国军网). Apr. 30, 2019. [http://www.81.cn/xue-xi/2019-04/30/content\\_9492869.htm](http://www.81.cn/xue-xi/2019-04/30/content_9492869.htm).
- Wang Yang and Zuo Wentao. *Recognize the Winning Elements of Intelligent Warfare* (认清智能化战争的制胜要素来源). China Military Online (中国军网). Jun. 20, 2020.  
[http://www.81.cn/theory/2020-06/20/content\\_9838385.htm](http://www.81.cn/theory/2020-06/20/content_9838385.htm).

- Xu Li. *Intelligent Warfare Will Not Let People Walk Away* (智能化战争不会让人走开). Seeking Truth (求是). Oct. 17, 2019. [http://www.qstheory.cn/defense/2019-10/17/c\\_1125117765.htm](http://www.qstheory.cn/defense/2019-10/17/c_1125117765.htm).
- Yang Wenzhe. *Exploring the Way to Victory in Intelligent Warfare Through Continuity and Change* (在变与不变中探寻智能化战争制胜之道). Seeking Truth (求是). Oct. 22, 2019. [http://www.qstheory.cn/llwx/2019-10/22/c\\_1125135285.htm](http://www.qstheory.cn/llwx/2019-10/22/c_1125135285.htm).
- Zhao Yun and Zhang Huang. *An Ethical Review of Intelligent Warfare* (智能化战争的伦理审视). Jul. 19, 2018. China Academy of Social Sciences. [http://news.cssn.cn/zx/bwyc/201807/t20180719\\_4505575.shtml](http://news.cssn.cn/zx/bwyc/201807/t20180719_4505575.shtml).

## Kingdom of Denmark

- *National Strategy for Artificial Intelligence*. Mar. 2019. The Danish Government, Ministry of Finance and Ministry of Industry, Business and Financial Affairs. [https://en.digst.dk/media/19337/305755\\_gb\\_version\\_final-a.pdf](https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf).

## Federal Republic of Germany

- Di Fabio, Udo, Commission Chair. *Ethics Commission: Automated and Connected Driving*. Jun. 2017. German Federal Ministry of Transportation and Digital Infrastructure. <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.pdf?blob=publicationFile>.

## Republic of India

- Mohanty, Bedavyasa. *Command and Ctrl: India's Place in the Lethal Autonomous Weapons Regime*. May 2016. Observer Research Foundation. Issue 143. [https://www.orfonline.org/wp-content/uploads/2016/05/ORF\\_Issue\\_Brief\\_143\\_Mohanty.pdf](https://www.orfonline.org/wp-content/uploads/2016/05/ORF_Issue_Brief_143_Mohanty.pdf).

## State of Israel

- Antebi, Liran. *Artificial Intelligence and National Security in Israel*. Feb. 2021. The Institute for National Security Studies. Memorandum No. 207. <https://www.inss.org.il/publication/artificial-intelligence-and-national-security-in-israel/>.

## Russian Federation

- *Abstracts of the Speech of Lieutenant General A.V. Gulyaev, Chief of the Main Armament Directorate of the Ministry of Defense of the Russian Federation, on the Topic of the International Military-Technical Forum Army-2020.* Apr. 10, 2020. Russian Federation Ministry of Defense.  
<http://mil.ru/army2020/statements/more.htm?id=12297890@egNews>.
- *Advanced Research Foundation Believes Robots Will Lead the Future Wars.* (Фонд перспективных исследований считает, что войны будущего поведут роботы). Ria.ru. Jul. 6, 2016 <https://ria.ru/20160706/1459555281.html>.
- *Artificial Intelligence (AI) in Russia.* 2019. Kingdom of the Netherlands.  
<https://www.rvo.nl/sites/default/files/2019/07/Artificial-intelligence-in-Russia.pdf>.
- *A Human Must be Removed from the Battlefield: Combat Robots are Pushing Soldiers Out.* (Человека с поля боя надо убирать»: боевые роботы теснят солдат). Moskovskij Komsomolets. Apr. 9, 2021.  
<https://www.mk.ru/politics/2021/04/09/cheloveka-s-polya-boya-nado-ubirat-boevye-roboty-tesnyat-soldat.html>.
- *Interview with Deputy Defense Minister Yuri Borisov,* Aug. 25, 2020. Vesti.ru.  
<https://www.vesti.ru/article/2449057>.
- *Meduza Interviewed Herbert Efremov, a Developer of Russian Hypersonic Weapons. His Name was Kept Secret Until September 2020.* (Спецкор «Медузы» Лилия Яппарова встретила с секретным конструктором ракет Гербертом Ефремовым и поговорила с ним о будущем оружия). Meduza. Oct. 7, 2020.  
<https://meduza.io/feature/2020/10/07/meduza-vzyala-intervyu-u-gerberta-efremova-razrabotchika-rossiyskogo-giperzvukovogo-oruzhiya-ego-imya-derzhali-v-sekrete-do-sentyabrya-2020-goda>.
- *Pantsir with Intellect: The System Can Counter Attacks Without Operator Input.* (Панцирь» с интеллектом: комплекс сможет отражать атаки без оператора). Izvetsia.ru. Jul. 28, 2020. <https://iz.ru/1040704/anton-lavrov-bogdan-stepovoi/pantcir-s-intellektom-komplekssmozhet-otrazhat-ataki-bez-operatora>.
- *Russian National Defense Management Center Uses Artificial Intelligence.* (Национальный центр управления обороной РФ применяет искусственный интеллект). Regnum.ru. Jan. 27, 2020.  
<https://regnum.ru/news/polit/2836730.html>.

- *Russian Scientists are Working on the Creation of a Neural Network that Will Control Robots During Mine Clearance Operations.* TvZvezda.ru. Aug. 26, 2020. <https://tvzvezda.ru>.
- Burenok, V.M., R.A. Durnev, and K.U. Kryukov. *Intelligent Armament: The Future of Artificial Intelligence in Military Affairs.* 2018. Weapons and Economics 1 (43). <http://www.viek.ru/43/4-13.pdf>.
- Davydov, Vitaly. *Soldiers will be Replaced by Terminators.* (Виталий Давыдов: живых бойцов заменят терминаторы). Apr. 21, 2020. Ria Novosti. <https://ria.ru/20200421/1570298909.html>.
- Dulnev, P.A., and S.A. Sychev. *Key Issues in Developing the Robotic System's Combat Formation.* (АКТУАЛЬНЫЕ ВОПРОСЫ ПОСТРОЕНИЯ БОЕВОГО ПОРЯДКА РОБОТОТЕХНИЧЕСКИХ ПОДРАЗДЕЛЕНИЙ). 2019. Vestnik Akademii Voennyh Nauk 3 (68): 48-53. <http://www.avnrf.ru/index.php/zhurnal-qvoennyj-vestnikq/arkhiv-nomerov/1162-vestnik-avn-3-2019>.
- Edmonds, Jeffrey, Samuel Bendett, Anya Fink, Mary Chesnut, Dmitry Gorenburg, Michael Kofman, Kasey Stricklin, and Julian Waller. *Artificial Intelligence and Autonomy in Russia.* May 2021. Center for Naval Analyses. DRM-2021-U-029303-Final. [https://www.cna.org/CNA\\_files/centers/CNA/sppp/rsp/russia-ai/Russia-Artificial-Intelligence-Autonomy-Putin-Military.pdf](https://www.cna.org/CNA_files/centers/CNA/sppp/rsp/russia-ai/Russia-Artificial-Intelligence-Autonomy-Putin-Military.pdf).
- Golovina, Svetlana. *Artificial Intelligence and Global Trends: The Ministry of Defense Discusses the Improvement of the Electronic Warfare System.* (Искусственный интеллект и мировые тенденции: в Минобороны рассказали о совершенствовании системы радиоэлектронной борьбы). Apr. 14, 2021. TVZVezda.ru. <https://tvzvezda.ru/news/2021414040-gISIA.html>.
- Karpov, Alexander, and Alena Medvedeva. *Intelligent Fighter: How New Onboard Systems Will Enhance the Capabilities of the MiG-35.* (Интеллектуальный истребитель: как новые бортовые системы повысят возможности МиГ-35). Apr. 10, 2021. Russian.rt.com. <https://russian.rt.com/russia/article/851472-istrebitel-mig-35-intellekt>.
- Kozyulin, Vadim. *Will the Drone Become More Humane than a Person - On Responsibility for Decisions on the Use of Weapons?* (Станет ли дрон гуманнее человека - Об ответственности за решения о применении оружия). Apr. 1, 2021. Nvo.ng.ru. [https://nvo.ng.ru/concepts/2021-04-01/6\\_1135\\_drone.html?print=Y](https://nvo.ng.ru/concepts/2021-04-01/6_1135_drone.html?print=Y).
- Martyanov, Oleg. *There Won't be an Army of Terminators, There Will be an Army of Markers.* (Олег Мартьянов: в будущем будет не армия терминаторов, а армия умных "Маркеров"). Jun. 29, 2020. Tass.ru. <https://tass.ru/interviews/8831445>.

- Maslenikov, O.V. et al. *Intellectualization is an Important Component of Digitization of the Armed Forces of the Russian Federation*. (Интеллектуализация - важная составляющая цифровизации вооруженных сил российской федерации). 2020. Voennaya Mysl 7. <https://vm.ric.mil.ru/upload/site178/RJvfqCrBxZ.pdf>.
- Pashentsev, Evgeney. *Artificial Intelligence and Security: What is Good and What is Bad?* (Искусственный интеллект и безопасность: что во благо, а что во зло?). 2019. International Affairs. <https://interaffairs.ru/news/show/24219>.
- Ramm, Alaksei, and Bogdan Stepovoi. *Sea-based Reconnaissance: AI Will Direct Ship-based Missiles*. (Разведка с моря: корабельные ракеты направит искусственный интеллект). Jul. 15, 2019. Iz.ru. <https://iz.ru/898018/aleksei-ramm-bogdan-stepovoi/razvedka-s-moria-korabelnye-rakety-napravit-iskusstvennyi-intellekt>.
- Stefanovich, Dmitry. *Artificial Intelligence and Nuclear Weapons*. May 6, 2019. Russian International Affairs Council. <https://russiancouncil.ru/en/analytics-and-comments/analytics/artificial-intelligence-and-nuclear-weapons/>.
- Zakvasin, Alexey, and Ekaterina Komarova. *Constantly Developing Line of Weapons: How Russian Army Robotic Systems are being Improved*. (Постоянно развивающаяся линейка вооружений: как в России совершенствуются армейские робототехнические комплексы). Apr. 9, 2021. Russian.rt.com. <https://russian.rt.com/russia/article/851123-shoigu-roboty-armiya-postavki>.
- Zarudnitsky, V.B. 2021. *The Nature and Content of Military Conflicts Today and in the Foreseeable Future*. (Характер и содержание военных конфликтов в современных условиях и обозримой перспективе). Voennaya Mysl 1. <https://vm.ric.mil.ru/Nomera>.

## Kingdom of Saudi Arabia

- *Realizing Our Best Tomorrow: Strategy Narrative*. Oct. 2020. Saudi Data and AI Authority, Kingdom of Saudi Arabia. [https://ai.sa/Brochure NSDAI Summit%20version EN.pdf](https://ai.sa/Brochure%20NSDAI%20Summit%20version%20EN.pdf).

## United Kingdom of Great Britain and Northern Ireland

- *UK Weapon Reviews*. UK Ministry of Defence. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/507319/20160308-UK\\_weapon\\_reviews.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507319/20160308-UK_weapon_reviews.pdf).
- *Human-Machine Teaming: Joint Concept Note 1/18*. UK Ministry of Defence. May 2018. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/507319/20160308-UK\\_weapon\\_reviews.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507319/20160308-UK_weapon_reviews.pdf).

[hment data/file/709359/20180517-concepts uk human machine teaming jcn 1 18.pdf](#).

## United States of America

### Office of the President

- *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. Jun. 2019. Executive Office of the President of the United States, Select Committee on Artificial Intelligence of the National Science & Technology Council. <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>.
- *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*. Dec. 3, 2020. Executive Order 13960. <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government>.

### United States Congress

- Schmidt, Eric, and Robert O. Work. *Final Report: National Security Commission on Artificial Intelligence*. Jan. 2021. <https://www.nscai.gov/2021-final-report/>.
- *Key Considerations for Responsible Development & Fielding of Artificial Intelligence*. Jul. 22, 2020. National Security Commission on Artificial Intelligence. <https://www.nscai.gov/wp-content/uploads/2021/01/Key-Considerations-for-Responsible-Development-Fielding-of-AI.pdf>.
- *Algorithmic Justice and Online Platform Transparency Act*. MUR21415 5NT. May 2021. <https://www.markey.senate.gov/imo/media/doc/ajopta.pdf>.
- Office of the Director of National Intelligence, Dan Coats, Director of National Intelligence, and Principal Deputy Director of National Intelligence Susan Gordon. *The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines*. Jan. 16, 2019. <https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf>.

### Department of Commerce

- Schwartz, Reva, Leann Down, Adam Jonas, and Elham Tabassi. *A Proposal for Identifying and Managing Bias in Artificial Intelligence*. Draft NIST Special Publication 1270. Jun. 2021. <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf>.
- Black, Paul E., Michael Kass, Michael Koo, and Elizabeth Fong. *Source Code Security Analysis Tool Functional Specification Version 1.1*. NIST Special Publication 500-268. Feb. 2011.

[https://www.nist.gov/system/files/documents/2021/03/23/source\\_code\\_security\\_analysis\\_spec SP500-268 v1.1.pdf](https://www.nist.gov/system/files/documents/2021/03/23/source_code_security_analysis_spec_SP500-268_v1.1.pdf).

## Department of Defense/Joint Staff/Office of the Secretary of Defense

- *Autonomy in Weapon Systems*. Department of Defense Directive 3000.09. May 8, 2017. <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>.
- *Joint Concept for Robotic and Autonomous Systems*. Joint Chiefs of Staff. Oct. 2016. (Approved for public release, but not available online).
- *Joint Fire Support*. Joint Publication 3-09. Apr. 10, 2019. [https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3\\_09.pdf](https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_09.pdf).
- *Joint Targeting*. Joint Publication 3-60. Jan. 31, 2013. [https://ifsc.ndu.edu/Portals/72/Documents/JC2IOS/Additional\\_Reading/1F4\\_jp3-60.pdf](https://ifsc.ndu.edu/Portals/72/Documents/JC2IOS/Additional_Reading/1F4_jp3-60.pdf).
- *Joint Targeting School Student Guide*. Mar. 1, 2017. Joint Targeting School. [https://www.jcs.mil/Portals/36/Documents/Doctrine/training/jts/jts\\_studentguide.pdf?ver=2017-12-29-171316-067](https://www.jcs.mil/Portals/36/Documents/Doctrine/training/jts/jts_studentguide.pdf?ver=2017-12-29-171316-067).
- *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity*. 2018. <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>.
- *Technical Assessment: Autonomy*. Feb. 2015. Office of Technical Intelligence, Office of the Assistant Secretary of Defense for Research & Engineering. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a616999.pdf>.
- Ahner, Dr. Darryl. *Test & Evaluation of Autonomous Systems*. Jul. 20, 2016. [https://www.afit.edu/stat/statcoe\\_files/TE%20Auto%20Syst%20Workshop%20DA\\_SDDTE%20Memo.pdf](https://www.afit.edu/stat/statcoe_files/TE%20Auto%20Syst%20Workshop%20DA_SDDTE%20Memo.pdf).
- Anderson, John, Marc Losito, and Sean Batir. *The Commander's AI Smartcard: Artificial Intelligence is Commanders' Business*. Feb. 8, 2021. War on the Rocks. <https://smallwarsjournal.com/jrnl/art/commanders-ai-smartcard-artificial-intelligence-commanders-business>.
- Hicks, Kathleen, Deputy Secretary of Defense. *Memorandum for Senior Pentagon Leadership, Commanders of the Combatant Commands, Defense Agencies and DOD Field Activities. Subject: Implementing Responsible Artificial Intelligence in the Department of Defense*. May 26, 2021. <https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF>.

- Hoehn, John R. *Joint All-Domain Command and Control: Background and Issues for Congress*. Mar. 18, 2021. Congressional Research Service. R46725.  
<https://crsreports.congress.gov/product/pdf/R/R46725>.
- Nickols, Wayne. *Autonomy and the National Defense Strategy*. Apr. 3, 2019.  
<https://ndiastorage.blob.core.usgovcloudapi.net/ndia/2019/set/Nickols.pdf>.
- Work, Robert O., Deputy Secretary of Defense. *Remarks by Deputy Secretary Work on Third Offset Strategy*. Apr. 28, 2016. Brussels, Belgium.  
<https://www.defense.gov/Newsroom/Speeches/Speech/Article/753482/remarks-by-deputy-secretary-work-on-third-offset-strategy/>.

## Department of the Air Force

- *The Targeting Cycle*. Mar. 15, 2019. Curtis E. LeMay Center for Doctrine Development and Education [https://www.doctrine.af.mil/Portals/61/documents/AFDP\\_3-60/3-60-D04-Target-Tgt-cycle.pdf](https://www.doctrine.af.mil/Portals/61/documents/AFDP_3-60/3-60-D04-Target-Tgt-cycle.pdf).
- Dahm, Werner J.A. *Killer Drones are Science Fiction*. Wall Street Journal. Feb. 15, 2012.  
<https://www.wsj.com/articles/SB10001424052970204883304577221590015475180>.
- Endsley, Mica R. *Autonomous Horizons: System Autonomy in the Air Force - A Path to the Future*. Jun. 2015. U.S. Air Force Office of the Chief Scientist. AF/ST TR 15-01.  
<https://www.af.mil/Portals/1/documents/SECAF/AutonomousHorizons.pdf>.
- Zacharias, Dr. Greg L. *U.S. Air Force Autonomous Horizons: The Way Forward Vol. II*. Mar. 18, 2019. Maxwell AFB, AL: Air University Press  
[https://www.airuniversity.af.edu/Portals/10/AUPress/Books/b\\_0155\\_zacharias\\_autonomous\\_horizons.pdf](https://www.airuniversity.af.edu/Portals/10/AUPress/Books/b_0155_zacharias_autonomous_horizons.pdf).

## Department of the Army

- *Targeting*. May 7, 2015. ATP 3-60.  
[https://armypubs.army.mil/epubs/DR\\_pubs/DR\\_a/pdf/web/atp3\\_60.pdf](https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/atp3_60.pdf).
- Brock, Major John W. II, US Army. *Why the United States Must Adopt Lethal Autonomous Weapon Systems*. Apr. 13, 2017. School of Advanced Military Studies, United States Army Command and General Staff College, Fort Leavenworth, Kansas.  
<https://apps.dtic.mil/dtic/tr/fulltext/u2/1038884.pdf>
- Bunker, Dr. Robert J. *Armed Robotic Systems Emergence: Weapon systems Life Cycles Analysis and New Strategic Realities*. Nov. 14, 2017. US Army War College, Strategic Studies Institute. Monograph 401.  
[https://press.armywarcollege.edu/monographs/401?utm\\_source=press.armywarcoll](https://press.armywarcollege.edu/monographs/401?utm_source=press.armywarcoll)

[ege.edu%2Fmonographs%2F401&utm\\_medium=PDF&utm\\_campaign=PDFCoverPages](https://www.armywarcollege.edu/monographs/F401&utm_medium=PDF&utm_campaign=PDFCoverPages).

- White, Samuel R. Jr., Project Director and Editor. *Closer Than You Think: The Implications of the Third Offset Strategy for the U.S. Army*. Oct. 2017. United States Army War College, Strategic Studies Institute.  
<https://csl.armywarcollege.edu/usacsl/Publications/Closer%20Than%20You%20Think%20-%20The%20Implications%20of%20the%20Third%20Offset%20Strategy%20for%20the%20US%20Army.pdf>.

## Department of the Navy

- *Autonomous and Unmanned Systems in the Department of the Navy*. Naval Research Advisory Committee. Sep. 2017. <https://www.onr.navy.mil/en/About-ONR/History/nrac/reports-and-executive-summaries/reports-chronological>.
- Berger, General David H., 38th Commandant of the Marine Corps. *Force Design 2030: Annual Update*. Apr. 2021.  
<https://www.marines.mil/Portals/1/Docs/2021%20Force%20Design%20Annual%20Update.pdf?ver=D8ZSD8j66Pci2kEsR4BYDw%3d%3d>.
- Berger, General David H., 38th Commandant of the Marine Corps. *Commandant's Planning Guidance*. 2019.  
[https://www.marines.mil/Portals/1/Publications/Commandant's%20Planning%20Guidance\\_2019.pdf?ver=2019-07-17-090732-937](https://www.marines.mil/Portals/1/Publications/Commandant's%20Planning%20Guidance_2019.pdf?ver=2019-07-17-090732-937).
- Berger, General David H., 38th Commandant of the Marine Corps. *Posture of the United States Marine Corps*. House Appropriations Committee - Subcommittee on Defense. Apr. 29, 2021.  
<https://www.hqmc.marines.mil/Portals/142/Users/183/35/4535/HHRG-117-AP02-Wstate-BergerD-20210429.pdf?ver=8Bg7j-jX894-lDwZuMLhQQ%3d%3d>.
- Braithwaite, Kenneth J., Secretary of the Navy. *Advantage at Sea: Prevailing with Integrated All-Domain Naval Power*. Dec. 2020.  
<https://media.defense.gov/2020/Dec/17/2002553481/-1/-1/0/TRISERVICESTRATEGY.PDF/TRISERVICESTRATEGY.PDF>
- Fox, Collin. *Taking Notes from Narcos: Semisubmersible Unmanned Ships for Great Power Competition*. CIMSEC: Center for International Maritime Security. May 1, 2020.  
<https://cimsec.org/taking-notes-from-narcos-semisubmersible-unmanned-ships-for-great-power-competition/>.
- Gilday, Admiral Michal M., Chief of Naval Operations. *Posture of the United States Navy*. House Appropriations Committee - Subcommittee on Defense. Apr. 29, 2021.

<https://docs.house.gov/meetings/AP/AP02/20210429/112497/HHRG-117-AP02-Wstate-GildayM-20210429.pdf>.

- Gilday, Admiral Michal M., Chief of Naval Operations. *CNO NavPlan*. Jan. 2021. <https://media.defense.gov/2021/Jan/11/2002562551/-1/-1/1/CNO%20NAVPLAN%202021%20-%20FINAL.PDF>.
- Harker, Thomas W., Secretary of the Navy (Acting). *Posture of the Department of the Navy*. House Appropriations Committee - Subcommittee on Defense. Apr. 29, 2021. <https://docs.house.gov/meetings/AP/AP02/20210429/112497/HHRG-117-AP02-Wstate-HarkerT-20210429.pdf>.
- Harker, Thomas W., Secretary of the Navy (Acting), Admiral Michal M. Gilday Chief of Naval Operations, and General David H. Berger Commandant of the Marine Corps. *Department of the Navy Unmanned Campaign Framework*. Mar. 16, 2021. [https://www.navy.mil/Portals/1/Strategic/20210315%20Unmanned%20Campaign\\_Final\\_LowRes.pdf?ver=LtCZ-BPIWki6vCBTdgtDMA%3d%3d](https://www.navy.mil/Portals/1/Strategic/20210315%20Unmanned%20Campaign_Final_LowRes.pdf?ver=LtCZ-BPIWki6vCBTdgtDMA%3d%3d).
- Johnson, Joan L., Deputy Assistant Secretary of the Navy, Research, Development, Test and Evaluation, and Chief of Naval Research, RADM Lorin C. Selby. *Department of the Navy Science & Technology Strategy for Intelligent Autonomous Systems*. Jul. 2, 2021. <https://nps.edu/web/slamr/-/intelligent-autonomous-systems-science-and-technology-strategy-issued>.
- Kraska, James. *Command Accountability for AI Weapon Systems in the Law of Armed Conflict*. 2021. *International Law Studies*. 2021. 97: 408-447. <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2958&context=ils>.
- Letendre, Linell A. *Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers*. 2020. *International Law Studies* 96: 274-294. <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2925&context=ils>.
- Sparrow, Robert. *Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender*. 2015. *International Law Studies* 91: 699-728. <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1413&context=ils>.
- Tangredi, CAPT Sam J., U.S. Navy (Ret.). *Sun Tzu Versus AI: Why Artificial Intelligence Can Fail in Great Power Conflict*. May 2021. *Proceedings of the U.S. Naval Institute* 147 (5): 1419. <https://www.usni.org/magazines/proceedings/2021/may/sun-tzu-versus-ai-why-artificial-intelligence-can-fail-great-power>.
- Vestner, Tobias, and Altea Rossi. *Legal Reviews of War Algorithms*. 2021. *International Law Studies* 97: 510-555. <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2963&context=ils>.

## Department of Energy

- *Using Artificial Intelligence to Advance the State of Multiple Industries*. Sept. 2019. [https://www.energy.gov/sites/prod/files/2019/10/f67/2019-09-30\\_Spotlight-Artificial\\_Intelligence\\_0.pdf](https://www.energy.gov/sites/prod/files/2019/10/f67/2019-09-30_Spotlight-Artificial_Intelligence_0.pdf).

## Department of Homeland Security

- *Artificial Intelligence Strategy*. Dec. 2, 2020. [https://www.dhs.gov/sites/default/files/publications/dhs\\_ai\\_strategy.pdf](https://www.dhs.gov/sites/default/files/publications/dhs_ai_strategy.pdf).
- *Leveraging Unmanned Systems for Coast Guard Missions: A Strategic Imperative*. The National Academies of Sciences, Engineering and Medicine. Special Report 335. 2002. doi: DOI 10.17226/25987. <https://www.nap.edu/read/25987/chapter/1#xi>.

## Department of Transportation

- *Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety*. National Highway Traffic Safety Administration. Sep. 2016. 12507-091216-v9. <https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf>.
- *Federal Aviation Administration: Notices to Airmen*. Order 7930.2S. Jan. 10, 2019. [https://www.faa.gov/documentLibrary/media/Order/7930.2S\\_Notices\\_to\\_Airmen\\_\(NOTAM\).pdf](https://www.faa.gov/documentLibrary/media/Order/7930.2S_Notices_to_Airmen_(NOTAM).pdf).
- *Federal Aviation Administration: Unmanned Aircraft Systems (UAS) Lost Link*. Nov. 10, 2016. Notice N JO 7110.724. [https://www.faa.gov/documentLibrary/media/Notice/N\\_JO\\_7110.724\\_5-2-9\\_UAS\\_Lost\\_Link\\_2.pdf](https://www.faa.gov/documentLibrary/media/Notice/N_JO_7110.724_5-2-9_UAS_Lost_Link_2.pdf).
- *Policy Statement Concerning Automated Vehicles 2016: Update to Preliminary Statement of Policy Concerning Automated Vehicles*. National Highway Traffic Safety Administration. <https://one.nhtsa.gov/Research/Crash+Avoidance/Automated+Vehicles>.

## State of Illinois

- *Illinois Biometric Information Privacy Act*, (740 ILCS 14/1). Oct. 3, 2008. <https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57>.

## State of Texas

- Texas Department of Public Safety. *Unmanned Aircraft System (UAS) Standard Operating Procedure*. Sep. 1, 2019. Chapter 4, Annex #11. <https://www.dps.texas.gov/sites/default/files/documents/docs/prch4anx11.pdf>.

## United Arab Emirates

- Omar Sultan Al Olama, Minister of State for Artificial Intelligence. *United Arab Emirates National Strategy for Artificial Intelligence 2031*. 2018. [https://ai.gov.ae/wp-content/uploads/resources/UAE National Strategy for Artificial Intelligence 2031.pdf](https://ai.gov.ae/wp-content/uploads/resources/UAE-National-Strategy-for-Artificial-Intelligence-2031.pdf).

## Think tanks and research centers

### Center for Naval Analyses

- Lewis, Larry. *Insight for the Third Offset: Addressing Challenges of Autonomy and Artificial Intelligence in Military Operations*. Sep. 2017. DRM-2017-U-016281-Final. [https://www.cna.org/CNA\\_files/PDF/DRM-2017-U-016281-Final.pdf](https://www.cna.org/CNA_files/PDF/DRM-2017-U-016281-Final.pdf).

### Center for a New American Security

- Horowitz, Michael C., and Paul Scharre. *AI and International Stability: Risks and Confidence-Building Measures*. Jan. 12, 2021. <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>.
- Work, Robert O. *Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities*. Apr. 2021. <https://www.cnas.org/publications/reports/proposed-dod-principles-for-the-combat-employment-of-weapon-systems-with-autonomous-functionalities>.

### Center for Security and Emerging Technology

- Mittelsteadt, Matthew. *AI Verification: Mechanisms to Ensure AI Arms Control Compliance*. Feb. 2021. [https://cset.georgetown.edu/wp-content/uploads/AI Verification.pdf](https://cset.georgetown.edu/wp-content/uploads/AI-Verification.pdf).

### Center for Strategic and Budgetary Assessments

- Clark, Bryan, Dan Patt, and Harrison Schramm. *Mosaic Warfare: Exploiting Artificial Intelligence and Autonomous Systems to Implement Decision-Centric Operations*. Feb. 11, 2020. <https://csbaonline.org/research/publications/mosaic-warfare-exploiting-artificial-intelligence-and-autonomous-systems-to-implement-decision-centric-operations/publication/1>.

## Center for Strategic & International Studies

- Lewis, James A., and William Crumpler. *Questions About Facial Recognition*. Feb. 2021.  
<https://www.csis.org/analysis/questions-about-facial-recognition>.

## MITRE

- *Biometric Face Recognition: References for Policymakers. An Informational Document Created by the FedID Community*. Dec. 2020.  
<https://www.mitre.org/publications/technical-papers/biometric-face-recognition-references-for-policymakers>.

## Appendix B: IAS risk elements

---

Before applying this list, refer to the previous sections, “The 13 DADs” and “How to Use the DADs and Their Elements”, for context and scope.

### DAD#1: Standard semantics and concepts

- Have all parties identified all the important terms being used in the development and use of the IAS that require definition?
- Are all parties (when they come from different organizations with different doctrine with respect to IAS use) using consistent and non-conflicting doctrinal terminology?
- Does IAS use require the use of rapidly emerging terminology that must be defined and agreed upon before use?
- Are all parties using the same definitions for “artificial intelligence,” “intelligent autonomous systems,” “autonomy,” “automatic,” and “autonomous functionality?”
- Are all parties using the same definitions for “peacetime status” and wartime status?”
- Are all parties using the same definitions for IAS “degree of autonomy?”
- Are all parties using the same definition for “realistic operational environment” for IAS developmental and operational test and evaluation purposes?
- Are all parties using the same definitions for “training data, input data and feedback data?”
- Are all parties using the same definitions for the several and distinct operational phases?<sup>27</sup>
- Are all parties using the same risk management framework?
- Are all parties using the same technical standards throughout the entire lifecycle of the IAS?
- Are all parties using the same metrics for quantitative analysis (e.g., analyzing confidence levels, comparing similarities, measuring differences)?

---

<sup>27</sup> This document uses a generic (and non-doctrinal) framework of: Search, Detection, Tracking, Identification, Cueing, Prioritization, Selection, Engagement Timing, Terminal Guidance, Engagement, Battle Damage Assessment,” but other constructs are just as appropriate, for example: F3EAD (Find, Fix Finish, Exploit, Analyze and Disseminate) can be used. The choice of framework is not as important here as the need for all parties to agree to and use a common framework.

## **DAD#2: Continuity of legal accountability**

### **Pre-operational phase considerations**

- Has the concept of operation been analyzed to verify that no temporal or spatial accountability gaps exist regarding the use of IAS?
- Has the concept of operation been analyzed to verify that no transfer of command and control over the IAS can occur without specific authorization by the person(s) designated to be accountable for the use of the IAS?
- Is/are the person(s) designated to be accountable for the use of an IAS the only person(s) with the physical ability to transfer decision-making capability to the IAS?
- Is/are the person(s) designated to be accountable for the use of an IAS the only person(s) able to authorize in situ changes to the IAS's configuration created by exposure to incoming data streams?
- Have all persons who may be designated to be accountable for the use of an IAS received training on the Law of Armed Conflict and ethics policies?
- Have all persons who may be designated to be accountable for the use of an IAS been briefed on the current and prevailing rules of engagement?
- Have all persons who may be designated to be accountable for the use of an IAS understand that transfer of decision-making capabilities to an IAS does not transfer accountability for the results of any decisions made by that IAS?
- Can IAS lethal capabilities (or other capabilities authorized only for use during wartime) be disabled during peacetime and only be activated after a verifiable transmission is received from an accountable (military or civilian) authority?
- Can the IAS engagement parameters be pre-set to either allow or prohibit it from developing its own target selection, discrimination, or engagement criteria?

### **Considerations for all eleven operational phases**

- |                   |                              |
|-------------------|------------------------------|
| 1) Search         | 7) Selection                 |
| 2) Detection      | 8) Engagement timing         |
| 3) Tracking       | 9) Terminal guidance         |
| 4) Identification | 10) Engagement               |
| 5) Cueing         | 11) Battle damage assessment |
| 6) Prioritization |                              |

- Does the IAS allow for human judgment to be exercised over the IAS during this phase while complying with the relevant rules of engagement?
- If one is required, has the commander designated someone to be accountable for exercising human judgement over any IAS in use during this phase?
- Is/are the accountable person(s) able to disable, redirect or recall the IAS if they obtain evidence that it may be operating in a manner contrary to law, policy, rules of engagement, or outside of expected technical parameters?
- Has every platform and asset that has autonomous functionality been identified?
- Has every subcomponent of the IAS that provides or contributes to autonomous functionality been documented and assessed?
- Are there multiple connected IAS involved in the mission?
- Have all systems assessed as having no autonomous functionality been certified to that affect?
- Is/are the person(s) accountable for executing the tasks in this phase aware of all IAS available for use?
- Is/are the person(s) accountable for executing the tasks in this phase trained and certified in the use of all IAS available for use?
- Will the IAS respond to instructions only from the person(s) designated to be accountable for the use of the IAS?
- Will the IAS respond only to instructions that follow the commander's intent or equivalent authoritative statement of the overarching mission objectives?

## Transfer of command and control

- Can the IAS transfer command and control between entities authorized to exercise this command and control without creating any temporal gaps in accountability?
- Is there a process by which the person(s) designated to be accountable for the use of an IAS in any particular phase are made aware of any decisions made by an IAS in a previous phase that they have "inherited" and are now accountable for?
- Does the IAS clearly communicate the transfer of command and control to and/or from the human operator?
- Is accountability for the use of the IAS, to include knowledge of what actions may transfer accountability to others or to themselves, understood by all participants in the operation?

- Can the human operator take command and control of just the use of force functions and allow the IAS to continue to execute other tasks (i.e., navigation, sensor data ingest) autonomously?
- Are there subsystems in the IAS that can continue to operate with full autonomy, while other subsystems need human judgment?
- Will the transfer of command and control between (culturally different) coalition and allied forces be subject to any cultural biases in training data sets that might affect IAS functions?
- Do the commander and human operators understand the implications of sharing accountability when transferring command and control in coalition and allied operations?
- Is the human operator sufficiently trained to appropriately judge when to take command and control even when the IAS does not recommend it?

## General considerations

- Are protections in place to prevent the problem of the “moral buffer,” where the user mentally transfers accountability for negative consequences to the machine?
- Has/have an accountable person(s) been identified for the inadvertent use of force (to include friendly-fire incidents) caused by the IAS?
- Are enemy actions that result in loss of IAS command and control and subsequent negative outcomes sensed, communicated, and recorded?
- Are communications between the IAS and the person(s) designated to be accountable for its use reliable enough (e.g., consistent, frequent) to support this accountability?
- Is command and control over the IAS limited to a human (mindlessly) pressing a “fire” button in response to indications from an IAS?
- Is there communication between the human operators who are each accountable for understanding their individual functions of the IAS that they operate?
- Are all human operators accountable for the outcome of IAS use, even if they were only involved in a subcomponent of the overall mission?

## DAD#3: Degree of autonomy

- Can the degree of IAS autonomy be adjusted?
- Can the degree of IAS autonomy be made known (through markings or public communications such as a notice to mariners) to supported friendly forces, to

noncombatants, and/or to enemy combatants at the discretion of the operator or commander?

- Can the IAS make it known to anyone who might be in a position to disable, board, or seize it that it has—and will exercise—its universal right to self-defense?
- Can the commander or operator throttle the IAS to dynamically increase or decrease the degree of autonomy to adjust to the dynamics of the operational situation?
- Is there an established and accepted threshold degree of autonomy, that when exceeded, the IAS becomes a LAWS, where a human no longer selects the target?
- Can the degree of IAS autonomy be conditionally changed, either through predetermined rule sets, or resulting from emergent information not covered by the predetermined rules, such as the elevation of the defense condition (DEFCON)?
- Can the degree of IAS autonomy for a multi-mission capable platform be chosen and/or adjusted independently for each mission?
- Can the degree of autonomy for each autonomous function of an IAS be adjusted individually?
- Can the degree of IAS autonomy be selected by the IAS such that it is the lowest degree of autonomy required to accomplish the mission?
- Can the degree of IAS autonomy be chosen to allow the IAS to calculate probable enemy losses for a candidate action, compare these losses to how the candidate action contributes to the success of the mission, and make a “return on investment” assessment before taking the action?
- Can the degree of IAS autonomy be adjusted downward for platforms with larger magazines in consideration of the increased level of risk that the larger magazine presumably entails?
- Can the degree of IAS autonomy be adjusted downward to reduce risk when operational conditions are such that there is a higher perceived risk to the IAS from either enemy action, environmental factors or malfunction?
- Is the human operator fully equipped and prepared to take over if the IAS malfunctions or breaks down mid-operation?
- Does the human operator fully understand the differences between an automatic system and an autonomous system?
- Can the degree of autonomy be adjusted based on various levels of abstraction (e.g., task, function, or mission)?

## DAD#4: Necessity of autonomy

### General considerations

- Can the contemplated operation be executed with a non-autonomous alternative?
- Do rules of engagement or policy require the commander to consider all available non-autonomous alternatives before employing an IAS?
- Will failure to transfer decision-making capabilities to the IAS result in a military disadvantage?
- Is the contemplated task monotonous or fatiguing (i.e., “dull”) to the point that human performance will degrade but IAS performance will not?
- Does the contemplated task put humans in harm’s way by exposing them to (dirty) hazardous materials?
- Does the contemplated task put humans in harm’s way by exposing them to (dangerous) enemy action or violent weather conditions?
- Does the use of the IAS deprive the commander of human ingenuity, creativity, flexibility, or (operational) art capabilities that makes the force less capable?
- Can the IAS conduct an operation more safely than a human can?
- Does the commander have any systems that provide the same capability as the IAS, but do not depend on AI or autonomy?
- Does the existing force structure and manning necessitate the use of IAS?
- Can the IAS help determine whether to use autonomous, semi-autonomous, or non-autonomous functional modes based on current circumstances and conditions?
- Can the use of IAS improve military capabilities?
- Can the use of IAS reduce loss of human life?

### IAS capabilities are clearly superior

- Is the perceptual space of the IAS clearly superior to that of the nonautonomous alternative (i.e., infrared or radio frequency emission detection)?
- Can the IAS consistently catch important clues that human operators would miss or misconstrue?
- Does the IAS consistently have greater operational range or longer time on station than the nonautonomous alternative?
- Can the IAS physically outmaneuver any enemy platforms it might face, for which the nonautonomous alternative cannot?

- Can the IAS consistently service targets that are inaccessible to the nonautonomous alternative?
- Can the IAS consistently conduct multiple tasks simultaneously, freeing up humans for other tasks that the human is better suited to than the IAS?

## Distinction

- Can the IAS create a model of a given object and differentiate the object (military versus civilian) as well or better than the nonautonomous alternative can?
- Can the IAS create a model of a person and differentiate enemy combatants, friendly combatants, and noncombatants as well as, or better than, than the nonautonomous alternative can?
- Is a level of confidence required to be established for the IAS to distinguish a target in different operational situations?
- Does the IAS know when a given object is outside of its training set (i.e., out-of-distribution detection)?

## Proportionality

- Would the use of IAS allow the commander to use more precise munitions to minimize collateral damage?
- Can the IAS determine the appropriate level of proportionality in attack during the conduct of a mission (e.g., when to continue, when to cease action)?
- Can the IAS provide better identification and more accurately strike a specific location on a target with smaller munition and reduced blast radius?
- Can the IAS discern whether or not the target is sufficiently valuable to risk collateral damage?

## Decision speed

- Is the speed of operation of the IAS in synch with the speed of the situational assessment?
- Does the superior decision speed of the IAS over the human justify the risk of an incorrect decision by the IAS that a human would not likely make incorrectly?
- Will the IAS be used for situations where human reactions are too slow for an effective response?
- Will IAS decision-making speed decrease if it must allow human operators to oversee and approve its actions?

- Does the mission require action in windows of opportunity too short for effective human intervention/action?
- Will the failure to transfer decision-making capabilities to the IAS result in the enemy having a faster OODA loop?

## Preventing cognitive overload

- Are battlefield information flows so fast as to justify a reliance on an IAS instead of a human decision maker?
- Are battlefield information flows so large as to justify a reliance on an IAS instead of a human decision maker?
- Are battlefield information flows so varied as to justify a reliance on an IAS instead of a human decision maker?
- Will the use of the IAS cause the pace of events on the battlefield to accelerate beyond the point that the commander can comprehend the events and take decisive action?
- Will the enemy use of the IAS cause the pace of events on the battlefield to accelerate beyond the point that the commander can comprehend and control these events?

## Imposing dilemmas on the enemy

- Can the decisions of the IAS overwhelm the adversary's decision-making process?
- Can the decisions of the IAS complicate the adversary's decision-making process by being less predictable than the decisions of a human operator?
- Can the IAS provide both speed and scale of action to impose multiple dilemmas on the enemy?
- Will the IAS enable superposition of multiple, fluidly composed and independent kill chains (or webs) that eliminate, or at least severely curtail, response options available to the enemy?
- Can the IAS create an operational tempo that does not permit the adversary to regroup or concentrate?

## DAD#5: Command and control

### Preventing loss of command and control

- Can the IAS determine if its (possibly changing) operational environment requires “human-in-the-loop,” “human-on-the-loop”, or “human-out-of-the-loop” control and make that requirement known to a human operator?
- Is the IAS prohibited from learning and executing new behaviors based on sensor inputs or data feeds received once decision-making capabilities have been transferred to it?
- If the IAS is allowed to modify its behavior, does it consult the human operator beforehand?
- Is the IAS prohibited from initiating operation in the absence of a control link to a human operator?
- Is the IAS prohibited from moving to a location where the control link to a human operator can be degraded or lost?
- Is the human operator still able to provide command and control over the actions of the IAS when it is employed as part of, and can act based on the conditions within, a swarm of other IAS?
- Is the human operator able to choose between a “human in, on, or out of the loop?”
- Does the IAS enable the commander to determine and/or manage the symmetries/asymmetries in the level of complexity between own forces and enemy forces?
- If a situation becomes too complex or unfolds too rapidly for the human operator to comprehend, can the human operator terminate an engagement?
- Are the number of human operators available sufficient to the number of IAS requiring command and control?
- Can the IAS detect enemy attempts to wrest command and control away and notify the human operator?
- Can the IAS detect malfunctions or out of tolerance performance conditions that could result in a loss of command and control and notify the human operator?
- Can the IAS constantly monitor the availability of a control link, even when not under the direct and immediate direction of the human operator?
- Can an IAS operating in a passive mode be prohibited from changing to an active mode of operation absent a direct instruction to do so by a human operator?

- Can the IAS engage an emergent target, not on any preplanned list, without validation from a human operator?
- Can the human operator choose to reject automated decision-making capabilities at their own discretion and at any time?
- Can the human operator select a minimum confidence level that must be attained before the IAS can act without approval?
- Can the IAS be made to have to check in with a human operator and be in receipt of an acknowledgement before proceeding with further actions?
- Can the IAS identify conditions that might cause it to lose access to the control link with the human operator?

## Relinquishing command and control to the IAS (but not accountability)

- Can the commander selectively limit the actions that an IAS can take before relinquishing command and control over it?
- Does the commander have a checklist of conditions that must be met, risk mitigation steps that have been taken, and an understanding of the risks being incurred, before relinquishing command and control over an IAS which possesses no lethal force capabilities, but whose autonomous operation could result in a lethal outcome via some secondary affect?
- Has the commander acknowledged to his or her supporting Staff Judge Advocate, before relinquishing command and control of the IAS, that by doing so he or she is not relinquishing accountability for the outcomes of its actions?
- Does the commander have a checklist of conditions that must be met, risk mitigation steps that have been taken, and an understanding of the risks being incurred, before relinquishing command and control over an IAS capable of selecting targets and applying lethal force to them?

## Transferring command and control

- Can the IAS verify that a request for transfer of command and control is from a human operator who is authorized to do so?
- Can the IAS determine when command and control has been transferred?
- Is the IAS notified when the commander delegates control, to include target engagement authority, to a subordinate commander?

- Can the IAS transfer control back to a commander, or between commanders safely and in a manner that does not result in any lapses?
- Can transfer of IAS control result in a loss of command and control, and can the transfer be delayed until a time when that possibility is at an acceptable minimum?
- Can the IAS default to a minimal risk operating condition in the event that transfer of command and control was not successful?
- Can the IAS detect the absence or unavailability of a human operator before transferring command and control and default to a minimal risk operating condition until the human operator becomes available?
- Can transfer of command and control of the IAS occur between coalition and allied forces?

## Recognizing (uncontrollable) emergent behavior

- Can the IAS recognize and cease the conduct of emergent behaviors?
- Is the IAS prohibited from learning during operations where there is no human in the loop?
- Can the IAS still learn but refrain from modifying behavior until evaluated by a human operator?
- Are commanders made aware of IAS under their command that have the ability to learn during operations and execute emergent behaviors based on incoming sensor data?
- Do commanders understand the potential occurrence of emergent behaviors with the IAS, based on its type and implementation of autonomous decisions?
- Are thresholds identified and set to alert the commander to the existence and nature of emergent behaviors?
- Can the commander limit the ability of the IAS to learn during operations and exhibit emergent behavior?
- Can the IAS notify the commander when it alters its operating code?
- Can the IAS retain a record of every action it has ever taken and notify the commander when it is about to take an action that deviates from previous actions by a predefined threshold or standard?
- Is the IAS susceptible to misread the environment and mismatch the appropriate action (e.g., come across a different environment but take the same action it has done before)?

## Failing safe when command and control is lost

- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), continue autonomously to execute the most recent mission or task instruction if that is the safest option?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), hold speed and course or loiter to await re-establishment of command and control?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), abort mission and return to base?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), initiate emergency (and possibly non-secure) communications to receive further instructions?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), destroy itself?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), be prohibited from autonomous selection and engagement of individual targets that have not been previously selected by an authorized human operator?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), be prohibited from autonomous selection and engagement of specific target groups that have not been previously selected by an authorized human operator?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), be prohibited from autonomously making any hostile and/or targeting decision whatsoever?
- Can the IAS, in the event that command and control is lost for any reason (and based on pre-determined criteria), follow the original human intent regarding target selection?
- Does the IAS have multiple redundancies in critical systems that enable graceful degradation of performance instead of a catastrophic loss of command and control?
- Does the IAS have internal sensors capable of detecting current or impending system failures or battle damage such that it can autonomously switch to an appropriate redundant backup system?

## Re-establishing command and control

- Can the human operator verify that he/she has re-established command and control to prevent an IAS from continuing to execute failsafe protocols?
- Does the IAS immediately alert emergency service providers and recovery crews in the event of a collision or malfunction who can verify safe operating condition before returning the IAS to service?

## DAD#6: Presence of persons and objects protected from the use of force

### General considerations pertaining to both persons and objects

- Can the IAS calculate a confidence factor for positive identification assessments of persons and objects protected from the use of force and prohibit the use of force if that confidence factor is below a minimum threshold set by the commander?
- Can the IAS task own platform or other platform sensors to gather additional data to improve positive identification assessments when the confidence factors are below the minimum threshold set by the commander?
- Can the IAS seek additional “reach back” intelligence information to improve positive identification assessments when the confidence factor is below the minimum threshold set by the commander?
- Can the IAS decision not to act based on a confidence factor below the minimum threshold set by the commander be overridden by that commander?
- Can the IAS decision not to act based on a confidence factor below the minimum threshold set by the commander be overridden by other commanders who have had that authority delegated to them by the commander?
- Can the IAS deployment be limited to a well-defined area where only enemy combatants or military objects are present?
- Can the IAS detect the presence of (uncategorized) humans near or within a targeted object?
- Can an IAS, even if unarmed, result in collateral damage in the event of a malfunction (such as a crash)?

### Distinction

- Can the IAS recognize symbols that designate persons and objects protected from the use of force, such as a Red Cross or Red Crescent?

## Proportionality

- Can the IAS estimate potential collateral damage and conduct collateral damage assessments before going forth with force decisions (from either the IAS or a human operator)?
- Can the IAS conduct collateral damage assessments and compare them to the commander's proportionality assessment for consistency?
- Can the IAS provide commanders and operators with the information needed to assess proportionality when the IAS cannot do so without human assistance?
- Can the IAS obtain, analyze and use information on the effects radii and patterns of munitions to determine if persons or objects protected from the use of force (to include friendly combatants) are within those radii and patterns?
- Can the IAS access the most up-to-date assessments of collateral effects radius assessments prepared by the commander's staff?
- Can the IAS obtain, analyze and use information on the potential for secondary explosions before delivering a munition?
- Can the IAS distinguish between actions that cause repairable/reversible harm and actions that cause irreparable/irreversible harm?
- Can the IAS assist the proportionality assessment summary by tracking assessments and the confidence levels thereof?
- Can the IAS incorporate outcomes of previous operations that resulted in unacceptable collateral damage to prevent a recurrence?
- Can the IAS ingest and utilize collateral effects analysis results prepared by the commander's staff?
- Can the IAS identify any temporal, spatial or technical limits on the accuracy of its collateral damage estimates and convey those to the operator/commander?
- Can the IAS determine when its actions might pose a collateral damage hazard to persons or objects protected from the use of force, cease the potentially hazardous action, and communicate its inability to take this action to the commander?
- Can the IAS seek approval to use a larger or less precise munition if its supply of smaller and/or more precise munitions is exhausted?

## Military necessity

- Can the IAS analyze the military necessity and appropriateness of an attack?
- Can the IAS assess the contribution that destruction of a target would make toward achieving a legitimate military objective?

- Can the IAS deconflict requirements to not intentionally harm persons or objects (in a manner that would violate laws, policies, or the rules of engagement) with its requirement to support achieving the military objective?
- Can the IAS cancel or suspend an attack if it becomes apparent that the objective is not a military one?
- Can the IAS abort an attack if the target no longer has military value because some other target that it depends on for its military value has been destroyed?

### Unnecessary suffering

- Can the IAS reduce or disable any prompts that would activate the use of force when it is operating in, or transiting through, an uncontested area?
- Can the IAS determine when the thresholds to authorize the use force by a human commander or operator have lowered over time, due to a growing indifference to taking human life or creating damage?
- Will repeated employment of the IAS as a LAWS numb the human operator or commander to the taking of human life?

### General considerations pertaining only to persons protected from the use of force

- Can the IAS determine who receives medical attention?
- Can the IAS make humans present in the operational environment aware of any hazards caused by interacting with it?

### Distinction

- Can the IAS distinguish between enemy combatants, friendly combatants, neutral combatants, and noncombatants?
- Can the IAS be prohibited from making determinations regarding who will be harmed—in a situation where harm to someone is unavoidable—based on any intrinsic characteristics of a person such as age, gender, or other physical or mental conditions?
- Can the IAS make determinations regarding who will be harmed—in a situation where harm to someone is unavoidable—based on an individual's status as an enemy combatant, a friendly combatant, or a noncombatant person?
- Can the IAS account for unknown transient noncombatant personnel in or near a target area?

- Can the IAS distinguish between adversarial human combatants and adversarial equipment and adjust the effect accordingly?
- Can the IAS default to categorizing a person as an “unknown,” and not engage in instances where positive identification is below the required confidence level?
- Can the IAS obtain, analyze and use information on civilian patterns of life to distinguish noncombatants from combatants?
- Can the IAS determine when enemy combatants have interspersed into a group of noncombatants that cannot be engaged with force?
- Can the IAS use signatures to distinguish combatants from noncombatants that are unique to either the combatant only or the noncombatant only?
- Can the IAS distinguish between benign and hostile intentions?
- Can the IAS distinguish between benign and hostile actions?

### **Military necessity**

- Can the IAS detect a “mission kill,” where a person has been incapacitated and rendered hors de combat but not killed, and cease further use of force?
- Can the IAS detect a “mission kill,” where a crewed vehicle has been damaged but not destroyed, rendering the crew hors de combat, and cease further use of force?

### **Unnecessary suffering**

- Can the IAS prohibit the use of any action, method or munition available to it that could conceivably cause collateral injuries or death or cause unnecessary suffering?
- Can the IAS prohibit the use of any action, method or munition available to it that could conceivably result in unacceptable amounts of noncombatant casualties?
- Can the IAS prohibit engagements that would cause widespread, long-term, and/or severe impact to the health of persons protected from the use of force?
- Can the IAS conduct an analysis to minimize the severity of injury, when causing some injury is unavoidable?

### **Perfidy**

- Can the IAS detect when a surrendering enemy combatant has violated the terms of surrender by attempting to escape?
- Can the IAS detect when a surrendering enemy combatant has violated the terms of surrender by committing a hostile act?
- Can the IAS detect and communicate perfidious acts to human operators?

## Specific classes of persons protected from the use of force

### Noncombatants

- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) military medical personnel?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) military religious personnel?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) human shields (persons placed around or within valid military target to hinder attack)?

### Surrendering enemy troops

- Can the IAS detect and perceive any signal of surrender that all combatant sides have previously agreed to use?
- Can the IAS detect and perceive indications of surrender from enemy combatants that have been trained on how to indicate surrender to the IAS?
- Can the IAS detect and perceive indications of surrender from enemy combatants that *have not* been trained on how to indicate surrender to the IAS?
- Can the IAS detect and perceive the changes in orientation and force posture of combatants that are commonly associated with the act of surrender?
- Can the IAS distinguish between retreating and surrendering enemy forces?

### Incapacitated persons

- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) wounded or sick persons?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) shipwrecked persons?

### Constitutionally protected US citizens

- Can the IAS detect and not intentionally harm US citizens in a manner that would violate Constitutional protections?
- Can the IAS detect and distinguish the presence of US citizens who are acting as enemy combatants?

### US and friendly force troops

- Can the IAS ingest data from all blue force tracker systems?

- Can the IAS obtain, analyze and use information to assess risk to friendly forces and calculate a confidence factor for each assessment?
- Can the IAS detect, distinguish and designate (crewed) vessels and aircraft as friendly, neutral, or adversarial and calculate a confidence factor for each designation?
- Can the IAS distinguish between enemy combatants, friendly force personnel, and noncombatant individuals and calculate a confidence factor for each designation?

## General considerations pertaining only to objects protected from the use of force

- Can the IAS detect, or access information regarding, the interdependencies that objects protected from the use of force have with other objects (that may or may not be protected from the use of force)?
- Can the IAS access an up-to-date and complete list of restricted targets and areas?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) objects whose destruction might damage relations with local noncombatant populations?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) objects whose destruction might provide propaganda value to enemy forces?
- Can the IAS provide information regarding objects protected from the use of force that will assist a human operator in making use of force decisions?

## Distinction

- Can the IAS determine when enemy combatants have entered into an area or structure that is protected from the use of force?
- Can the IAS use signatures to distinguish military from civilian objects that are unique to either the military object only or the civilian object only?
- Can the IAS account for unknown transient civilian or noncombatant personnel and/or equipment near a targetable military object?

## Military necessity

- Can the IAS detect a “mission kill,” where an object has been damaged but not destroyed, to the point that it no longer has military utility and cease further use of force?

## Unnecessary suffering

- Can the IAS prohibit engagements that would cause widespread, long-term, and/or severe impact to the health of general population (such as the destruction of a power plant that services a hospital or a refugee camp)?

## Perfidy

- Can the IAS detect perfidious use of objects protected from the use of force (such as using an ambulance for military transport)?
- Can the IAS detect perfidious use of facilities protected from the use of force (such as using a house of worship as a sniper position)?

## Specific objects

- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) diplomatic offices, foreign missions, and the sovereign nonmilitary properties of other nations?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) religious, cultural, historical institutions, cemeteries, and structures?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) fixed medical facilities?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) mobile medical facilities?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) ambulances and clearly marked medical ground transport vehicles?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) air ambulances and clearly marked medical aircraft?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) hospital ships?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) public education facilities?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) civilian refugee camps?

- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) prisoner of war camps and government detention facilities/prisons?
- Can the IAS identify and not intentionally initiate (in a manner that would violate laws, policies, or the rules of engagement) engagements that may result in pollution or have potential to release toxic chemicals?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) dams or dikes whose engagement may result in flooding of civilian areas?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) facilities whose engagement may threaten astronauts and/or manned space flight missions?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) civilian meeting places?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) public utilities and facilities?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) facilities and/or structures with unknown functionality/purpose?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) agricultural processing and storage facilities?
- Can the IAS identify and not intentionally harm (in a manner that would violate laws, policies, or the rules of engagement) facilities that provide products or services for both civilian and military use?

## **DAD#7: Pre-operational audit logs**

### **General considerations**

- Is all pre-operational documentation made visible and accessible to the greatest extent possible, limited only by the potential creation of operational vulnerabilities?
- Is a formal procedure in place whereby individuals with concerns over unethical or illegal application of an IAS can (and are required to) document those concerns without fear of retaliation?

- Is positive control of the individual IAS and all of its subcomponents documented as they move from one organization to the next in the research, development, and test and evaluation process?
- Are all commercial vendor marketing materials and claims about acquired IAS available for inspection and oversight?

## Training data provenance

- Is the chain of custody for all IAS training data documented from creation to final use?
- Are all requests for sharing of IAS training data documented?
- Are there access controls and processes in place to ensure that only authorized persons may make modifications to IAS training data?
- Is the identity of all persons who accessed, or who authorized access to the IAS training data documented?
- Are the current, and all previous versions of the IAS training data retained for inspection?
- Are all modifications to the IAS training data documented?
- Is the IAS training data available for inspection immediately prior to use?
- Is the IAS training data retention policy documented?
- Is the nature and origin of all IAS training data documented?
- Are the descriptions and labels that describe the IAS training data documented?
- Does the IAS use any open-source training data with absent or suspect provenance documentation?
- Can a higher-level IAS verify the reliability of inputs from its subsystem components that may or may not be autonomous?
- Is the training data for the overall IAS the same training data that was used to design and verify the component autonomous subsystems?

## Algorithm provenance

- Is the chain of custody for all IAS algorithm's source code documented from creation to final use?
- Are all requests for sharing of IAS algorithms documented?
- Are there access controls and processes in place to ensure that only authorized persons may make modifications to IAS algorithms?

- Is the identity of all persons who accessed, or who authorized access to IAS algorithms documented?
- Are the current, and all previous versions of the IAS algorithms retained for inspection?
- Are all modifications to the IAS algorithms documented?
- Are the IAS algorithms available for inspection immediately prior to use?
- Is the IAS algorithm retention policy documented?
- Is the nature and origin of all IAS algorithms documented?
- Are the descriptions and labels that describe the IAS algorithms documented?
- Do the IAS algorithms use any open-source code with absent or suspect provenance documentation?
- Is the design documentation describing the IAS algorithms technical architecture available for inspection?
- Are code scanning tools used that can identify poor coding practices and common security vulnerabilities?

## Computer hardware provenance

- Is the chain of custody for all IAS computer hardware components documented from creation to final use?
- Is the identity of all persons who accessed, or who authorized access to IAS computer hardware components documented?
- Are there access controls and processes in place to ensure that only authorized persons may make modifications to hardware or software?
- Are all modification and repairs to the IAS computer hardware components documented?
- Are the IAS computer hardware components available for inspection immediately prior to use?
- Is the nature and origin of all IAS computer hardware components documented?
- Does the IAS use any commercial off the shelf computer hardware components with absent or suspect provenance documentation?
- Is the design documentation describing the IAS computer hardware components technical architecture available for inspection?

## Acquisition and development

- Has the IAS been acquired using any waivers allowing circumvention of safety policies or regulations?
- Has the IAS been acquired using any waivers allowing circumvention of ethics policies or regulations?
- Has the IAS been acquired using any waivers allowing circumvention of any developmental or operational testing, verification, and validation?
- Has the IAS been acquired using any accelerated/rapid/other acquisition authorities that dismiss or delay risk mitigating steps otherwise present in the deliberate acquisition system?
- Are all acquisition and development-related waivers documented and available for inspection?
- Is the IAS test and evaluation plan made public to the greatest extent possible (without divulging operational vulnerabilities)?
- Have the IAS developers produced ethical risk mitigation documentation for the IAS similar to the technical risk mitigation documentation already required by acquisition regulations?
- Do IAS approval documents record any conditions of use under which the approval would be invalid or rescinded?
- Are all conditions that would require re-testing of a fielded IAS documented and made available to the end user?
- If the IAS has the potential to be considered a LAWS by DoD Directive 3000.09, has the IAS undergone the required review and approval process outlined in this policy?

## Authorization

- Are any justifications for not conducting a legal review of a new or upgraded IAS documented and publicized?
- Is the independence of any and all required IAS oversight bodies or regulatory authorities documented and publicized?
- Have conflict of interest screenings been conducted for all persons involved in the research, development, test and evaluation processes?
- Have all identified conflicts of interest been either mitigated or waived?
- Have the justifications for conflict of interest mitigations and/or waivers been documented and publicized?

## DAD#8: Operational audit logs

### General considerations

- Is the collection of IAS operational data fields and their retention in an IAS operational audit log mandatory?
- Is the IAS instrumented with the internal and external environmental sensors needed to support the collection of all required operational audit log data fields?
- Is the IAS equipped with sufficient data storage and/or transmission capabilities to capture and retain data at the frequency required for post-operational uses and for the entire duration of the mission?
- Are all data descriptions understandable to non-technical persons who may use the data logs for post-operational reconstruction and analysis?
- Has the data in the operational audit log been verified to be of sufficient quality, quantity, and provenance to support any potential post-operational legal proceeding?
- Are access controls in place to limit access to the IAS operational audit logs to authorized personnel only?
- Are challenges to the use of an IAS based on examination of the IAS operational audit log, and the results of those challenges recorded by the audit log data steward?
- Are all requests for explanations or inspections of the IAS operational audit log, and the results of those inspections recorded by the audit log data steward?
- Is there a procedure in place to request, and approve requests, to examine the operational audit log?
- Is a procedure in place whereby individuals who have examined IAS operational audit logs and have concerns regarding the potential unethical or illegal use, can make those concerns known without fear of retaliation?
- Are transfers of IAS command and control documented?

### Operational audit log data fields

- Does the IAS collect metadata associated with all sensed objects in its operation environment that includes object identification, confidence metrics associated with its state values (i.e., position, velocity, etc.), and historical state information?
- Does the IAS collect time stamp metadata for *all* data fields and during the *entire* mission duration?
- Does the IAS collect data on the operating conditions of all critical internal systems?

- Does the IAS record all commands it receives?
- Does the IAS record the identities of the accountable operator(s) and commander(s)?
- Does the IAS record actions taken, and decisions made by the IAS and by the human operators?
- Does the IAS record the sensed actions of friendly force personnel, adversary force personnel and noncombatant persons?
- Does the IAS record all algorithms in use (to include any version number) that are utilized to produce an action/decision?
- Does the IAS record all the training data dependencies of all algorithms in use (to include any version number) that are utilized to produce an action/decision?
- Does the IAS record its connections to, data exchanges with, and interdependencies with, other external systems to include other IAS?
- Does the IAS record the loss of data signal connections?
- Does the IAS record connections, data exchanges and interdependencies between its own internal subsystems?
- Does the IAS record data from battle damage assessments?
- Does the IAS record data regarding technical malfunctions and degradations of performance?
- Does the IAS record its position relative to all battlespace management areas, including both exclusion and inclusion areas (e.g., Notices to Airman and Mariners)?
- Does the IAS record the results of any risk assessments made on board during a mission and the resultant mitigation actions taken?

## Traceability of audit log data elements

- Are all decisions made by the IAS traceable back to the sensor data, operator commands, algorithms and training data used?
- Can all actions/decisions made by an IAS and the human operator who authorized the action/decision be traced back to the military objective or mission orders they were intended to support?
- Can human operators retrace decision steps and interactions with the environment that resulted in an IAS action/decision of interest?
- Are the inputs from component autonomous subsystems that were used by the IAS to make decisions available for review by a human operator?

## Operational audit log retention

- Does the IAS operational audit log have a metadata field identifying the date after which retention is no longer required?
- Is IAS operational audit log data deleted automatically or manually with human oversight when its retention date is reached?
- Are IAS operational audit logs retained beyond the statute of limitations date for any illegal acts that could possibly have been committed during an operation?
- Do acquisition professionals who may wish to use IAS operational audit logs to improve the performance of this, or some other IAS, have a role in deciding when the audit logs are no longer retained?
- Do military commanders, trainers, and doctrine writers who may wish to use IAS operational audit logs to improve future tactics, techniques and procedures have a role in deciding when the audit logs are no longer retained?
- Are data storage and transmission resources sufficient to retain all of the IAS operational audit log data created during the contemplated life cycle of the IAS?
- Do the IAS operational audit logs require any additional handling caveats above and beyond existing document retention and destruction policies?
- Is retained IAS operational audit log data that is subsequently reused, reassessed first for its quality and continued relevancy and applicability?
- Are all accesses to retained IAS operational audit log data, either granted or denied, recorded by the data steward?

## Operational audit log uses

- Does the IAS operational audit log enable training data generation for subsequent IAS algorithm development?
- Does the IAS operational audit log enable accountability for adverse actions?
- Does the IAS operational audit log enable further IAS refinement and testing?
- Does the IAS operational audit log enable tactics techniques and procedures development?
- Does the IAS operational audit log enable human operator training?
- Does the IAS operational audit log enable post mission hot washes for corrective action?
- Does the IAS operational audit log enable acquisition of IAS not related to the current IAS or its mission(s)?

- Are the IAS operational audit logs made available for internal oversight?
- Are the IAS operational audit logs made available for external third-party oversight?

## **DAD#9: Human-machine teaming**

### **General considerations**

- Can the IAS alert the humans that they are interacting with an IAS and not another human?
- Can the IAS distinguish between blue forces with human operators and other blue force IAS?
- Will continued reliance on the IAS cause human operator skills to diminish such that they can no longer back up the IAS in the event that the IAS becomes unavailable?
- Can the IAS make decisions when incapacitated human operators cannot?
- What types of decisions is the IAS allowed to make or delegate?
- Can the IAS detect when the human operator is under too much stress to make a sound decision and unilaterally make decisions without human oversight that normally would require it?
- Can the IAS provide explanations of its actions and/or predictions that are understandable by the human operator?
- Can the IAS detect human user complacency?
- Is the rate of IAS false positive results known and below a previously defined and acceptable value?
- Is the rate of IAS false negative results known and below a previously defined and acceptable value?
- Is there an option to turn off the IAS if there are too many false alarms?

### **Provision of shared situational awareness**

- Can the IAS communicate with all systems from which it must ingest data to achieve maximum situational awareness during the fog and friction of war?
- Is the IAS equipped with system self-monitoring sensors that can detect and communicate (internal) conditions that may lead to a negative outcome if not corrected?

- Can the IAS predict when operational design domain conditions and constraints might be violated and communicate the future time of the violations, the probability of occurrence and the severity to a human operator?
- Can the IAS and the human teammates effectively communicate their state, intent, and current problems to the other teammate?
- Can the IAS communicate the current and predicted future status of fuel and ammunition levels to the human operator?
- Can the IAS communicate its current and predicted future availability for mission execution to the human operator?
- Can the IAS communicate (previously defined) system malfunctions to the human operator?
- Can the IAS communicate critical messages to the human operator?
- Can the IAS detect both sharp and gradual changes in its operational environment and provide alerts to the human operator and other systems based on criticality?
- Can the IAS detect both sharp and gradual changes in its own performance and provide alerts to the human operator and other systems based on criticality?
- If an IAS's autonomous functionality fails and it must transfer command and control of itself to a human operator, will that operator have been sufficiently engaged and be able to understand the IAS's current situation in time to safely repurpose the IAS or avert danger/loss of the IAS?
- Can the IAS determine when it is inappropriate to interrupt a human operator with a query for assistance?
- Can the IAS recognize the difference between rules-based decisions and value/judgement-based decisions and query the human operator for guidance in the latter case?
- Can the IAS recognize the difference between decisions based on deductive versus inductive vs abductive reasoning and query the human operator for guidance in the latter case?
- Can the IAS upload and implement any doctrine or rules of engagement that identify predetermined situations where human guidance is a prerequisite to taking further action?
- Can the IAS identify ambiguous or incomplete instructions and request clarification from the human operator in sufficient time to be operationally effective or to prevent mistakes/mishaps?
- Can the IAS detect anomalies and seek human guidance when its programming says that it is under attack, but the political situation is completely benign?

## Conforming and nonconforming human behavior

- Can the IAS detect and respond appropriately to conforming and non-conforming human behaviors?
- Can the IAS identify an instruction from a human operator that violates *its preprogrammed definitions* of the Law of Armed Conflict, policy, or rules of engagement, not carry out the instruction, and provide a reason to the human operator for not doing so?
- Can an IAS decision to not carry out an instruction that violates *its preprogrammed definition* of the Law of Armed Conflict, policy, or rules of engagement be overridden by the human operator or some other commander?
- Can the IAS request mission re-planning if planned mission execution violates mission success criteria, or other operational constraints (such as the crossing of a geographic border), and report detected violations to the human operator?
- Can the IAS perform arbitration between competing mission and navigation objectives based on constraints configured by the human operator?
- Can the IAS detect situations that fall beyond its narrow application domain that subsequently require it to contact a human operator for additional guidance?

## Managing interactions

- Can the IAS determine the optimal frequency of how often they must query humans for commands and ensure that they do not miss any commands?
- Can the IAS communicate with the human using natural language, gestures or haptics to ease and to accelerate information exchanges with human operators?
- Can the IAS monitor the attentional focus, cognitive load, and task status of the human operator, and only communicate information that does not overburden the human operator with unnecessary or irrelevant information or tasks?
- Can the IAS detect the emotional and physical state of the human operator to maximize efficiency of communications and activities?
- Can the IAS support human-machine joint training to allow an understanding to develop between the human operator and the IAS regarding team objectives, platform roles, co-dependency relationships, and mutual expectations for competence, dependability, predictability, and timeliness?
- Is the delegation of actions explicitly represented in the dialogue between the human and the machine?

- Can the human modify the conditions placed on the delegated actions and/or decisions of the machine?
- Is there monitoring and documentation that the machine is acting within its prescribed boundaries?

## DAD#10: Test and evaluation adequacy

### General considerations

- Did the IAS test and evaluation procedure receive an already-trained IAS, or is the IAS to be tested also to be trained in the replicated operational environment?
- Did the IAS test and evaluation procedure test all components for the presence of autonomous capabilities?
- Did the IAS test and evaluation procedure record all test results to make them available for inspection and oversight?
- Did the IAS test and evaluation procedure use traditional model checking that may be unsuited for the IAS?
- Did operational testing provide authorized and certified conditions for use of IAS?
- Did the IAS test and evaluation procedure use simulations to push the system to its breaking point so that the boundaries between success and failure were tested and verified in a representative operational environment?
- Did the IAS test and evaluation procedure utilize an environment (e.g., *Live Virtual Constructive environment*) that best captures the behaviors that may emerge under real-world operational conditions?
- Were components or sub-components of the IAS available for testing and fleet concept of employment and experimentation?
- Has the threshold been identified for adequate test and evaluation of autonomous capabilities with respect to their manned counterparts, and with respect to the risk of not having any autonomous systems?
- Is there ongoing test and evaluation after fielding to evaluate the ability of the IAS to adapt to new, unexpected circumstances, or new input data?
- Does the IAS possess measurable and testable autonomous capabilities?
- Have members of the AI and safety engineering research communities with expertise in the potential safety or failure risks from loss of command and control been consulted?

- Was the operational testing and verification witnessed by an independent third party (non-DOD) such as the Underwriter's Laboratory or operational test and evaluation personnel from another federal agency?

## Simulated replication of the operational environment

- Did the IAS test and evaluation procedure use simulations adequately representative of the contemplated operational environment and document all operational parameters that would cause it to be used in an environment that the simulation did not faithfully replicate?
- Did the IAS test and evaluation procedure use simulated or otherwise artificially generated data to replicate the contemplated operational environment?
- Has the relevance of the IAS training data to the contemplated operational environment been validated?
- Has any use of simulated data been validated against real-world data?

## Breadth of testing in the replicated environment

- Did the IAS test and evaluation procedure subject the system to adversarial machine learning attacks to identify potential attack vectors?
- Did the IAS test and evaluation procedure limit the replicated environment to what the IAS will encounter in the contemplated operational environment and document all parameters that would cause it to be used in an environment test and evaluation did not faithfully replicate?
- Did the IAS test and evaluation procedure consider interactions with human operators using a fully representative sample of all potential human operators?
- Did the IAS test and evaluation procedure use computer models to generate thousands of data-based scenarios to anticipate and analyze any emergent behavior?
- Did the IAS test and evaluation procedure consider how an IAS might overwrite all or part of its original software as a result of exposure to new sensor data, such that the procedure must be repeated on the new software baseline?
- Did the IAS test and evaluation procedure consider how an IAS might overwrite all or part of its original software as a result of exposure to new training data, such that the procedure must be repeated on the new software baseline?
- Did the IAS test and evaluation procedure consider whether the data stream providing instructions to the IAS is vulnerable to hacking or "in flight" hijacking?

- Did the IAS test and evaluation procedure consider infiltration into the industrial supply chain?
- Did the IAS test and evaluation procedure utilize virtual twin models to test the IAS in simulated environments to predict how it will interact with IAS and non-autonomous systems from neutral nations, allied nations and from other services?
- Did the IAS development include an AI-developed capability to conduct planning where there will be a need to design test plans in a timely manner to assure thorough testing while being able to predict certain behaviors with these systems (e.g., *Autonomous Systems Test (AST) Planning*)?
- Did the IAS development include an AI-developed capability that will guarantee safety during testing through the use of immersion technologies to adapt tests to best suit IAS cognitive capabilities (e.g., conducting *AST Execution & Control*)?
- Did the IAS development include an AI-developed capability to conduct assessments (e.g., *Autonomous System Performance Assessment*) to capture how IAS interact with existing and newly fielded manned and unmanned systems in realistic operational environments?
- Can functional components be tested in the replicated environment while assessing reliability and identification of vulnerabilities?
- Can human-IAS interaction be measured in the replicated environment while assessing reliability and identification of vulnerabilities?

## Depth of testing in the replicated environment

- Did the IAS test and evaluation procedure push the system to its breaking point so that this/these points are known to the operators?
- Did the IAS test and evaluation procedure utilize red team attacks (kinetic and cyber) to capture the full range of behaviors that might emerge under real-world operational conditions?
- Did the IAS test and evaluation procedure utilize red team attacks specifically designed to drive the system into inappropriate behaviors?

## DAD#11: Autonomy training and education

### Ensuring legal and ethical use

- Have all senior policy and decision makers, on scene commanders, scientists, engineers, acquisition officials, legal counsels, operators and supported troops

received training in the Law of Armed Conflict, IAS ethics principles, and current rules of engagement?

- Can human operators with the power to override IAS use during an operation recognize legally and ethically questionable decisions and actions?
- Do all commanders and users that possess current proficiency certifications to operate the IAS, fully understand how the IAS processes ethical dilemma scenarios?

## Certifications

- Do training and education requirements exist for IAS as part of the certification process for the deploying forces that may employ them?
- Are IAS training, education, and qualification certifications required and recorded for all personnel?
- Does the commander's legal staff monitor the laws, ethics policies and regulations that govern the use of the IAS to ensure current training, education, and qualification certifications are compliant with them?
- Does the commander's legal staff have a sufficient technical understanding of IAS?
- Are all legal reviews double checked by technical and procurement officials to ensure that they are based on a sound technical understanding of the IAS?

## Recognizing IAS limitations

- Do all commanders and users understand the difference between general and narrow artificial intelligence?
- Do all commanders and users understand the concept of "brittleness" with respect to narrow artificial intelligence applications?
- Do all commanders and users know how to evaluate when IAS are, and are not, necessary to achieving the military objective?
- Do all commanders and users know the technical limitations of IAS?
- Do all commanders and users know the limitations of IAS imposed by the operational environment?
- Do all commanders and users know the limitations of IAS imposed by the data environment?
- Do all commanders and users know how to distinguish between components that do and do not possess autonomous capabilities?
- Do all commanders and users understand the degrees of autonomy and their impact on IAS capabilities and limitations?

- Do all commanders and users know how to conduct a risk assessment for the use of IAS?
- Can all commanders and users determine the degree of autonomy needed in different operational environments and understand the risk mitigation tactics, techniques, procedures and technologies available to them?
- Can the commanders discern the consequences of the IAS limitations?

## Understanding the role of data

- Do all commanders and users understand the dependence of IAS on training data and the new attack vectors this dependency opens up to the enemy?
- Do all commanders and users understand the difference between training data, input data and feedback data?

## Understanding the algorithms

- Do all commanders and users understand the difference between inductive, deductive, and abductive reasoning, and the limitations of algorithms in each?
- Do all commanders and users understand the different types of machine learning (supervised, unsupervised, reinforcement) used by the IAS and the operational environments amenable to each?

## Recognizing AI-specific failure modes

- Can all commanders and users recognize a data poisoning attack?
- Can all commanders and users recognize an instance of “reward function gaming?”
- Can all commanders and users recognize an instance of “adversarial spoofing?”
- Can all commanders and users recognize an instance of “catastrophic forgetting?”
- Can all commanders and users recognize an instance of “concept drift?”
- Can all commanders and users recognize an instance of “model inversion?”
- Can all commanders and users recognize a “deep fake?”
- Can all commanders and users recognize an instance of reprogramming by unauthorized users?
- Can all commanders and users recognize a system that is “tightly coupled?”

## **DAD#12: Mission duration and geographic extent**

### **Temporal subdivision**

- Can the commander reduce the mission duration such that the IAS risk mitigation conditions and planning factors will be, or can reasonably be expected to be, uniformly consistent?
- Does the target selection that the human operator provides the IAS have an expiration time within the duration of the mission, such that the human operator will be alerted to select and verify new targets after the expiration time has expired?
- Is the commander able to alter the degree of autonomy of the IAS during long missions based on whether it is in a loitering mode, or an active task execution mode?
- Is the reliability of communications links commensurate with both the duration and the geographic extent of the mission?
- Can the IAS compare risk mitigation and planning factors provided to it at the beginning of a long mission, and compare these to sensed conditions just prior to an action/decision and abort the action if they are different?

### **Spatial subdivision**

- Can the commander reduce the expanse of the operational environment to an area where risk mitigation conditions and planning factors are known to be, or are reasonably expected to be, uniformly consistent?
- Can the commander subdivide a large operational area into smaller ones where risk mitigation conditions and planning factors are known to be, or are reasonably expected to be, uniformly consistent, and assign one IAS to each area, with each IAS specifically programmed to address the conditions unique to each subsection?
- Is the IAS equipped with a capability that tailors its approved degree of autonomy and authorized actions based on geographical location?
- If a target is verified within a geographic location but transits outside it later in time, can the IAS still engage, or does it immediately cease action?
- Is the IAS equipped with a capability that tailors its approved degree of autonomy and authorized actions based on its distance from a human operator?
- Can the IAS recognize when it is assigned an area of regard too large for reliable autonomous operation?

- Can the IAS recognize when it has been assigned to an area of regard where the fundamental risk mitigation conditions are not uniformly consistent?

## Preprogrammed disablement

- Is the IAS equipped with a switch that will disable its autonomous functionalities after a predetermined set time has elapsed?
- Is the IAS equipped with a geo-fence switch that will disable its autonomous functionalities if it strays from an authorized operational area?
- If risk is reduced with increased mission time or expanded geographic range, can the human operator increase autonomous capabilities of the IAS as needed?
- Is the commander notified if they have an IAS with a disablement switch that has been overridden by higher authority?
- Is the IAS equipped with a disablement switch that can selectively disable only certain functions or subsystems?
- Is the IAS equipped with a “soft” disablement switch that disables autonomous functionality but can be reversed if certain predetermined conditions are detected or an override authorization is received?
- If the IAS is disabled, can it safely and securely return to a predetermined location?

## DAD#13: Civil and natural rights

### Rights of potentially affected persons

- Does the IAS potentially affect the rights enumerated in the first ten amendments to the US Constitution (the Bill of Rights)?
- Does the IAS potentially affect the rights enumerated in the Civil Rights Act of 1964, the Equal Protection Clause of the U.S. Constitution and all subsequent enactments?
- Does the IAS potentially affect the rights enumerated in the European Union General Data Protection Regulation or the US Privacy Statutes regarding collection, transmission and safeguarding of personal data?
- Does the IAS potentially affect due process?
- Does the IAS potentially affect agreements to not retain personally identifiable information after termination of IAS use?
- Does the IAS potentially affect the “dictates of public conscience?”

- Does the IAS potentially affect the establishment of mass surveillance capabilities in an area where there is an expectation of privacy?

## Rights of persons affected by the actions or decisions of an IAS

- Are potentially affected persons notified that the IAS is in use?
- Are potentially affected persons allowed to monitor the use of the IAS?
- Are potentially affected persons allowed to opt in or opt out of the use of the IAS?
- Are potentially affected persons made aware of the opportunities to, and the procedures for, requesting access to IAS source code for inspection purposes?
- Are potentially affected persons aware of the opportunities to, and the procedures for, requesting access to IAS concepts of operations and standard operating procedures for inspection purposes?
- Are potentially affected persons aware of the opportunities to, and the procedures for, requesting access to IAS training data for inspection purposes?
- Are potentially affected persons advised that IAS use can have adverse impacts based on age, race, color, ethnicity, sex, religion, national origin, gender, gender identity, sexual orientation, familial status, biometric information, or disability status or economic class?
- Are potentially affected persons advised that IAS use can have adverse impacts based on membership in a group that enjoys legally protected status?
- Are potentially affected persons advised that IAS use can adversely impact their access to housing, education, employment, insurance, credit, or access to places of public accommodation?
- Are potentially affected persons made aware when sensitive personal information is collected and retained by the IAS?
- Are potentially affected persons notified when their personal data is destroyed after IAS use?
- Are potentially affected persons made aware of retention policies and schedules for personal data?
- Are potentially affected persons allowed to deny the disclosure of their personal data to third parties?
- Are potentially affected persons made aware of personal data protection policies and measures?
- Are potentially affected persons notified of the purpose of the IAS that is trained using their personal data?

- Are potentially affected persons informed when they are interacting with an IAS and not a human being?
- Are potentially affected persons made aware of the demographic, professional and intellectual diversity of the staff that developed the IAS?
- Are potentially affected persons made aware of negative determinations made by an IAS and given an opportunity to challenge those determinations to a human arbitrator?
- Are potentially affected persons informed as to what policies, laws, regulations, and/or rules governed the use of the IAS?
- Are potentially affected persons informed as to the test and evaluation procedures used to validate the IAS?
- Are potentially affected persons informed when the IAS is repurposed for use from some other, unrelated original purpose?
- Are potentially affected persons made aware when the perpetuation of a bias by an IAS has been discovered?
- Are potentially affected persons advised that IAS use can have adverse effects, even though they are unintended?

## Guarding against bias and rights violations

### Development phase search for sources of bias

- Does the demographic makeup of the IAS development team reflect that of the population(s) potentially affected by use of the IAS?
- When it is not possible or practicable to assemble a development team with a demographic makeup reflective of the potentially affected population, do team members have access to training materials that educate them on the cultures and characteristics of those populations?
- Do development teams look for unintended biases in the training data?
- Do development teams look for unintended biases in the IAS algorithms?
- Do development teams consider cultural biases in coalition and allied IAS and across joint services?
- Do development teams look for “data hubris” by attempting to identify the underlying causal mechanisms in their models so that spurious correlations identified by the IAS are not taken to be actual causal relationships?
- Do development teams look for bias introduced by IAS interaction with human operators?

- Do development teams look for bias introduced by IAS ingest of sensor data in the operational environment?
- Do development teams look for “black box” decision-making algorithms that make it difficult or impossible to detect bias?
- Do development teams look for synthetic training data generated from models that are potentially subject to bias?
- Do development teams validate potentially biased synthetic data against real data?
- Do development teams look for bias seeping into the training data during the data collection process?
- Do development teams look for “Simpson’s paradox,” where fusing data introduces new bias to the IAS?
- Do development teams look for “technological solutionism,” the perception that technology will lead to only positive solutions?
- Do development teams look for *representation bias* where the training data is not a representative reflection of the affected population?
- Do development teams look for *omitted variable bias* where the operational environment contains a variable not present in the models used to develop the IAS?
- Do development teams look for *social or behavior bias* where the affected person’s reactions to the IAS influences IAS actions/decisions?
- Do development teams look for *ranking bias* where the IAS actions or decisions are based on a small number of heavily weighted results and not on all available information?
- Do development teams look for *aggregation bias* where the IAS requires multiple models to fairly represent all sub-groups within the parent group population?
- Do development teams look for *evaluation bias* where IAS models may be erroneously validated?
- Do development teams look for *temporal bias* where the operational context changes, perhaps imperceptibly, so that initial assumptions about the environment, and the training data chosen to represent it, are no longer valid?
- Do development teams look for *automation bias* where human operators tend to reflexively accept decisions made by an automated system?
- Do development teams look for *assimilation bias* where human operators tend to modify information to fit into pre-existing analytical frameworks?
- Do development teams look for *confirmation bias* where developers gather training data that reaffirm conclusions they’ve already made?

- Do development teams look for *simplification bias* where human operators tend to simplify phenomena they encounter?
- Do development teams look for *activity bias* where training data comes from a system's most active users, rather than less active or inactive users?
- Do development teams look for *annotator bias* where developers rely on automation as a heuristic replacement for their own information seeking and processing?
- Do development teams look for *content production bias* where nonexistent differences in the data result from structural, lexical, semantic or syntactic differences in use by developers?
- Do development teams look for *exclusion bias* where specific groups of user populations are excluded from testing and subsequent analyses?
- Do development teams look for *feedback loop bias* where an algorithm learns from user behavior and feeds that behavior back into the model?
- Do development teams look for *funding bias* where biased results are reported in order to support or satisfy a funding agency or financial supporter?
- Do development teams look for *historical bias* where models are trained on past and potentially biased decisions?
- Do development teams look for *mirror imaging bias* where an IAS is developed to counter an adversary force (such as in a military application) and the developers assume that members of the adversary force have the same cultural, ethical, and cognitive characteristics of members of their own force?
- Do development teams look for *inherited bias* where tools that are built with machine learning are used to generate inputs for other machine learning algorithms?
- Do development teams look for *interpretation bias* where users interpret algorithmic outputs according to their internalized biases and views?
- Do development teams look for *linking bias* where network attributes obtained from user connections, activities, or interactions differ and misrepresent the true behavior of the users?
- Do development teams look for *loss of situational awareness bias* where human-machine teaming leads to humans being unaware of their situation such that, when command and control of a system is given to them, they are unprepared to assume their duties?
- Do development teams look for *modal confusion bias* where modal interfaces confuse human operators, who misunderstand which mode the system is using, taking actions which are correct for a different mode but incorrect for their current situation?

- Do development teams look for *ranking bias* where top-ranked information is perceived to be more important than lower ranking information?
- Do development teams look for *sampling bias* where non-random sampling of subgroups, causing trends estimated for one population to not be generalizable to data collected from a new population?
- Do development teams look for *training data bias* where algorithms are trained on one type of data and do not extrapolate beyond those data?
- Do development teams look for *uncertainty bias* where predictive algorithms favor groups that are better represented in the training data, since there will be less uncertainty associated with those predictions?

### **Post-operational human-on-the-loop oversight**

- Are IAS results revisited using human-on-the-loop oversight to look for biased results and rights violations that the IAS is not capable of identifying?
- Are human-on-the-loop oversight functions conducted by a group of persons representative of the population(s) potentially affected by the IAS?
- Is IAS use eliminated or curtailed when oversight groups discover bias or rights violations?
- When oversight groups discover bias or rights violations in one IAS, do they then look for the same bias in other IAS that have used all or part of the training data or algorithms used by the biased IAS?
- When oversight groups discover bias or rights violations in one IAS, do they ensure that a replacement or revision does not reproduce the same negative outcome, or introduce a new negative outcome?
- Are oversight groups familiar with “reviewer’s bias,” where the reviewers may be susceptible to introducing their own sets of expertise, experiences, and biases into evaluations?

# Abbreviations

---

|        |                                                  |
|--------|--------------------------------------------------|
| AI     | artificial intelligence                          |
| APB    | acquisition program baseline                     |
| CDD    | capability development document                  |
| CNA    | Center for Naval Analyses                        |
| DAD    | dimension of autonomous decision-making          |
| DEVRON | Development Squadron                             |
| DHS    | Department of Homeland Security                  |
| DOD    | Department of Defense                            |
| DON    | Department of the Navy                           |
| DRAID  | Joint AI Center's Data Readiness for AI Data     |
| FFRDC  | federally funded research and development center |
| IAS    | intelligent autonomous system                    |
| JAREL  | joint autonomy risk elements list                |
| KPP    | key performance parameter                        |
| KSA    | key system attribute                             |
| LAWS   | lethal autonomous weapons system                 |
| LOAC   | Law of Armed Conflict                            |
| METL   | mission essential task list                      |
| MOP    | measure of performance                           |
| R&D    | research and development                         |
| ROE    | rules of engagement                              |
| SJA    | Staff Judge Advocate                             |
| SME    | subject matter expert                            |
| SYSCOM | system command                                   |
| T&E    | test and evaluation                              |
| TTP    | tactics, techniques, and procedures              |
| UJTL   | universal joint task list                        |
| UxS    | unmanned systems                                 |

# References

---

- [1] DOD Directive 3000.09. May 8, 2017. *Autonomy in Weapon Systems*.  
<https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>.
- [2] DOD Directive 5000.01. Sept. 9, 2020. *The Defense Acquisition System*. Office of the Under Secretary of Defense for Acquisition and Sustainment.  
<https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500001p.pdf?ver=2020-09-09-160307-310>.
- [3] DOD Directive 2311.01. July 2, 2020. *DOD Law of War Program*.  
<https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/231101p.pdf?ver=2020-07-02-143157-007>.
- [4] DOD Directive 3000.03E. August 31, 2018. *DOD Executive Agent for Non-Lethal Weapons (NLW) and NLW Policy*.  
<https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300003p.pdf?ver=2017-09-27-125836-647>.
- [5] Kathleen Hicks, Deputy Secretary of Defense. May 26, 2021. Memorandum for Senior Pentagon Leadership, Commanders of the Combatant Commands, Defense Agencies, and DOD Field Activities. Subject: Implementing Responsible Artificial Intelligence in the Department of Defense. <https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF>.
- [6] "DOD Adopts Ethical Principles for Artificial Intelligence." Department of Defense Press Release. Feb. 24, 2020.  
<https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/>.
- [7] Scharre, Paul. 2018. *Army of None: Autonomous Weapons and the Future of War*. New York: W.W. Norton & Company.
- [8] "Campaign to Stop Killer Robots." <https://www.stopkillerrobots.org/learn/>.
- [9] Work, Robert O. "Keynote Address to the Policy and Ethics of Intelligent Autonomous Systems: Technical Exchange Meeting." (Used with permission), Autonomy Community of Interest, Office of the Secretary of Defense, March 4, 2021.
- [10] "Key Performance Parameter." Defense Acquisition University Glossary.  
<https://www.dau.edu/glossary/Pages/Glossary.aspx#!both|K|27793>
- [11] Federal Acquisition Regulations Sec. 35.017. *Federally Funded Research and Development Centers*. <https://www.acquisition.gov/far/35.017>.
- [12] *Federally Funded Research and Development Centers*.  
<https://www.law.cornell.edu/cfr/text/48/35.017>.
- [13] Joan L. Johnson, Deputy Assistant Secretary of the Navy, Research, Development, Test and Evaluation, and Chief of Naval Research RADM Lorin C. Selby. July 2, 2021. *Department of the Navy, Science and Technology Strategy for Intelligent Autonomous Systems*.  
<https://nps.edu/web/slamr/-/intelligent-autonomous-systems-science-and-technology-strategy-issued>.
- [14] Thomas W. Harker, Secretary of the Navy (Acting), Admiral Michal M. Gilday Chief of Naval Operations, and General David H. Berger Commandant of the Marine Corps. March 16, 2021. "Department of the Navy Unmanned Campaign Framework."  
<https://www.navy.mil/Portals/1/Strategic/20210315%20Unmanned%20Campaign%20Final%20Res.pdf?ver=LtCZ-BPIWki6vCBTdgtDMA%3d%3d>.

- [15] Joint Publication 1-02. Nov. 8, 2010 (as amended through Feb. 8, 2016). *Department of Defense Dictionary of Military and Associated Terms*.  
<https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf?ver=QkmPX3lFZqhMjdEGeSoB4A%3d%3d>.
- [16] Joint Publication 5-01. Dec. 1, 2020. *Joint Planning*.  
[https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp5\\_0.pdf?ver=us\\_fQ\\_pGS\\_u65ateysmAng%3d%3d](https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp5_0.pdf?ver=us_fQ_pGS_u65ateysmAng%3d%3d).
- [17] Secretary of the Navy Instruction 5000.2F. March 26, 2019. *Defense Acquisition System and Joint Capabilities Integration and Development System Implementation*.  
<https://www.secnnav.navy.mil/doni/Directives/05000%20General%20Management%20Security%20and%20Safety%20Services/05-00%20General%20Admin%20and%20Management%20Support/5000.2F.pdf>.
- [18] Sparrow, Robert. 2015. "Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender." *International Law Studies* 91: 699-728. <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1413&context=ils>.
- [19] Office of the General Counsel, Department of Defense. June 2015. *Department of Defense Law of War Manual*. [https://dod.defense.gov/Portals/1/Documents/law\\_war\\_manual15.pdf](https://dod.defense.gov/Portals/1/Documents/law_war_manual15.pdf).
- [20] "Defense Acquisition Guidebook, Chapter 8." Defense Acquisition University.  
<https://www.dau.edu/guidebooks/Shared%20Documents/Chapter%208%20Test%20and%20Evaluation.pdf>.
- [21] Behler, Robert F. *FY 2020 Annual Report to Congress*. Director, Operational Test and Evaluation.  
[https://www.dote.osd.mil/Portals/97/pub/reports/FY2020/other/2020DOTEAnnualReport.pdf?ver=rvLsaCQ\\_njLmPDrNIFJBWQ%3d%3d](https://www.dote.osd.mil/Portals/97/pub/reports/FY2020/other/2020DOTEAnnualReport.pdf?ver=rvLsaCQ_njLmPDrNIFJBWQ%3d%3d).
- [22] "Surface Development Squadron One (SURFDEVRON 1)."  
<https://www.surfpac.navy.mil/surfdevron1/>.
- [23] "Unmanned Underwater Vehicle Squadron One (UUVRON 1)."  
<https://www.csp.navy.mil/csds5/Detachments/Detachment-Unmanned-Undersea-Vehicles/>.
- [24] OPNAV Notice 5400. August 20, 2020. *Establish Commanding Officer, Unmanned Carrier Launched Multi-role Squadron One Zero*.  
[https://www.secnnav.navy.mil/doni/Directives/05000%20General%20Management%20Security%20and%20Safety%20Services/05-400%20Organization%20and%20Functional%20Support%20Services/5400.2288%20\(20\).pdf](https://www.secnnav.navy.mil/doni/Directives/05000%20General%20Management%20Security%20and%20Safety%20Services/05-400%20Organization%20and%20Functional%20Support%20Services/5400.2288%20(20).pdf).
- [25] Bosch, Karel van den, and Adelbert Bronkhorst. *Human-AI Cooperation to Benefit Military Decision Making*. North Atlantic Treaty Organization. STO-MP-IST-160.  
<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj3qcTugtTyAhWCv54KHTbyCIIQFnoECBwQAQ&url=https%3A%2F%2Fwww.sto.nato.int%2Fpublications%2FSTO%2520Meeting%2520Proceedings%2FSTO-MP-IST-160%2FMP-IST-160-S3-1.pdf&usq=AOvVaw2kiRBQmXTFyA19rlG1ZZaE>.
- [26] DARPA News. "Creating Cross-Domain Kill Webs in Real Time." *CHIPS: The Department of the Navy's Information Technology Magazine*. Sept. 21, 2020.  
<https://www.doncio.navy.mil/chips/ArticleDetails.aspx?ID=13872>.
- [27] Dalton, Andy. "The Navy's New Cloud Network Forms a Tactical 'Kill Web'." Engadget. May 19, 2016. <https://www.engadget.com/2016-05-19-us-navy-tactical-cloud-network-kill-web.html>.
- [28] Osborn, Kris. "Kill Web: Why the U.S. Military Sees Speed as the Ultimate Weapon." *The National Interest*. Dec. 20, 2020. <https://nationalinterest.org/blog/buzz/kill-web-why-us-military-sees-speed-ultimate-weapon-174769>.
- [29] Murphy, Robin, and James Shields. "Task Force on the Role of Autonomy in the DoD Systems." Jun. 2012.  
[https://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga\\_082152.pdf](https://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_082152.pdf).

- [30] Joint Publication 1-04. August 2, 2016. *Legal Support to Military Operations*. [https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp1\\_04.pdf](https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp1_04.pdf).
- [31] Dan Coats, Director of National Intelligence, and Principal Deputy Director of National Intelligence Susan Gordon. Jan. 16, 2019. *The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines*. Office of the Director of National Intelligence. <https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf>.
- [32] Cummings, M. L. *The Surprising Brittleness of AI*. Women Corporate Directors. <https://www.womencorporatedirectors.org/WCD/News/JAN-Feb2020/Reality%20Light.pdf>.
- [33] Chairman of the Joint Chiefs of Staff Instruction (CJCSI) 5123.01H. August 31, 2018. *Charter of the Joint Requirements Oversight Council (JROC) and Implementation of the Joint Capabilities Integration and Development System (JCIDS)* <https://www.jcs.mil/Portals/36/Documents/Library/Instructions/CJCSI%205123.01H.pdf?ver=2018-10-26-163922-137>.
- [34] Miller, Edward S. 1991. *War Plan Orange: The U.S. Strategy to Defeat Japan, 1897–1945*. Annapolis, Maryland: United States Naval Institute Press.
- [35] Letendre, Linell A. 2020. “Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers.” *International Law Studies* 96: 274-294. <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2925&context=ils>.
- [36] JAIC Public Affairs. “Enabling AI Data Readiness in the Department of Defense.” April 1, 2021. [https://www.ai.mil/blog/04\\_01\\_21\\_enabling\\_ai\\_data\\_readiness\\_in\\_the\\_dod.html](https://www.ai.mil/blog/04_01_21_enabling_ai_data_readiness_in_the_dod.html).
- [37] Ruane, Kate (American Civil Liberties Union), Face Recognition Technology Moratorium Coalition Letter, Feb. 16, 2021. [https://www.aclu.org/sites/default/files/field\\_document/02.16.2021\\_coalition\\_letter\\_requesting\\_federal\\_moratorium\\_on\\_facial\\_recognition.pdf](https://www.aclu.org/sites/default/files/field_document/02.16.2021_coalition_letter_requesting_federal_moratorium_on_facial_recognition.pdf).
- [38] Jakubowska, Ella, and Diego Narajo. May 13, 2020. *Ban Biometric Mass Surveillance: A Set of Fundamental Rights Demands for the European Commission and EU Member States*. European Digital Rights. <https://edri.org/our-work/blog-ban-biometric-mass-surveillance/>.
- [39] Richardson, Rashida. Dec. 2019. *Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*. AI Now Institute. <https://ainowinstitute.org/ads-shadowreport-2019.pdf>.
- [40] Stumborg, Michael F. “See You in a Month: AI’s Long Data Tail.” *War on the Rocks*. Oct. 17, 2019. <https://warontherocks.com/2019/10/see-you-in-a-month-ais-long-data-tail/>.
- [41] Joint Chiefs of Staff. May 24, 2021. “Universal Joint Task List.” <https://www.jcs.mil/Doctrine/joint-Training/UJTL/>.
- [42] Zhang, Baobao, and Allan Dafoe. Jan. 2019. *Artificial Intelligence: American Attitudes and Trends*. Center for the Governance of AI, Future of Humanity Institute, University of Oxford. <https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/executive-summary.html>.
- [43] Mitchell, Billy. “Google’s Departure from Project Maven Was a ‘Little Bit of a Canary in a Coal Mine’.” *FedScoop*. Nov. 5, 2019. <https://www.fedscoop.com/google-project-maven-canary-coal-mine/>.
- [44] Corderre, Michael, and Michael Register. “Fighting Back Against IEDs.” *Police Magazine*. Sept. 17, 2009. <https://www.policemag.com/340191/fighting-back-against-ieds>.
- [45] “Improvised Explosive Device Lexicon.” United Nations Mine Action Service. [https://unmas.org/sites/default/files/unmas\\_ied\\_lexicon\\_0.pdf](https://unmas.org/sites/default/files/unmas_ied_lexicon_0.pdf).
- [46] International Committee of the Red Cross. March 2016. *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*. [https://icrcndresourcecentre.org/wp-content/uploads/2017/11/4283\\_002\\_Autonomous-Weapon-Systems\\_WEB.pdf](https://icrcndresourcecentre.org/wp-content/uploads/2017/11/4283_002_Autonomous-Weapon-Systems_WEB.pdf).

- [47] International Panel on the Regulation of Autonomous Weapons. May 2020. *A Path Toward the Regulation of LAWS*. <https://www.ipraw.org/wp-content/uploads/2020/05/iPRAW-Briefing-Path-to-Regulation-May2020.pdf>.
- [48] Work, Robert O. April 2021. *Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities*. Center for a New American Security. <https://www.cnas.org/publications/reports/proposed-dod-principles-for-the-combat-employment-of-weapon-systems-with-autonomous-functionalities>.
- [49] Cook, Adam. 2019. *Taming Killer Robots: Giving Meaning to the "Meaningful Human Control" Standard for Lethal Autonomous Weapons Systems*. Air University Press, Curtis E. LeMay Center for Doctrine Development and Education. JAG School Paper, no. 1. [https://media.defense.gov/2019/Jun/18/2002146749/-1/-1/0/IP\\_001\\_COOK\\_TAMING\\_KILLER\\_ROBOTS.PDF](https://media.defense.gov/2019/Jun/18/2002146749/-1/-1/0/IP_001_COOK_TAMING_KILLER_ROBOTS.PDF).

**This report was written by CNA's Operational Warfighting Division (OPS).**

OPS focuses on ensuring that US military forces are able to compete and win against the nation's most capable adversaries. The major functional components of OPS work include activities associated with generating and then employing the force. *Force generation* addresses how forces and commands are organized, trained, scheduled, and deployed. *Force employment* encompasses concepts for how capabilities are arrayed, protected, and sustained at the operational level in peacetime and conflict, in all domains, against different types of adversaries, and under varied geographic and environmental conditions.

CNA is a not-for-profit research organization that serves the public interest by providing in-depth analysis and result-oriented solutions to help government leaders choose the best course of action in setting policy and managing operations.



Dedicated to the Safety and Security of the Nation

DRM-2021-U-030642-1Rev-2

3003 Washington Boulevard, Arlington, VA 22201

[www.cna.org](http://www.cna.org) • 703-824-2000