IEEE SA
STANDARDS ASSOCIATION

INDUSTRY CONNECTIONS REPORT

IEEE SA INDUSTRY CONNECTIONS RESEARCH GROUP ON ISSUES OF AUTONOMY AND AI IN DEFENSE SYSTEMS

# A FRAMEWORK FOR HUMAN DECISION-MAKING THROUGH THE LIFECYCLE OF AUTONOMOUS AND INTELLIGENT SYSTEMS IN DEFENSE APPLICATIONS

IEEE

Documents
Document Sharing Hub

sed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.

Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

**How to cite this document:** IEEE SA Research Group on Issues of Autonomy and AI in Defense Systems. (2024). *A Framework for Human Decision Making Through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications*. New York, NY: IEEE SA.

## AUTHORS

This report has been prepared by the following individuals (listed alphabetically), all participating in their individual capacities:

Sten Allik
Rachel Azafrani
David Barnes
Ingvild Bode
Ariel Conn
Eileen M. Lach
Craig Lennon
Rain Liivoja
John (Jack) Shanahan
Melodena Stephens
Lisa Titus
Alan R. Wagner

## ACKNOWLEDGMENTS

Special thanks are given to the following current and former members and affiliates of the Industry Connections Research Group, all participating in their individual capacities, who attended and contributed to meetings and helped review this paper in its various iterations (in alphabetical order):

Greg Adamson, Chad Bieber, Justin Bronk, Simon Chesterman, Di Cooke, Christian Ellis, Noah Ford, Charles Ovink, Jonathon Parry, Giacomo Persi Paoli, Signe Redfield, Brendan Schulman, Yvette Stevens, Catherine Tessier, and Wendell Wallach.

On behalf of the Research Group, we extend our gratitude to IEEE SA and the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. We are particularly thankful to Konstantinos Karachalios for his vision, to Anja Kaspersen for her guidance, and to John C. Havens for his support. We also thank Andrew Wright for his editorial assistance. We would also like to express our sincere thanks to the dedicated reviewers who took the time and lent us their expertise to offer feedback on the various iterations of the document. On behalf of the Industry Connections Research Group, special thanks are also given to Ariel Conn for her significant contributions and commitment to the development process of this document. The Chairs for this process were Ingvild Bode and Rachel Azafrani. For questions about the report, please reach out to: iaaids@ieee.org.

This document was prepared using an open, collaborative, and consensus-building approach, following the processes of the Industry Connections program, a program of the IEEE SA.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

# TERMS OF USE

Users may access, download, print, and/or retain one (1) copy of the document for individual use only, including job-related functions.

Compliance with this framework does not constitute compliance with applicable legal and regulatory requirements.

Except as allowed by the copyright laws of the United States of America or applicable international treaties, or as explicitly allowed in these Terms of Use, copies of the document may not be further copied, prepared, and/or distributed without prior written permission from IEEE.

Report for training, teaching, or any further distribution or work product of the document, in whole or in part, must be licensed by IEEE, and a licensing fee will be applied.

Please contact us by completing the IEEE Standard Permissions Use Form.

*The Institute of Electrical and Electronics Engineers, Inc. 3 Park Avenue, New York, NY 10016-5997, USA*

*Copyright © 2024 by The Institute of Electrical and Electronics Engineers, Inc.*

*All rights reserved. 4 October 2024. Printed in the United States of America.*

*PDF: STDVA27266 979-8-8557-1141-7*

*IEEE is a registered trademark in the U. S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated. All other trademarks are the property of the respective trademark owners.*

*IEEE prohibits discrimination, harassment, and bullying. For more information, visit http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.*

*No part of this publication may be reproduced in any form, in an electronic retrieval system, or otherwise, without the prior written permission of the publisher.*

*Find IEEE standards and standards-related product listings at: http://standards.ieee.org.*



Copyright © 2024 IEEE. All rights reserved.

Documents Document Sharing Hub

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.

Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

# NOTICE AND DISCLAIMER OF LIABILITY CONCERNING THE USE OF IEEE SA INDUSTRY CONNECTIONS DOCUMENTS

This IEEE Standards Association ("IEEE SA") Industry Connections publication ("Work") is not a consensus standard document. Specifically, this document is NOT AN IEEE STANDARD. Information contained in this Work has been created by, or obtained from, sources believed to be reliable, and reviewed by members of the IEEE SA Industry Connections activity that produced this Work. IEEE and the IEEE SA Industry Connections activity members expressly disclaim all warranties (express, implied, and statutory) related to this Work, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; quality, accuracy, effectiveness, currency, or completeness of the Work or content within the Work. In addition, IEEE and the IEEE SA Industry Connections activity members disclaim any and all conditions relating to: results; and workmanlike effort. This IEEE SA Industry Connections document is supplied "AS IS" and "WITH ALL FAULTS."

Although the IEEE SA Industry Connections activity members who have created this Work believe that the information and guidance given in this Work serve as an enhancement to users, all persons must rely upon their own skill and judgment when making use of it. IN NO EVENT SHALL IEEE OR IEEE SA INDUSTRY CONNECTIONS ACTIVITY MEMBERS BE LIABLE FOR ANY ERRORS OR OMISSIONS OR DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS WORK, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Further, information contained in this Work may be protected by intellectual property rights held by third parties or organizations, and the use of this information may require the user to negotiate with any such rights holders in order to legally acquire the rights to do so, and such rights holders may refuse to grant such rights. Attention is also called to the possibility that implementation of any or all of this Work may require the use of subject matter covered by patent rights. By publication of this Work, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. The IEEE is not responsible for identifying patent rights for which a license may be required, or for conducting inquiries into the legal validity or scope of patent claims. Users are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. No commitment to grant licenses under patent rights on a reasonable or non-discriminatory basis has been sought or received from any rights holder. The policies and procedures under which this document was created can be viewed at http://standards.ieee.org/about/sasb/iccom/.

This Work is published with the understanding that IEEE and the ICCom members are supplying information through this Work, not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought. IEEE is not responsible for the statements and opinions advanced in this Work.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

# TABLE OF CONTENTS


EXECUTIVE SUMMARY ..... 6

1. BACKGROUND ..... 10

1.1. ABOUT THIS PROJECT ..... 10
1.2. A BRIEF OVERVIEW OF THE DEBATE AROUND AIS IN DEFENSE SETTINGS ..... 12
1.3. WHY A LIFECYCLE FRAMEWORK ..... 14

2. INSIGHTS FROM THE PROCESS ..... 16

2.1. WORKING BACKWARD FROM AN EXAMPLE SCENARIO ..... 16
2.2. WHERE THE GROUP AGREED AND DISAGREED ..... 18
2.2.1. A NOTE ON GENERATIVE AI ..... 19

3. INTRODUCING THE LIFECYCLE FRAMEWORK ..... 20

3.1. LIFECYCLE STAGES AND ONGOING ACTIVITIES ..... 20
3.2. CROSS-CUTTING FACTORS TO BEAR IN MIND ..... 22

4. FIVE ONGOING ACTIVITIES ..... 24

4.1. EVALUATION OF LEGAL, ETHICAL, AND POLICY CONCERNS ..... 24
4.2. RESPONSIBILITY, ACCOUNTABILITY, AND KNOWLEDGE TRANSFERS ..... 25
4.3. CONSIDERING THE HUMAN: TRAINING, EDUCATION, AND HUMAN-SYSTEM INTEGRATION ..... 26
4.4. TEVV, MONITORING, HARDWARE SYSTEM OR SOFTWARE UPDATES AND INTEROPERABILITY, MAINTENANCE ..... 27
4.4.1. TESTING, EVALUATION, VERIFICATION, AND VALIDATION (TEVV) ..... 27
4.4.2. MONITORING ..... 28
4.4.3. HARDWARE SYSTEM OR SOFTWARE UPDATES AND INTEROPERABILITY ..... 28
4.4.4. MAINTENANCE ..... 28
4.5. RISK ASSESSMENTS ..... 28

5. NINE STAGES OF THE LIFECYCLE FRAMEWORK ..... 29

5.1. BEFORE DEVELOPMENT ..... 29
5.1.1. GENERAL LEGAL, ETHICAL, AND RELATED TECHNICAL CONCERNS IDENTIFIED AND ADDRESSED ..... 29
5.1.2. RATIONALE FOR MILITARY DEVELOPMENT AND USE OF AIS, FORMULATION OF SYSTEM REQUIREMENTS, AND CONSIDERING THE ROLE OF RESEARCHERS ..... 34
5.2. RESEARCH AND DEVELOPMENT ..... 37
5.3. PROCUREMENT AND ACQUISITION ..... 42
5.4. TESTING, EVALUATION, VERIFICATION, AND VALIDATION (TEVV) ..... 44
5.5. CONSIDERING THE HUMAN: EDUCATION, TRAINING, AND HUMAN-SYSTEM INTEGRATION ..... 48
5.6. POLITICAL AND STRATEGIC CONSIDERATIONS ..... 52
5.7. OPERATIONAL-LEVEL COMMAND AND CONTROL ..... 55
5.8. TACTICAL EMPLOYMENT ..... 57
5.9. REVIEW, REUSE, AND/OR RETIRE ..... 60

6. ACRONYMS ..... 62



Documents
Document Sharing Hub

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.

Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

# A FRAMEWORK FOR HUMAN DECISION MAKING THROUGH THE LIFECYCLE OF AUTONOMOUS AND INTELLIGENT SYSTEMS IN DEFENSE APPLICATIONS

## EXECUTIVE SUMMARY

The framework set out in this document addresses stakeholders involved in policy, design, testing, procurement, decision-making, deployment, and evaluation processes related to autonomous and intelligent systems (AIS), especially in public-sector decisions about defense applications.

This document highlights that there is no such thing as a completely autonomous system. Humans are always involved somewhere along the way. The framework, therefore, aims to inform discussions about manifold technical, ethical, and legal issues arising out of pursuing a human-centric approach across the entire lifecycle of AIS, especially with regard to human responsibility and accountability.

The scope of this report is broad, encompassing diverse insights that reflect on technical, ethical, and legal considerations inherent in AIS with a focus on military settings. The framework supports stakeholders in raising and offering first steps towards applying existing sets of broad ethical principles and standards in the context of AIS, including but not limited to those associated with Article 36 of Additional Protocol I to the Geneva Conventions.

Considering the full lifecycle of systems—from development to procurement, testing, deployment, use, sustainment, and decommissioning—the framework offers a granular view of stakeholder involvement, clarifying when and how challenges should be addressed and identifying who should be responsible and accountable at each lifecycle stage.

Decision-makers should consider that AIS may not necessarily be appropriate for use across defense settings, whether because the system itself is immature or the level of human knowledge is insufficient. They need to comprehensively assess whether the system is ethical and legal. This may include considering issues, such as the potential benefits, risks, and harms of developing and deploying any system; identifying potential blind spots; effects on and gaps in human skills; technical or scientific shortfalls; and issues, such as supply chain reliability.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

The framework also calls attention to how many risks are introduced and must be addressed at the early stages of the lifecycle, including at the design stage of AIS.

The framework is the outcome of discussions among the IEEE SA Industry Connections Research Group on Issues of Autonomy and AI in Defense Systems, an independent group with members drawn from various countries, academic disciplines, and professional backgrounds. It can serve as a first step towards the potential development of standards processes around AIS in defense with a focus on ethically aligned design.

**Nine lifecycle stages and five ongoing activities**

Members of the Research Group concluded that there are nine stages in the lifecycle of AIS in defense contexts (illustrated in FIGURE 1), in which human decision-makers play especially key roles, and between which there is likely to be a transfer of responsibility and accountability to new human decision-makers:

a) Before AIS development
   - General legal, ethical, and related technical concerns identified and addressed
   - Rationale for military development and use of AIS, formulation of system requirements, and considering the role of researchers

b) Research and development

c) Procurement and acquisition

d) Testing, evaluation, verification, and validation (TEVV)

e) Considering the human: Education, training, and human-system integration

f) Political and strategic considerations

g) Operational-level command and control

h) Tactical employment

i) Review, reuse, and/or retire

Additionally, the group identified five “ongoing activities” in which human decision-making elements occur repeatedly throughout the lifecycle:

1) Evaluation of legal, ethical, and policy concerns

2) Responsibility, accountability, and knowledge transfers

3) Considering the human: training, education, and human-system integration

4) TEVV, monitoring, hardware system or software updates and interoperability, maintenance

5) Risk assessments


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

Some of the nine lifecycle stages are necessarily sequential—for example, a system must be developed before it can be deployed. Many, however—especially during the first half of the lifecycle—may have significant overlap, and may start or end at around the same time. These stages may also be repeated many times after a system has been deployed, when it received new hardware or software updates. Redundancy is built into the framework by design.

The focus throughout is on the humans involved in decisions. The framework is designed to help identify the relevant decision-makers, their role at each stage, the knowledge and documentation they need from previous decision-makers, and the types of questions and issues to be addressed at each stage.

Everyone associated with deploying an AIS for military purposes needs to have a common understanding of its capabilities and limitations. Individual, incremental steps in design and development may seem unproblematic, but then trigger cascading effects when combined with other technological components. A system cannot be understood in isolation from other technical components, the logistical environment of use, the human environment, or the personnel it will interact with for a given task.

**Lessons from the process**

The Research Group approached the task of developing the lifecycle framework by first creating a fictional example scenario involving an uncrewed underwater vehicle deployed to protect ships carrying food in a conflict situation. They defined possible outcomes of this deployment and worked backward to explore, test, and validate stages in the lifecycle as they developed it.

This approach proved valuable in helping to differentiate stages that were more chronological from stages that needed to be addressed throughout most of the lifecycle. It clarified the need to address questions at appropriate stages, the trade-off between apparent simplicity of operation at later stages and complexity of development at earlier stages, and the importance of providing adequate technical and legal documentation at each stage for the use of people involved at later stages.

The process of developing the framework made clear the need to involve technical people from civilian sectors in discussions with experts in legal and ethical issues, policymakers, and military leadership. Different experts have different understandings of terminology and capabilities. Different stages of the lifecycle tend to have a preponderance of one type of expert rather than the diversity of experts needed for a more sufficient awareness of the multiple concerns AIS raises.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

Given how contentious discussions about AIS in military contexts can be, there was an encouraging degree of agreement among Research Group members on many issues. Areas of broad agreement included the value of considering a risk-based framework—accounting for differences in factors such as predictability, type of target, duration, and environment of use—and the importance of commanders and operators having sufficient contextual understanding of how a system works and what it can and cannot do.

**A contribution to an ongoing debate**

Debates about the legality—and, to a more limited extent, the ethics—of AIS in military settings have been taking place for around a decade. Some uses are relatively uncontroversial—such as navigation, management, analytics, and logistical assistance—although the systems' accuracy, interoperability, ability to mitigate against algorithmic bias, and verifiable maturity can still be questioned. Other capabilities remain highly contentious, especially those that would allow a system, as defined by the International Committee of the Red Cross (ICRC), to determine its own objectives: identifying, selecting, and applying force to targets without human intervention.

Nothing in this framework should be taken as precluding that new regulations, standards, policies, or laws may need to be developed to address concerns about AIS in defense contexts. The framework is intended to advance and contribute to these discussions and efforts. Although this framework was developed for AIS in defense, it is broad and adaptable enough that it should prove useful for stakeholders in the development and use of most AIS.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

# 1. BACKGROUND

## 1.1. ABOUT THIS PROJECT

The framework set out in this document is intended to help inform a range of stakeholders involved in various types of decision-making related to autonomous and intelligent systems (AIS), especially in the military context. The framework provides a granular means for identifying the technical considerations that are necessary to meet legal requirements, to inform and guide ethical discussions, and to determine human responsibility and accountability across the entire lifecycle of AIS.

Article 36 of Additional Protocol I to the Geneva Conventions requires states to review the legality, under international law, of all new weapons, means, and methods of warfare—which increasingly incorporate AIS. In practice, given the new and technical complexity of these systems, states often do not have the requisite technical knowledge and procedures in place to conduct meaningful, standardized reviews. The framework set out in this document is intended, in part, to contribute to filling that gap. However, its scope is much broader.

Many AIS used in defense contexts are not codified as weapons, means, or methods of warfare, and in many cases, may not be different from commercial versions of AIS. Yet states will nonetheless need to address various legal, ethical, and technical challenges of their use. It is, for example, possible for systems to be compliant with international law but still harmful. Existing sets of principles on ethics often use terms that can be difficult to operationalize, such as systems being “responsible” or “trustworthy.” The way such terms are commonly understood in relation to human processes does not always map easily onto computational systems, especially but not only in military contexts.

The framework is intended to offer more technical guidance around putting principles into practice. While focused on military contexts, the framework can also be used beyond. It helps identify and address issues related to AIS, notwithstanding whether the systems are for military or civilian use. Therefore, it is intended for all stakeholders and decision-makers across the lifecycle stages, including those in research and development and procurement, policymakers, and leaders engaged in decisions about the use of force and the use of AI in public sector operations, and participants in international governance and treaty discussions.

The framework is the outcome of discussions among the IEEE SA Industry Connections Research Group on Issues of Autonomy and AI in Defense Systems (“the Research Group”). With members drawn from a range of nations and disciplines, the Research Group’s mandate included defining the next steps for addressing the challenges


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

identified in a 2021 document produced by a predecessor independent research group under IEEE SA auspices, entitled *Ethical and technical challenges in the development, use, and governance of autonomous weapons systems*.[1]

The Research Group agreed on the approach of developing a framework that could more clearly define the points of human decision-making throughout the lifecycle of AIS in defense contexts. The Research Group provided a technical platform for diverse stakeholders to share insights and develop common understanding.

In developing the framework, the Research Group had in mind systems that are in development now or may be developed in the next few years. The framework is intended for all public sector systems that incorporate autonomous or AI capabilities. For example, the development of self-driving, self-maneuvering, or self-navigation capabilities would benefit from adherence to this framework. The Research Group also designed the framework so that it could adapt as technology changes in the future.

### 1.1.1. A NOTE ON TERMINOLOGY

The title of the Research Group—"On Issues of Autonomy and AI in Defense Systems"—reflects its wide scope. Discussions about such issues can often be complicated by the use of terminology that is defined differently by different organizations and can be interpreted too narrowly.[2]

This document eschews, where possible, the use of "autonomous weapons systems (AWS)" to reduce the risk of excluding relevant defense and military systems. The preferred terminology is AIS in defense or military contexts. This has the advantage that AIS—autonomous and intelligent systems—is now an industry conformity discourse, and this document aims to encourage discussion of conformity as it applies to public sector settings.

The Research Group also considered alternatives to the word "lifecycle," which some consider inappropriate for systems designed to injure and kill. As "lifecycle" suggests a biological entity that can be born, live and die, it may contribute to the problem of anthropomorphizing technology and exaggerating the extent to which AI systems process information in an analogous way to humans. Suggested alternatives included "idea to

[1] After several years of detailed expert consultations, IEEE SA was invited to provide insights from industry practices, ontological standards for autonomous and intelligent systems, processes, relevant research, and strategies for incorporating safety and ethics by design. This was initiated by the Director of Disarmament in Geneva and the Chair of the Group of Governmental Experts for Emerging Technologies in the area of Lethal Autonomous Weapons Systems.
[2] The Research Group considered putting forward its own definition of AWS. As many other groups have already offered varying definitions, it is not clear there is value in adding yet another definition to the discussion. In international treaty discourse, for instance, the acronym "LAWS"—lethal autonomous weapon system—is often used. However, this can exclude AIS systems not considered to be "lethal," which is a matter of interpretation. Other processes address this concern by referring instead to "AWS," or autonomous weapon systems, but this may also be overly restrictive. Some interpret "autonomous" as implying a total lack of human input, rather than the presence of some degree of autonomous or AI-enabled capabilities. Similarly, the term "weapons" can be narrowly read as excluding other relevant military or defense systems.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

decommission stages" and "conception to decommission stages." Ultimately, however, the Group decided to stick with "lifecycle" as it is so commonly used in the technology and military fields.

In many of the lifecycle stages described below, the use of some specialist technological and military language is unavoidable. In general, however, the document aims to use terminology that is as accessible as possible to stakeholders who may not have a background in either the technology or defense sectors, such as politicians, policymakers, ethicists, and lawyers.

## 1.2. A BRIEF OVERVIEW OF THE DEBATE AROUND AIS IN DEFENSE SETTINGS

Nothing in this framework should be taken as precluding that new regulations, policies, or laws may need to be developed to address concerns about AIS in defense contexts. The framework is intended to advance and contribute to ongoing debates on this subject while addressing the reality of AIS developed for military purposes. Even for relatively uncontroversial uses of AIS, stakeholders should follow this framework so that development and use address legal and ethical concerns, e.g., that such systems are not misused later in ways that could have been anticipated. The Research Group also recognizes that new issues will arise as technology changes.

AIS can offer the same kind of attractions to political and military leaders as such systems do to industry leaders. For example, AIS promises to enable more efficient operations, improve cost-effectiveness, take over routine work, and free up skilled humans for other tasks. Systems are increasingly becoming smaller, more affordable, and more expendable. However, the hidden costs associated with developing AIS are frequently overlooked, including, for example, environmental costs. AIS is developed to become more capable of undertaking a wider range of tasks in various domains of warfare: air, space, ground, sea, and non-kinetic, such as information and communications.

Some uses of AIS in defense settings are relatively uncontroversial—such as navigation, management, analytics, and logistical assistance—although the systems' accuracy, interoperability, ability to mitigate against algorithmic bias, and verifiable maturity can still be questioned. Other capabilities remain highly contentious, especially those that would allow a system, as defined by the International Committee of the Red Cross (ICRC), to determine its own objectives: identifying, selecting, and applying force to targets without human intervention.

Some military personnel and political leaders argue that such capabilities could, in theory, lead to fewer civilian casualties by lowering the possibility of error in high-stress environments. However, as AIS can fail in different


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

ways than humans, it cannot be assumed that a reduction in errors would necessarily imply that a system is safer. Concerns include:

- Unanticipated failure modes are possible as systems move from lab testing to use in real environments. Systems may react unexpectedly to natural variations in data, such as blur in images; have unforeseen vulnerabilities to attacks against data, networks, or models; or be susceptible to "reward hacking," which results in subtle changes to what a system is trying to achieve.
- Systems may prove less able than humans to anticipate potential breaches of International Humanitarian Law (IHL) or the Law of Armed Conflict (LOAC). Some AI systems have been shown to exacerbate existing human biases and discrimination. It can be difficult to understand how AI arrived at a decision (the "black box" problem). Humans tend to place too much trust in machines ("automation bias").
- Loss of human control over the use of force raises ethical concerns related to human dignity and moral agency. When a system behaves unexpectedly, neither those who developed its component parts nor those who decided to deploy it may feel morally responsible. It may be difficult to hold specific humans accountable for actions performed by machines.

Debates about the legality—and, to a more limited extent, the ethics—of AIS in military settings have been taking place for around a decade across national and international bodies, and within governments and militaries. These debates are becoming more urgent. As states anticipate the growing use of such systems by opposing forces—including non-state actors and individuals, as the technology becomes more accessible—there is potential for arms-race dynamics and unregulated proliferation.

Systems are at significant risk of being designed, developed, and deployed without sufficient human and regulatory oversight. One reason is that the secrecy and sensitivity of such projects limit willingness to seek feedback. Another is that individual, incremental steps in design and development may seem unproblematic, but then trigger cascading effects when combined with other technological components.

TABLE 1 summarizes examples of the main governance initiatives to date, which have resulted in various sets of guidelines and principles. So far, these are all voluntary commitments. The ICRC, regarded as the guardian of IHL, has called for new binding international rules. It advocates prohibiting some kinds of systems—including those with effects that cannot be sufficiently understood, predicted, and explained, and those designed or used to apply force against persons directly—and "strict restrictions" on other kinds.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

**TABLE 1. Governance initiatives on AIS in military contexts**

| Level | Examples |
|---|---|
| Global | (2023). General Assembly First Committee Resolution: Lethal Autonomous Weapons Systems (UN Document No. A/C.1/78/L.56)<br>(2019) 11 Guiding Principles, Convention on Certain Conventional Weapons (CCW) |
| Cross-Regional | (2023) REAIM Call to Action<br>(2023) Political Declaration on Responsible of AI (US-led) |
| Regional | (2024) Freetown Communiqué (Economic Community of West African States)<br>(2023) Belen Communique (Latin America and the Caribbean)<br>(2023) CARICOM Declaration<br>(2021) NATO Principles of Responsible Use of AI in Defense<br>(2021) European Parliament Resolution on Artificial Intelligence incl military use |
| State | (2022) Ambitious, Safe, Responsible: Our Approach to the Delivery of AI-Enabled Capability in Defense (UK)<br>(2022). Provisions contained in the Arms Control and Disarmament Strategy (Switzerland)<br>(2020) AI Principles: Recommendations on the Ethical Use of AI (US)<br>(2020) A Method for Ethical AI in Defense (Australia)<br>(2019) AI For Defense (France) |

In October 2023, the First Committee of the UN General Assembly adopted a resolution stressing "the urgent need for the international community to address the challenges and concerns raised by autonomous weapons systems." The Secretary General has likewise recommended, "a legally binding instrument to prohibit lethal autonomous weapon systems that function without human control or oversight, and which cannot be used in compliance with international humanitarian law, and to regulate all other types of autonomous weapon systems." At the same time, there is a recognition that being compliant with IHL may not be sufficient to demonstrate that a system is responsibly deployed throughout its lifecycle.

## 1.3. WHY A LIFECYCLE FRAMEWORK

Considering the full lifecycle of systems—from development to procurement, testing, deployment, use, sustainment, and decommissioning—the framework offers a granular view of stakeholder involvement, clarifying when and how challenges should be addressed, and identifying who should be responsible and accountable at each lifecycle stage. It can also allow leaders to consider each project holistically, and gain insights at every stage about potential problems, conflicting guidance, or legal jeopardy.

Research Group members agreed that most of the challenges identified in the 2021 document, although seemingly technical in nature, remain human-centric despite the autonomous nature of the systems under discussion. Similarly, there was agreement that most of the challenges are not one-off problems to be solved but instead need to be addressed by humans repeatedly—when a system is first considered, then developed, tested, deployed, sustained, and retired.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

Most discussions around AIS in defense settings tend to focus on two aspects: First, discussing what happens if the system functions in unintended or unpredictable ways, and how to assign responsibility and accountability for failure. Second, discussing what IHL permits, prohibits, and requires. Both discussions require addressing human decisions throughout the entire lifecycle of the system, including well before it is even considered for deployment.

Everyone associated with the deployment of an AIS for military purposes needs to have a common understanding of its capabilities and limitations: how it works, what criteria will trigger the use of force, the risks and potential consequences of failures, and how geopolitical context could affect its impact. Reaching a common understanding is not always easy given the professional diversity of stakeholders involved, such as lawyers, ethicists, and technologists, an issue further explored in 2.2.

A system cannot be understood in isolation of the human personnel it will interact with for a given task or its intended environment of use. Do the humans involved understand what the system will do? Are there well-defined, well-understood, and well-motivated transition points and feedback loops between human and machine activities, actions, and decisions? Is the environment dynamic or static, observable or not, communication-denied or not?

This lifecycle framework can better enable all involved to define and design around intent as early as possible and maintain that throughout the lifecycle. This should help human operators understand the context within which they are working. It should also lead to revisiting aspects of the project before use if the understanding of context and intent has changed.

In other words, the lifecycle framework makes clear that a system should not be used unless it has documented redundancies; clear documentation about contexts in which it can reasonably be deployed; an intuitive user interface; processes that take into account human psychology and cognitive limits; and people who understand how the technology works. Addressing these and other specific issues identified in the framework offers a way to operationalize general principles such as being reliable, trustworthy, and responsible.

This work builds on previous thinking about an AIS lifecycle and the role of humans during stages of development and use. The AIS lifecycle developed by the Research Group is more comprehensive, fine-grained, and modular than previous models. The greater the potential for autonomy in a military system, the more critical it becomes to consider human-centric design principles as early and as often as possible.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

The work of the Research Group is a step towards operationalizing and clarifying such a human-centric approach to AIS. Often, it raises challenges and questions rather than providing definite answers. These answers could be spelled out in a formal standards process around AIS in the military context. Such an approach could build on the proven track record of industry standards as consensus documents drafted by expert teams to provide operational-level guidance and a scientific basis for regulation.

# 2. INSIGHTS FROM THE PROCESS

## 2.1. WORKING BACKWARD FROM AN EXAMPLE SCENARIO

The Research Group developed the lifecycle framework over two years of group and individual meetings. They approached the task by first developing an example scenario. They then worked backward from this scenario to explore, test, and validate stages in the lifecycle as they developed it. This document is available as a separate record online and shares notes from discussions about the scenario and the framework to provide insight into the thought processes involved.[3] An abridged version of the fictional scenario is as follows:

> As food insecurity increases worldwide, an international armed conflict results in threats to ships carrying grain. Researchers in academia and industry develop an uncrewed underwater vehicle (UUV) with capabilities to defend grain ships. It is called the Food Defender (FD). When a national military deploys the FD, one of the following occurs:
>
> 1) It attacks and sinks an adversarial ship while ignoring friendly ships.
> 2) It fails to attack an adversarial ship, and, as a result, a grain ship is attacked.
> 3) It is hacked by an adversarial system and disabled or used to attack a grain ship.
> 4) It sinks a ship of refugees, having mistaken it for an adversarial ship.

The scenario was intended to describe a fairly straightforward combination of capabilities that exist today, based on unclassified information, but without resembling any existing system too closely. The hypothetical system was designed to limit the risk of civilian interaction so the Research Group's attention could be focused on other issues that could otherwise be under-addressed.

[3] For a full version of the example scenario, please see IEEE SA Research Group on Issues of AI and Autonomy in Defense Systems. (2024). Human Decision-Making Through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications: Example Scenario with Comments, available at: https://doi.org/10.5281/zenodo.13837345


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

Working backward from the scenario proved to be a valuable approach. It enabled members of the group to remain focused. It clarified which issues need to be addressed at each stage of the lifecycle and how the same issues might need to be addressed at different stages by different people. This helped to differentiate stages that were more chronological from stages that needed to be addressed throughout most of the lifecycle. The result was a well-rounded framework that can aid in considering the potential impacts of actions at later stages, problems that might arise, and questions of who should be held responsible and accountable.

Four general lessons emerged from the process of developing the framework.

**1) Addressing questions at appropriate stages:** Many conversations about AIS in defense contexts focus on the moment when the system is used. However, the group found that questions arising at that stage tended to lack focus and specificity. Considering the same questions at earlier stages provided a better sense of how to address them before the system was deployed, identify who should address them, and establish a chain of responsibility and accountability for when something went wrong. This could be addressed by following an ethics-by-design approach.

**2) Awareness of trade-offs and risks in complexity:** Autonomous and AI-enabled capabilities can appear to simplify the operation of a system at the point of use. However, the development and oversight of such capabilities can—and, at least for the foreseeable future, will—add complexity at an earlier stage. Militaries and governments considering AIS for military purposes need to gain a clear understanding of the costs, benefits, tradeoffs, and risks of developing or adding autonomous and AI-enabled capabilities to a system, and be prepared to make greater investments in people as well as technology.

**3) Involvement of technical people from civilian sectors:** It was challenging to find people with technical backgrounds to be part of the Research Group. Many who worked in related fields, such as aviation, autonomous vehicles, or medical diagnostic software, suggested that their work was unrelated to defense. However, there is uncertainty about which technologies might be available and weaponized or militarized in the future.

Technical experts can help non-technical people—such as those with legal and ethical backgrounds, policymakers, and military leadership—understand how technology works, its capabilities and limitations, and what questions to ask. Advancements in controls, standards, and policies for AIS can be achieved more effectively if technical and non-technical people are involved in the conversations. Different stages of the lifecycle tend to have a preponderance of one type of expert—e.g., technical experts in research and development and procurement and acquisition—rather than the diversity of experts needed for a more


Documents
Document Sharing Hub
Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

sufficient awareness of the multiple concerns AIS raises.

**4) Importance of documentation:** Technical and legal documentation at each stage of the lifecycle is critical to give context and understanding to people interacting with a system at later stages of the lifecycle. It is the most effective means for providing commanders and operators with the information they need to understand what the system can and cannot do.

Documentation can also help verify that testing regularly confirms the system is functioning as intended and could be used in compliance with IHL and that human users understand how and when to use the system. It helps to establish responsibility and accountability throughout all stages of the lifecycle and during transitions between stages. For example, legal reviews conducted at different stages of the lifecycle should be clearly documented and known to legal advisers and commanders on a particular mission.

## 2.2. WHERE THE GROUP AGREED AND DISAGREED

Observations about the Research Group's experience working on issues related to AIS in military contexts may help other multistakeholder groups of experts to better collaborate and overcome misunderstandings. Participants had different ideas and understandings about the types of technologies to be considered and their current capabilities.

For example, when developing the scenario, military experts described a UUV they thought would be legal. Legal experts assessed aspects of it to be in violation of IHL. They offered suggestions for improvement, but roboticists pointed out technical limitations. Only after extensive and nuanced discussions could all members agree on the scenario. Challenges such as these highlight how critical it is to bring together experts with different backgrounds and expertise to consider the full complexity of AIS in defense contexts.

Different experts have different understandings of terminology and capabilities. When the scenario was tested with wider groups of experts, about half agreed that a system like the FD was realistic to imagine within a timeframe of around five years. A quarter believed that the technology was further off, while the other quarter said it already existed.

Given how contentious discussions about AIS in military contexts can be, another significant finding was how much agreement there was among the group on a variety of issues. The Research Group comprised people from very different backgrounds. The disagreements centered on definitional issues, limits of acceptability for the use


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications

of AIS in military contexts, and the question of whether or not such systems should be held to a higher standard than humans. These areas of disagreement were relatively limited and outweighed by many points of agreement, including the following:

- There is value in considering a risk-based framework for decisions about the development, deployment, and employment of AIS in defense contexts.
- It is beneficial to categorize AIS into risk categories, accounting for factors such as predictability, type of target, duration, and environment of use. For example, all agreed that AIS for nuclear weapons would fall into an "unacceptable" risk category.
- General purpose robots, especially those designed for civilian or non-military use, should not be weaponized. (Although members of this Research Group agreed on this point, many reviewers expressed uncertainty and suggested the statement was too broad.)
- A system cannot be acceptable for use unless commanders and operators have sufficient contextual understanding of how it works and what it can and cannot do. A system would, for example, be unacceptable if the user cannot foresee whether the effect of the attack may be unlawful.
- When considering what to allow or regulate, it may be more helpful to think about autonomous and AI-enabled capabilities—e.g., targeting—rather than try to broadly define and regulate AIS themselves.
- Systems should not be able to learn and update in the field without proper testing, evaluation, verification, and validation (TEVV), legal review, and addressing how the update might affect the contextual understanding of the human operators. Because of the substantial technical challenges to such fast and flexible TEVV, this means AIS should not be able to learn and update in the field.
- AIS should be reliable, predictable, and understandable in their functioning and effects. The use of a military AIS that functions erratically could be indiscriminate and, therefore, in potential violation of IHL.

### 2.2.1. A NOTE ON GENERATIVE AI

Rapid progress in generative AI came to public attention only after the example scenario had been developed. Widespread publicity of its associated problems demonstrates the need for much more transparency, maturity, understandability, and reliability before it should be considered for use in highly consequential applications. This is especially important given the expected increase in the employment of AI agents—software entities that perform some tasks autonomously—and the potential use of such agents in future AIS.


Copyright © 2024 IEEE. All rights reserved.
Authorized licensed use limited to: IEEE Xplore. Downloaded on October 09,2024 at 14:48:28 UTC from IEEE Xplore. Restrictions apply.
Full document details: IEEE A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications