ARL-TN-1297 ● FEB 2026

Methodologies to Facilitate AI/ML Assurance for Safety-Critical Systems

by James Michaelis

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TN-1297 ● FEB 2026

Methodologies to Facilitate AI/ML Assurance for Safety-Critical Systems

James Michaelis
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
February 2026	Technical Note	10/01/2024	09/30/2025

4. TITLE AND SUBTITLE
Methodologies to Facilitate AI/ML Assurance for Safety-Critical Systems

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
James Michaelis

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-IB Adelphi, MD 20783-1138	ARL-TN-1297

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCID: James Michaelis, 0000-0003-3732-2145

14. ABSTRACT
In line with U.S. Army research and modernization priorities, AI/ML systems have been acknowledged as imperative to enhance operational effectiveness and maintain strategic advantages on the battlefield. Since these systems commonly operate under mission-critical settings, methods to facilitate trust and risk assessment represent an essential Army requirement. To address such needs, ARL initiated an investigation into methodologies for definition and usage of safety argumentation over U.S. Army AI/ML technologies. FY25 program efforts have included the following activities: 1) investigating approaches for designing argument structures to establish claims of AI/ML safety backed up by corresponding evidence gathered through test, evaluation, validation, and verification (TEVV) activities; 2) defining approaches to enable encoding of assurance cases to facilitate their review by Army stakeholders and corresponding usage by AI/ML TEVV support services; and 3) conducting a set of engagements with research teams at ARL to obtain information on AI/ML technologies under development, which are intended to support assessment of the assurance case methodologies for the Army's S&T needs.

15. SUBJECT TERMS
Military Information Sciences, artificial intelligence, machine learning, system safety, risk management framework, ontology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	24

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
James Michaelis	(520) 691-6388

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Executive Summary

In line with U.S. Army research and modernization priorities, AI/ML systems have been acknowledged as imperative to enhance operational effectiveness and maintain strategic advantage on the battlefield. Since these systems commonly operate under mission-critical settings, methods to facilitate trust and risk assessment represent an essential Army requirement.

Prior efforts to establish AI/ML risk management frameworks have identified safety argumentation as a key requirement for AI/ML trust and risk assessment. To address such needs, ARL initiated an investigation into methodologies for definition and usage of safety argumentation over Army AI/ML technologies. The 2025 program efforts have included the following activities:

- **Assurance Case Design:** An investigation into approaches for designing argument structures to establish claims of AI/ML safety backed up by corresponding evidence gathered through test, evaluation, validation, and verification (TEVV) activities.
- **Assurance Case Encoding:** Definition of approaches to enable encoding of assurance cases to facilitate their review by Army stakeholders and corresponding usage by AI/ML technology TEVV support services.
- **AI/ML Systems Cataloging:** A set of engagements conducted with research teams at ARL to obtain information on AI/ML technologies under development, which is intended to support assessment of the assurance case methodologies for the Army's S&T needs.

1. Introduction and Mission Description

In line with U.S. Army research and modernization priorities,[1] AI/ML systems have been acknowledged as imperative to enhance operational effectiveness and maintain strategic advantage on the battlefield. Recent advances in AI/ML S&T have focused on a variety of use cases, including support for autonomous platforms,[2] target recognition,[3] and course of action analysis.[4] Since these systems commonly operate under mission-critical settings, methods to facilitate trust and risk assessment represent an essential Army requirement.[5]

Prior efforts to establish AI/ML risk management frameworks (RMFs), such as the National Institute of Standards and Technology (NIST) RMF,[6] have identified safety argumentation as a key requirement for AI/ML trust and risk assessment. To address such needs, the U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) has initiated investigation into methodologies for definition and usage of safety argumentation over Army AI/ML technologies.

Our program efforts have centered specifically on approaches for definition and assessment of assurance cases and argument structures to establish claims of AI/ML safety based on corresponding evidence gathered through test, evaluation, validation, and verification (TEVV) activities. Three sets of activities were carried out in support of 2025 efforts:

- **Assurance Case Design:** An investigation into approaches for designing argument structures intended to support Army-relevant AI/ML technology design and safety requirements.
- **Assurance Case Encoding:** Definition of approaches to enable encoding of assurance cases to facilitate their review by Army stakeholders and corresponding usage by AI/ML TEVV support services.
- **AI/ML Systems Cataloging:** A set of engagements conducted with research teams at ARL to obtain information on AI/ML technologies under development, which is intended to support assessment of the assurance case methodologies for Army S&T needs.

2. Program Objectives

In support of the 2025 program activities, corresponding objectives included the following:

- Conducting technical assessments of existing methodologies for assurance case definition and corresponding evidence gathering for supporting claims

of AI/ML safety. These activities were conducted as part of ARL's participation in The Technical Cooperation Program (TTCP) AI Strategic Challenge (AISC) and focused on conducting assessments of existing assurance methodology applicability to both Army- and Five Eyes (FVEY)-developed AI/ML technologies.

- Defining approaches to enable storage, querying and editing of Goal Structuring Notation (GSN)[7]–based assurance cases (Figure 1) via ontology-based encodings. These approaches were intended to support 1) assurance case review by Army stakeholders, 2) assurance case usage within AI/ML TEVV support services, and 3) future establishment of a library of assurance case templates intended to support encoding of assurance cases for Army AI/ML technologies.
- Identifying and cataloging key design and usage details on a set of reference AI/ML technologies under development at ARL, which are intended to support ARL-focused assessment of assurance case methodologies under investigation through the program.

3. Major Partners

Key partners for 2025 efforts were established based on prior collaborations initiated in 2022 via TTCP AISC and included the Naval Information Warfare Center Pacific as well as the U.S. Naval Air Warfare Center Weapons Division (NAWCWD). Both organizations provided access to subject matter experts in assurance methodologies and design practices for safety-critical AI/ML systems, which in turn have helped establish and scope research tasks conducted in the 2025 program efforts at ARL.

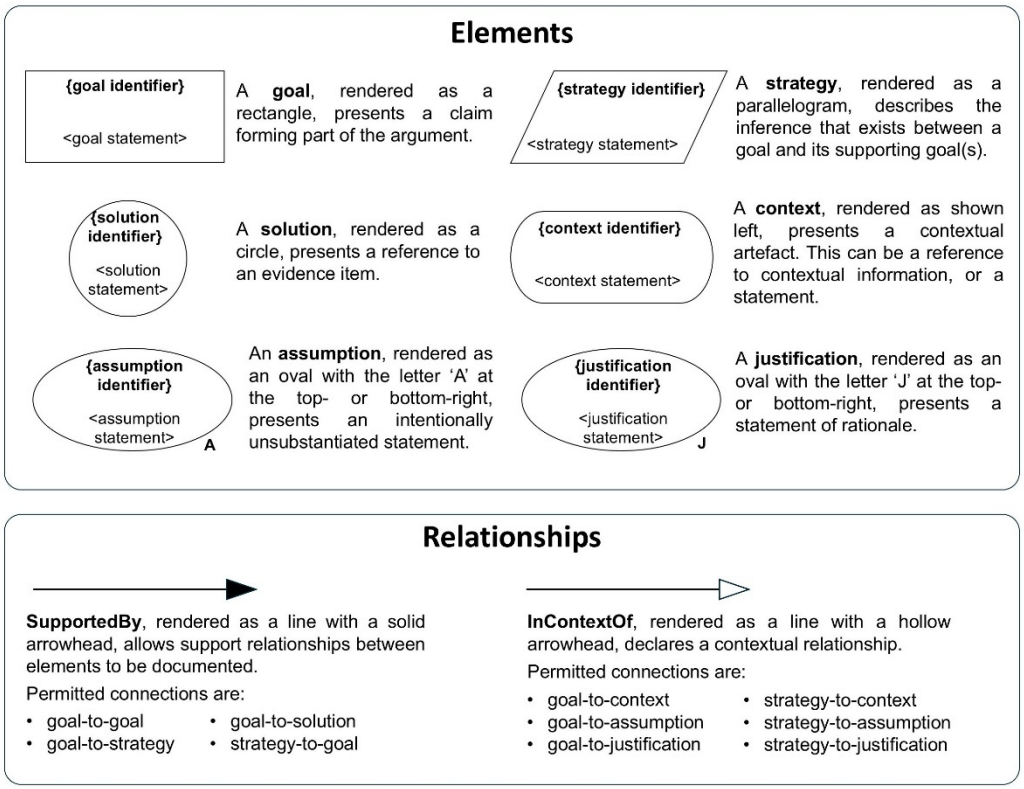

Figure 1. GSN for assurance argument structuring.

4. Current State of the Art

Program activities conducted in 2025 aimed to build upon prior research conducted outside ARL in the following three areas:

- **RMFs:** intended to catalog design and usage factors of AI/ML technology required to both conduct risk assessments and establish trustworthiness.
- **Assurance Case Frameworks:** intended to support definition of assurance cases for both AI/ML technologies and autonomous platforms.
- **TEVV Frameworks:** intended to both outline TEVV requirements for AI/ML technologies and capture relevant information to establish AI/ML technology compliance.

RMFs, such as the NIST RMF,[6] catalog factors required to establish AI/ML trustworthiness. For trustworthiness of an AI/ML technology to be established within the NIST RMF, the technology must be safe, valid and reliable, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed. Program activities in 2025 were focused primarily on investigating assurance methodologies to address safety

aspects of trustworthiness with respect to Army technology and operating requirements.

Assurance Case Frameworks, which are built upon work conducted in the safety-critical systems community, have included the following approaches developed in academic settings:

- The assurance of machine learning in autonomous systems (AMLAS)[8] methodology for assurance case definition over ML-based technologies.
- The safety assurance of autonomous systems in complex environments (SACE)[9] methodology for assurance case definition over autonomous systems.

Although AMLAS and SACE have been previously demonstrated in academic settings for systems such as supervised learning components of driverless vehicles, their applicability beyond these use cases, specifically to the Army system with AI/ML components, has previously remained unestablished. The 2025 efforts have aimed to assess whether these methodologies could be applicable to establishing assurance over Army and FVEY AI/ML technologies.

TEVV frameworks, intended to both outline TEVV requirements for AI/ML technologies and capture relevant information to establish AI/ML technology compliance, include the following externally developed efforts:

- The Levels of Rigor (LoR) methodology,[10] developed by the NAWCWD, which defines sets of testing requirements for ML technologies intended to establish varying levels of operational readiness.
- Google Model Cards,[11] which capture details on ML model training and testing procedures, as well as intended use cases for the models.

Program efforts, conducted in coordination with collaborators from the TTCP AISC, have aimed to assess whether these methodologies could be applicable to generation of evidence in support of assurance cases for Army and FVEY AI/ML technologies.

5. Research Activities and Engagements

To advance the defined 2025 program objectives (Section 2), a series of corresponding activities and engagements were conducted.

As part of ARL engagements in the TTCP AISC, we surveyed a set of existing methodologies for assurance case definition and corresponding evidence gathering aimed at supporting claims of AI/ML safety. Existing methodologies that were evaluated included the following:

- The AMLAS[8] methodology for assurance case definition over ML-based autonomous systems.
- The LoR methodology,[10] developed by NAWCWD, which defines sets of testing requirements for ML-based systems intended to provide evidence in support of assurance cases.
- Google Model Cards,[11] which capture details on ML model training and testing procedures, as well as intended use cases for the models.

This survey was conducted to assess current practices in ML technology documentation under investigation outside ARL, with the intent of informing best practices for assurance case management in Army AI/ML systems.

In parallel, a set of engagements were initiated with research teams in the ARL Military Information Sciences (MIS) and Science of Intelligent Systems (SIS) divisions to obtain information on AI/ML technologies under development, which are intended to support assessment of the assurance case methodologies in the context of Army S&T needs. Examples of ARL technologies under investigation during the 2025 efforts included the following:

- **Multi-agent Coordination in Deep Reinforcement Learning[4]:** This effort aims to formulate novel models at the intersection of Multi-Agent Reinforcement Learning, foundation models, and graph-based learning methods for improved coordinated decision-making and course-of-action generation.
- **Archangel[3]:** A supervised learning system intended to support human pose identification via imagery analysis, with the intent of supporting unauthorized vehicle (UAV)-based identification of ground personnel in need of assistance.
- **Hierarchical Vehicle Classification:** This effort leveraged supervised learning to support classification of vehicles using imagery gathered from UAV platforms, with the intent of identifying potential enemy targets to be intercepted.

To support engagement with the listed AI/ML teams, a technology questionnaire (see Appendix A for a full copy of the questionnaire) was defined and distributed to gather the following information:

- Factors that correspond to the AI/ML technology design, test, and evaluation requirements.
- Intended use cases for the AI/ML technology.
- Potential risks posed by the AI/ML technology during operational use.
- Associated risk mitigations strategies applicable to the AI/ML technology.

As of this report, a review of the listed ARL AI/ML technologies remains underway to support qualitative assessment of the previously listed assurance case and TEVV support frameworks for assurance case definition and management.

In addition, to facilitate encoding of GSN-based assurance cases, an ontology was defined to support encoding of argument structures as well as linking of arguments to supporting evidence. In turn, this ontology was applied toward the definition of templates to support encoding of assurance cases corresponding to Army AI/ML technologies. An overview of the ontology design is provided in Appendix B.

6. Key Findings

Based on outcomes from the 2025 program activities, key findings were observed in two areas. First, qualitative assessments of the assurance support methodologies identified in the TTCP AISC, performed in the context of cataloged ARL AI/ML technologies, have revealed a set of capability gaps in AMLAS and LoR with respect to meeting the following Army AI/ML assurance needs: 1) lack of support for assuring reinforcement learning-based systems; 2) limited capability to assure AI/ML technologies requiring continuous model retraining; and 3) limited support for assuring AI/ML technologies featuring interactions between multiple ML models.

Second, initial findings from the ARL AI/ML technology review indicated that AI/ML technologies at different Technology Readiness Level (TRL) stages will require distinct processes to establish safety-based assurance, with lower TRL technologies potentially requiring supplemental testing and risk assessment activities.

7. Recommendations

Based on the initial program findings established through 2025 research activities and engagements, the following recommendations are made for subsequent program efforts:

1) Upon completion of technology questionnaire responses by the ARL AI/ML development teams identified in the technology cataloging activity, identify and document corresponding requirements to enable safety-based assurance of each system. Apply these identified requirements toward the definition of assurance case management guidelines.
2) Continue to identify and catalog AI/ML technologies under investigation at ARL, focusing on systems featuring emerging technologies including large language models.

3) Apply identified guidelines for assurance case management toward the corresponding definition of an assurance management framework for Army AI/ML technologies, which are intended to support the cataloged AI/ML technologies and should be extensible based on assessments of additional AI/ML technologies.
4) Build upon the designed assurance management framework and continue work toward design and prototyping of services for enabling automated verification of assurance arguments for AI/ML technologies against real-time conditions in the operating environment.

8. References

1. Feickert A. 2025 Army Transformation Initiative (ATI) force structure and organizational proposals: background and issues for Congress. 2025 July. CRS Report No.: R48606. https://www.congress.gov/crs-product/R48606
2. Adipudi V et al. Accelerating robotics test and evaluation with a streamlined simulation process. Proceedings of Combined Symposium on Mechanical Systems and Robotics; Springer; c2024. p. 15–23.
3. Shen Y et al. Archangel: a hybrid UAV-based human detection benchmark with position and pose metadata. IEEE Access. 2023;11(2023):80958–80972.
4. Sur I et al. Data-driven distributed common operational picture from heterogeneous platforms using multi-agent reinforcement learning. Proceedings of 29th International Command and Control Research and Technology Symposium (ICCRTS 2024); 2024 Nov. London, UK.
5. Werner BD et al. Roadmap development to reduce risk associated with the deployment of artificial intelligence enabled systems. 2023 Annual Reliability and Maintainability Symposium (RAMS); 2023 Jan; Orlando, FL. IEEE. p. 1–6
6. NIST AI 100-1. Artificial intelligence risk management framework (AI RMF). National Institute of Standards and Technology (NIST); 2023; Technical Report No.: NIST.AI.100-1. https://doi.org/10.6028/NIST.AI.100-1
7. Goal Structuring Notation Community Standard - Version 3. Assurance Case Working Group (ACWG). Report No.: SCSC-141C. https://scsc.uk/scsc-141c
8. Hawkins R et al. Guidance on the assurance of machine learning in autonomous systems (AMLAS). arXiv; preprint 2021. arXiv:2102.01564. https://arxiv.org/abs/2102.01564
9. Hawkins R et al. Guidance on the safety assurance of autonomous systems in complex environments (SACE). arXiv; preprint 2022. arXiv:2208.00853. https://arxiv.org/abs/2208.00853
10. Nagy B. Level of rigor for artificial intelligence development. Naval Air Warfare Center Weapons Division (US); 2022. Report No.: NAWCWD TP 8864.
11. Mitchell M et al. Model cards for model reporting. Proceedings of the 2019 ACM Conference on Fairness, Accountability, and Transparency; 2019 Jan; Atlanta, GA. p. 220–229.

Appendix A. ML Technology Questionnaire

Building upon the methodology survey conducted via engagements in The Technical Cooperation Program (TTCP) AI Strategic Challenge (AISC), a technology questionnaire was defined and distributed to AI/ML development teams. This survey consisted of seven sections, each corresponding to different information categories previously identified as necessary to support assurance case definition and management for AI/ML technologies.

SECTION 1: General Information

FOCUS: Information on project points-of-contact as well as available resources (e.g., publications and code repositories).

QUESTIONS:

- What is the name of the Machine Learning (ML) project?
- Who is the technical point of contact for the project?
- Where can more information about the project be found?

SECTION 2: Information on System

FOCUS: Details on system in which the AI/ML component is to be incorporated, covering system use cases, components, and operational workflow.

QUESTIONS:

- Summarize at least one use case for the system.
- List any ML components required to support the given use cases.
- List additional system components required to support the use cases.
- List the steps taken by the system to support the use cases, referencing the given ML components and additional system components as needed.

SECTION 3: Operating Environment

FOCUS: Details on the intended operating environment for the system.

QUESTIONS:

- Indicate whether the operating environment for the system described in Section 2 is real-world, digital, or hybrid.
- Summarize the conditions under which the system will be expected to operate.

SECTION 4: Information on ML Component

FOCUS: Details on the ML component, including architecture type and performance metrics.

QUESTIONS:

- Indicate whether the ML approach is based on Supervised Learning, Reinforcement Learning, or an alternate approach.
- Summarize the ML model used.
- Summarize one or more performance metrics for assessing model performance against the identified use cases.

SECTION 5: ML Component Risks and Safety Requirements

FOCUS: Summary of potential risks the model poses during usage, as well as safety requirements intended to minimize risks.

QUESTIONS:

- List any potential risks posed by failure of the ML component described in Section 4 to performance of the broader system described in Section 2.
- For each potential risk, list one or more corresponding ML safety requirements identified to support risk mitigation.

SECTION 6: Model Development and Testing

FOCUS: Covers details on conditions applied to model development and testing.

QUESTIONS (For Supervised Learning Technologies):

- Summarize steps taken to obtain and process datasets used in model training and testing. These steps may cover data collection activities, including 1) real-world data collections, 2) procedures for generation of synthetic data, and 3) obtaining third-party datasets.
- Was each ML safety requirement identified in Section 5 evaluated during model testing?

QUESTIONS (For Reinforcement Learning Technologies):

- Summarize steps taken to obtain and process datasets used in model development.
- Summarize any safety constraints applied to the model aligned to ML safety requirements identified in Section 5.
- Summarize testing or analysis steps conducted to establish that the defined constraints are followed by the model.

SECTION 7: Model Integration with System

FOCUS: Summary of testing activities applied to track ML component compatibility with system as well as monitor and mitigate erroneous behavior.

QUESTIONS:

- Has compatibility of the model with the system been verified? Provide a summary of any necessary testing or analysis steps conducted to establish compatibility.
- Is the system described in Section 2 tested for erroneous behavior once the ML component(s) are integrated?
- If erroneous system behavior occurs, what mitigation steps are taken to address it?
- Are there any conditions in which the model needs to be updated after integration into the system?

Appendix B. Assurance Case Ontology

To facilitate encoding of Goal Structuring Notation (GSN)–based assurance cases, an ontology was defined to support encoding of argument structures as well as linking of arguments to supporting evidence.

Assurance cases were defined through the ontology center on four foundational classes: Elements, Statements, Identifiers, and Groups. Definitions for each foundational class are provided in Table B-1.

Table B-1. Foundational classes of GSN ontology.

Name	Description
Element	Encompasses the set of non-relationship components defined within the GSN standard (see Figure 1 in the body of the paper).
Statement	Statement describing associated Element instance, which can be textual or reference external information sources (e.g., datasets and log files).
Identifier	Unique identifier for each Element instance within an assurance case.
Group	Defines a grouping of one or more Identifiers intended to support grouping of subsections in an assurance case.

In turn, Table B-2 lists a series of subclasses of the Element class, which are established based on definitions in the GSN standard[1]:

Table B-2. Element subclasses of GSN ontology.

Name	Description
Strategy	Describes the nature of the inference that exists between a goal and its supporting goal(s).
Context	Presents a contextual artefact, which can either be a reference to contextual information or a statement.
Solution	Presents a reference to an evidence item or items.
Justification	Presents a statement of rationale.
Statement	Statement describing associated Element class, which can be textual or reference external information sources (e.g., datasets and log files).
Assumption	Presents an intentionally unsubstantiated statement.
Goal	Presents a claim forming part of the argument.

[1] Goal Structuring Notation Community Standard - Version 3. Assurance Case Working Group (ACWG). Report No.: SCSC-141C. https://scsc.uk/scsc-141c

In turn, the following properties are additionally defined to support establishment of relationships between Element instances, which in turn are based upon the GSN standard[1] (Table B-3):

Table B-3. Properties of GSN ontology.

Name	Domain / range	Description
hasGroup	Identifier / Group	Links unique Identifier to an identifier grouping.
hasIdentifier	Element / Identifier	Links Element instance to a unique Identifier.
isUninstantiated	Element / *Boolean* (true/false)	Indicates whether node is instantiated in GSN argument. Takes a Boolean value (true/false).
hasStatement	Element / *String*	Links GSN Element to corresponding statement.
supportedBy	Element / Element	Allows inferential or evidential relationships to be documented. Permitted connections are *goal-to-goal*, *goal-to-strategy*, *goal-to-solution*, and *strategy-to-goal*.
inContextOf	Element / Element	Declares a contextual relationship. Permitted connections are *goal-to-context*, *goal-to-assumption*, *goal-to-justification*, *strategy-to-context*, *strategy-to-assumption*, and *strategy-to-justification*.

List of Symbols, Abbreviations, and Acronyms

AI	Artificial Intelligence
AISC	AI Strategic Challenge
AMLAS	Assurance of Machine Learning for Use in Autonomous Systems
ARL	Army Research Laboratory
DEVCOM	U.S. Army Combat Capabilities Development Command
FVEY	Five Eyes (Australia, Canada, New Zealand, the United Kingdom, and the United States)
FY	Fiscal Year
GSN	Goal Structuring Notation
LoR	Levels of Rigor
MIS	Military Information Science
ML	Machine Learning
NAWCWD	Naval Air Warfare Center Weapons Division
NIST	National Institute of Standards and Technology
NIWC	Naval Information Warfare Center
RMF	Risk Management Framework
SACE	Safety Assurance of Autonomous Systems in Complex Environments
SIS	Science of Intelligent Systems
S&T	Science and Technology
TEVV	Test, Evaluation, Validation, Verification
TRL	Technology Readiness Level
TTCP	The Technical Cooperation Program
UAV	Unmanned Aerial Vehicle

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CI TECH LIB