INSTITUTE FOR DEFENSE ANALYSES

Artificial Intelligence & Autonomy Test & Evaluation Roadmap Goals

Heather Wojton, Project Leader

Brian Vickers
Daniel Porter
Rachel Haga

July 2021

Approved for Public Release.
Distribution Unlimited.

IDA Document NS D-22750

Log: H 2021-000268

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

Approved for public release; distribution is unlimited.

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Task BD-9-2299(90), "Test Science Applications," for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The IDA Technical Review Committee was chaired by Mr. Robert R. Soule and consisted of Dr. Matthew Avery, Dr. Jonathan Bell, Dr. Han Yi, and Dr. John Haman from the Operational Evaluation Division, and Mr. Brian Williams from the Joint Advanced Warfighting Division.

For more information:
Heather Wojton, Project Leader
hwojton@ida.org • (703)845-2482

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-22750

Artificial Intelligence & Autonomy Test & Evaluation Roadmap Goals

Heather Wojton, Project Leader

Brian Vickers
Daniel Porter
Rachel Haga

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Executive Summary

As the Department of Defense acquires new systems with artificial intelligence (AI) and autonomous (AI&A) capabilities, the test and evaluation (T&E) community will need to adapt to the challenges that these novel technologies present. As part of an effort to confront the AI&A systems now being developed, a prior document, "Test & Evaluation of AI-Enabled and Autonomous Systems: A Literature Review," extracted challenges and solutions that have been documented in other communities. This product continues that effort by focusing on department-relevant solutions that the T&E community needs to achieve in order to successfully execute tests of AI&A systems in the future.

The goals listed in this AI Roadmap address the broad range of tasks that the T&E community will need to achieve in order to properly test, evaluate, verify, and validate AI-enabled and autonomous systems. It includes issues that are unique to AI and autonomous systems, as well as legacy T&E shortcomings that will be compounded by newer technologies.

The seven goals are as follows:

1. Make progress toward solutions for unsolved AIES T&E challenges.
2. Integrate T&E throughout the entire life cycle of AI-enabled systems.
3. Improve and develop new test methods so that T&E results adequately characterize the performance, risk, and uncertainty of AIES for stakeholders.
4. Institute systemic processes and common architectures where they will facilitate cooperation and coordination over the testing life cycle.
5. Determine and invest in the necessary infrastructure, data tools, and other common solutions required to rigorously and efficiently characterize AIES performance.
6. Advocate for and invest in a workforce that is mission-ready at the intersection of AI, operational realities, and T&E.
7. Establish long-term collaborations among DOD, other government entities, industry, and academia to efficiently pursue solutions to shared T&E challenges.

Approved for public release; distribution is unlimited.

We aim for this document to serve as a starting point for various stakeholders in the acquisitions community to have a common understanding of the issues. Once all of the goals are agreed upon, stakeholders can be identified to clarify courses of action, draft implementation plans, and begin taking the steps necessary to ensure that we can continue rigorous planning for tests, collect appropriate data, and use valid statistical methods for evaluations.

T&E helps reduce program risk and inform decision-makers of performance, suitability, and other performance parameters across the operational envelope it will perform against in the field. AI&A technologies challenge our capabilities to accomplish this mission. The T&E community should act now to modernize our approaches for AI&A by acting on these AI&A Roadmap Goals.

IDA

Artificial Intelligence & Autonomy Test & Evaluation Roadmap Goals

Heather Wojton, Brian Vickers, Daniel Porter, Rachel Haga

July 2021

Institute for Defense Analyses
4850 Mark Center Drive ● Alexandria, Virginia 22311-1882

Approved for public release; distribution is unlimited.

BLUF: In this brief we are seeking agreement on goals.
We are not discussing detailed COAs/recommendations/policy here

We want to get to concrete steps overcoming T&E challenges for AI&A systems.

- Example COA: Update Title X and DOD 5000 series for AI&A systems.
- Example participants: DOT&E, USD(R&E), USD(A&S), USD(P)


Not discussing at this level today.

First, we need agreement on the goals to accomplish.

- Example: Integrate T&E throughout the entire lifecycle of AI&A systems.

AI&A: Artificial Intelligence & Autonomy

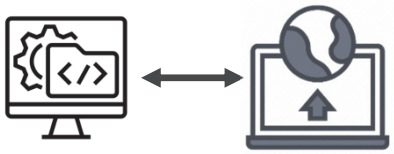
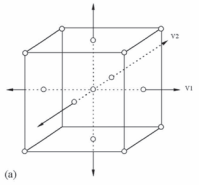
Approved for public release; distribution is unlimited.

AI Roadmap Goals cover the full range of T&E topics

A. Path Selection

Select Path Forward on Unresolved Challenges

Selecting paths creates goals in B-level buckets; will be completed in parallel with pursuing existing B-level goals.

B1. Lifecycle	B2. Methods	B3. Coordination
		
Shift T&E Left and Right	Overcome Methodological Challenges	Streamline Testing Across Stakeholders
B4. Infrastructure	**B5. Workforce**	**B6. Collaboration**
		
Modernize T&E Tools, Facilities	Mission-Ready at Intersection of AI, T&E, Ops	Best Practices from Industry, Gov., Academia

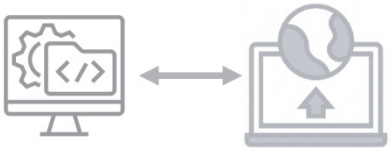
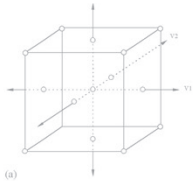
Approved for public release; distribution is unlimited.

AI Roadmap Goals cover the full range of T&E topics

A. Path Selection

Select Path Forward on Unresolved Challenges

Selecting paths creates goals in B-level buckets; will be completed in parallel with pursuing existing B-level goals.

B1. Lifecycle	B2. Methods	B3. Coordination
 Shift T&E Left and Right	 Overcome Methodological Challenges	 Streamline Testing Across Stakeholders
B4. Infrastructure	**B5. Workforce**	**B6. Collaboration**
 Modernize T&E Tools, Facilities	 Mission-Ready at Intersection of AI, T&E, Ops	 Best Practices from Industry, Gov., Academia

Approved for public release; distribution is unlimited.

TESTERS NEED TO

Have rigorous, robust T&E approaches that are agreed upon across the community.

AI&A PROBLEM(s)

Currently, there is no consensus across the T&E community on how to solve new issues arising for AI&A, such as learning systems, emergence, adversarial AI, M&S strategy, and RAI

AI&A: Artificial intelligence & autonomy

RAI: Responsible AI

Approved for public release; distribution is unlimited.

1. Gain consensus and make progress toward solutions for unsolved AI&A T&E challenges.

- Determine how to get sufficient coverage of the expansive and ill-defined operational spaces of AI&A systems within realistic budgets and timelines.
- Determine a strategy for recertifying systems that continue to update their behavior and "learn" after fielding.
- Agree on a strategy for M&S environment development that will allow DOD to test emergent behaviors, including with humans.
- Establish a framework for adversarial (behavioral, algorithmic exploitation, TTPs), cyber, and EMSO testing and determine the vulnerability, research, and staffing investments that are necessary.
- Resolve how T&E can support RAI efforts.

AI&A: Artificial intelligence & autonomy
EMSO: Electromagnetic Spectrum Operations
M&S: Modeling and Simulation
TTPs: Tactics, techniques, and procedures
T&E: Test & Evaluation

Approved for public release; distribution is unlimited.

TESTERS NEED TO

Test against requirements, operational considerations.
Carry out effectiveness, suitability, and survivability evaluations and make recommendations about fielded performance.

AI&A PROBLEM(s)

REQUIREMENTS: Systems might not have testable requirements that define mission success.
ACQUISITION: AI&A components on different acq. pathways may mature at different rates.
SUSTAINMENT: Systems may "learn" after certification/deployment; fielded system changes.

AI&A: Artificial intelligence & autonomy

Approved for public release; distribution is unlimited.

2. Integrate T&E throughout the entire life cycle of AI&A systems.

- Take steps toward an adaptive, cyclic continuum of contractor, developmental, and operational testing, including HSI, RAM, and robustness & resilience.
- Develop processes that enable programs to follow data best practices for curation, training, robustness, security, sharing, and other evolving needs.
- Shift testing left: Integrate T&E perspectives into analysis of alternatives, requirements setting, contracting, developmental processes, and shared events.
- Shift testing right: Incorporate T&E as an assumed component of post-fielding activities for AI&A.

AI&A: Artificial intelligence & autonomy
RAM: Reliability, availability, & maintainability
HSI: Human-Systems Integration
T&E: Test & Evaluation

Approved for public release; distribution is unlimited.

TESTERS NEED TO

Design and execute scoped, efficient, implementable system tests.
Make inferences from test performance to fielded performance.
Characterize risk for decision-makers in reliable, robust ways.

AI&A PROBLEM(s)

Testers can't explore as many operational states
as they can with legacy systems.
Also, the causal factors driving effectiveness, suitability, and survivability
during test are less clear than with traditional systems.

AI&A: Artificial intelligence & autonomy

Approved for public release; distribution is unlimited.

3. Improve and develop new test methods so that T&E results adequately characterize performance and uncertainty for operational evaluations and other stakeholder needs.

- Adapt and develop methods in areas where existing methods for selecting test factors and metrics do not meet mission needs.
- Improve test point selection approaches that enable efficient and useful tests about AI&A system performance.
- Create novel or update existing analysis techniques where current methods are insufficient for the size and/or complexity of AI&A system data.
- Outline practices for clearly communicating the effectiveness, suitability, and survivability of AI&A systems under different conditions in order to support acquisition decisions and operational employment.
- Develop methods and standards for ensuring that a system is sufficiently safe for its current level of maturity.
- Productize the research and lessons learned to improve the ease and quality of T&E across the community.
- Advance and develop the methods that allow for effective evaluation of human-systems integration with AI&A systems.

AI&A: Artificial intelligence & autonomy

Approved for public release; distribution is unlimited.

TESTERS NEED TO

Plan and execute tests of systems.
Execute tests between components and systems to learn about integration, possible emergent behaviors.

AI&A PROBLEM(s)

Proprietary technologies and classification requirements make test execution difficult if not impossible.
Unclear when, where, and how emergent behaviors will manifest in integration tests, system-system tests, or elsewhere.
Custom-built modules may not reach acceptable levels of transparency, traceability, extensibility, etc.

AI&A: Artificial intelligence & autonomy

Approved for public release; distribution is unlimited.

4. Institute systemic processes and common architectures that facilitate cooperation and coordination among testers, programs, and stakeholders.

- Establish a security clearance process for testers to reduce barriers to rapid and adaptive joint system testing.
- Create tools and processes to enable the tracking and reuse of data and evidence as programs move across the continuum of testing.
- Agree on common safety standards for AI&A systems.
- Prevent proprietary barriers from hindering the testing of individual and multiple systems.
- Establish a standard joint responsibility and authority for testing emergent behavior between agents.
- Encourage architecting AI&A systems using common frameworks and modular subsystems that increase predictability, testability, traceability, extensibility, and governability.
- Identify areas where Joint Offices (or similar entities) can be established to prioritize efforts, reduce redundancy, and streamline support of operational needs.

AI&A: Artificial intelligence & autonomy

Approved for public release; distribution is unlimited.

TESTERS NEED TO

Generate cross-domain, reusable solutions whenever possible.
Leverage necessary levels of M&S data across operational space to supplement planned levels of live data.

AI&A PROBLEM(s)

Infrastructure, test beds, [program] instrumentation may be inadequate for new capabilities – e.g., VV&A of M&S, traceability, novel surrogate threats.

AI&A: Artificial intelligence & autonomy

M&S: Modeling and simulation

VV&A: Verification, validation, and accreditation

Approved for public release; distribution is unlimited.

5. Identify and invest in the needed infrastructure, data tools, and other common solutions required to rigorously and efficiently characterize AI&A performance.

- Ensure that AI&A decision processes are instrumented to produce the data necessary for T&E; e.g., cognitive instrumentation and "explainers."
- Create common data pipelines and networks to take heterogenous system data, then collate, transfer, store, and redistribute to stakeholders while maintaining the security and integrity of the data.
- Upgrade and invest in new test assets and ranges to ensure that AI&A systems face sufficiently realistic threats and common operational challenges during testing.
- Invest in digital modernization; e.g., reusable digital test beds for common mission environments.
- Develop test harnesses for automated testing that can be scaled and tailored for recurring T&E applications.
- Invest in developing the VV&A tools, techniques, and standards to determine whether simulation and testing environments are sufficiently realistic for AI&A systems.
- Invest in infrastructure and tools that help ensure the safety of AI&A systems in both traditional and continuous test paradigms.

AI&A: Artificial intelligence & autonomy

VV&A: Verification, validation, & accreditation

Approved for public release; distribution is unlimited.

TESTERS NEED TO

Have a workforce that can effectively engage with varied T&E work for AI programs.

AI&A PROBLEM(s)

Widespread need for AI skillsets across industry, government, and academia makes recruiting and retention difficult because of supply and demand.

AI&A: Artificial intelligence & autonomy

Approved for public release; distribution is unlimited.

6. Advocate for and invest in a workforce that is mission-ready at the intersection of AI, operational realities, and T&E.

- Increase the number of applicants with AI-relevant skillsets to the T&E workforce.
- Improve retention of technical skillsets that are highly desired by industry.
- Increase the technical literacy and upskill the current T&E workforce.
- Certify and track AI talent to enable the necessary T&E teams, integrated product teams, and other teams.
- Instill a culture of robust testing, ethical approaches to AI, and realism about AI capabilities throughout all levels of the T&E workforce.

AI: Artificial Intelligence

T&E: Test & Evaluation

Approved for public release; distribution is unlimited.

TESTERS NEED TO

Work within DOD and with other government entities, industry, academia, and international partners to learn about and adopt best practices as they become available.

AI&A PROBLEM(s)

Standards and established communication lines do not yet exist and best practices are rapidly developing and updating for T&E of AI&A.

AI&A: Artificial intelligence & autonomy

Approved for public release; distribution is unlimited.

7. Establish long-term collaborations among DOD, other government entities, industry, academia, and allies to efficiently pursue solutions to shared T&E challenges.

- Advocate for establishing a National Lab for AI, in line with the NSCAI Final Report recommendation.
- Create channels for knowledge transfer between government entities.
- Increase investment in both basic and applied research for AI&A T&E.
- Foster international partnerships with allied T&E communities.

AI&A: Artificial intelligence & autonomy
T&E: Test & Evaluation
NSCAI: National Security Commission on Artificial Intelligence

Approved for public release; distribution is unlimited.

Recap of Top-Level Roadmap Recommendations

1. Gain consensus and make progress toward solutions for unsolved AI&A T&E challenges.
2. Integrate T&E throughout the entire life cycle of AI&A
3. Improve and develop new test methods so that T&E results adequately characterize performance and uncertainty for operational evaluations and other stakeholder needs.
4. Institute systemic processes and common architectures where they facilitate cooperation and coordination among testers, program offices, and stakeholders.
5. Determine and invest in the needed infrastructure, data tools, and other common solutions required to rigorously and efficiently characterize AI&A performance.
6. Advocate for and invest in a workforce that is mission-ready at the intersection of AI, operational realities, and T&E.
7. Establish long-term collaborations among DOD, other government entities, industry, academia, and allies to efficiently pursue solutions to shared T&E challenges.

Approved for public release; distribution is unlimited.

NEXT STEPS: IDA will itemize COAs, participants

- Shift testing right: Incorporate T&E as an assumed component of post-fielding activities for AI&A.
  - COA1: Establish a joint, executive-level working group to address barriers to implementing T&E in the field
    DOT&E | USD(R&E) | JAIC | Services
  - COA2: Set standards and/or triggers for AI&A systems to determine the need for additional testing and recertification of fielded systems.
    DOT&E | JAIC | Service Test Orgs.
  - COA3: Identify candidate programs for pilot testing of a more continual post-fielding model of T&E.
    DOT&E | USD(R&E) | JAIC | Services
  - COA4: Identify or develop technologies and processes that would allow for automated data collection from fielded units.
    DOT&E | USD(R&E) | JAIC | Service Test Orgs. | Service Labs
  - COA5: Develop a testing and deployment strategy for graded autonomy moving from limited to full AI&A capabilities.
    DOT&E | JAIC

COA: Course of action

Approved for public release; distribution is unlimited.

Questions?

Brian Vickers, bvickers@ida.org
Daniel Porter, dporter@ida.org

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
08-07-2021	Final	November 2020 – July 2021

4. TITLE AND SUBTITLE

Artificial Intelligence & Autonomy Test & Evaluation Roadmap Goals

5a. CONTRACT NUMBER

HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Brian D. Vickers; Daniel J. Porter; Rachel A. Haga; Heather M. Wojton

5d. PROJECT NUMBER

BD-09-2299

5e. TASK NUMBER

2299(90)

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER

NS-D-22750

H 2021-000268

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Director, Operational Test and Evaluation
1700 Defense Pentagon
Washington, DC 20301

10. SPONSOR/MONITOR'S ACRONYM(S)

DOT&E

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release. Distribution Unlimited.

13. SUPPLEMENTARY NOTES

Project Leader: Heather Wojton

14. ABSTRACT

As the Department of Defense (DoD) acquires new systems with increased artificial intelligence (AI) & autonomous (AI&A) capabilities, the test and evaluation (T&E) community will need to adapt to the challenges that accompany these new technologies. In order to overcome hurdles to T&E and assure that DoD is fielding trustworthy systems we have created the AI&A Roadmap Goals to share with stakeholders. This set of goals encompasses issues across the full program lifecycle, identifies methodological challenges, looks at coordination and infrastructure improvements, and highlights open challenges that require agreement on path(s) forward across the T&E community. This document illustrates both the scope and depth of the challenges facing the T&E community for AI&A systems, and acts as a call to action for all relevant stakeholders to work together in pursuit of solutions adequate for the T&E of AIES.

15. SUBJECT TERMS

artificial intelligence (AI); autonomy; Roadmap; test & evaluation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Heather Wojton
Unclassified	Unclassified	Unclassified	Unlimited	29	19b. TELEPHONE NUMBER *(Include area code)* (703) 845-6811