REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
06/15/2023	Conference Briefing	

4. TITLE AND SUBTITLE

DTE&A: Systems Engineering Processes to Test AI Right DTE&A Update to the Industrial Committee on Test and Evaluation (ICOTE)

5a. CONTRACT NUMBER

W56KGU-18-D-0004

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Carol L Pomales; Jim D Lockett

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102

8. PERFORMING ORGANIZATION REPORT NUMBER

PRS-23-2085

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

Office of the Director for Developmental Test, Evaluation, and Assessments

10. SPONSOR/MONITOR'S ACRONYM(S)

OSD DTE&A

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

12. DISTRIBUTION/AVAILABILITY STATEMENT

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

13. SUPPLEMENTARY NOTES

Submission to ICOTE -Industrial Committee on Test and Evaluation

14. ABSTRACT

Differences in AIES T&E Tasks from Conventional Software. The SEPTAR paper was developed by a MITRE team supporting Developmental Test Evaluation and Analysis (DTE&A) with input from DOD Stakeholders and Academia • It presents the benefits for proactive planning for Test and Evaluation (T&E) activities for AIES.

15. SUBJECT TERMS

Systems Engineering (General); Artificial Intelligence, SEPTAR; systems engineering; AI

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		9	Susan Carpenito
					19b. TELEPHONE NUMBER *(Include area code)* 781-271-7646

DTE&A: Systems Engineering Processes to Test AI Right

DTE&A Update to the Industrial Committee on Test and Evaluation (ICOTE)

OSD DTE&A MITRE Support Team

26 June 2023

Sponsor: OUSD DTE&A
Project No.: 101074.23.401.D320.P04

McLean, VA

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD

Differences in AIES T&E Tasks from Conventional Software

The challenges and differences between conventional software versus AIES testing include:

- AIES are inherently complex where possible outcomes are vast; some AI methodologies are probabilistic or have a level of uncertainty associated with results
- Conventional software is expected to produce the exact right answer for every possible input. For AIES, the requirement is to produce answers that are close enough, often enough, over a wide enough range of inputs where "enough" can vary
- Variation in AIES outputs can result from both changes in training and testing data and code whereas in conventional software the output variations result only from code changes
- Data can have bias, which can become embedded in several ways within an AIES
- Traditional software quality models do not align neatly to AIES. While an AIES may provide a correct answer, it may give same or different results from the same inputs

We must re-think test coverage to revise our aperture for completeness of testing over the continuum of the lifecycle

SEPTAR: Objective

- The DoD is making sizable investments ($14.7 billion) in Artificial Intelligence (AI) R&D and acquiring AI through programs
- The SEPTAR paper was developed by a MITRE team supporting Developmental Test Evaluation and Analysis (DTE&A) with input from DOD Stakeholders and Academia
- It presents the benefits for proactive planning for Test and Evaluation (T&E) activities for AIES
- SEPTAR recommendations were determined to deliver the following goals:
  - Ensure AIES are more likely to be delivered on time
  - Meet budgetary goals,
  - Perform effectively to meet mission expectations
- Assuring and understanding the processes used to build the AIES informs on the later T&E


MITRE

MTR230044 MITRE TECHNICAL REPORT

Systems Engineering Processes to Test AI Right (SEPTAR)

DRAFT

Sponsor: OUSD DTE&A
Project No.: 101074.23.401.D320.P04

The views, opinions and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official government position, policy, or decision, unless designated by other documentation.

This document was prepared for authorized distribution only. It has not been approved for public release.

McLean, VA

Authors:

Carlos Balhana
Ivy Chen
Dr. Ronald W. Ferguson
Jim Lockett
Dr. Danny Moore
Carol Pomales
Dr. Flo Reeder

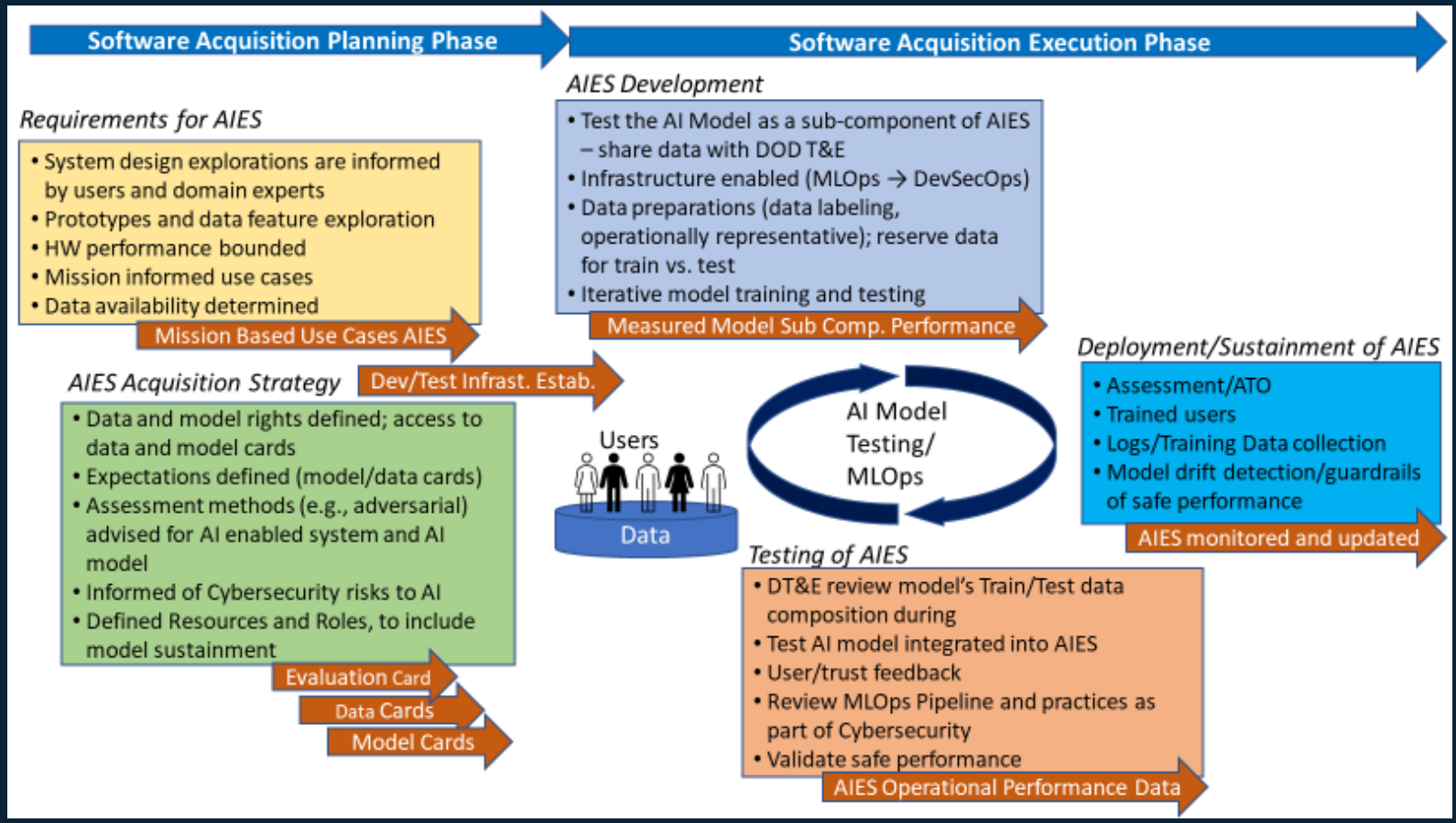
April 2023

SE Process to Test AI Right (SEPTAR)

Broadening the T&E continuum promote the paradigm of both "shifting left *and* right" for effective performance implementation and management of an AIES.

Defining data needs for AIES up front T&E personnel can avert potential adverse SELC outcomes by assessing whether current data practices do indeed bear positively on system development, model performance, and operational requirements.

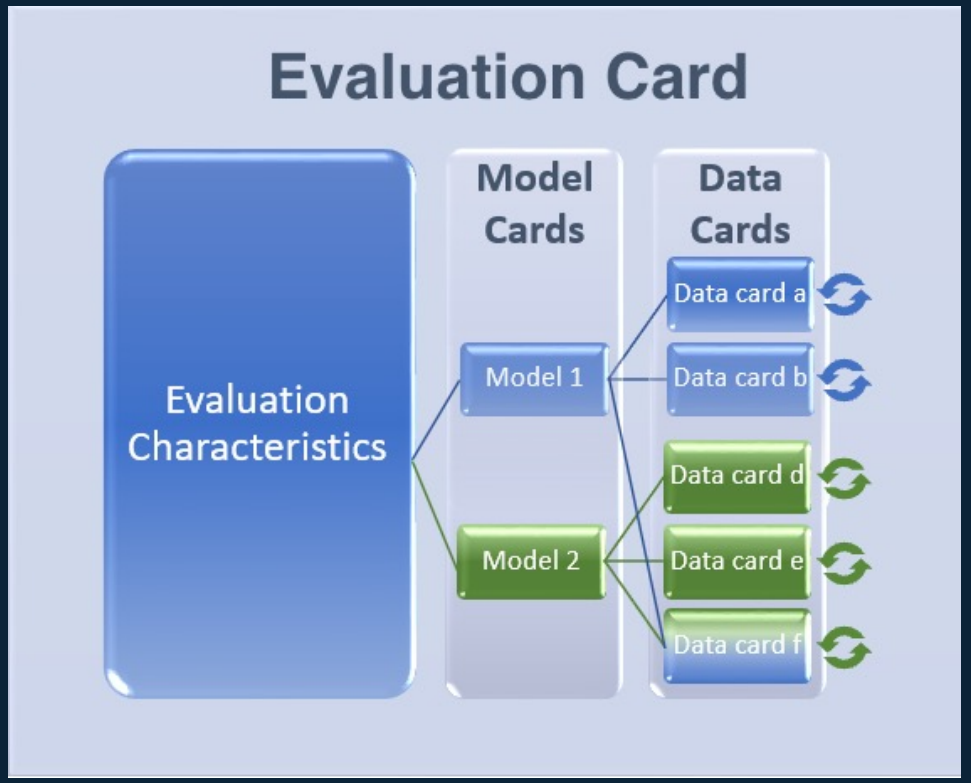
Leveraging insights on AIES project execution to inform the T&E of an AIES system Assessing the process that was followed to design, train and integrate the AIES can be used to assess inform the necessary scope for the overall AIES.



MITRE

The Cards

- Model Cards define the model: description of the model and usage, owner/license, measures of quality/performance, data used to train the model, intended users, limitations (including ethical)
- Data Cards come from the data owner/provider. It defines the data: source, collection, licenses, sanitization, labeling
- An AI model's evaluation card documents the methods and conditions under which the AI models will be evaluated and tracks results of these AI-subcomponent level assessments. The evaluation card approach ensures a consistent and well understood T&E methodology is used.
- These cards should begin during the earlier planning phases (requirements, acquisition strategy) and be updated frequently throughout the SDLC.
- Data and model cards must be reviewed and verified by the T&E team throughout the SDLC process and any associated evaluation cards must be updated to reflect any changes made to the data and model cards over time.



MITRE

How can Industry help enable effective T&E of AIES?

- Enable DOD to put agreements in place to ensure AIES performance is understood in the most effective ways
- Enable the Data (key to AI/ML) by developing data enabling capabilities (validation operationally representative, synthetic data, data labeling, data safeguarding)
- Design for testability - designing AIES so that we can automate aspects of the testing process
- Enable the collection of independent performance data on model training and testing in order to reduce the formal testing costs and make solutions easier for DOD to rapidly deploy
- Harden and protect your MLOPs pipeline to reduce cybersecurity risks
- For AIES to be sustainable, we need to have the ability to capture operationally relevant data for retraining as missions and environment change; industry perspective and enablement on how we close this feedback loop
- Provide feedback on SEPTAR content and emerging DOD Guidance and Policy collaboratively produced by DOT&E, DTE&A, CDAO

MITRE

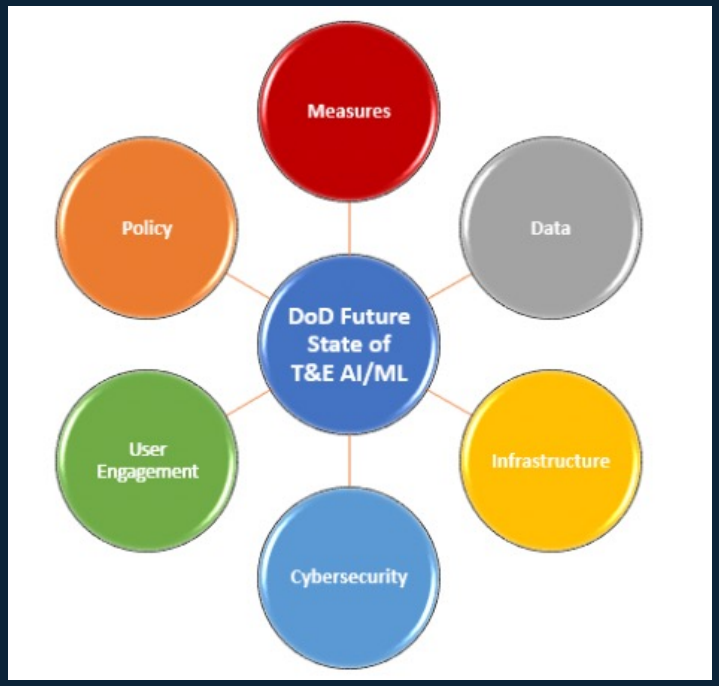
BACKUP

Projected ‘Future State’ for the T&E of AIES

Policy
- Need to mature practices and drive practitioners by establishing guidance and policy

User Engagement
- Need to standardize HMT practices through requirements, design and test; ensuring appropriate and informed trust

Cybersecurity
- All stakeholders must be informed and address AIES Cybersecurity risks
- Tools and methods to conduct threat informed T&E need to emerge



Measures
- Standard metrics aligned to mission are established and integrated to assess AI as a subcomponent during Development and DT&E.

Data
- Is a critical resource that must be planned, designed (operationally representative) and allocated (Model Train/Test.) to inform T&E

Infrastructure
- MLOps tools are a key enabler to produce independent test evidence and protect the AIES during vulnerable phases

Artificial Intelligence (AI) is a critical technology that the DoD must use to meet certain mission needs to keep pace with its near-peer adversaries. The Undersecretary for Research and Engineering cited “trusted AI” as among the top research priorities.* DOD must define holistic T&E methods for AI models and AI enabled systems that assure trust. Coordinated efforts are required to achieve this future state.

“AI, Networks, Hypersonics Are the Pentagon's Top Research Priorities,” Tirpak, John A., Air Force Magazine, Jan 2022

Roles across the SELC to Enable the T&E of AIES

		DOD PMO							DOD: Independent Test			Contractor		
	Tasks	Contract Officer	Functional Sponsor	User	Program Manager	Data Engineer	Dev Ops Engineer	Cyber Security	Dev Test Engr	Cyber-Security Engineer	OTA	Lead Developer	Data Engr/ CTR Lead	Lead Test
Requirements	AI Prototyping and Exploration	I	R	C	A	C	C	C	C	I	I			
	Requirements/ Use Case Development		R	C	A	C	C	C	C	C	C	I	I	I
Contracting	Ensures Contract Language includes AIES considerations	R	I	C	A	C	C	C	C	C	C			
	Collect, Assess, Label, Prep, Split, Curate Model Training Data	I	C	C	A	R*	C	C	C	C	I	I	R*	C
	Model Dev/Test Infrastructure	I	I	C	A	C	R*	C	C	I	I	R*	C	I
	Model Training/Test infrastructure	I	I	C	A	R*	C	C	C	I	I	I	R*	I
Development	Evaluation Card	I	C	C	A	C		C	R	C	C	C	C	C
	Model Cards	I	I	I	A	C		C	C	C	I	C	R	C
	Data Cards	I	I	C	A	R*		C	C	C	C	C	R*	C
	AI Model Testing	I	C	C	A	C		C	R	C	C	C	C	C
Testing	AIES Testing (and Model Training Process Review)	I	I	C	A	C		C	R (DT&E)	R (CVPA / AA)	R (OT&E)	C	C	R (Unit)
Deployment and Sustainment	Deployed System, Model Monitoring	I	I	C	A	R*		C	I	I		I	R*	C
	Deployed System, Model retraining and redeployment	I	I	C	A	R*		C	C	C	I	C	R*	C

* Indicates a decision must be made if the activity is DoD or Contractor Led, and/or any split of responsibilities should be clearly defined.

Responsible (R): Does the work to complete the task. **Accountable (A):** Delegates work and performs final review of completed deliverable. **Consulted (C):** Provides review and consultation. **Informed (I):** Kept informed of project process.

MITRE