

REPORT DOCUMENTATION PAGE					<i>Form Approved</i> OMB No. 0704-0188				
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.									
1. REPORT DATE (DD-MM-YYYY) 06/14/2022		2. REPORT TYPE Technical Report			3. DATES COVERED (From - To)				
4. TITLE AND SUBTITLE DoD Enterprise Test & Evaluation (T&E) of Artificial Intelligence (AI) Enabled Systems Strategic Goals				5a. CONTRACT NUMBER W56KGU-18-D-0004/W56KGU20F0009					
				5b. GRANT NUMBER					
				5c. PROGRAM ELEMENT NUMBER					
6. AUTHOR(S)				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-22-4191				
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of the Director for Developmental Test, Evaluation, and Assessments, Office of the Undersecretary of Defense					10. SPONSOR/MONITOR'S ACRONYM(S) OUSD 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									
14. ABSTRACT This document complements "The Future State of Test and Evaluation of Artificial Intelligence-Enabled Systems in the Department of Defense" by providing a set of strategic goals to achieve this future state. The goals address AI-enabled systems, not including autonomous systems. This document contains the strategic DoD Enterprise Test and Evaluation (T&E) of AI1 goals and a high level goals timeline developed in June 2022 by a group of DoD T&E of AI SMEs and representatives from: CDAO, USD(R&E)-DTE&A, DOT&E, TRMC, USD(R&E)-CT, AFOTEC, MITRE, IDA, CMU SEI.									
15. SUBJECT TERMS Artificial Intelligence; Systems Development; test; evaluation; Artificial Intelligence; AI									
16. SECURITY CLASSIFICATION OF: <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%; padding: 2px;">a. REPORT</td> <td style="width: 33%; padding: 2px;">b. ABSTRACT</td> <td style="width: 33%; padding: 2px;">c. THIS PAGE</td> </tr> </table>			a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 21	19a. NAME OF RESPONSIBLE PERSON Susan Carpenito	
a. REPORT	b. ABSTRACT	c. THIS PAGE							
			19b. TELEPHONE NUMBER (Include area code) 781-271-7646						



MITRE Technical Report (#MTR220583)

DoD Enterprise Test & Evaluation (T&E) of Artificial Intelligence (AI) Enabled Systems Strategic Goals

Sponsor: OUSD DTE&A
Project No.: 0721D320

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 approved for public release, case number 22-4191. Distribution unlimited.

©2023 The MITRE Corporation.
All rights reserved.

McLean, VA

Developed by DoD T&E of AI SMEs from:

**DTE&A
DOT&E
CDAO
TRMC
AFOTEC
USD(R&E)
MITRE
IDA
CMU SEI**

**Content Assembled by:
Sue Murray, Chuck Meier, Carol Pomales**

June 14 – 15, 2022

Table of Contents

1	Purpose of this Document.....	3
2	DoD Enterprise T&E of AI Goals Formation Approach.....	4
3	DoD Enterprise T&E of AI Gaps Identified by SMEs.....	5
4	Long-Term (5+ Year) Goals	9
5	Medium-Term (2 to 5 Year) Goals.....	10
6	Near-Term (1 to 1.5 Year) Goals.....	12
7	Goals High-Level Timelines (Swimlanes).....	15
8	Way Forward	20

1 Purpose of this Document

This document complements “The Future State of Test and Evaluation of Artificial Intelligence-Enabled Systems in the Department of Defense” by providing a set of strategic goals to achieve this future state. The goals address AI-enabled systems, not including autonomous systems.

This document contains the strategic DoD Enterprise Test and Evaluation (T&E) of AI¹ goals and a high-level goals timeline developed in June 2022 by a group of DoD T&E of AI SMEs and representatives from:

- Chief Digital and Artificial Intelligence Office (CDAO)
- Undersecretary of Defense for Research and Engineering – Developmental Test, Evaluation, and Assessments (USD(R&E)-DTE&A)
- Office of the Director of Operational Test and Evaluation (DOT&E)
- Undersecretary of Defense for Research and Engineering – Test Resource Management Center (TRMC)
- Undersecretary of Defense for Research and Engineering – Critical Technologies (USD(R&E)-CT)
- Air Force Operational Test and Evaluation Center (AFOTEC)
- MITRE Corporation
- Institute for Defense Analysis (IDA)
- Carnegie Mellon University Software Engineering Institute (CMU SEI)

This group formulated a set of strategic goals that consists of long-term goals (5+ years), medium-term goals (2 – 5 years), and near-term goals (1 to 1.5 years) to achieve the T&E of AI Future State, address supply and demand for T&E of AI (assessed in a DoD-wide data call in May to June 2022), and address gaps in DoD’s T&E of AI capabilities. Each near-term and medium-term goal contributes to one or more long-term goal. Some goals were additionally identified to address a specific gap or gaps that the group identified. The Gaps section of this document contains the list of major gaps, identified by the cross-DoD group, to be overcome to achieve the T&E of AI Future State.

This set of forward-looking goals and high-level goals timeline provide DoD a way ahead to improve the Department’s T&E of AI capabilities and achieve the desired DoD AI Future State. Together, these goals are instrumental in shaping a coherent and actionable way forward for T&E of AI for the Department.

The DoD Enterprise T&E of AI near-term goals are being used by DTE&A to develop an overall roadmap with activities, owners, and timelines to make progress in improving DoD T&E of AI capabilities and capacity. A near-term goals roadmap will be the start of paving the way forward towards achieving the long-term goals and the T&E of AI Future State.

¹ The term “T&E of AI” used in this document means T&E of AI-enabled systems. Where the term “AIES” is used, it means AI-enabled systems.

2 DoD Enterprise T&E of AI Goals Formation Approach

To formulate goals for improving DoD's capacity and capability in T&E of AI-enabled systems,, CDAO hosted a two-day in-person DoD Enterprise T&E of AI Goals Formation Workshop (Figure) with subject matter experts representing multiple DoD organizations involved in T&E of AI: CDAO, DTE&A, DOT&E, TRMC, AFOTEC, and USD(R&E). Partners from MITRE, IDA, and CMU SEI also participated in the goals formation.

The purpose of this workshop was to define broad, strategic goals and a high-level goals timeline for the DoD T&E of AI enterprise, rather than goals for individual DoD organizations involved in T&E of AI. Additional scoping and direction for the workshop called for all participants to:

- Take a DoD Enterprise T&E of AI perspective when defining goals, setting aside organizational boundaries.
- Formulate goal statements that are broad statements of aim or purpose to achieve the desired future state and to address gaps between supply and demand. Goals are written at a level above objectives, which include specific measures and targets.
- Focus goals formation on AI-enabled systems, not including autonomous systems. The CAPE Study report (done at the same time as this goals formation) includes data call findings about autonomous systems.
- Use six AI T&E domains—Data, Infrastructure, Cybersecurity, User Engagement, Policy, Measures—as frames for ideation and discussion, and synthesize up to overarching goals not constrained by these domains.
- Use the following timeframes when ideating goals:
 - Long-term: 5+ years
 - Medium-term: 2 – 5 years
 - Near-term: 1 – 1.5 years

In the DoD Enterprise T&E of AI Goals Formation Workshop, the cross-DoD AI T&E representatives did the following:

- Reviewed the DoD T&E of AI current state summary and refreshed the future state description.
- Ideated on a first pass set of long-term goals to achieve the future state (taking a top-down approach first).
- Reviewed and achieved common understanding of the DoD T&E of AI demand and supply data call results as of the workshop date.
- Identified major gaps and enablers to achieving the future state, as well as gaps between demand and supply.
- Ideated, grouped, and synthesized near-term, medium-term, and long-term goals within each of the six T&E of AI domains—Data, Infrastructure, Cybersecurity, User Engagement, Policy, Measures—and also identified Workforce as an additional domain needed.
- Developed the high-level goals timeline by bucketing goals into near-, medium-, and long-term.
- Reviewed and refined the whole goals timeline to consolidate similar ideas, identify “line-of-sight” connections from near- to long-term goals and from medium- to long-term goals, and organize the goals.
- Presented the high-level goals timeline to SESs from various DoD organizations in a wavetop review.

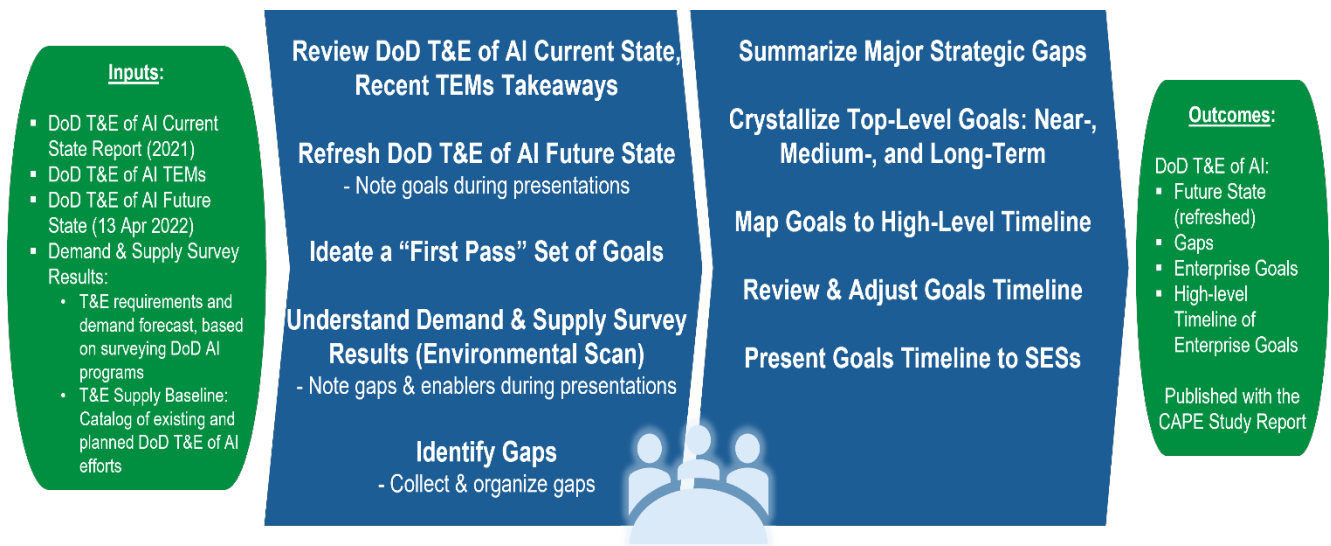


Figure 1: Enterprise Goals Formation Approach Taken by Cross-DoD T&E of AI Representatives

The DoD T&E of AI SMEs and representatives in the Goals Formation Workshop identified major gaps between the current state and the desired future state for T&E of AI. The group also successfully developed a comprehensive set of long-, medium-, and near-term DoD Enterprise T&E of AI goals, as well as a high-level goals timeline with relationships between the near- and long-term goals and between medium- and long-term goals. The following sections of this document provide these gaps, goals, and timeline.

3 DoD Enterprise T&E of AI Gaps Identified by SMEs

Gaps are holes that the DoD T&E of AI enterprise needs to fill to achieve the desired future state and can be either:

- Gaps between supply and demand or
- Gaps or obstacles between current state and future state.

In the DoD Enterprise T&E of AI Goals Formation Workshop, the cross-DoD group of T&E of AI subject matter experts identified gaps based on the group’s review of the current state, refresh of the future state, first-pass identification of long-term goals, and review of the demand and supply data call results.

The **strategic gap statements** developed by the cross-DoD T&E of AI representatives in the Goals Formation Workshop are listed below. The detailed gaps that support each strategic gap statement are listed underneath. In the workshop, the group was in consensus that Workforce should be added as a new T&E of AI domain; while related to the Policy and Guidance domain, building the needed T&E of AI

workforce and access to users goes beyond policy/guidance. The domain(s) for each gap are identified in brackets at the end of each gap statement.

1. Lack of sufficient workforce with AI literacy, knowledge, and skills within DoD, including operators and stakeholders at all levels (from leaders down).

- a. Need more in-house AI-qualified practitioners at both Program Management Offices (PMO) and Operational Test Agencies (OTAs). [Workforce]
- b. Need to increase DoD workforce literacy, knowledge, and skills in AI. Need more AI education and training at all levels of our organizations, from leaders through all levels of personnel. Need to increase general workforce awareness and understanding of what AI is, what AI can and cannot do, how AI can help in operational/mission processes, etc. [Workforce]
- c. Need dedicated T&E of AI staff to be able to move from current state to future state. There is a shortage of trained and experienced personnel in T&E of AI and a high demand for and shortage of AI experts. [Workforce, Infrastructure]
- d. Need users of AI-enabled systems for conducting both developmental test (DT) and operational test (OT). Lack of access to users when needed to test and evaluate AI. [Workforce, Policy/Guidance]
- e. Gap in current T&E professionals' ability to test AI. Need training on AI for the T&E workforce. Specific knowledge and skills are needed for T&E of AI compared to traditional systems T&E. [Workforce, Infrastructure]
- f. Need opportunity to gain exposure to use of enterprise T&E of AI tools that CDAO is building. [Workforce, Infrastructure]

2. Lack of sufficient, relevant, and accessible data, data management, and data rights.

- a. Lack of an enterprise-level system for collecting, storing, sharing, and accessing data for AI development. [Data]
- b. Need for metadata with all data. [Data]
- c. Lack of data transparency, accessibility, and reusability. [Data]
- d. Lack of data taxonomy and labelling standards. [Data]
- e. Need a more replicable, reliable processes to extract, reduce, and share system data with testers. PMOs need accessible data for system development. [Data, Infrastructure]
- f. Need AI hubs for shared developmental models and datasets. [Infrastructure]
- g. Lack of availability of needed data. [Data]
- h. Proprietary data and metadata; stove piped and restricted access to data. We have limited, siloed data. Title 10 authority to request data. [Data]
- i. Need synthetic data methods. [Data]

- 3. Lack of measures for AI model/AI-enabled system performance, mission effectiveness (performance of mission pre- and post-AI), and trustworthiness during program planning, design, development, test, and continuous evaluation and sustainment.**
 - a. Difficulty connecting mission effectiveness to an AI model. [Measures]
 - b. Lack of metrics/understanding for how well a human performs a mission, which would enable T&E to evaluate if AI performed the mission better. [Measures]
 - c. Lack of methods to connect mission effectiveness to AI model performance during development. [Measures]
 - d. Difficult to gauge at program design phase the AI benchmarks and abilities to support missions to needed extents. [Measures]
 - e. Difficult to articulate clear AI use cases. [Measures]
 - f. Lack of measures/metrics to inform AI trustworthiness. [Measures]
- 4. Cultural resistance/gap to doing human-machine teaming and to assessing effective use.**
 - a. Lack of human-machine teaming and human factors T&E for AI (methods, policies, etc.). [Infrastructure]
 - b. Resistance to having users assess effective use. [User Engagement]
- 5. Lack of policies that are needed for T&E of AI across the AI lifecycle.** (The items below are examples)
 - a. Lack of policy and guidance for continuous evaluation of AI across the whole lifecycle, to include continuous ongoing evaluation after fielding. Lack of policy clarity on AI assurance requirements throughout the lifecycle. This is gap/obstacle to moving from current to future state. [Policy/Guidance]
 - b. Lack of a formal mandate for assessment of AI capability experimentation activities by Services and reporting of assessment results to the DoD decision-makers (e.g., DOT&E). (This includes progression between TRL levels). [Policy/Guidance]
 - c. Integrated Test Collaboration for experimentation with and assessment of AI systems needs to expand their mandate with reporting to Capability Development decision-makers. [Policy/Guidance]
 - d. Inability to get software installed on PMO computer networks due to IT policies (the policy is the obstacle). [Policy/Guidance]
 - e. Lack of policies relating to privacy and confidentiality of T&E practices. [Policy/Guidance]
 - f. Lack of policy when AI-enabled systems/capabilities are added to legacy platforms, which leads to systems that might evade T&E. [Policy/Guidance]

6. Lack of modeling and simulation (virtual testing) environments that are suitable and available for T&E of AI.

- a. Need modeling and simulation (M & S) to enable AI. Need clear understanding of how to use M & S for T&E of AI. [Infrastructure]
- b. Need to tie digital engineering and model-based systems engineering (MBSE) of AI-enabled systems into other DoD systems engineering methods. [no Domain identified; Infrastructure is closest tie]
- c. Lack of clarity on where to invest in modeling and simulation (M & S) that can support program and/or T&E needs. [no Domain identified; Infrastructure is closest tie]

7. Range safety and range operations constraints limit the ability to test AI for safety assurance in a representative operational environment.

- a. Lack of formal T&E/assurance methods. [Policy/Guidance]
- b. Difficulty establishing safety assurance for AI systems on test “range.” [Domain not identified; Policy/Guidance is closest tie]
- c. Lack of a systems engineering approach to influence AI systems design during all stages of the AI lifecycle, including T&E. [Policy/Guidance]
- d. Need new kinds of safety arguments for AI-enabled systems. [Domain not identified]
- e. Need to build AI expertise for test planning vice AI capability development. Need to grow AI operators through the development of AI capabilities and send them to T&E of AI. [User Engagement]

8. Lack of standards, access, awareness, operationalization etc. of tools for T&E of AI.

- a. Lack of bias detection/quantification tools and techniques. [Infrastructure – tools]

9. Lack of adversarial and cyber assessments for AI-enabled systems.

- a. Lack of understanding of threats to AI. [Cybersecurity]
- b. Research and development (R&D) still needed on AI cybersecurity. [Cybersecurity]
- c. Lack of processes to test against adversarial cyber TTP adaptation, CCD, EMSO for AI. [Cybersecurity]
- d. Need cyber and resilience measures/metrics. [Cybersecurity]

10. Difficulty articulating AI use cases, on which thorough T&E of AI depends. Lack of ways to predict and test for pernicious emergent behavior.

- a. Difficulty articulating AI use cases in the DoD. T&E needs depend strongly on the AI use case. [Measures]
- b. Lack of ways to predict and test for pernicious emergent behavior. [Domain not identified; Measures and/or Cybersecurity are best ties]

11. Lack of clear DoD T&E of AI organizational roles and responsibilities and communication/awareness of these roles and responsibilities.

- a. Need clearly defined and assigned T&E of AI roles and missions across DoD organizations. T&E of AI lacks organizational roles and missions understanding. Who is responsible for what? [Policy/Guidance]
- b. Difficulty understanding what is occurring within and across the DoD Components with respect to AI. [Domain not identified; Policy/Guidance and Roles/Workforce are best ties]
- c. Lack of awareness of resources and agencies for both program offices and test agencies. [Domain not identified; Policy/Guidance and Roles/Workforce are best ties]
- d. Lack of awareness of AI capabilities in development that the T&E community needs to flight follow/influence and/or support via experimentation. [Situational Awareness; Data; Policy/Guidance is best tie]
- e. Need a strategy/method to scope and resource program test events with assurance cases. [Domain not identified; Infrastructure and Policy/Guidance are best ties]

12. Lack of common/standard T&E of AI processes and methodology across the AI lifecycle (acquisition, planning, design, development, test, sustainment).

- a. Need DoD contracting guidelines for AI development and testing [Policy/Guidance]
- b. Lack of clear description of how DoD T&E functions within the DevSecOps processes. [Domain not identified; Infrastructure and Policy/Guidance the closest ties]
- c. Need T&E of AI processes and methods incorporated into DoD Acquisition (e.g., Software Pathway). [Policy/Guidance]

4 Long-Term (5+ Year) Goals

The long-term goals are broad statements of the key actions the DoD Enterprise T&E of AI needs to accomplish in five years to achieve the desired T&E of AI Future State.

The long-term goals are:

- **Long-Term Goal 1 (LT1):** Develop a DoD T&E **workforce** that has knowledge of AI as a capability and is skilled in the tools, methods, and best practices for testing and evaluating AI models and AI-enabled systems.
- **Long-Term Goal 2 (LT2):** Provide sufficient, operationally relevant, usable **datasets** that DoD AI developers and T&E personnel can find and access to independently develop and test AI models and AI-enabled systems that are ultimately interoperable.
- **Long-Term Goal 3 (LT3):** Provide interoperable **tools** and reusable measures in repositories accessible across DoD to develop, train, test, evaluate, manage, and sustain AI models and AI-enabled systems across the AI lifecycle.

- **Long-Term Goal 4 (LT4):** Involve users and operators early and often throughout the AI lifecycle to ensure effective generation of AI use cases, requirements, T&E plans, AI-enabled work processes with human-machine teaming, and AI trustworthiness assurance methods, all resulting in suitable, verified, and validated AI capabilities and outcomes.
- **Long-Term Goal 5 (LT5):** Establish a T&E of AIES ecosystem across DoD that provides **integrated and interoperable infrastructure, processes, and practices** for development and T&E of AI-enabled systems to streamline evaluations of effectiveness, safety, suitability, cybersecurity, trustworthiness (responsible, fair/impartial, explainable, etc.) and improve test range capabilities.

5 Medium-Term (2 to 5 Year) Goals

The group of cross-DoD T&E of AI representatives defined a set of medium-term DoD Enterprise T&E of AI goals to achieve the Enterprise long-term goals and to bridge from the near-term to the long-term goals. Medium-term goals are efforts the DoD T&E of AI enterprise should aim to complete within two to five years.

The medium-term goals are listed below. Each individual medium-term goal's mapping to T&E of AI domain(s) and specific long-term goal(s) is noted in brackets at the end of each medium-term goal statement. Some medium-term goals map to more than one long-term goal and/or more than one T&E of AI domain.

In addition, in some T&E of AI domains the cross-DoD workshop group identified connections between near-term goals and medium-term goals. These near-term to medium-term connections indicate which goals must be accomplished first, or at least started first, to achieve the subsequent goal. Mappings from near-term goals to medium-term goals are noted in the brackets in instances where these mappings were identified in the workshop by the cross-DoD representatives.

In some cases, the DoD T&E of AI representatives defined a goal to address a specific gap(s). Where identified by the workshop group, mappings to gap(s) are also noted in the brackets after each goal statement. The Gaps section of this document contains the list of major gaps identified by the cross-DoD group to achieving the enterprise DoD T&E of AI Future State.

- **MT01: Reinforce clear roles, responsibilities, and authorities in the DoD** with respect to T&E of AI-enabled systems [Policy/Guidance, Workforce; maps to LT1; maps from NT1]
- **MT02: Stand up red teams that are trained in T&E of AI-enabled systems.** [Cybersecurity; maps to LT1, maps from NT4]
- **MT03: Resource, develop, and implement AI versions of software factories** (like Kessel Run) that can augment DevSecOps processes with capabilities and tools to enable AI. [Workforce, Infrastructure, Policy/Guidance; maps to LT1, LT3]
- **MT04: Break down classification barriers and improve processes to enable access to independent datasets** and information needed to develop, train, and test AI-enabled systems. [Infrastructure and Policy/Guidance; maps to LT2]

- **MT05:** Establish flexible guidelines to **improve data sharing across the DoD, government, academia, and the private sector**. [Data; maps to LT2; maps from NT8 and NT9]
- **MT06:** Engage senior DoD leaders to communicate and enforce policy and guidance that drives cultural and organizational changes to **improve cross-DoD sharing of data and tools** and other DoD-wide collaboration. [Policy/Guidance; maps to LT2 and LT3]
- **MT07:** Establish and **disseminate frameworks and practices for data** management, characterization, and taxonomies that aid AI developers and T&E personnel in tagging data and validating operational representation (reality) of data, increase transparency and reusability, and enable relevant test measures. [Data, Policy/Guidance; maps to LT2 and LT3; maps from NT11]
- **MT08:** Establish and **disseminate policy, infrastructure, and technology to collect data at the tactical edge** and to label, store, transport, and enable access by developers and T&E personnel across the enterprise (and where applicable, support development of synthetic data). [Data, Infrastructure, Policy/Guidance; maps to LT2, LT3, and LT5]
- **MT09:** **Expand and equip test range infrastructure** to accommodate T&E of AI-enabled systems, including T&E range safety and assurance use case development to enable more realistic and flexible testing of AI-enabled systems on DoD ranges. [Infrastructure; maps to LT3]
- **MT10:** **Provide the right tools to cybersecurity T&E of AIES teams** (e.g., red teams) to enable use of improved cybersecurity practices for T&E of AI-enabled systems, including capabilities to accurately represent AI threats for testing realism. (Examples: basic AI, fuzzing, and tool capabilities brought into one cyber-focused T&E event; AI vulnerability assessment framework such as ATLAS.) [Cybersecurity; maps to LT3; maps from NT15, NT16, NT4]
- **MT11:** Build, establish, and **disseminate interoperable T&E of AIES tools and AI model repository(ies) that shift T&E left** (upstream into planning and requirements definition), are straightforward to integrate across the toolset, are platform-agnostic, can be deployed rapidly, address trustworthiness, and can be used to evaluate AI-enabled systems post-fielding. [Infrastructure; maps to LT3 and LT5; maps from NT16]
- **MT12:** **Create a more rigorous, repeatable T&E of AIES assurance framework** and assurance cases that: explicitly address user (operator and commander) concerns; involve users; and use multiple sources of test evidence collection including open air T&E, modeling and simulation, human-machine teaming, and formal methods. [User Engagement; maps to LT4, LT5, and Gap 4]
- **MT13:** Develop, establish, and **disseminate repeatable processes and practices for measuring human-machine teaming and user interaction** with AI-enabled systems. [User Engagement; maps to LT4]
- **MT14:** Create and **provide end-to-end infrastructure with adequate storage and bandwidth** that enables/supports developers and T&E professionals across the entire AI lifecycle (including experimentation, prototyping, data preparation and tagging, development, T&E, configuration

management, protection, and ongoing evaluation and maintenance). [Infrastructure; maps to LT5]

- **MT15: Define and regularly update standards for interoperability** of data, models, and tests. [Policy/Guidance, Infrastructure; maps to LT5]
- **MT16: Make available a “playbook” of reusable exemplar measures** for T&E of AI models and AI-enabled systems (system performance and operation/mission performance), based on type of AI model and application, that can be tailored or augmented for specific operational T&E requirements and user trustworthiness throughout the full AI lifecycle and that can be integrated to support independent system test and sustainment monitoring. [Measures; maps to LT5; maps from NT17]
- **MT17: Inform Program Management Offices about relevant threats to AI** models and AI-enabled systems to drive choices and risk assessments. [Cybersecurity; maps to LT5; maps from MT21 and NT3]
- **MT18: Ensure safety policies provide flexibility and rigor** for AI development and T&E without generating unnecessary safety incidents. [Policy/Guidance, maps to LT5]
- **MT19: Define and disseminate exemplar requirements for cybersecurity in AI-enabled systems** for AI development and T&E and integrate these requirements into the Risk Management Framework (RMF) and DoD software acquisition. [Cybersecurity; maps to LT5; maps from NT5]
- **MT20: Craft effective T&E policy** that accounts for whether a given system uses “artificial intelligence.” [Policy/Guidance; maps to LT5]
- **MT21: Develop threats information and characterize cybersecurity threats distinctive to AI-enabled system types** (including distinction between ML and fixed-code approaches) to enable relevant, informed AI adoption. [Cybersecurity; maps to LT5, MT17, and Gap 9; maps from NT3]
- **MT22: Refine DoD AI development and T&E processes** used throughout the AI lifecycle and “certify” what the processes are used for. [Infrastructure; maps to LT5]

6 Near-Term (1 to 1.5 Year) Goals

The group of cross-DoD T&E of AI representatives defined a set of near-term enterprise goals to contribute to and set the stage for achievement of the long-term goals. Near-term goals are efforts the DoD AI T&E enterprise should aim to complete within one to 1.5 years.

The near-term goals are listed below. Each individual near-term goal’s mapping to T&E of AI domain(s) and specific long-term goal(s) is noted in brackets at the end of each near-term goal statement. Some near-term goals map to more than one long-term goal and/or more than one T&E of AI domain.

In addition, in some T&E of AI domains the cross-DoD workshop group identified connections between near-term goals and medium-term goals. These near-term to medium-term connections indicate which

goals must be accomplished first, or at least started first, to achieve the subsequent goal. Mappings from near-term goals to medium-term goals are noted in the brackets in instances where these mappings were identified in the workshop by the cross-DoD representatives.

In some cases, the DoD T&E of AI representatives defined a goal to address a specific gap(s). When it was identified by the workshop group, the goal's mappings to gap(s) are also noted in the brackets after each goal statement. The Gaps section of this document contains the list of major gaps identified by the cross-DoD group to achieving the enterprise DoD T&E of AI Future State.

- **NT01: Define and communicate clear organizational roles, responsibilities, authorities, and dependencies for T&E of AI-enabled systems across DoD.** As part of assigning roles and responsibilities, determine level of acceptable risk for DoD T&E of AI policies and identify organizational leadership responsible for assuming this risk. Communication, stakeholder engagement, and leadership mobilization of these roles and responsibilities are essential to making them enforceable and sustainable. [Policy/Guidance, Workforce. Maps to LT1 and MT1. Informs NT5, NT6, and NT7 (NT7: teams should do CONOPS, mission engineering, and HMT on AI efforts). Start in near-term and continue establishing and reinforcing in medium-term.]
- **NT02: Build organizational understanding (literacy) about AI as a capability and T&E of AI-enabled systems across the DoD T&E enterprise, from senior leaders throughout all staff levels.** This includes understanding of what AI can and cannot do, how AI can augment/support human work and mission operations, unique aspects of T&E for AI, etc. [Policy/Guidance; Workforce. Maps to LT1. Start in near-term and continue ongoing. Informed/supported by NT18 (Go-To Platform).]
- **NT03: Create communication methods to provide information about threats and vulnerabilities of AI-enabled systems to requirements development organizations (COCOMs, PMOs) and T&E organizations.** This includes threats of various types, such as cybersecurity and environmental threats. [Cybersecurity. Maps to LT1 and MT17. Informed/supported by NT18 (Go-To Platform). Informs/supplies NT18.]
- **NT04: Identify missing skillsets, tools, and capabilities of cyber teams for cyber testing of AI-enabled systems and capabilities.** [Cybersecurity. Maps to LT1 and MT2. Informs NT15.]
- **NT05: Establish and communicate processes, practices, and guidance to include the T&E of AI community in defining and scoping specifications/expectations/requirements for AI-enabled systems** (i.e., shift T&E to the left), to ensure all AI models and AI-enabled systems are designed, developed, and tested for their intended operational/mission use. [Policy/Guidance, Cybersecurity. Maps to LT1, LT4, LT5. Informed by NT1 and informed/supported by NT18 (Go-To Platform). Informs/supplies NT18.]
- **NT06: Create and communicate guidance and policy to facilitate involvement of end-users, operators, and units in testing and evaluation of AI models and AI-enabled systems.** [Policy/Guidance; User Engagement. Maps to LT1, LT4, LT5. Informed/supported by NT18 (Go-To Platform). Informs/supplies NT18.]
- **NT07: Create and disseminate guidance and practices for AI-enabled systems development teams to document and refine a CONOPS and AI-enabled work process(es) with human-**

machine teaming (HMT) from project start throughout AI-enabled systems development and maturation. [Policy/Guidance, User Engagement. Maps to LT1, LT4, and LT5. Informs NT18 (Go-To Platform) and MT13. Informed by NT1 and NT18.]

- **NT08: Obtain DoD stakeholders’ (Services, acquisition, operational commands, etc.) agreement that data should be discoverable and agreement on the controls necessary.** [Data. Maps to LT2 and MT5. Informs NT9.]
- **NT09: Increase access to and availability of high-quality training, validation, and test data, including synthetic data.** [Data. Maps to LT2. Informs MT5, MT7, MT11, and NT18. Informed by NT8, NT11, NT12, NT18, and NT19.]
- **NT10: Build data architecture to support tailorable tech stacks for AI pipelines** (data engineering, modeling, test, implementation, monitoring, and ongoing updates). [Data, Infrastructure. Maps to LT2 and MT7. Informed by NT12 (identify and fill connectivity gaps/needs).]
- **NT11: Define and develop frameworks and practices for data management, characterization, and taxonomies** that aid AI developers and T&E personnel in tagging data and validating operational representation (reality) of data, increase transparency and reusability, and enable relevant test measures. (Define the practices in near-term and establish/disseminate the practices in medium-term.) [Data. Maps to LT2, MT4, MT7, and MT8. Informs NT9 and NT18 (Go-To Platform). Informed/supported by NT18.]
- **NT12: Identify site connectivity needs/gaps and expand connectivity** between sites and platforms/systems for data sharing, modeling and simulation, user/operator participation, tools access, and other capabilities needed to conduct T&E of AI models and AI-enabled systems. [Infrastructure. Maps to LT2 and LT3. Informs NT9 (increase access to data).]
- **NT13: Provide increased visibility and access to T&E tools** for AI developers and T&E personnel, including all existing tools and industry-standard tools (e.g., FedRAMP and security guidance slow and limit tools), define tool requirements, and establish standards for what the tools are used for. Establish policies that enable DoD to acquire and use industry tools more easily and effectively. [Policy/Guidance. Maps to LT3. Informed by NT14 and NT15, and informed/supported by NT18 (Go-To Platform). Informs/supplies NT18.]
- **NT14: Integrate modeling and simulation (M&S) for T&E into DoD T&E of AI practices** early in the AI lifecycle and ensure LVCA (live, virtual, constructive) environments and M&S strategies account for varied program needs and technical limitations. [Infrastructure. Maps to LT3, MT11, Gap 6, and Gap 7. Informed/supported by NT18 (Go-To Platform). Informs NT13 (tools awareness) and NT16 (build/obtain tools) and informs/supplies NT18.]
- **NT15: Identify gaps, define requirements, and begin development of cybersecurity T&E tools, capabilities, and a repeatable process for AI-enabled systems**, informed by threats, across the whole lifecycle (start in near-term with pilots, continue to grow the tools in medium-term). [Cybersecurity. Maps to LT3 and MT10. Informed by NT3 (types of cyber threats) and NT4 (missing cyber tools/capabilities/skillsets). Informs NT16 (build/obtain tools).]

- **NT16: Identify T&E of AI-enabled systems tool gaps and requirements, research tools to fill gaps, prototype tools, and build and/or acquire tools and capabilities for T&E of AI-enabled systems** that are cloud-agnostic and can be used across a wide variety of environments including the data and AI development environments. [Cybersecurity, Infrastructure. Maps to LT3, LT5, and MT11. Informed/supported by NT18 (Go-To Platform). Informed by NT14 (M&S tools) and NT15 (cyber tools). Informs/supplies NT18.]
- **NT17: Define exemplar, reusable measures for T&E of AI models and AI-enabled systems** (system performance and operation/mission performance), based on type of AI model and application, that can be tailored or augmented for specific operational T&E requirements and user trustworthiness throughout the full AI lifecycle and that can be integrated to support independent system test and sustainment monitoring. Assess cross-domain, cross-mission suitability of the exemplar measures. [Measures. Maps to LT3, LT5, MT16, Gap 12. Form a working group/taskforce to do this in the near term. Informed/supported by NT18 (Go-To Platform). Informs/supplies NT18.]
- **NT18: Iteratively collect, create, share, and refine T&E of AI-enabled systems best practices, tools, guidance, policies, measures, learning opportunities, and other content specific to AI models and applications.** Achieving this will be facilitated by creating a “Go-To (one-stop-shop) T&E of AI Platform (repository) and DoD T&E of AI Stakeholder Forum to make best practices, guidance, tools, measures, courses, learning opportunities, threats/vulnerabilities information, and other reusable content readily available. (Start in near-term, continue in medium-term.) [Policy/Guidance. Maps to LT3 and LT5. Informed/supplied by content from NT3, NT5, NT6, NT7, NT9, NT11, NT13, NT14, NT16, NT17, NT20. Informs/supports NT3, NT5, NT6, NT7, NT9, NT11, NT13, NT14, NT16, NT17, NT20.]
- **NT19: Clarify data, model, and process characterization and reporting** across program types and lifecycles. This includes finding a standard way to communicate about AI and associated entities such as the data, models, and process (e.g., using data cards and model cards). “Process” refers to how AIES move through development and testing cycles (e.g., CI/CD pipelines, DevOps, etc.). [Policy/Guidance, Infrastructure, Data. Maps to LT3 and LT5. Informs NT9 (increase data availability) and NT11 (data characterization).]
- **NT20: Develop policies to advance interoperability between DoD T&E of AI-enabled systems personnel and model developers** to gain deeper insights on model behavior, including predictability and interpretability. [Policy/Guidance. Maps to LT5. Informed by NT1 (DoD T&E of AI ecosystem roles) and informed/supported by NT18 (Go-To Platform). Informs/supplies NT18.]

7 Goals High-Level Timelines (Swimlanes)

Based on the long-term, medium-term, and near-term goals and connections between these goals, the group of cross-DoD T&E of AI subject matter experts developed a high-level goals timeline mapping. The group mapped near-term goals and medium-term goals to long-term goals. All goals—near-term and medium-term—contribute to the achievement of the long-term goals. For DoD to achieve the desired future state, the DoD needs to take a coordinated, multi-faceted approach across all T&E domains that

starts in the immediate near-term. An ad-hoc, piece-meal approach to building DoD Enterprise T&E of AI capacity and capability will further encumber the effort to achieve the future state and will not serve the Department well.

The “line-of-sight” relationships between the near-term goals and medium-term goals to long-term goals are depicted in DoD Enterprise T&E of AI Goals High-Level Timeline (swimlane) diagrams, below. In these High-Level Timeline diagrams:

- Each long-term goal has its own swimlane that contains the contributing near-term goals and medium-term goals.
- In many cases, a near-term goal or a medium-term goal may contribute to more than one long-term goal. In these instances, the near-term goal or medium-term goal appears in each applicable long-term goal swimlane, and dashed lines indicate the multiple uses of that near-term goal or medium-term goal across multiple swimlanes.

There are six DoD Enterprise T&E of AI Goals High-Level Timelines:

- One **Overall DoD Enterprise T&E of AI Goals High-Level Timeline (Figure 2)**, showing all long-term, medium-term, and near-term goals together in one big picture, with the long-term goal swimlanes
- Five **Individual Long-Term Goal High-Level Timelines (Figure 3)**, each showing a long-term goal’s swimlane with the text for contributing near-term and medium-term goals

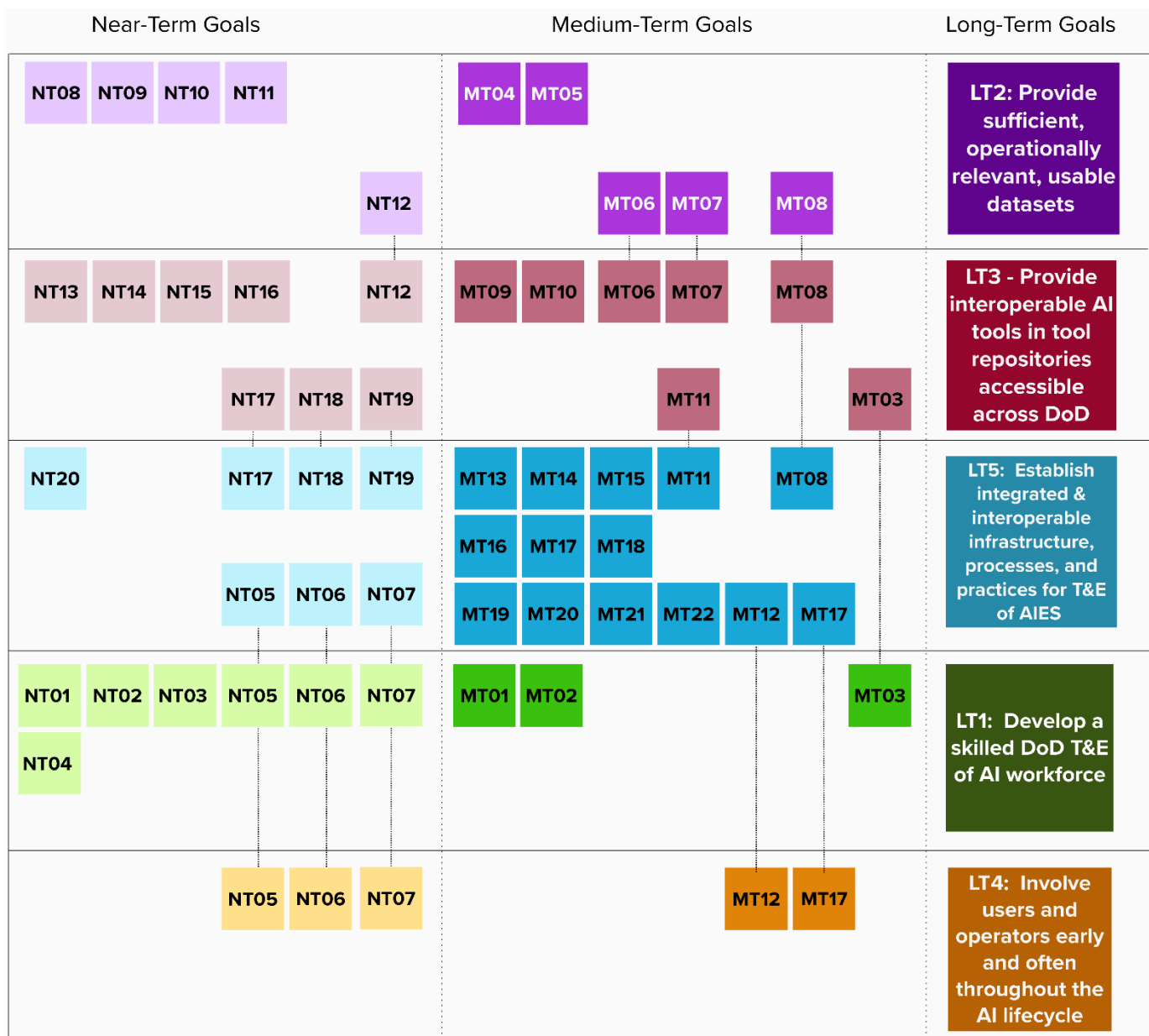


Figure 2: Overall DoD Enterprise T&E of AI Goals High-Level Timeline

Near-Term Goals			Medium-Term Goals			Long-Term Goal
NT01: Define and communicate clear roles, responsibilities, and authorities for T&E of AIES across DoD	NT02: Build organizational understanding (literacy) of AI as a capability & T&E of AI across the DoD T&E enterprise	NT03: Create comm. methods to provide info on threats & vulnerabilities of AIES to requirements & T&E organizations	MT01: Reinforce clear roles, responsibilities, and authorities in DoD with respect to T&E of AIES	MT02: Stand up red teams that are trained in T&E of AIES	MT03: Resource, develop, & implement AI versions of software factories to augment DevSecOps processes with capabilities and tools to enable AI	LT1: Develop a DoD T&E workforce that has knowledge of AI as a capability and is skilled in the tools, methods, and best practices for testing and evaluating AI models and AI-enabled systems
NT04: Identify missing skillsets, tools, & capabilities of cyber teams for cyber testing of AIES	NT05: Establish processes/practices to include the T&E of AI community in defining & scoping requirements for AIES	NT06: Create & communicate guidance/policy to facilitate involvement of end-users, operators, & units in T&E of AIES				
NT07: Create & disseminate guidance for AI dev teams to document a CONOPS & processes with human-machine teaming (HMT)						

Near-Term Goals			Medium-Term Goals			Long-Term Goal
NT08: Obtain DoD stakeholders' agreement that data should be discoverable, with corresponding controls	NT09: Increase availability of and access to high-quality training, validation, & test data, including synthetic data	NT10: Build data architecture to support tailorable tech stacks for AI pipelines	MT07: Establish & disseminate frameworks & practices for data management, characterization, & taxonomies that aid developers & T&E personnel in tagging data & validating operational representation of data, increase transparency & reusability, & enable relevant test measures	MT04: Break down classification barriers & improve processes to enable access to independent datasets & information to develop, train, & test AIES	MT05: Establish flexible guidelines to improve data sharing across DoD, government, academia, & the private sector	LT2: Provide sufficient, operationally relevant, usable datasets that DoD AI developers & T&E personnel can find and access to independently develop & test AI models & AI-enabled systems that are ultimately interoperable
NT11: Develop frameworks & practices for data management, characterization, & taxonomies to support T&E of AIES	NT12: Identify needs/ gaps & expand connectivity between sites & platforms/ systems for data sharing, tools, etc. for T&E of AIES			MT06: Enforce policy & guidance that drive cultural & organizational changes to improve cross-DoD sharing of data, tools, etc.	MT08: Establish & disseminate policy, infrastructure, & technology to collect data at the tactical edge & label, store, transport, and enable access by developers and T&E personnel across DoD	

Near-Term Goals			Medium-Term Goals			Long-Term Goal
NT12: Identify needs/ gaps & expand connectivity between sites & platforms/ systems for data sharing, tools, etc. for T&E of AIES	NT13: Provide increased visibility and access to T&E tools for AI developers and T&E personnel, define tool reqmts, & establish standards for what the tools are used for	NT14: Integrate modeling and simulation (M&S) for T&E into DoD T&E of AI practices early in the AI lifecycle & ensure LVCA environments & M&S strategies account for program needs & technical limitations	MT03: Resource, develop, & implement AI versions of software factories to augment DevSecOps processes with capabilities and tools to enable AI	MT06: Enforce policy & guidance that drive cultural & organizational changes to improve cross-DoD sharing of data, tools, etc.	MT07: Establish & disseminate frameworks & practices for data management, characterization, & taxonomies that aid developers & T&E personnel in tagging data & validating operational representation of data, increase transparency & reusability, & enable relevant test measures	LT3: Provide interoperable tools & reusable measures in repositories across DoD to develop, train, test, evaluate, manage, & sustain AI models and AI-enabled systems across the AI lifecycle
NT15: Identify gaps, define requirements, and begin development of cybersecurity T&E tools, capabilities, & repeatable process for AIES, informed by threats, across the whole lifecycle	NT16: Build & acquire cloud-agnostic T&E of AI tools & capabilities that can be used across a variety of environments, including the data and AI development environments	NT17: Define exemplar reusable measures for T&E of AI models & AIES that can be tailored or augmented for specific operational T&E requirements & user trustworthiness throughout the full AI lifecycle.	MT08: Establish & disseminate policy, infrastructure, & technology to collect data at the tactical edge & label, store, transport, and enable access by developers and T&E personnel across DoD	MT09: Expand & equip test range infrastructure to accommodate T&E of AIES, including T&E range safety & assurance use case development for more realistic testing of AI on DoD ranges		
NT18: Iteratively collect, create, share, & refine T&E best practices, tools, measures, guidance, policies, & other content specific to AI models & applications	NT19: Clarify data, model, and process characterization and reporting across program types and lifecycles.		MT10: Provide the right tools to enable use of improved cybersecurity practices for T&E of AIES	MT11: Build, establish, & disseminate interoperable T&E of AIES tools & AI model repository(ies) that shift T&E upstream, are platform-agnostic, can be used to evaluate AIES post-fielding, etc.		

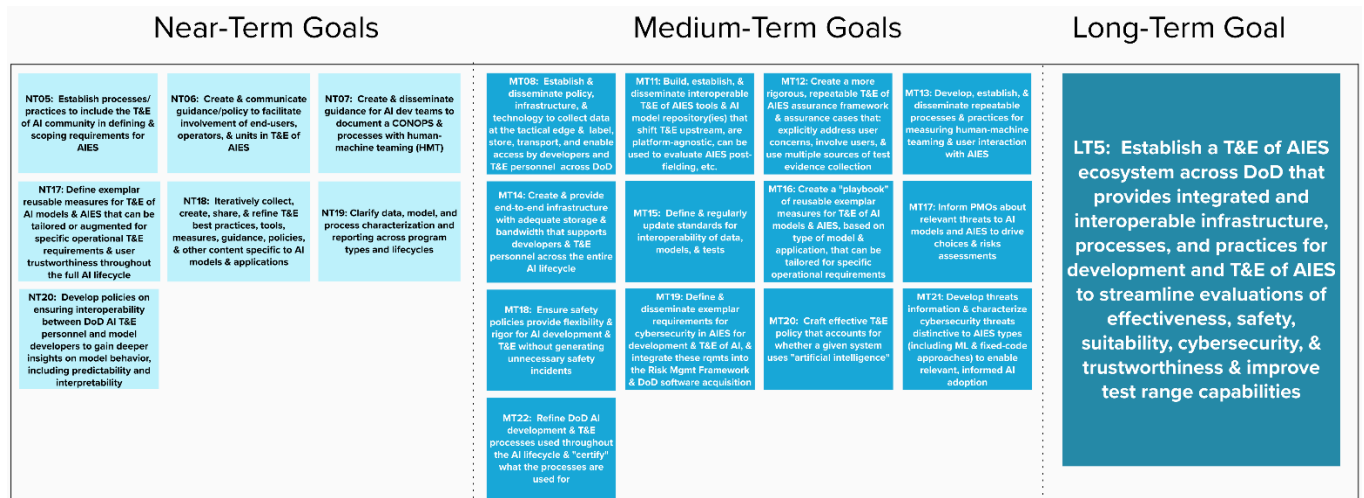
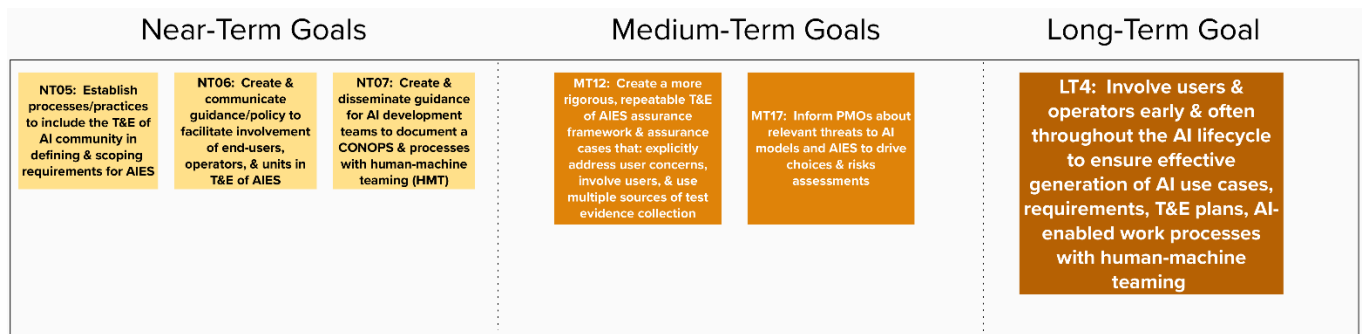


Figure 3: Individual Long-Term Goal High-Level Timelines (Swimlanes)

8 Way Forward

The set of forward-looking goals and high-level goals timeline developed by the DoD T&E of AI subject matter experts in the Enterprise Goals Formation Workshop provide DoD a way ahead to achieve the desired DoD T&E of AI Future State. The goals and timeline reflect an understanding of the current state, a refresh of the future state, data from the DoD-wide T&E of AI data call, identification of key gaps and enablers, and suggested courses of action from subject matter experts from the DoD T&E of AI community. Together, these goals are instrumental in shaping a coherent and actionable way forward for T&E of AI for the Department.

Some additional recommendations for DoD from the DoD T&E of AI SMEs and representatives in the Goals Formation Workshop include:

- Add “Workforce” as a T&E of AI domain, to be addressed alongside the other AI T&E domains.
- Build and increase a T&E-of-AI-ready workforce including AI developers, T&E personnel, users, and operators to participate in T&E, as well as their leaders, and support this workforce with T&E of AI T&E training, tools, and capabilities. This relates to long-term goal LT1 about workforce.
- Add and enrich existing cybersecurity processes and practices for AI in lieu of creating or reinventing a new AI cybersecurity pipeline.

Important next steps for the DoD T&E of AI enterprise to make the goals, and therefore the Future State, become reality include:

- **Assign an overall DoD Enterprise T&E of AI Goals “Steward”** with leadership authority to manage, oversee, track, and communicate these goals and their progress throughout the DoD T&E of AI community. Identify senior leader champions to provide visible senior leader support to attain the goals.
- **Establish a cross-DoD Enterprise T&E of AI Goals Working Group** to work with the Goals Steward and provide a unified, enterprise approach to assigning, working on, reporting on, and refreshing the enterprise goals. Key activities of the Goals Working Group include: develop a goals roadmap; track goals execution; help remove obstacles; communicate progress; engage senior leaders when needed; and periodically review and adjust the goals and roadmap.
- **Develop and communicate a DoD Enterprise T&E of AI Near-Term Goals Roadmap**, involving stakeholders from across the DoD T&E of AI ecosystem. A Near-Term Goals Roadmap would contain activities, owning and supporting organizations, products, dependencies, and timelines to achieve the near-term goals. Existing, currently underway, and planned activities and products should be leveraged to the extent possible. The cross-DoD Goals Working Group, with coordination and direction from the Goals Steward, will develop the Near-Term Goals Roadmap.
 - Start with a near-term goals roadmap to work on what can be done first.
 - Consider prioritizing the near-term goals to focus limited resources.
 - For each high priority near-term goal:
 - Identify owner and support organizations:
 - Owner/lead organization(s), preferably one organization and no more than two organizations, that would orchestrate and coordinate the activities to achieve the near-term goal.
 - Support organizations that would participate in and assist activities to achieve the near-term goal.

- Identify existing, currently in-progress, and future planned products and activities that contribute to achievement of the near-term goal. There may be several products that already exist and activities underway and scheduled for the future that help achieve the goal.
 - Identify the fiscal year and quarter in which each product/activity will be completed. Some products may already exist.
 - Identify additional products and activities needed, in addition to existing, in-progress, and future products and activities.
- Create a roadmap of the high-priority near-term goals. The High-Level Goals Timelines (swimlanes) developed by DoD T&E of AI representatives in the Goals Formation Workshop provide a starting point with line-of-sight from near-term goals to medium-term and long-term goals.
- Involve additional DoD T&E of AI stakeholders in development, review, and communication of the goals roadmap.
- **Ensure DoD senior leaders communicate the enterprise DoD T&E of AI goals and roadmap(s)** to stakeholders early and periodically to build awareness, understanding, and commitment to the goals, roadmap, and DoD organizational ownership of goals.
- **Build commitment among organizations involved in T&E of AI to work on achieving the goals.** Involve DoD senior leaders routinely to address any issues with goal assignments and to approve resources needed for owners to accomplish their goals.