CDAO

Chief Digital & Artificial Intelligence Office

# JATIC: Joint AI T&E Infrastructure Capability



**Controlled by:** CDAO, Assess & Assure Division
**Distribution/Dissemination Control:** FEDCON
**POC:** *david.jin5.civ@mail.mil*

# Program Background

- Funded by Congress as a Program of Record
- Started in 2023, funded through 2028
- An investment of nearly $200 MM for AI Test & Evaluation (T&E)

**Program objective:**
Develop software to **accelerate** and **enable** AI test & evaluation for DoD testers

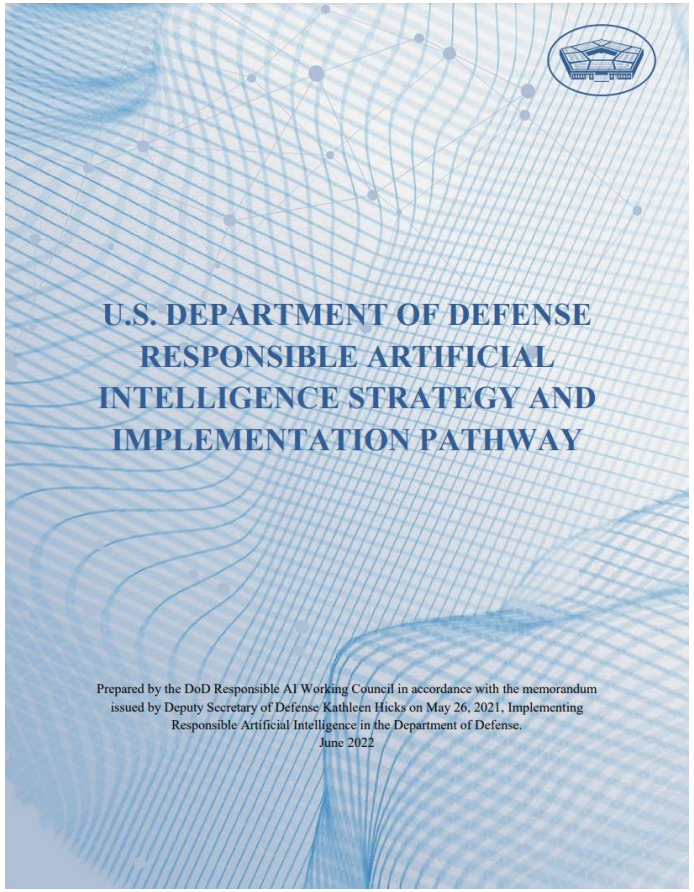
**In order to…**
Provide rigorous assurance of the effectiveness, robustness, and safety of the DoD's AI-enabled systems

CDAO

# Program Background

## RAI Strategy & Implementation Pathway

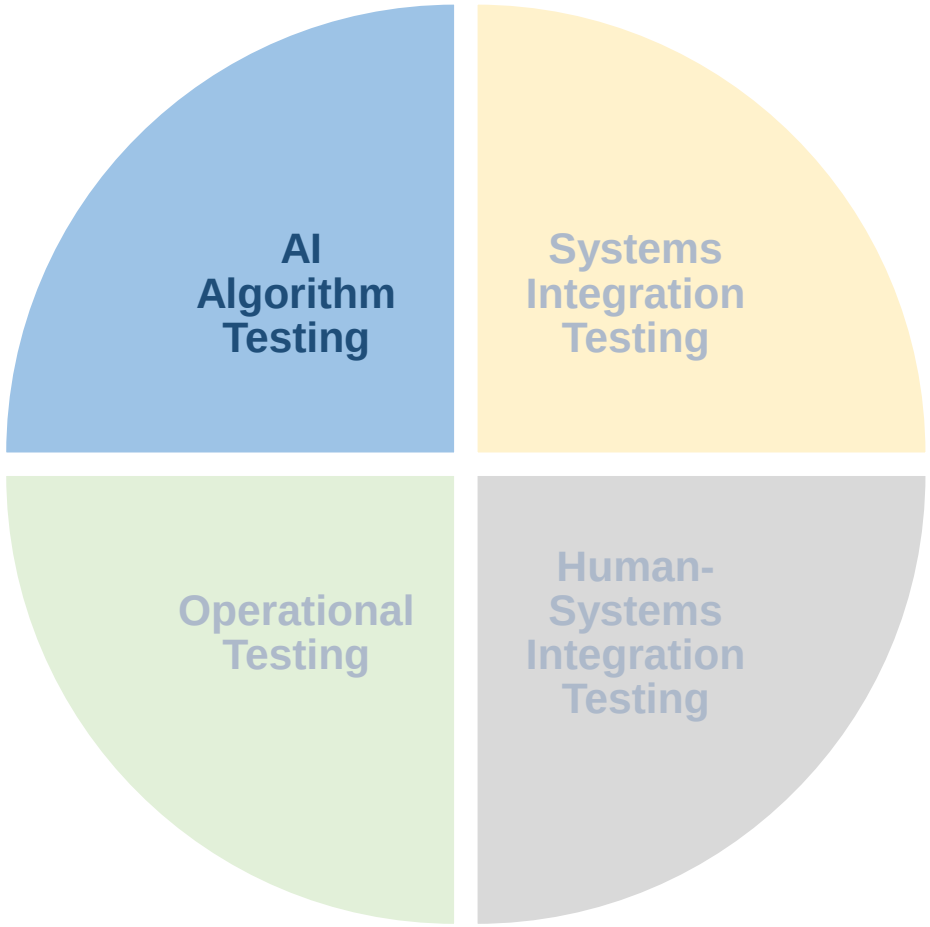
- Signed by Deputy Secretary of Defense, Kathleen Hicks, June 2022
- **LOE 2.1.2:** Develop or acquire AI-related Test & Evaluation (T&E) tools to be used as a resource for AI developers and testers… drawing upon best practices and innovative research from industry and the academic community, as well as commercially available technology
- **LOE 2.1.3:** Create a central repository of tools for T&E of AI… that enables easy and continuous testing for DoD testers




https://www.ai.mil/docs/RAI_Strategy_and_Implementation_Pathway_6-21-22.pdf

CDAO

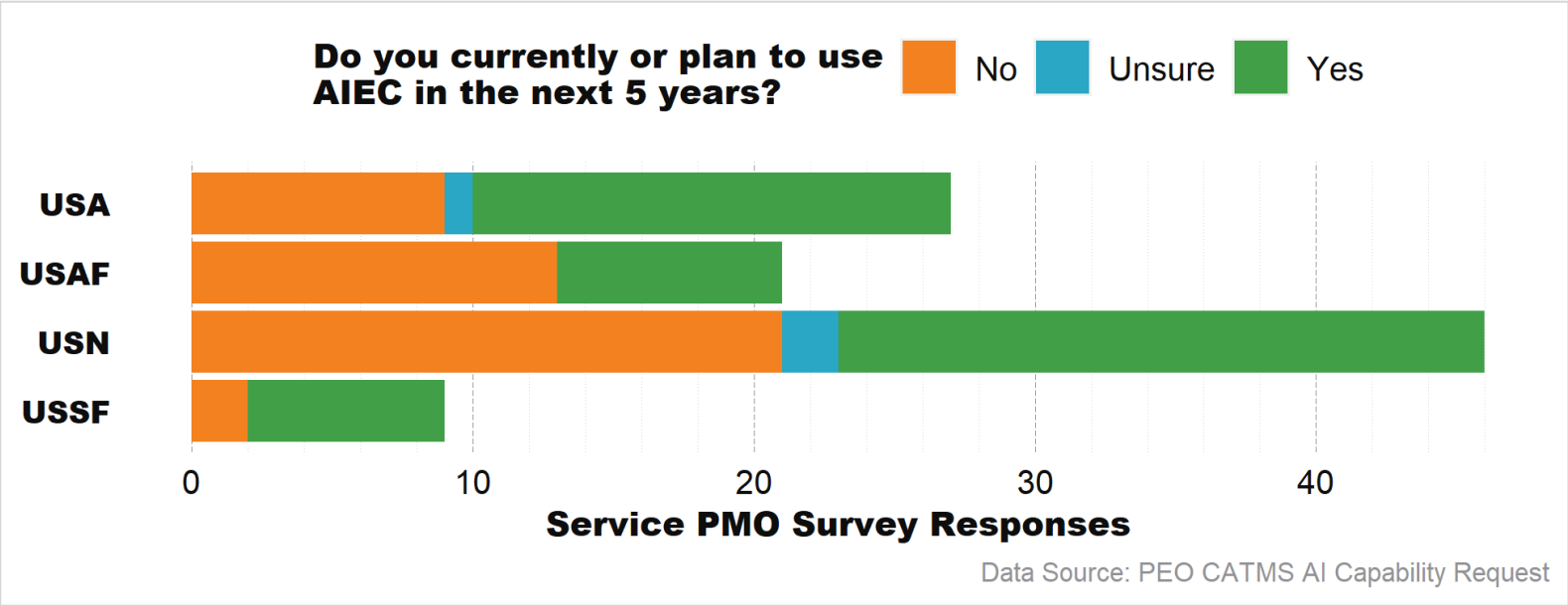
# Scope




- Among the four quadrants of AI T&E, we are focused entirely on **AI Algorithm Testing**
- Why?
  - Applicability of tools across multiple missions and systems
  - Required domain knowledge for further stages of testing
- Within that, our initial focus is **CV Classification & Object Detection**

CDAO

# AI Adoption DoD-wide

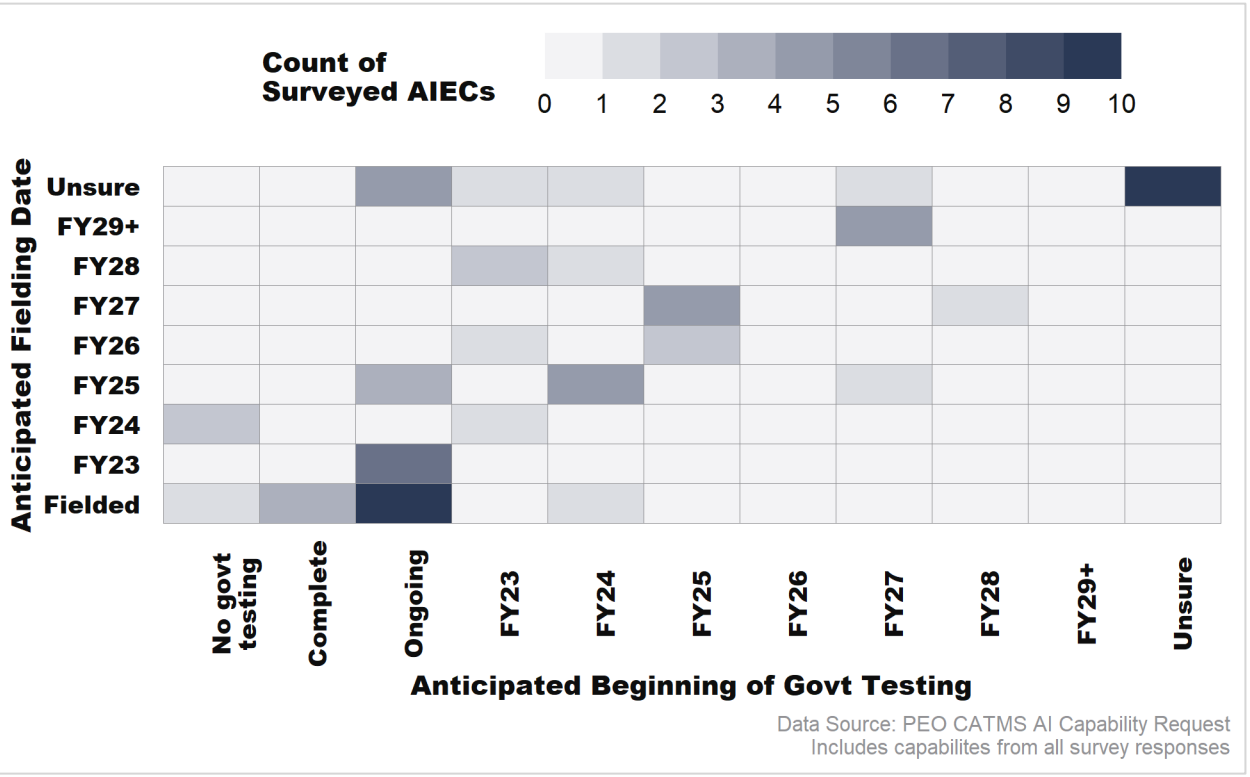


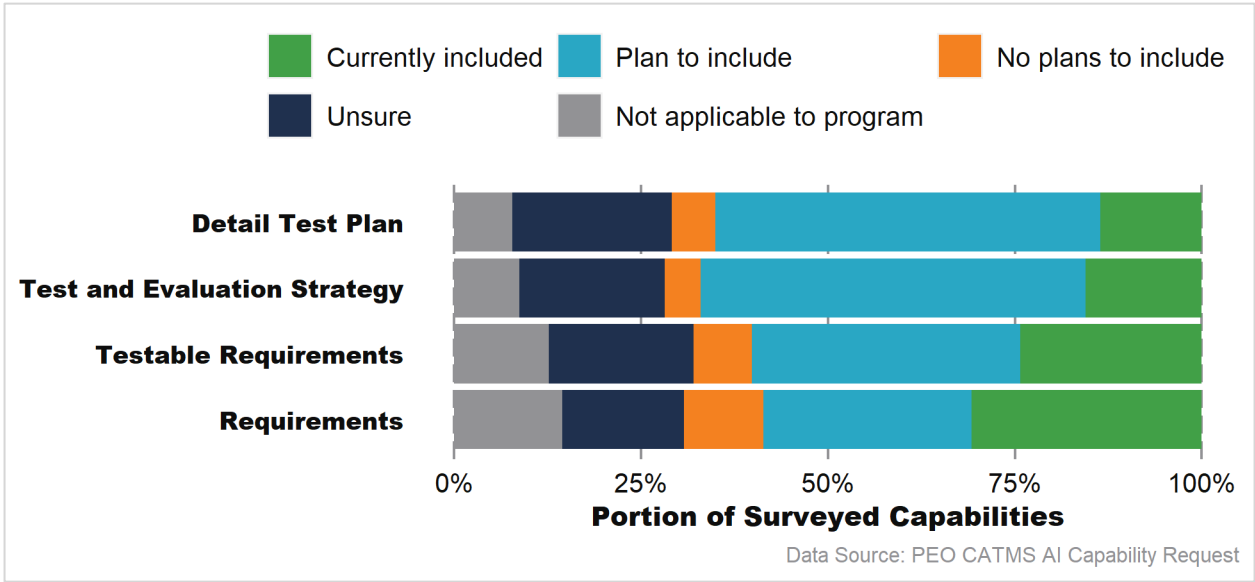

Among surveyed offices, 55% of Service Program Management Offices (PMOs) indicate they currently or plan to use AI-enabled capabilities (AIEC)

*CDAO Report: The National Artificial Intelligence Test and Evaluation Infrastructure Capability, July 2023

CDAO

# AI Test & Evaluation Maturity DoD-wide




Despite wide interest:

- Huge amount of uncertainty across the DoD in performance of AI T&E
- Demand for guidance and resources
- Uncertainty on novel risks and impacts on existing systems

*CDAO Report: The National Artificial Intelligence Test and Evaluation Infrastructure Capability, July 2023

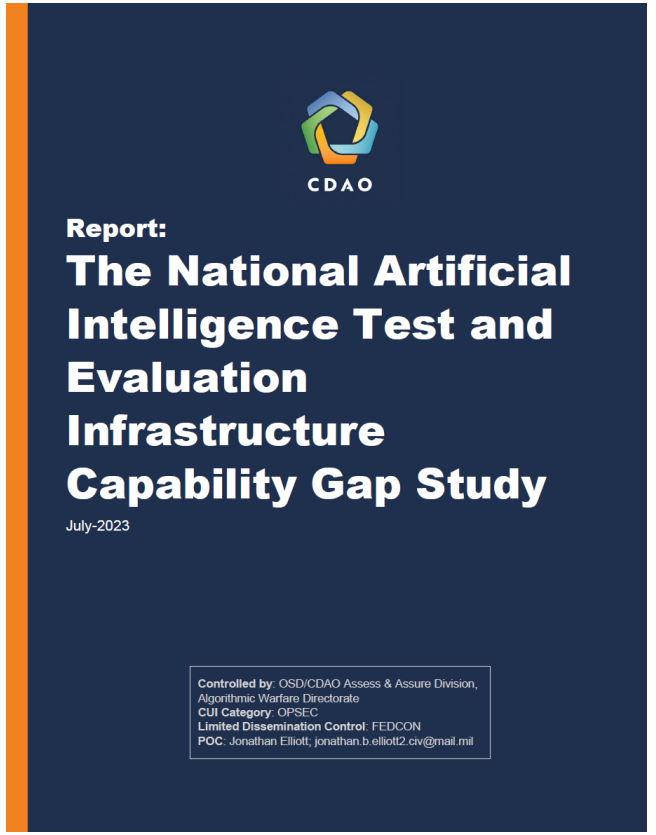
CDAO

# What are the key problems?

## Report on AI T&E demand and gaps

- “There is widespread interest for DoD enterprise-level T&E infrastructure to address the novel and exacerbated challenges posed by the T&E of [AI].”

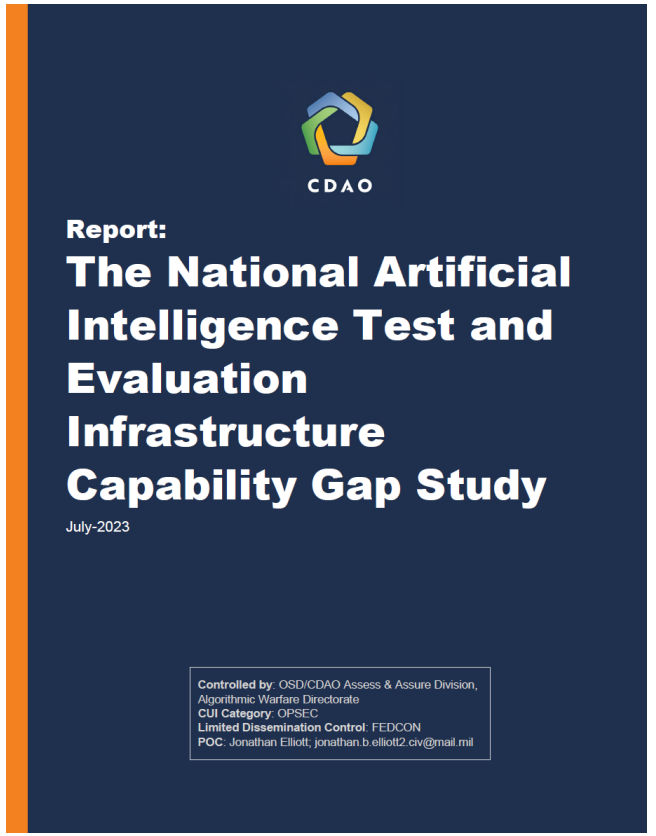
- “While programs are currently investing locally in T&E resources… there is still a consistent desire across survey programs for DoD enterprise support.”




CDAO

# What are the key problems?

1. Lack of maturity and domain knowledge in DoD AI testers
2. Difficulty in scaling tools across various DoD environments, platforms, and missions
3. Lack of tools for operationally-realistic conditions




CDAO

# Key Problems

**Problem 1:**

Lack of maturity and domain knowledge in DoD AI testers

**Our focus:**

- Education on advanced AI testing
- Accessibility of tools
- Ease of use of tools

CDAO

# Key Problems

**Problem 2:**

Difficulty in scaling tools across varied environments, platforms, and missions

**Our focus:**

- Ease of deployment to DoD environments
- Interoperability of tools with each other and AI/ML platforms

CDAO

# Key Problems

**Problem 3:**

Lack of tools for operationally-realistic conditions

**Our focus:**

- Tailor functionality to apply to operationally-realistic DoD use cases

CDAO

# Bridging the gap

## Research & Engineering

Research into advanced applications of AI for DoD-unique modalities:

- T&E of operator-AI performance
- T&E of AI in systems of systems
- Realistic adversarial attacks
- Use of simulation for DoD modalities, e.g., sonar, radar
- AI monitoring at the edge

## CDAO JATIC

- Transition existing AI T&E work into DoD by increasing maturity and usability
- Increase speed and rigor of AI T&E by providing common tools, standards, infrastructure
- Inform future DoD research investments

## DoD Service PEOs, PMOs

- Testing and fielding AI-enabled capabilities across:
  - Logistics
  - Intelligence
  - Operations
  - Health
  - …
- Huge interest and demand to employ AI
- Lack of knowledge, expertise, or centralized investment
- Uncertainty on novel risks and impacts on existing systems

CDAO

# AI Assurance Toolbox

A set of **python libraries** to enable rigorous AI T&E, designed for interoperable usage, easy deployment, and wide integration

- Straightforward deployment, setup, and use within variety of development or testing environments
- Seamless integration with key MLOps platforms and capabilities
- Using standardized model, data, and metrics protocols which are:
  - widely compatible
  - easy-to-satisfy
  - informative
  - dependency-free


Amazon SageMaker



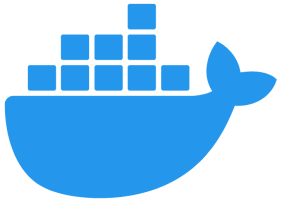

W&B

# AI T&E Platform

An orchestrated MLOps solution composed of open-source capabilities, specifically tailored for AI/ML testing

- JATIC python libraries are ideal for organizations who have *already adopted* an enterprise MLOps platform, such Databricks or Sagemaker
- For those without infrastructure, the **JATIC AI T&E Platform** provides best-of-breed open-source tools to **jumpstart AI T&E from Day 1**
  - Deployable quickly to commercial cloud, on-prem, local machines, or HPC using Infrastructure as Code

- JATIC AI T&E Platform will provide capabilities for :
  - Workflow orchestration
  - Model registry, experiment tracking
  - Database / object store
  - Visualization dashboard
  - Jupyter lab / IDE
  - Multi-GPU resource management


# Capabilities

| Tool | AI T&E Capability | Developer |
|---|---|---|
| **Adversarial Robustness Toolbox*** | State-of-the-art library of **adversarial attacks** and **defenses** | IBM |
| **Armory*** | Testbed for scalable evaluations of **adversarial attacks** and **defenses** | TwoSix Tech |
| **Data-analysis metrics library*** | Evaluate datasets for similarity, drift, and complexity | ARiA |
| **XAI Toolkit*** | Generate **visual saliency maps** on AI predictions using black-box and white-box techniques | Kitware |
| **Natural Robustness Toolkit*** | Generate **operationally realistic data perturbations** and **augmentations** in-silico using **sensor-model** based techniques to test model robustness | Kitware |
| **MAITE*** | A source of common types, protocols, and utilities to enable synergistic and streamlined AI T&E workflows | MIT |
| **rAI Toolbox*** | Generate **data perturbations** and **augmentations** in-silico to test model robustness | MIT |
| **RAVEN*** | Open-source AI & data science platform, designed for collaboration, scalability, and rapid deployment | Quansight |
| **RealLabel** | Using model inferences, identify potential **ground label errors** within data | MORSE Corp |
| **Gradient** | Develop standard **AI T&E reports** in Powerpoint, directly from python | MORSE Corp |

CDAO

***indicates open-source capability**

# Deployment

The **AI Assurance Toolbox** and **AI T&E platform** will be freely distributed and easily portable to DoD environments


**Enterprise-wide availability**

- Many JATIC python libraries will be made **publicly available** on GitHub, PyPI, and conda
- DoD-specific source code and containers images will be hosted on Repo1 and IronBank


**Secure & hardened**

- JATIC software will be hardened for deployment onto IL-2, IL-6 and other classified information systems
- CDAO will work closely with DoD organizations to support deployment of JATIC tools into environments
- JATIC will never collect testing data or results – your models & data, your results

CDAO

# Collaboration & Access

**CDAO T&E is actively seeking key government partners leading AI/ML to:**

- *Transition research or S&T technologies for AI T&E and AI Assurance*
- *Support developmental testing of AI technologies to be integrated and fielded into larger systems*
- *Understand your AI T&E requirements, building AI T&E tools within JATIC to support*
- *Obtain feedback from you to iterate and mature our capabilities*

**Join at https://gitlab.jatic.net with a *.mil*, *.gov*, or FFRDC/UARC email to get access to our current tools!**

CDAO