CDAO

Chief Digital & Artificial Intelligence Office

JATIC: Joint AI T&E Infrastructure Capability

DISTRIBUTION STATEMENT A
Approved for public release; distribution is unlimited.

Controlled by: CDAO, Assess & Assure Division
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

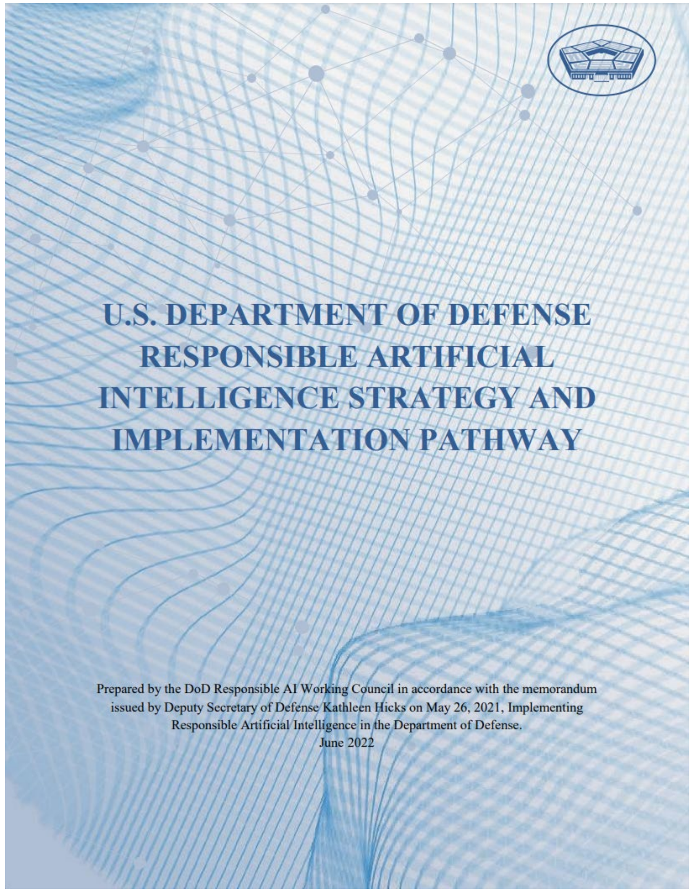
In order to…
Provide rigorous assurance of the effectiveness, robustness, and safety of the DoD's AI-enabled systems

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



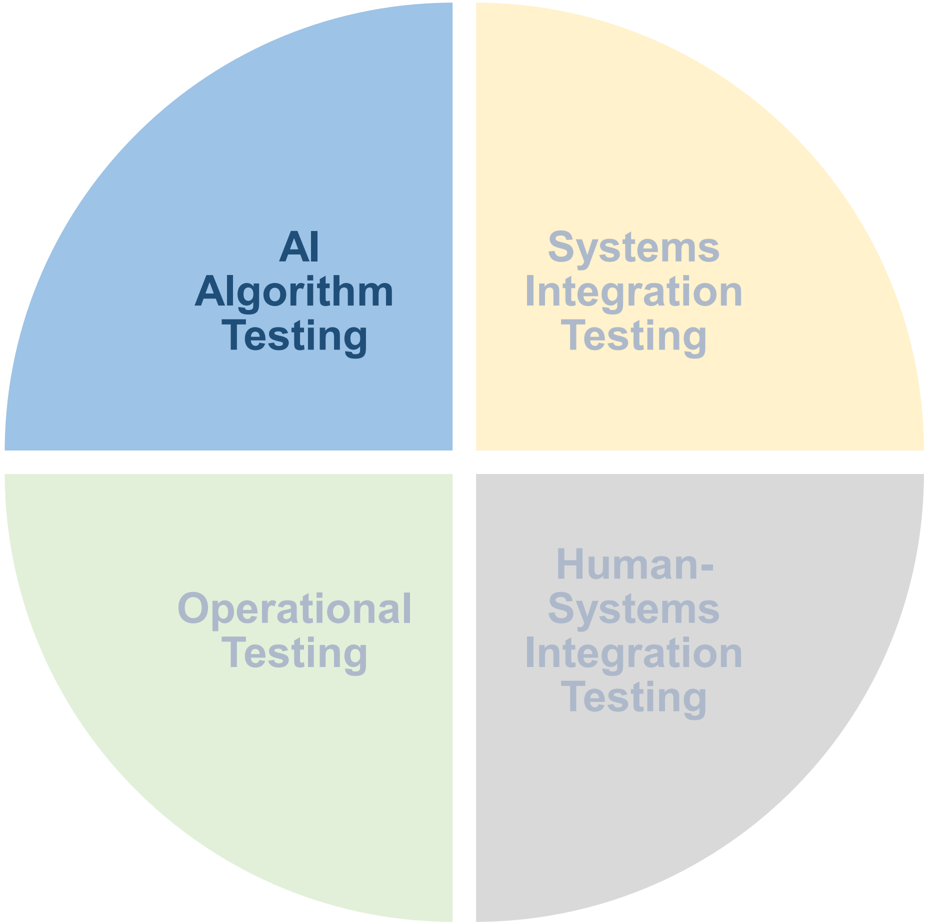
U.S. DEPARTMENT OF DEFENSE RESPONSIBLE ARTIFICIAL INTELLIGENCE STRATEGY AND IMPLEMENTATION PATHWAY

Prepared by the DoD Responsible AI Working Council in accordance with the memorandum issued by Deputy Secretary of Defense Kathleen Hicks on May 26, 2021, Implementing Responsible Artificial Intelligence in the Department of Defense.
June 2022

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

CDAO

Scope


AI Algorithm Testing
Systems Integration Testing
Operational Testing
Human-Systems Integration Testing

- 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**

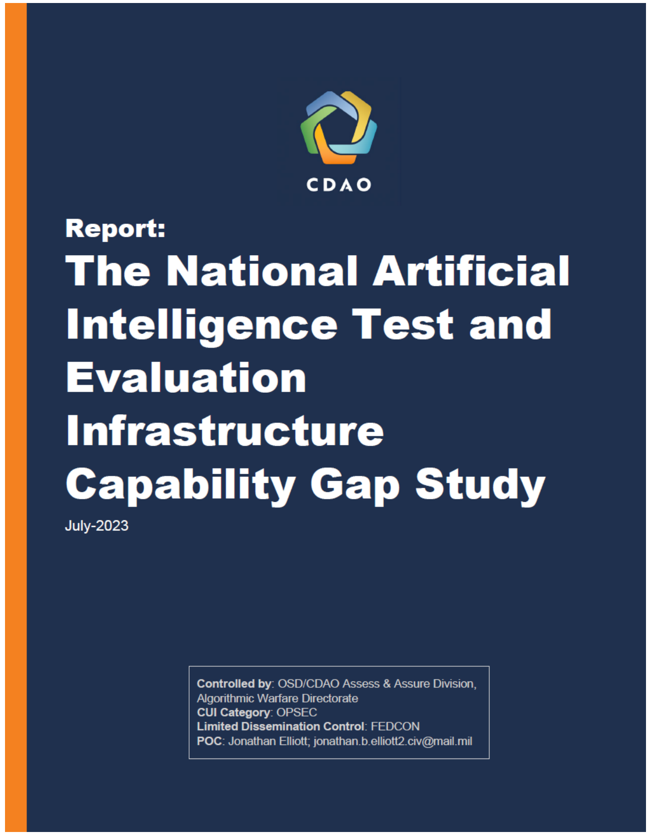
CDAO

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

Report:

The National Artificial Intelligence Test and Evaluation Infrastructure Capability Gap Study

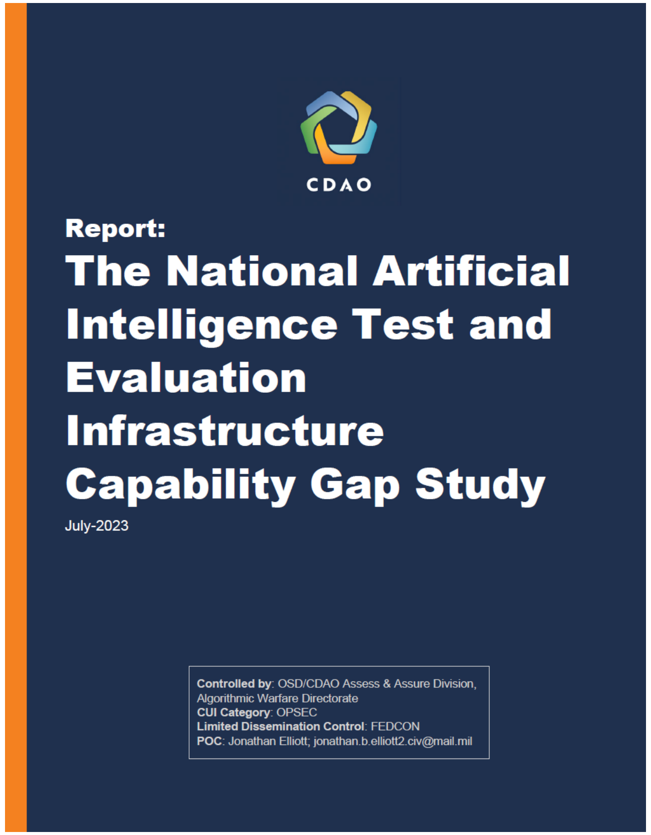
July-2023

Controlled by: OSD/CDAO Assess & Assure Division, Algorithmic Warfare Directorate
CUI Category: OPSEC
Limited Dissemination Control: FEDCON
POC: Jonathan Elliott; jonathan.b.elliott2.civ@mail.mil

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

Report:

The National Artificial Intelligence Test and Evaluation Infrastructure Capability Gap Study

July-2023

Controlled by: OSD/CDAO Assess & Assure Division, Algorithmic Warfare Directorate
CUI Category: OPSEC
Limited Dissemination Control: FEDCON
POC: Jonathan Elliott; jonathan.b.elliott2.civ@mail.mil

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

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

Key Problems

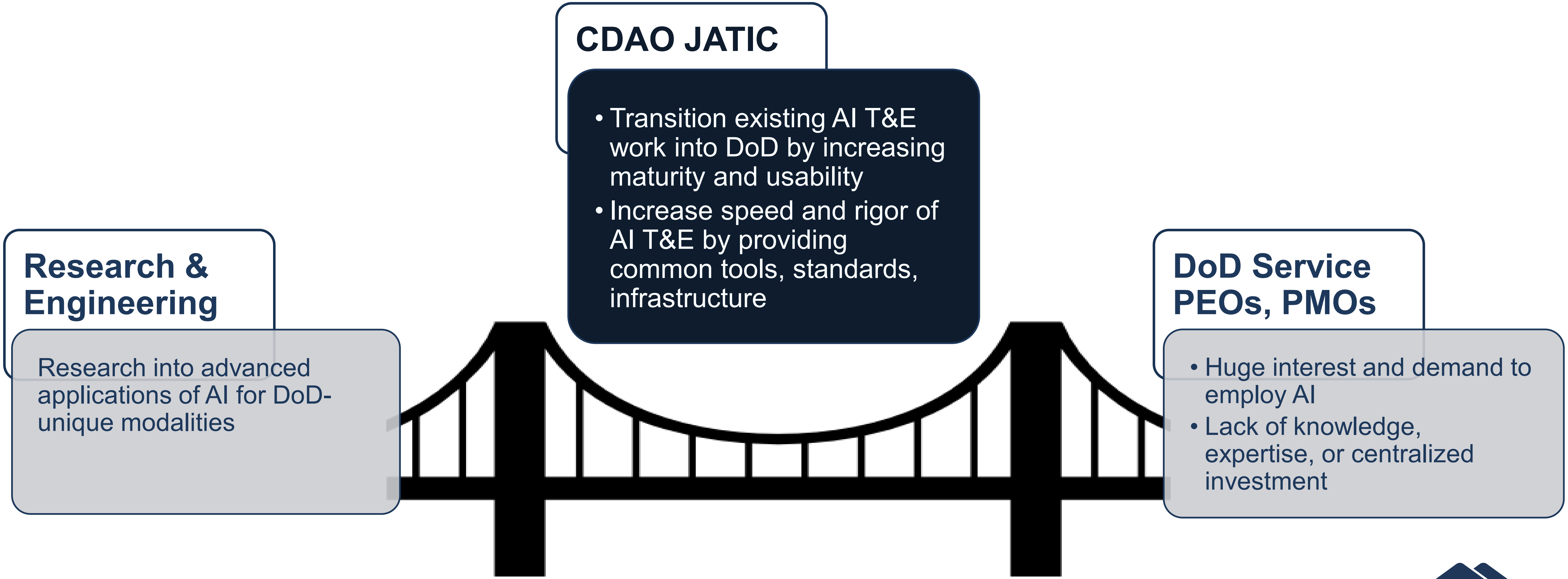
Problem 3:

Lack of tools for operationally-realistic conditions

Our focus:

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

Bridging the gap



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

Research & Engineering

Research into advanced applications of AI for DoD-unique modalities

DoD Service PEOs, PMOs

- Huge interest and demand to employ AI
- Lack of knowledge, expertise, or centralized investment

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

databricks

Amazon SageMaker

mlflow

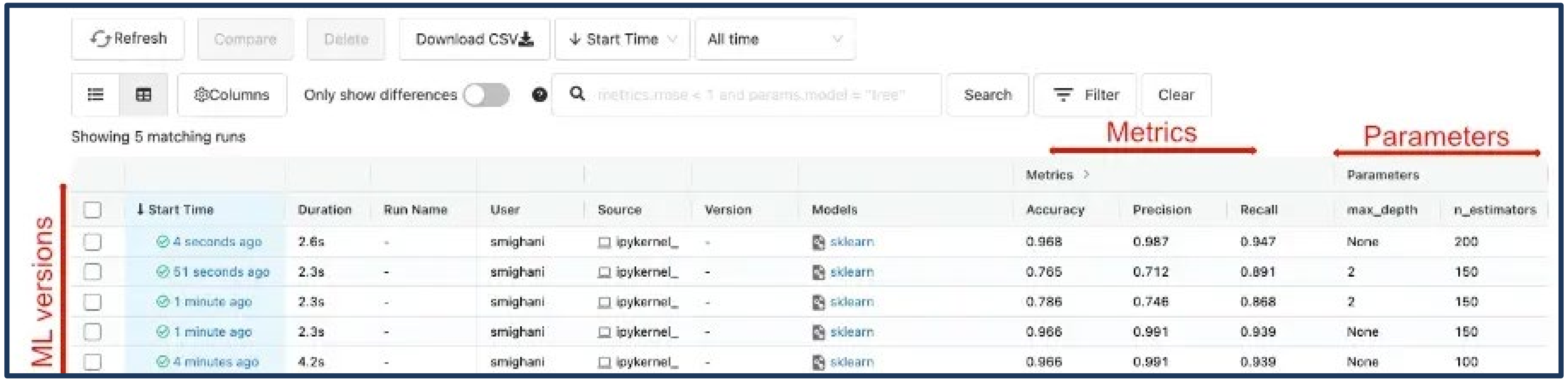
W&B

CDAO

T&E + MLOps

To be effective, AI T&E capabilities **must** *integrate seamlessly* with MLOps pipelines

- *Continuous testing* of AI models **requires** this close integration, especially as AI models are retrained more frequently
- Integration into MLOps provides incredible synergies between T&E and other AI/ML capabilities:
  - T&E + Workflow orchestration -> automated execution of model test plans
  - T&E + Model registries & experiment tracking -> improved T&E traceability and enhanced model metadata
  - T&E + Visualization dashboards -> seamless comparison between many models across test cases
  - T&E + Hyper-parameter optimization -> optimize model hyperparams for robustness, explainability, etc.
  - T&E + Labeling -> model T&E inference results inform potential errors in ground truth labels



Showing 5 matching runs

							Metrics >			Parameters	
□	↓ Start Time	Duration	Run Name	User	Source	Version	Models	Accuracy	Precision	Recall	max_depth	n_estimators
□	✓ 4 seconds ago	2.6s	-	smighani	□ ipykernel_	-	sklearn	0.968	0.987	0.947	None	200
□	✓ 51 seconds ago	2.3s	-	smighani	□ ipykernel_	-	sklearn	0.765	0.712	0.891	2	150
□	✓ 1 minute ago	2.3s	-	smighani	□ ipykernel_	-	sklearn	0.786	0.746	0.868	2	150
□	✓ 1 minute ago	2.3s	-	smighani	□ ipykernel_	-	sklearn	0.966	0.991	0.939	None	150
□	✓ 4 minutes ago	4.2s	-	smighani	□ ipykernel_	-	sklearn	0.966	0.991	0.939	None	100

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
Dataset 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	Evaluate models against physics-based, operationally-realistic perturbations	Kitware
rAI Toolbox*	Generate **data perturbations** and **augmentations** in-silico to test model robustness	MIT
jatic toolbox	A source of common types, protocols, and utilities to enable synergistic and streamlined AI T&E workflows	MIT
Gradient	Develop standard **model cards, data cards, and AI T&E reports** in Powerpoint, directly from python	MORSE Corp
Nebari*	Open-source AI & data science platform, designed for collaboration, scalability, and rapid deployment	Quansight

***indicates existing 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

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