



DOT&E Initiatives for Adequate Test and Evaluation of AI-Enabled Systems

18 October 2023

Dr. Kristen Alexander
Chief Learning and Artificial Intelligence Officer
Director, Operational Test and Evaluation

DOT&E Mission

The DOT&E is the **senior advisor to the Secretary of Defense** on operational test and evaluation (OT&E) and Live Fire Test and Evaluation (LFT&E) in the Department of Defense (DoD).

**Policy and
Guidance**

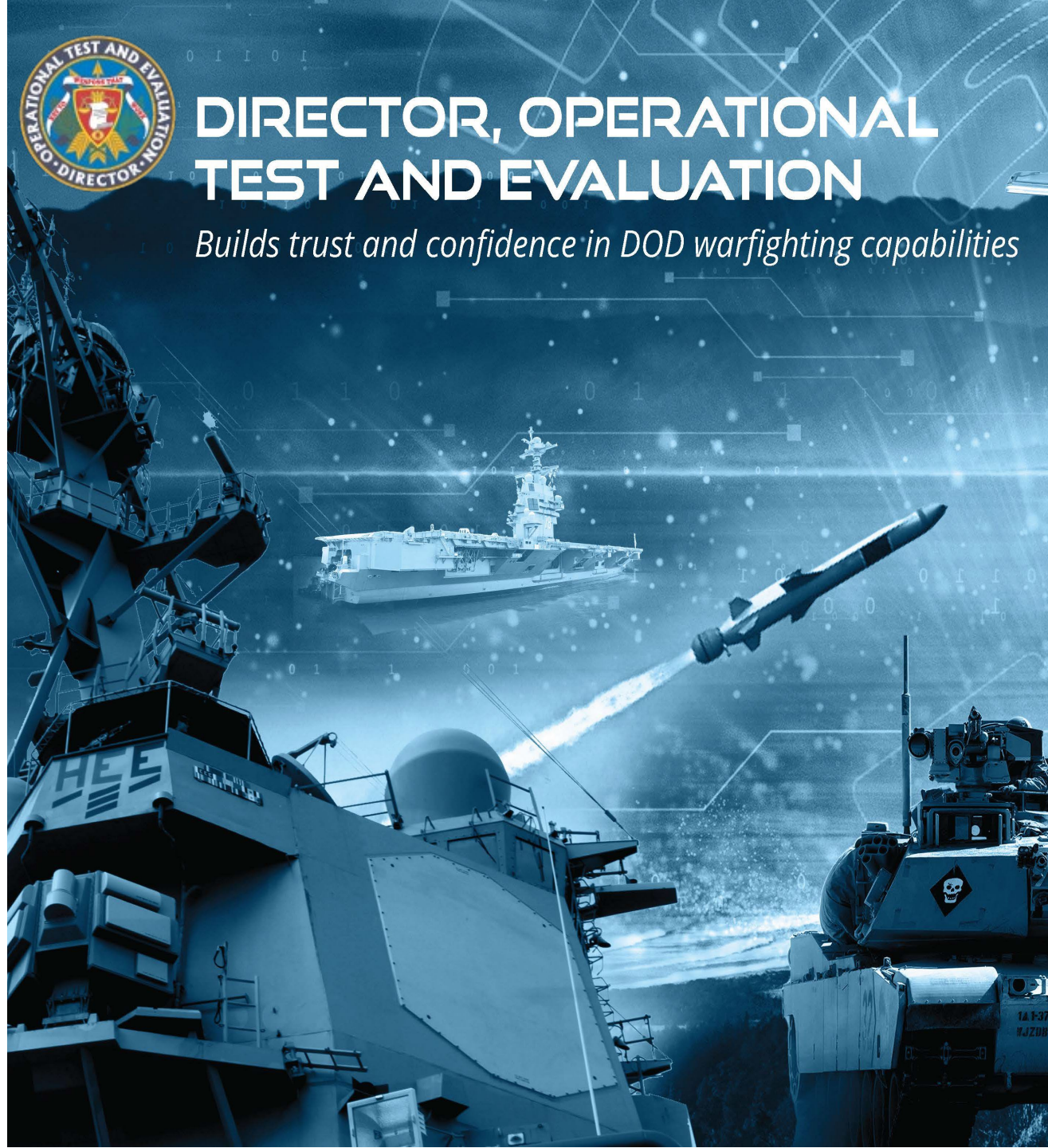
Reporting

**Strategic
Initiatives**

Oversight

**Congressional
Taskings**

**Field Activities
Oversight**








DIRECTOR, OPERATIONAL TEST AND EVALUATION

Builds trust and confidence in DOD warfighting capabilities

| DOT&E Statement of Intent

We will transform test and evaluation to enable delivery of the world's most advanced warfighting capabilities at the speed of need.

HON Nickolas Guertin
Director, Operational Test and Evaluation

-
-  **Test the way we fight**
 -  **Accelerate the delivery of weapons that work**
 -  **Improve the survivability of DOD in contested environments**
 -  **Pioneer T&E of weapon systems built to change over time**
 -  **Foster an agile and enduring T&E enterprise workforce**

Strategic Drivers

- 🚀 Engineering of Software-Reliant Systems
- 🚀 Artificial Intelligence / Machine Learning
- 🚀 Joint All-Domain Operations
- 🚀 Data
- 🚀 Speed to Field
- 🚀 Culture
- 🚀 Talent Management



Pioneer T&E of weapon systems built to change over time

- **LOE 4.1:** Increase the use of credible digital twins in T&E
- **LOE 4.2:** Evaluate the operational and ethical performance of AI-based systems
- **LOE 4.3:** Advance the evaluation of software-reliant systems' operational performance



4.2 Evaluate the operational and ethical performance of AI-enabled systems

Why this matters?

It is important to:

- Monitor and evaluate the drift in deployed AI models' behavior, which could occur when **real-world data deviate from the training data** used to create the model
- Demonstrate that **AI-enabled systems** are responsible, ethical, equitable, traceable, reliable, and governable

| Key Actions

4.2.1	Identify areas that require further research to support adequate operational and ethical performance of AI-enabled systems
4.2.2	Develop policies, standards, metrics, and a risk-based framework for assessing performance of emerging AI-enabled systems

The desired end state for **4.2** is an **adequate assessment of operational and ethical performance of AI-enabled systems.**



Research Review and Best Practices

AI research occurring at the model development and testing level

FY23

FY24

FY25

FY26

Beyond

Integration of Research via Use Cases

Near Term Operational Risk / Sense of Urgency

Baselining

Institutionalization of Research into T&E Processes

T&E Preplanning

Model T&E

IT&E (Across LVC)

Operational T&E

Operations and Sustainment T&E

T&E and IP considerations

Operational metrics

Integrated info, synthetic data usefulness

Updates to Processes

Monitoring, domain shift detection, re-engineering

T&E for AAF

Model T&E → OT&E

Best practices across acquisition lifecycle

Integrated T&E Planning

Multi-modal monitoring for responsible AI

Enhanced planning

DevSecOps/SW Factory

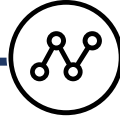
Predictive capability of IT&E in OT&E

OT feedback → Operations

Monitoring



Material Solution Analysis



Technology Maturation & Risk Management



Engineering & Manufacturing Development



Production & Deployment



Operations & Support

Start Early

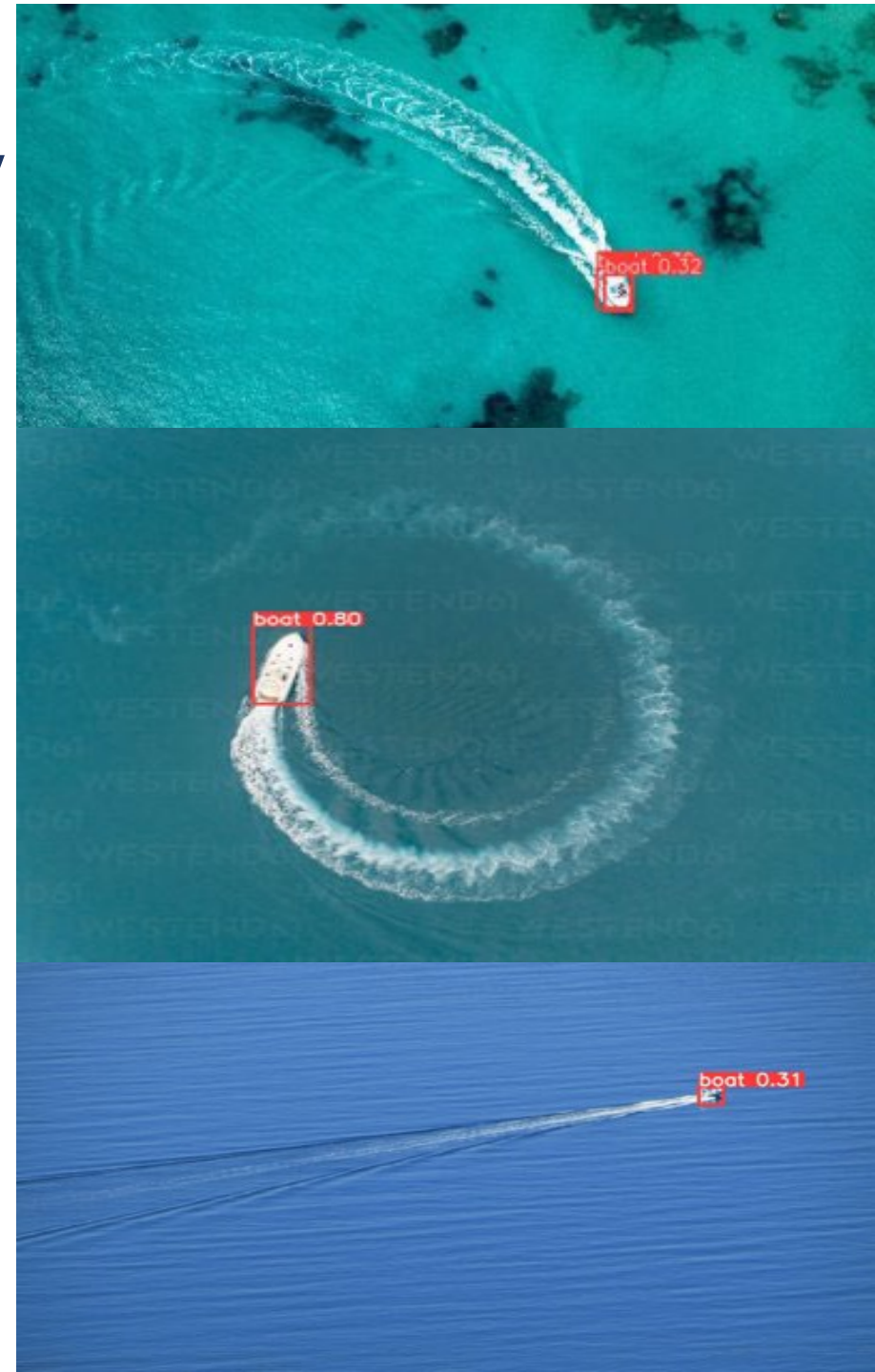
Integrated DT/OT & Iterate Continuously

Research Priorities to Inform Policy

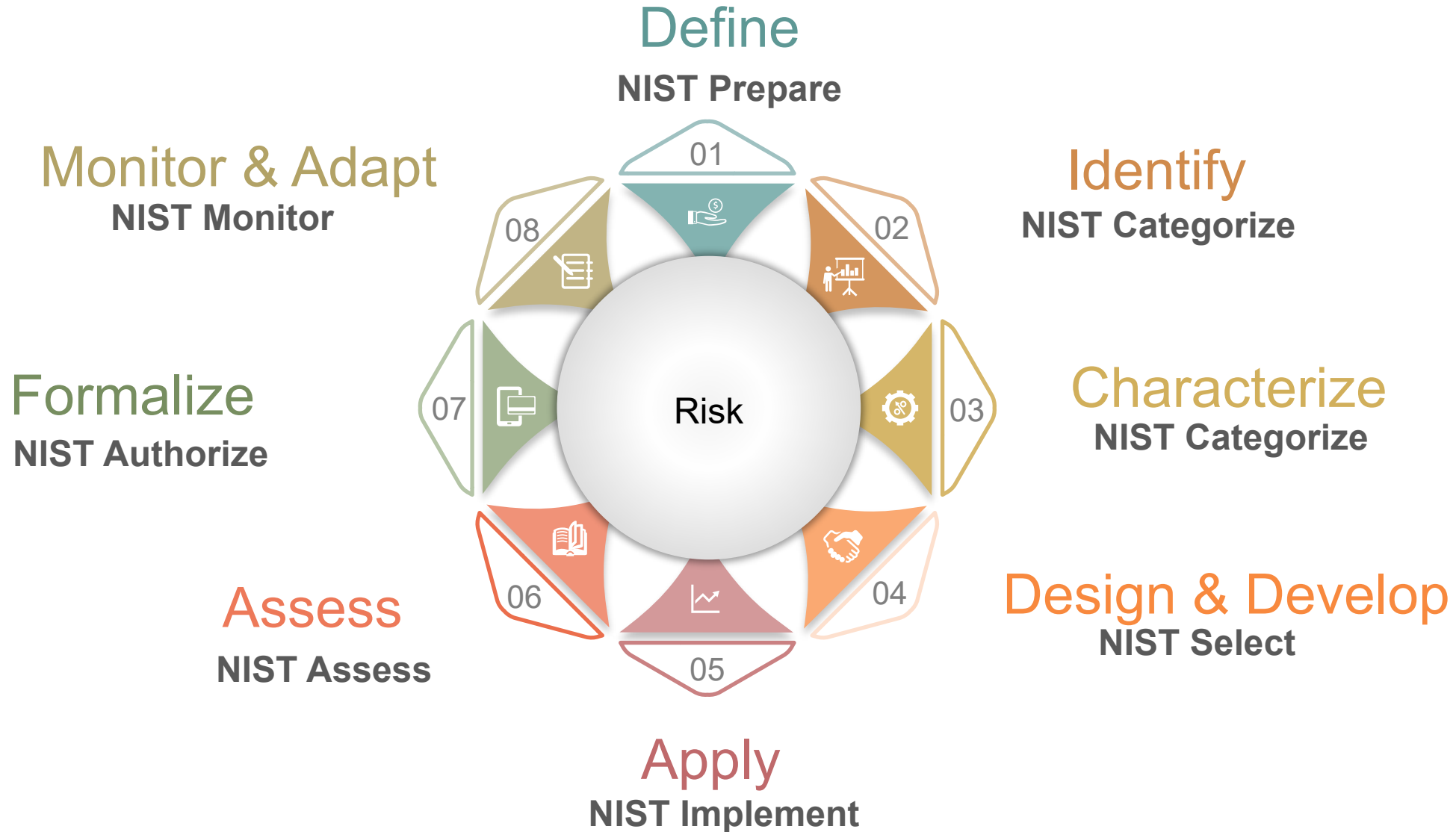
Risk: The Services field AI-enabled systems before the T&E Enterprise can define and execute adequate testing to measure capabilities and limitations to the Warfighter.

DOT&E focus :

- **Defining test adequacy** for AI enabled systems
- **Defining T&E requirements** for systems build to change over time (potentially in real time)
- Defining **new performance metrics and assessments methods** for AI that assess ethical, responsible and accountable aspects of DoD guidance
- Defining cyber-T&E and **new requirements for AI red teams** for AI enabled systems



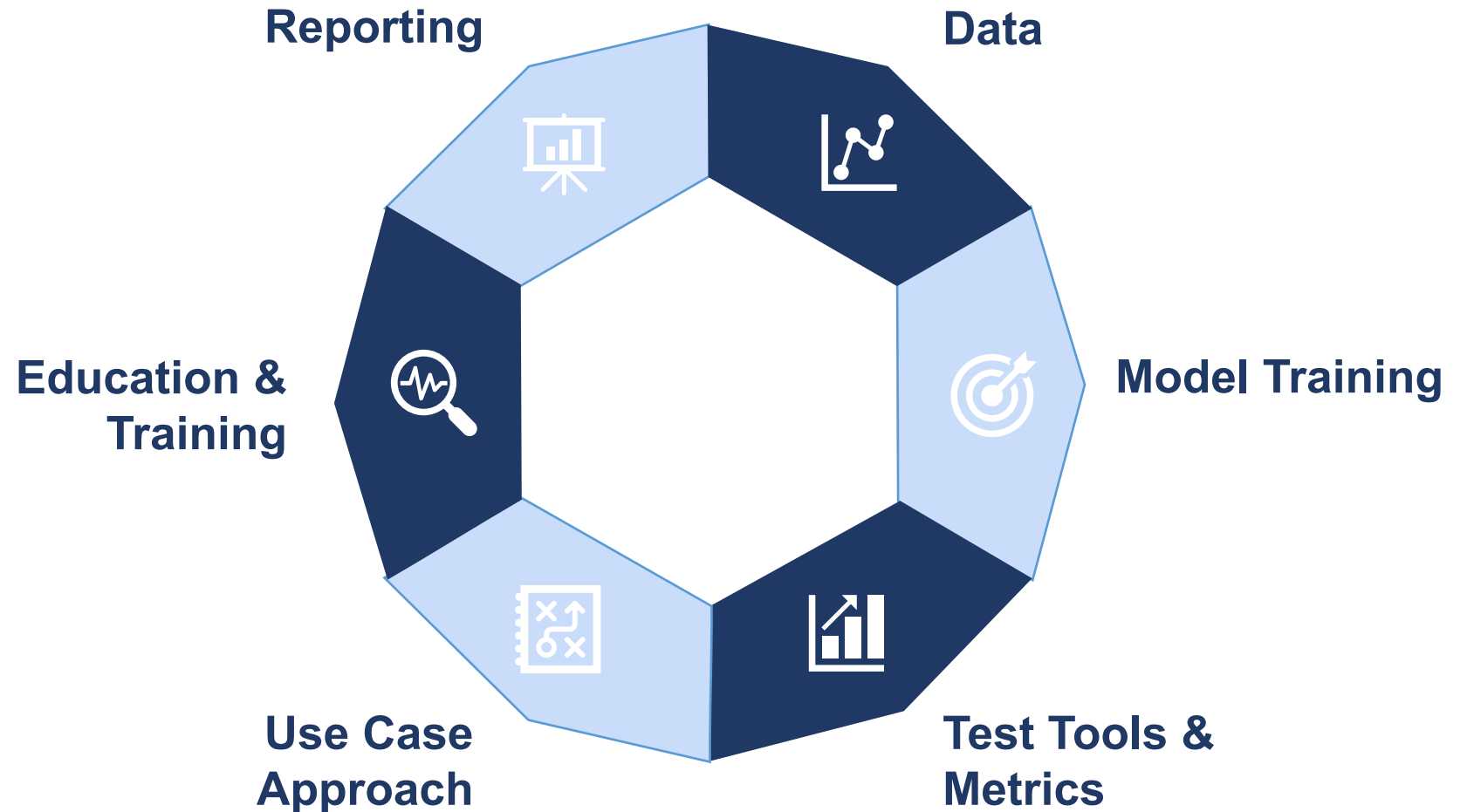
Risk Based Framework Development



| AI T&E Harness

T&E of AI Prototype environment objectives:

- Accelerate the transition of research and methods into T&E tools
- Prototype policy, standards, metrics, and risk frameworks
- Accelerate education and training of T&E practitioners



Use Case Development

Objective

Methodology for creating use cases that can be applied in research, framework development, etc.

Solution

Repeatable approach to use case development scalable across applications, domain and mission



Strategies

- Research
- Develop & refine
- Validate methods
- Apply frameworks
- Disseminate & Iterate



Benefits

- Repeatable process
- Domain scalability
- Utility validation
- Formalized process



Results

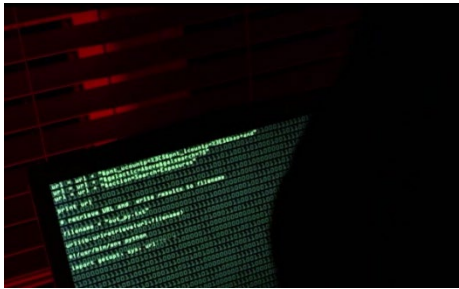
- Assessment of existing capabilities
- New use case development plan
- Validation measures



Additional value added

- ◇ Formal methodology for generating use cases to apply to various frameworks or methodologies
- ◇ Identification of redundant efforts and capability gaps within the domain and missions for a given use case
- ◇ Requirements for future work to answer capability gaps
- ◇ Repository of existing efforts across DoD relating to use case domain mission area through research

Piloting Use Cases



**Cognitive
Electronic Warfare**



Adversarial AI



**M&S & Unmanned
Underwater
Vehicles**



**Robotic Combat
Vehicle - Light**

Framework for AI/Autonomy T&E

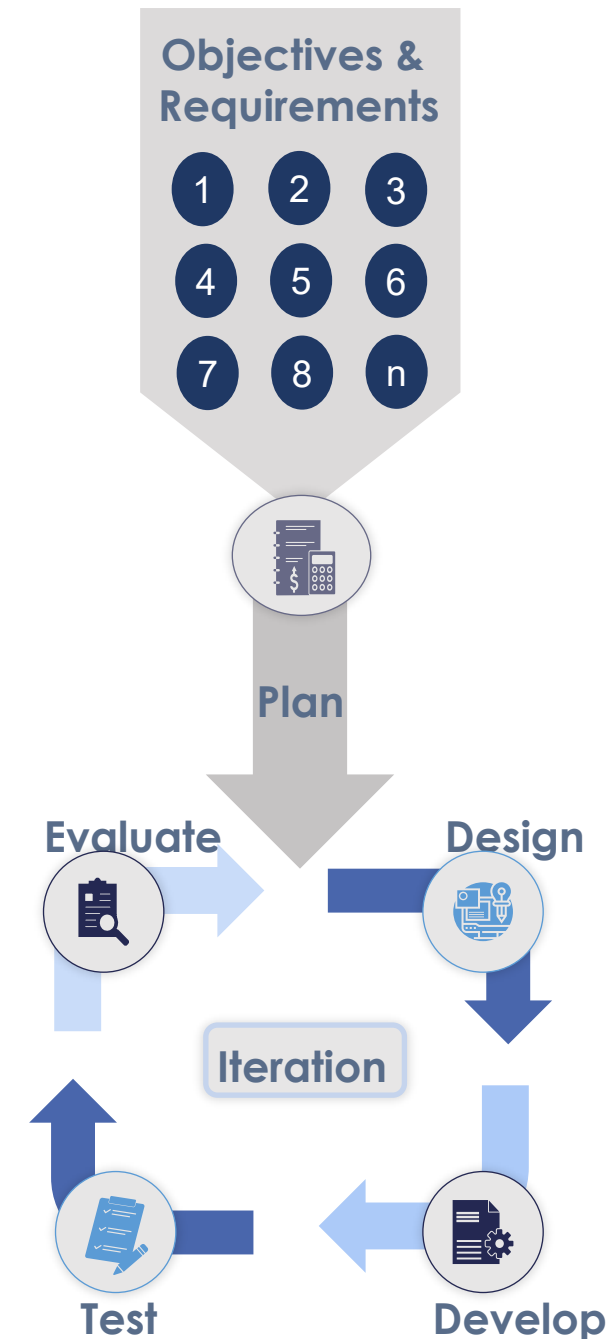
Objective: Provide a single, flexible, scalable Risk Based Framework for T&E of AI-enabled/autonomous systems within the Defense Acquisition System

Considerations:

1. Digital Engineering (DE)
2. Stakeholder identification and support
3. Scalability with future advancements

Approach:

1. Investigate existing frameworks
2. Identify success and failures of each framework
3. Utilize best practices of each framework
4. Draft initial framework approach
5. Apply framework initially to UUV Use Case
6. Iterations of Assess → Refine → Apply



Partnership and collaboration with AI T&E Community



Regular collaboration across stakeholder entities to ensure aligned vision and unified efforts

Next Steps

➤ Continue community-wide engagement for collaboration and partnership

➤ Stand up Action Teams to execute key initiatives, identify gaps, and develop mitigation strategies

➤ Partner with CDAO, DTE&A, TRMC, and the Services to develop use cases and inform guidance and application

Think and Act Like an Enterprise

To fight as a Joint Force, we must T&E as a Joint Force.

HON Nickolas Guertin
Director, Operational Test and Evaluation

