



SERC RESEARCH REVIEW 2024 | NOVEMBER 12, 2024

# Digital Engineering for Test and Evaluation

WRT-2024, WRT-1070, WRT-1071

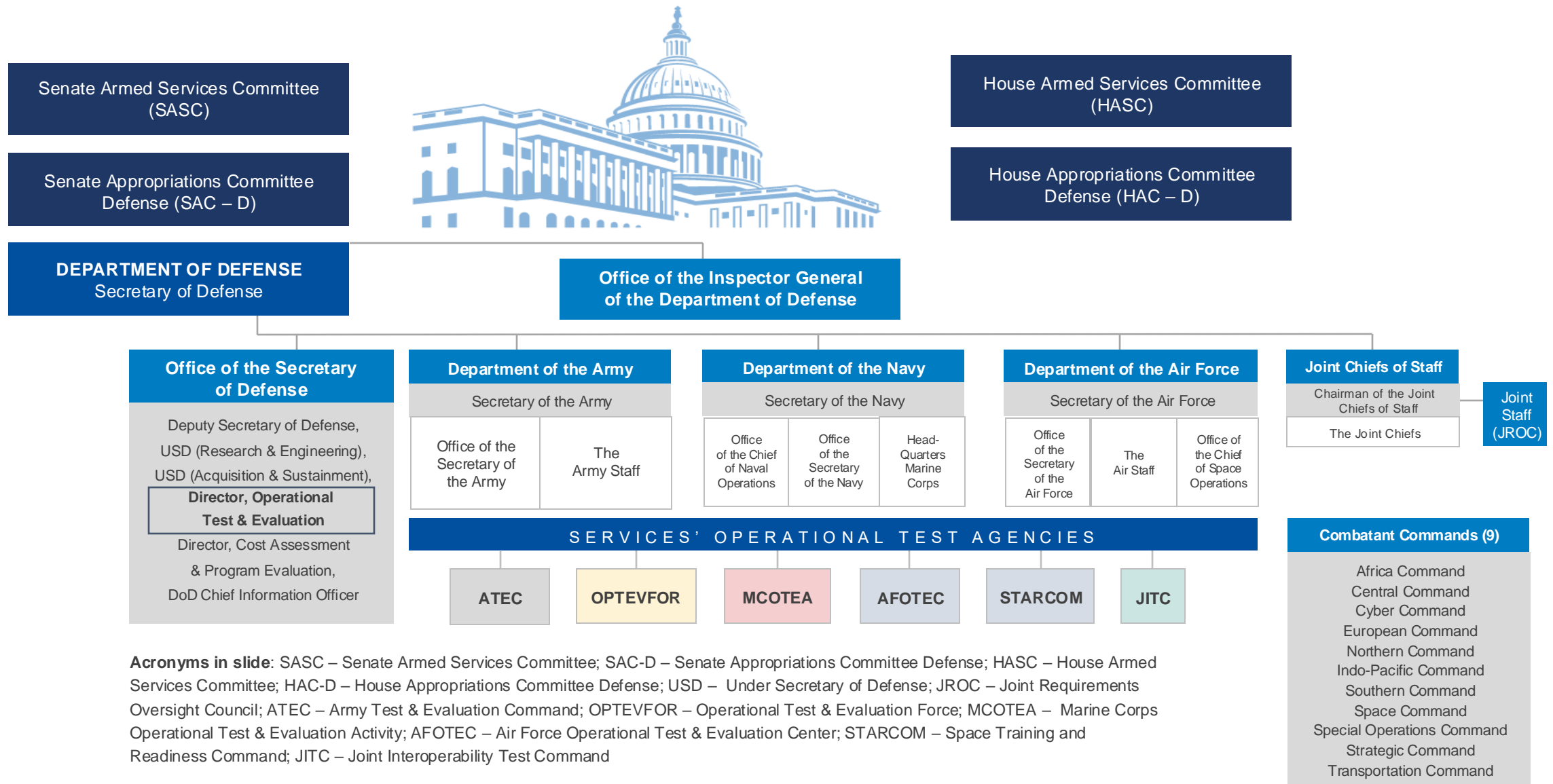
Sponsor: Director Operational Test and Evaluation

Laura J. Freeman

# Intro to DOT&E



# DOT&E in DoD Organizational Chart



# What is the Director Operational Test & Evaluation?



**POLICY AND  
GUIDANCE**



**OVERSIGHT**



**REPORTING**



**CONGRESSIONAL  
TASKING AND  
MANAGEMENT**



**STRATEGIC  
INITIATIVES**

# Why Transform T&E with Digital Engineering?

---



# Transformation of Today's Battlefield...



**SEAMLESS COLLABORATION ACROSS ALL DOMAINS**

**ATTRITABLE SYSTEMS AT SCALE**

**ENABLED BY AUTONOMY & AI**

# Transformation of Today's Battlefield...



**LONG-RANGE LETHALITY**

**ATTRITABLE SYSTEMS AT SCALE**

**CONTESTED ELECTROMAGNETIC SPECTRUM**

**ANTI-ACCESS / AREA DENIAL (A2/AD)**

# Mission Statement

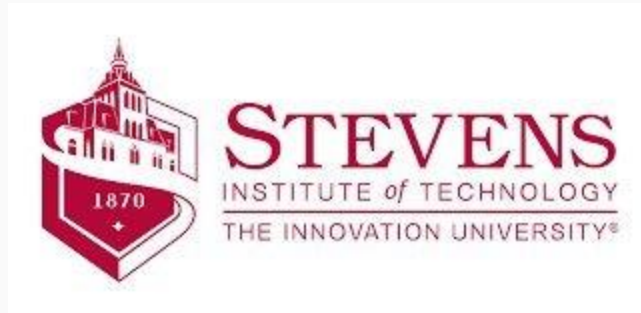
---

- Transform T&E state-of-the-art to address:
  - Rapidly changing and technologies and systems that continually evolve over their lifespan
  - Support the DoD in rapidly providing warfighting capabilities to counter advanced threats and new technologies
- Areas of emphasis:
  - Digital Transformation
    - Transforming current processes (e.g., digital TEMPs), and
    - Developing new methods for T&E that leverage digital transformation.
  - Speed to Fielding
    - Middle Tier Acquisition - “is used to rapidly develop fieldable prototypes within an acquisition program to demonstrate new capabilities and/or rapidly field production quantities of systems with proven technologies that require minimal development.”
    - Rapid Acquisition
    - “Speed of need”
  - Theme: data & models as the universal translator of information



# SERC/AIRC Partners

---



# Challenges to the Faster and Cheaper Delivery of Weapons that Work

There are several challenges with our current state that may inhibit our goal, often having cascading effects between each.



## Cultural

***In our T&E Enterprise – CT, DT, LF, OT, and M&S can operate in separate silos***

- No exposure to operational context until late in the T&E campaign  
⇒ Late deficiency discovery, cost overruns, delays



## Technological

***We use tools and processes designed for hardware capabilities, versus those defined for software***

- Current data collection, storage, and analysis is largely uncoordinated across different test phases, and often exists across disparate retrograde spreadsheets  
⇒ Data is not VAULTIS nor suitable for modern data analytic and AI  
Lost linkages → Decisions not necessarily data-drive and T&E gaps can be difficult to identify



## Methodological

***We do not systematically obtain nor analyze a holistic “whole body of evidence” that considers output from all phases of test***

- Test resources are not optimized  
⇒ Redundant or unnecessary testing or knowledge gaps



**Acronyms:** T&E = Test and Evaluation, CT = Contractor Testing, DT = Developmental Testing, LF = Live Fire Testing, OT = Operational Testing, M&S = Modeling and Simulation

# Research Approach & Resulting Products

---

- The AIRC research team will support DOT&E and their associated Strategy and Implementation Plan through:
  - research and technology
  - process
  - policy advancement
- Resulting products will include:
  - **Workshops** and draft **supporting reports**
  - **Technical papers** and **articles**
  - **Exemplar** products (models, software, tools, data structures, etc.)
  - **Frameworks** and **architectures** to shape T&E practices
  - **Courseware** for teaching and advancing workforce knowledge in proposed subject areas

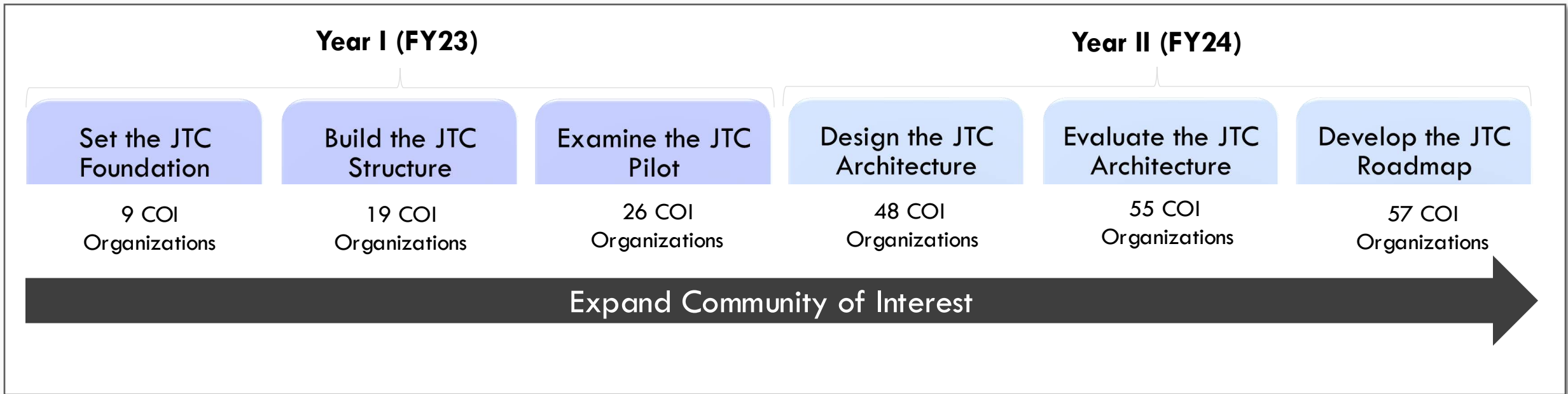
# Topic Overview

---

Research topics to include:

- Joint Test Concepts
- Integrated Testing
- Model Based Test and Evaluation Master Plan (TEMP)
- Integrated Decision Support Key (IDSK)
- Operation Safe Passage
- Verification, Validation and Uncertainty Quantification (VVUQ)
- T&E for AI-enabled systems

# Joint Test Concept: Study Design



## Year I Outcomes:

- **Mapped** complex T&E-enterprise challenges
- **Produced** JTC Pilot
- **Integrated** change leadership considerations
- **Created** diverse community of interest (COI)

## Year II Goals:

- **Create** a JTC reference architecture
- **Assess** the architecture through simulation
- **Develop** a JTC Implementation Roadmap that includes quick win opportunities



# Integrated Test and Evaluation



- Integrate T&E in model-based engineering to operationalize and optimize the Shift Left approach
  - Advanced performance inference techniques (e.g., **Bayesian** or similar) can be used to **carry forward data** from early prototypes through evaluation of production-representative systems.
  - Accelerate the development of tools that enable adequate performance **inference from a growing body of evidence**
  - Key approach: Carry knowledge across test phases via Bayesian priors
- Outcomes
  - Demonstrate how Bayesian methods can leverage all information to characterize systems' operational performance



# Operation Safe Passage

- Mission

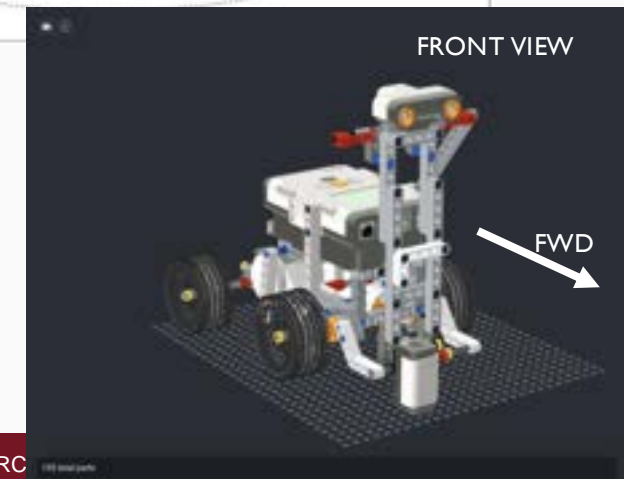
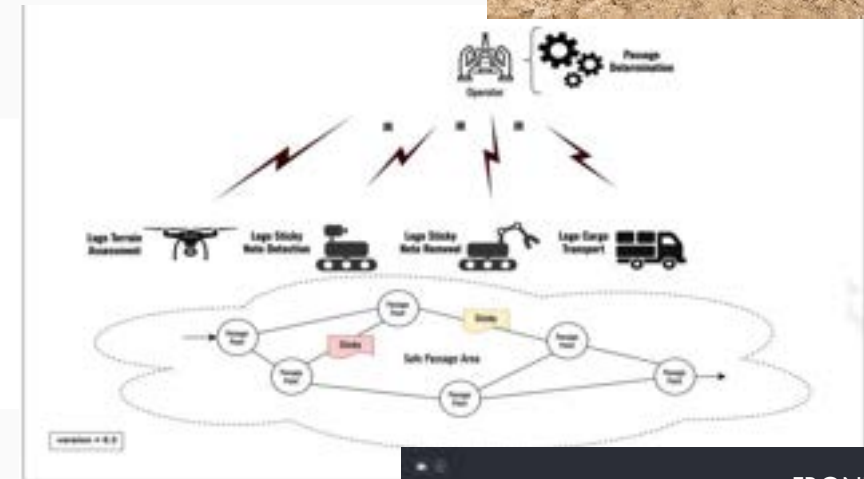
Generate a testbed and reference architecture for a proxy program to mature digital engineering/model-based products and methods

- Products

- Mission Model
- System Model
- Computational Model
- Physical Test Bed
- MB Temp
- IDSK

- Results

- Successful Integration of T&E Results through Mission/System Model to Inform Test Planning and Management



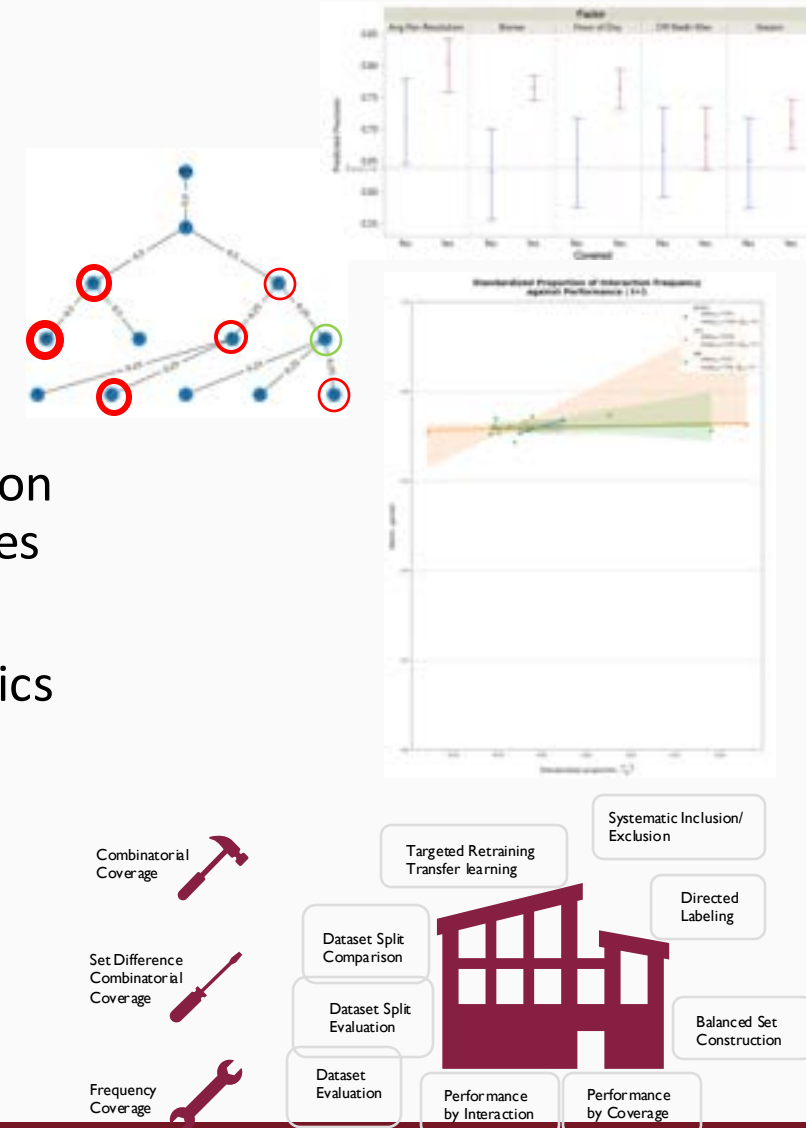
# VVUQ and Digital Twins

---

- Detailed Literature Review of 84 Publications on the Application of Digital Twin methods for Uncertainty Quantification
  - UQ Not widely adopted for migration across modeling environments
  - Focus has been on Physics-Based Modeling and Simulations
  - IP Protection with Digital Twin Implementation a Primary Concern for adoption
- Developed Representative “Digital Twin” for Hypersonic Platform
  - Refined Model Based on Test Results for Mission Evaluation
  - Quantified Uncertainty for Focused Flight Performance Characteristics
  - Integrated Model into an Uncertainty Quantification Workflow
  - Lacked Full System For Closed Loop Maturation

# T&E for AI-Enabled Systems (8 projects)

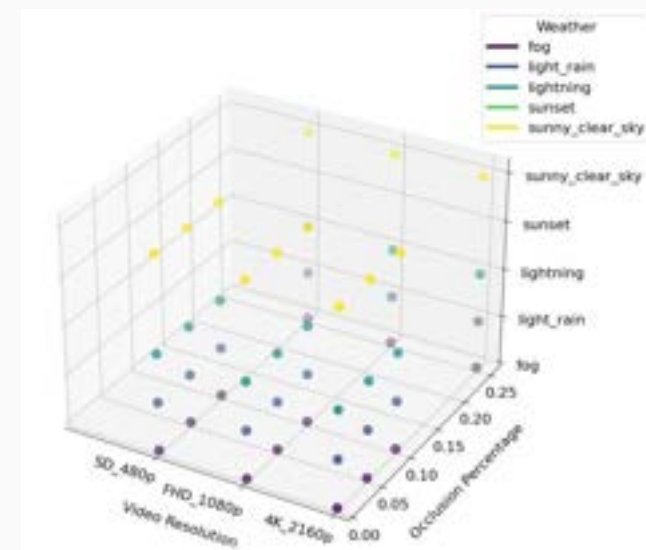
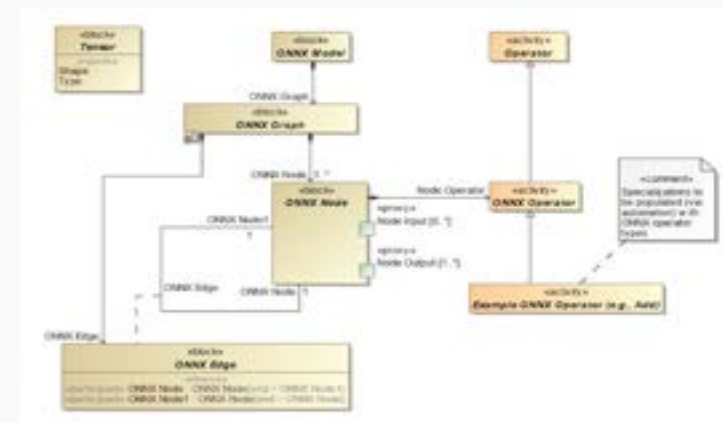
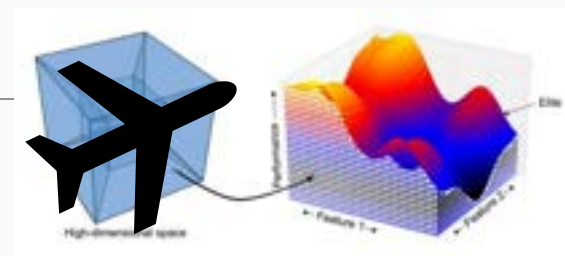
- *T&E for AI/ML Best Practices*: provided comprehensive understanding of current T&E practices, methods and challenges across the ML-enabled system's lifecycle broken into 1) component phase, 2) integration and deployment phase, and 3) post-deployment phase.
- *Hierarchical scoring*: method for fine-grained evaluation of errors for classification/object detection given structure over class labels and operator-relevant distances between labels
- *Systematic inclusion/exclusion (SIE)*: an experimental framework drawing on design of experiments and combinatorial testing to identify critical features defining dimensions of a model's operating envelope
- *Coverage of data explorer (CODEX)*: Python tooling that implements metrics and algorithms for data coverage applied to AI/ML T&E
- *Systems Theoretic Processes Analysis (STPA) for AI Ethics Assessment*: developed a loss-driven engineering framework to evaluate ethical risks of ML deployment concerning disengagement within UAV operations





## T&E for AI-Enabled Systems (8 projects)

- *Cognitive Electronic Warfare (CogEW)*: research evolved statistical processes for collection of most informative data points and develop CogEW exemplar use case. Core challenges to adequately test CogEW: 1) scale, 2) accessibility, and 3) Efficient test processes and execution. Recommend 1) adaptative sequential experimental design (ASED) vs. larger, static designs and 2) conformal prediction to add uncertainty quantification information to test points vs. just error/variance driving ASED process.
- *Model-Based Test of AI using Systems Modeling Language (SysML) and Open Neural Network Exchange (ONNX)*: ONNX captures essential concepts needed to produce a useful system model, SysML constructs sufficient to describe the structure and associated behavior of a learning system.
- *T&E of Multi-fidelity AI models*: developed a sequential multi-fidelity T&E method integrating information from multiple levels of representativeness of system and environment and demonstrated on computer vision object detection in autonomous driving use case.



# T&E Contributions 2024 – By the Numbers!

---

- **87** Products
  - **14** Events coordinated/participated
  - **73** Publications and products – 3 best paper awards!
- **17** research topics
- **7** universities/institutions
- **59** researchers/contributors
  - **7** Undergraduate researchers
  - **11** Graduate researchers

# Where the Research Leads: Compare and Contrast

## Our T&E Culture



### CURRENT STATE

- CT, DT, LF, OT, and M&S operate in separate silos
  - Operational contexts not introduced until late in T&E campaigns → late deficiency discovery, cost overruns, delays



### FUTURE STATE

- Integrated T&E that collects and analyzes targeted data to maximize operational performance knowledge early and often, while optimizing test resources
- Deficiencies discovered early, when they are easiest and least expensive to fix



Coordination, Cooperation,  
and Teamwork



# Where the Research Leads: Compare and Contrast

## Our T&E Technology – Data Repositories and Linkages



### CURRENT STATE

- Scattered spreadsheets
- Data is not VAULTIS
- Unsuitable for modern data analytics and AI
- No linkage → Decisions not necessarily data-driven and T&E gaps can be difficult to identify



### FUTURE STATE

- Streamlined databases
- Data is VAULTIS
- Suitable for modern data analytics and AI
- Transparent linkage across acquisition decisions, operational performance metrics, test events, test data, and M&S  
→ Data-driven decision-making and straightforward T&E gap identification



Geographically Distributed Enterprise Data and Analysis Architecture and Tools – Digital Transformation that meets all users where they are

**Acronyms:** T&E = Test and Evaluation, VAULTIS = Visible-Accessible-Understandable-Linked-Trustworthy-Interoperable-Secure, AI = Artificial Intelligence, M&S = Modeling and Simulation



# Where the Research Leads: Compare and Contrast

## Our T&E Methodology – Experimental Design and Body of Evidence



### CURRENT STATE

- Early 1900s classical design of experiments
  - Static test design
  - Data collected during different phases of test typically cannot be combined
- Test resources are not optimized, resulting in redundant testing or knowledge gaps



### FUTURE STATE

- We invent dynamic, adaptive design of experiments and implement them in easy-to-use, real-time software
  - Data collected during all phases of test is combined and analyzed as a “whole body of evidence”
  - Dynamically optimized test resources



Bayesian Inference + Modern  
Computer Hardware and Software



# Where the Research Leads: Compare and Contrast

## Our T&E Methodology – M&S and Its V&V



### CURRENT STATE

- The use and V&V of M&S is sporadic and unsynchronized
    - It is often an afterthought to apply M&S to T&E after it was developed for another purpose
    - V&V is typically not in accordance with the lifecycle approach detailed in DoDI 5000.09 “M&S VV&A”
- Lost opportunities to credibly apply M&S more broadly for the purposes of T&E across program lifecycles

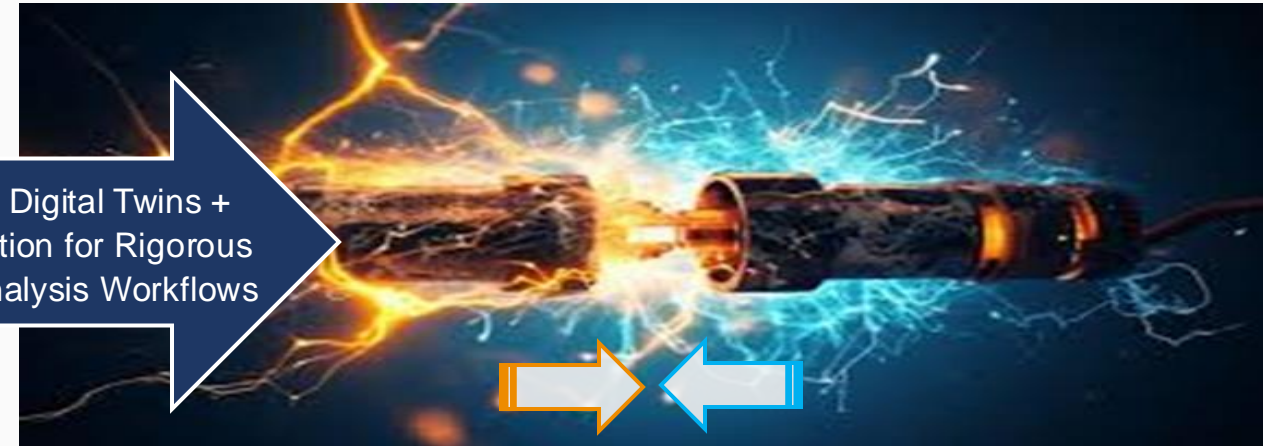


### FUTURE STATE

- A deliberate lifecycle approach to developing and “V&V’ing” M&S for the purposes of T&E, that is synchronized with live test data in accordance with DoDI 5000.09 “M&S VV&A” to generate a “predict, live test, refine” feedback loop
  - Credibly interrogate complex scenarios that might not be possible to orchestrate exclusively in the live domain
  - Expose early prototypes to operational contexts in digital worlds



Digital Engineering + Digital Twins +  
Uncertainty Quantification for Rigorous  
V&V Integrated Into Analysis Workflows



**Acronyms:** T&E = Test and Evaluation, M&S = Modeling and Simulation, V&V = Verification and Validation, VV&A = Verification Validation and Accreditation, DoDI = DoD Instruction

# Next Steps

---

- Supporting the rollout of updated DoD Instructions and Manuals to promote change in T&E practice
- Updated research topics
  - Kill Webs VVUQ
- Continue to emphasize research transition through:
  - **Workshops** and draft **supporting reports**
  - **Technical papers** and **articles**
  - **Example** products (models, software, tools, data structures, etc.)
  - **Frameworks** and **architectures** to shape T&E practices
  - **Courseware** for teaching and advancing workforce knowledge in proposed subject areas



# Thank you!


Stay connected with SERC Online:



Email the presenter: Laura Freeman

 [laura.freeman@vt.edu](mailto:laura.freeman@vt.edu)

Email the research team:

 [Peter Beling \(beling@vt.edu\)](mailto:beling@vt.edu)  
[Geoff Kerr \(geoffreykerr@vt.edu\)](mailto:geoffreykerr@vt.edu)