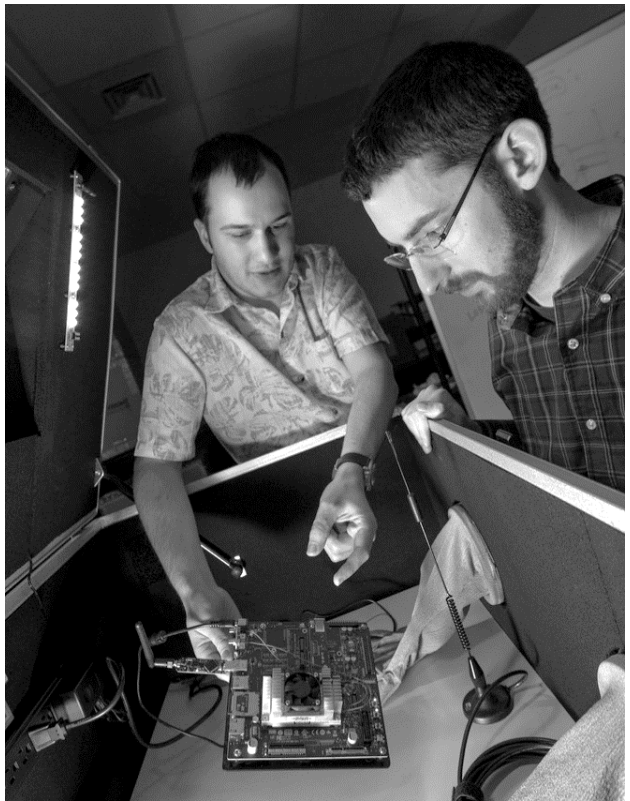


Tailorable Risk-Informed AI Test and Evaluation Strategy (TRAITES) Framework

Systems Engineering Research Center
AI4SE & SE4AI Workshop
September 17, 2025

Erin Lanus, Ph.D.	Research Associate Professor
Emma Meno	Research Associate
Laura Freeman, Ph.D.	Deputy Director

POC: lanus@vt.edu

NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

Motivating Risk-Based Testing for AI T&E

How do we test AI enabled systems to buy down risk?

What is risk? [1]

What can testing tell us?

When shouldn't we use testing?

What system knowledge is a precursor to test?

POV 2: AI different due to data and uses

POV 1: Know how to test complex systems

POV 3: AI T&E currently for fixing engineering issues, not pre-deployment validation

[1] K. L. Dempsey, L. A. Johnson, M. A. Scholl, K. M. Stine, A. C. Jones, A. Orebaugh, N. S. Chawla and R. Johnston, "Information security continuous monitoring (ISCM) for federal information systems and organizations," Gaithersburg, 2011.

CogEW Example

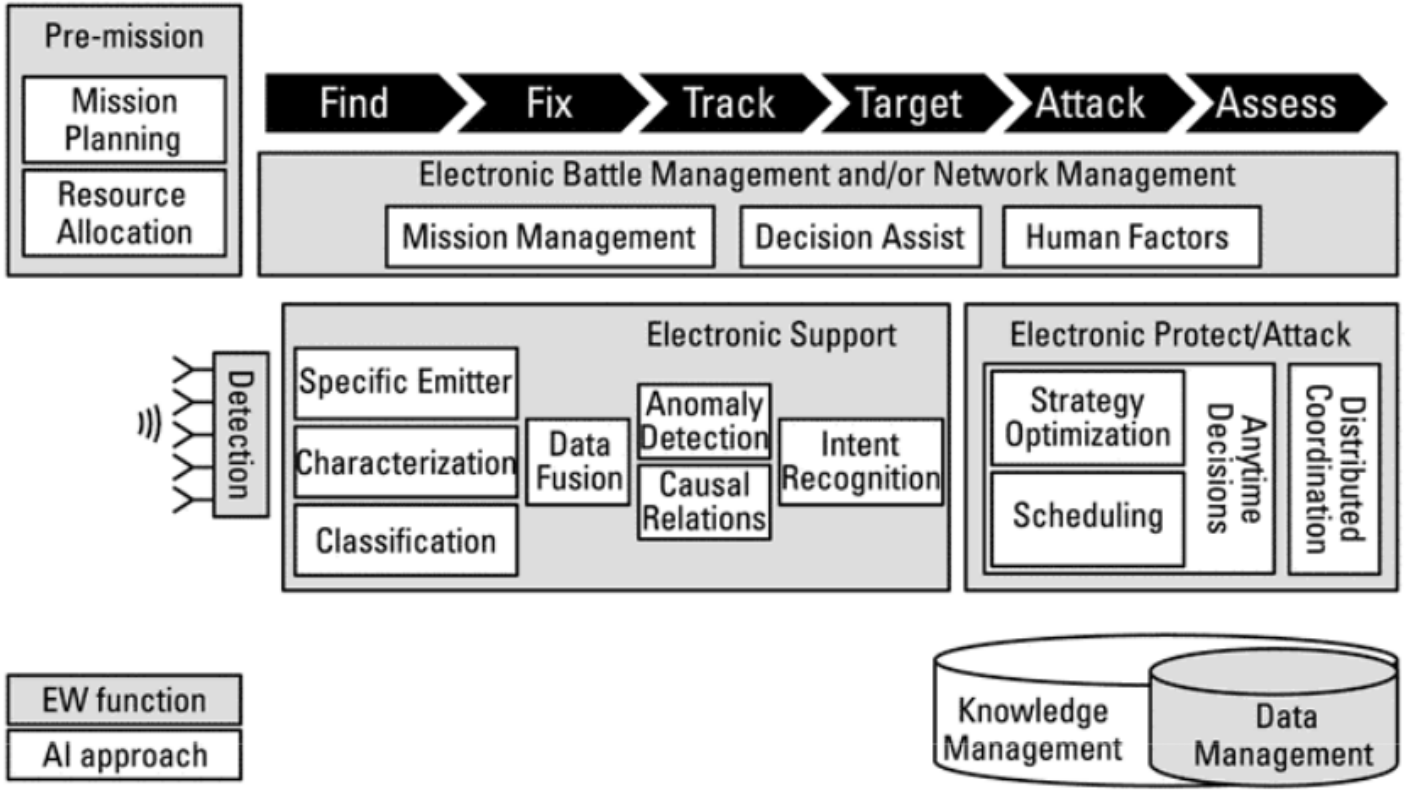


Figure 1.4 AI situation assessment, decision making, and learning capabilities are relevant for all EW functions.

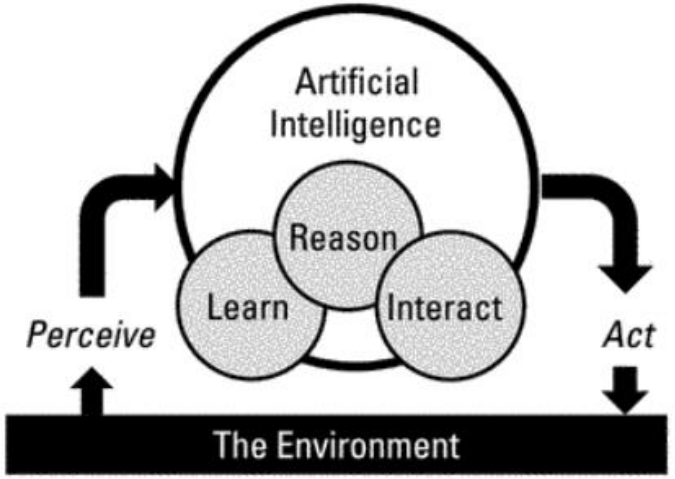


Table 1.1
EW Activities and AI Counterparts

AI Term	EW Term
Situation assessment	Electronic support
Decision making	Electronic protect and electronic attack Electronic battle management
Execution monitoring	Electronic warfare battle damage assessment
Learning	Electronic warfare reprogramming (of data and software)

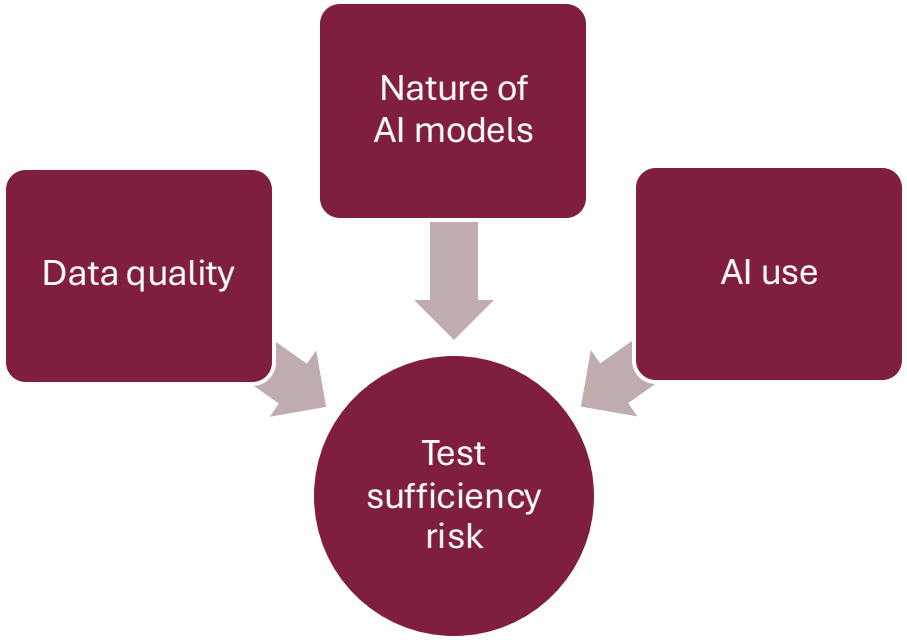
Graphics from Karen Haigh; Julia Andrusenko, Cognitive Electronic Warfare: An Artificial Intelligence Approach, Artech, 2021.

Current State of AI/ML in Open-Source CogEW Research

AI/ML	CURRENT STATE of AI/ML		CogEW RESEARCH APPLICATIONS
	Models and Algorithms	Toolkits and Libraries	
Deep Learning	• ANNs • CNNs • RNNs	• Tensorflow • Keras	• RF Signal Filtering and Denoising • RF Signal Identification and Classification • CogEW System Threat Assessment, Strategy, and Behavior Recognition
Computer Vision	• CNNs	• You-Only-Look-Once (YOLO) • OpenCV • PyTorch	• Visually Detecting and Deinterleaving RF Signals • LPI Radar Waveform Recognition • Target Tracking and Detection
Reinforcement Learning	• Q-Learning • DQN and DDQN • DDPG	• Open AI Gym • Stable Baselines • TensorForce	• Anti-Jamming Decision Making for Cognitive Radio/Radar • Jamming Decision Making Against Cognitive Radio/Radar • Resource Allocation and Task Scheduling • Electronic Reconnaissance and Target Searching

[2] K. Haigh, T. Nguyen and Test Resource Management Center, "Challenges of Testing Cognitive EW Systems," in 2023 IEEE AUTOTESTCON, National Harbor, MD, 2023.

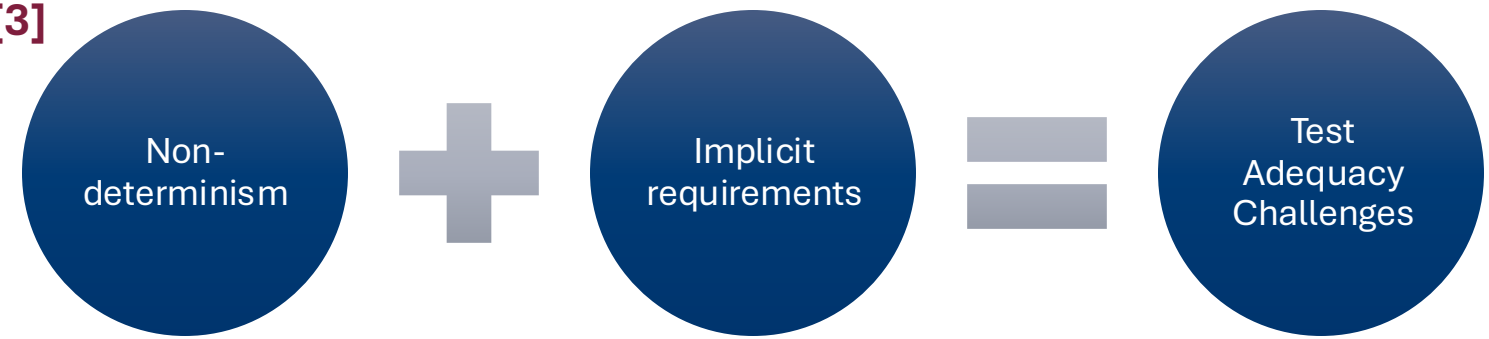
Existing Frameworks



What can we learn from existing frameworks?

AI risks differ from traditional software [3]

Differences in T&E between AI and traditional software [4]



[3] E. Tabassi, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD, 2023.
[4] C. Balhanalvy, I. Chen, R. W. Ferguson, J. Lockett, D. Moore, C. Pomales and F. Reeder, "Systems Engineering Processes to Test AI Right (SEPTAR) Release 1," McLean, 2023.

Distillation: Testing Traditional Systems vs. AI

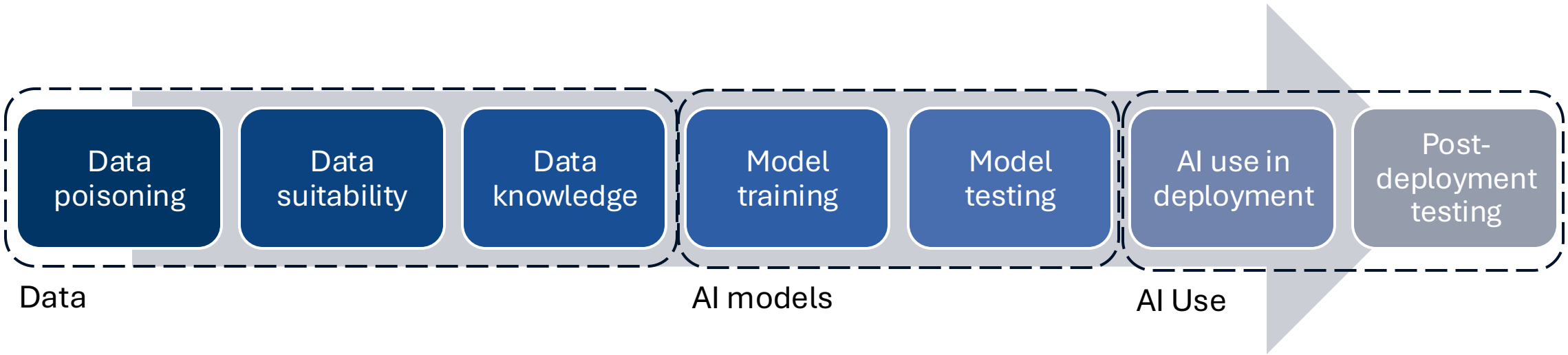
Same as Traditional System

- Fidelity/# points dependent on level of test
- Test decomposed system AND integration
- Metrics chosen based on system requirements
- Domain SMEs involved in test planning
- Security and safety testing apply
- DOE informs test budgeting

Unique to AI

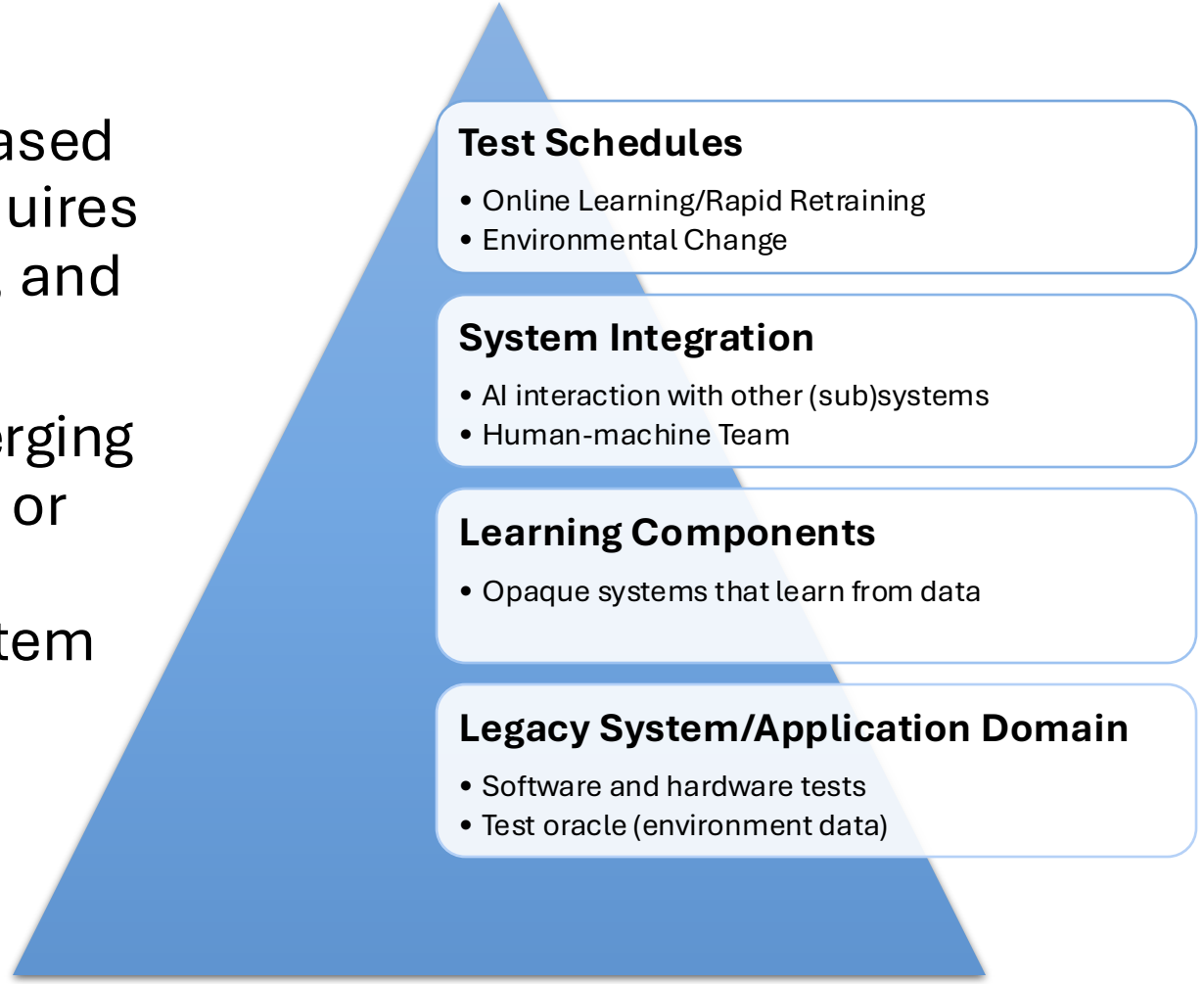
- How requirements are specified (from data)
- How functionality is created (from data)
- Interfaces between AI and other components create new attack vectors
- May not be decomposable (emergent behavior)
- Dynamically varying state of system (online learning)

Positioning Risks Along AI Lifecycle

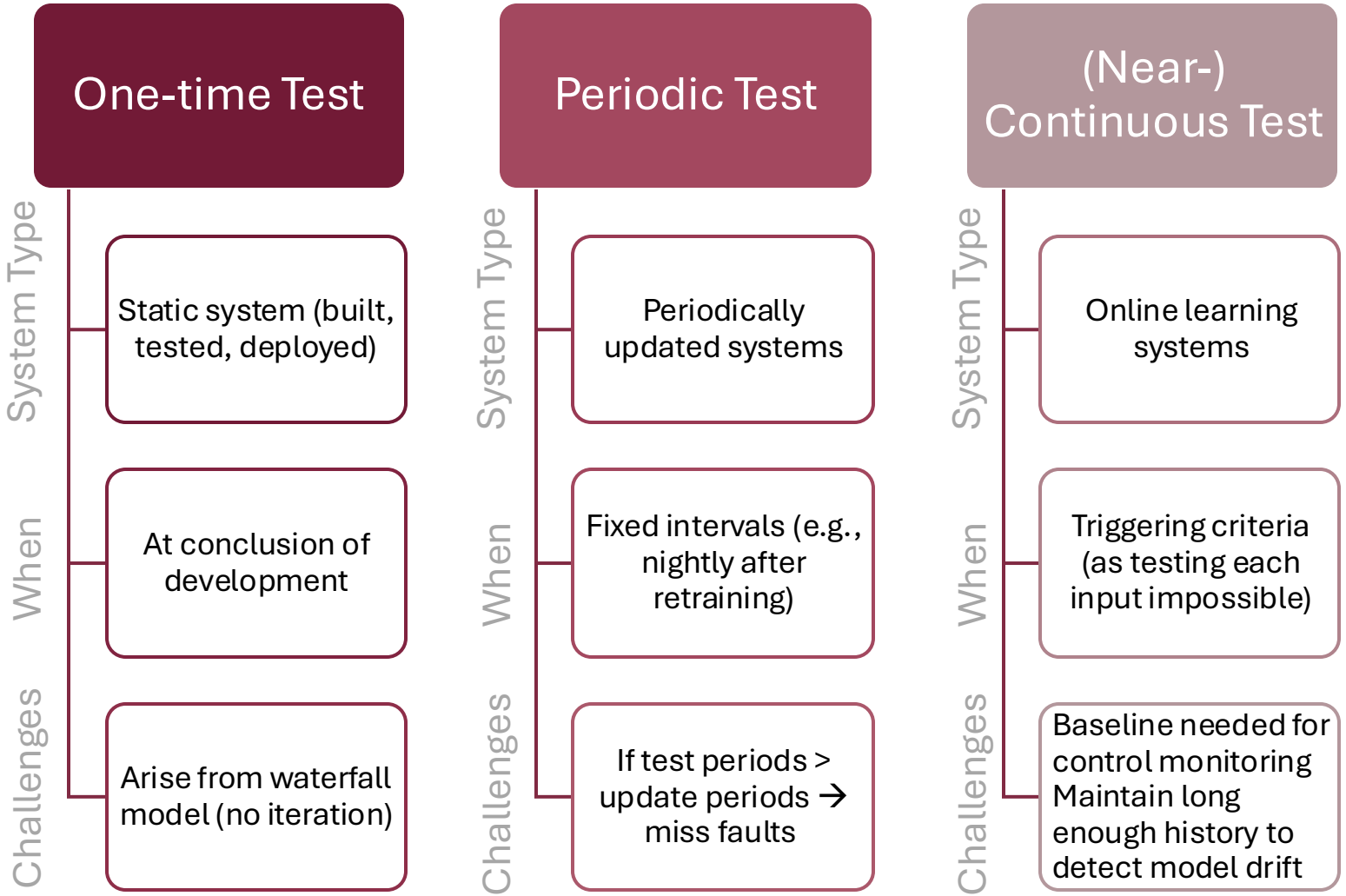


Hierarchy of Dependencies for Risk-based Level of Test

- Testing to buy down mission-based risk in emerging technology requires appropriate methods, metrics, and test designs for each level
- Focus on what changed in emerging technology without reinventing or forgetting best practices for application domain/legacy system



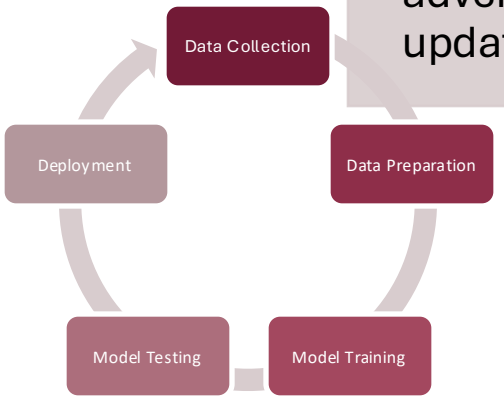
Test Schedule Appropriate to System Change



Recommendation: TRAITES Framework

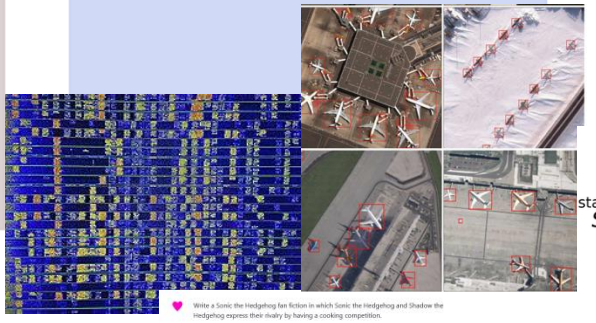
AI Lifecycle Phase

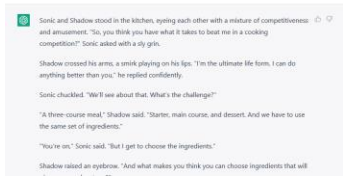
- data poisoning in initial data collection
- backdoor triggers in post-deployment
- post-deployment adversarial model updates



Problem Domain

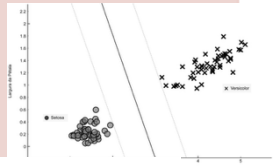
- Vehicle Detection
- RF signal Detection
- Misinformation Detection

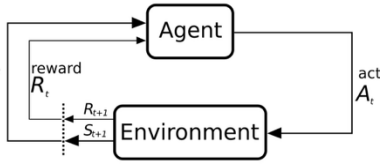


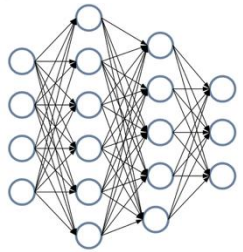


Learning Algorithm

- RL
- CNN
- LLM







AI <> System Interactions

- human-AI interactions
- AI-other software/hardware
- AI-environment

Tailorable Risk-informed AI T&E Strategy (TRAITES) Framework to

1. Categorize and enumerate risk introduced by AIES
2. Identify test methods/metrics to reduce risk