Operational Test & Evaluation Considerations for Emerging Technologies

Honorable Dr. Douglas C. Schmidt
Director, Operational Test & Evaluation,
Office of the Secretary of Defense

Operational & Live Fire Test & Evaluation (OT&E/LFT&E)

Mission: Evaluate the operational *effectiveness, suitability, survivability,* & (when necessary) *lethality* to defend our homeland & prevail in conflict

DOT&E in Context of the Department of Defense

Senate Armed Services Committee (SASC)

Senate Appropriations Committee Defense (SAC – D)

House Armed Services Committee (HASC)

House Appropriations Committee Defense (HAC – D)

Department of Defense
Secretary of Defense

Office of the Inspector General of the Department of Defense

Office of the Secretary of Defense

- Deputy Secretary of Defense,
- USD (Research & Engineering),
- USD (Acquisition & Sustainment),
- Director, Operational Test & Evaluation (DOT&E)
- Director, Cost Assessment & Program Evaluation,
- DoD Chief Information Officer

Department of the Army

Secretary of the Army

- Office of the Secretary of the Army
- The Army Staff

Department of the Navy

Secretary of the Navy

- Office of the Chief of Naval Operations
- Office of the Secretary of the Navy
- Head-Quarters Marine Corps

Department of the Air Force

Secretary of the Air Force

- Office of the Secretary of the Air Force
- The Air Staff
- Office of the Chief of Space Operations

Joint Chiefs of Staff

Chairman of the Joint Chiefs of Staff

The Joint Chiefs

Joint Staff (JROC)

Combatant Commands

- Africa Command
- Central Command
- Cyber Command
- European Command
- Northern Command
- Indo-Pacific Command
- Southern Command
- Special Operations Command
- Space Command
- Strategic Command
- Transportation Command

Services' Operational Test Agencies

- ATEC
- OPTEVFOR
- MCOTEA
- AFOTEC
- STARCOM
- JITC

Acronyms in slide: SASC – Senate Armed Services Committee; SAC-D – Senate Appropriations Committee Defense; HASC – House Armed Services Committee; HAC-D – House Appropriations Committee Defense; USD – Under Secretary of Defense; JROC – Joint Requirements Oversight Council; ATEC – Army Test & Evaluation Command; OPTEVFOR – Operational Test & Evaluation Force; MCOTEA – Marine Corps Operational Test & Evaluation Activity; AFOTEC – Air Force Operational Test & Evaluation Center

What is the Director, Operational Test & Evaluation (DOT&E)?

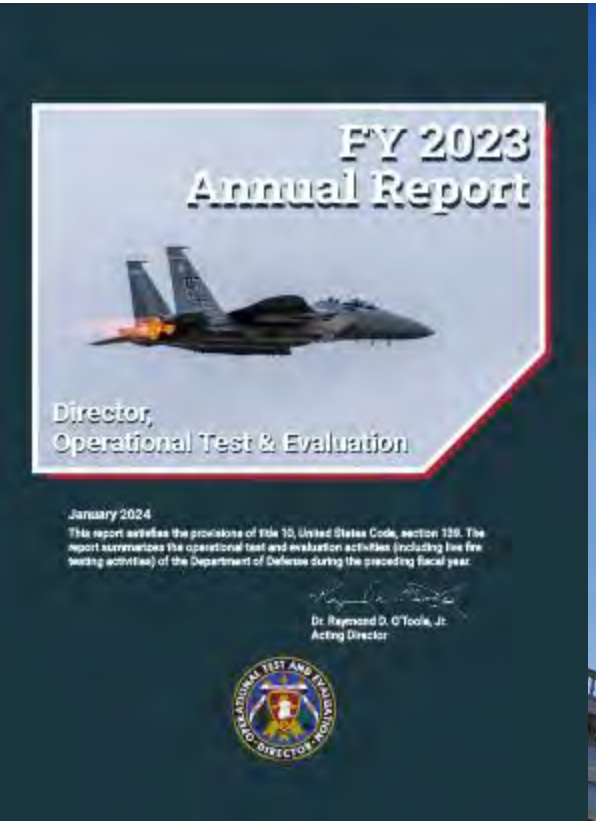
Policy & Guidance

Oversight

Reporting

Congressional Tasking

Strategic Initiatives



Test the way we fight	Architect T&E around validated mission threads & demonstrate the operational performance of the Joint Force in multi-domain operations
Accelerate the delivery of weapons that work	Embrace digital technologies to deliver high-quality systems at more dynamic rates
Improve survivability of DoD in a contested environment	Identify, assess, & act on cyber, electromagnetic spectrum, space, & other risks to DOD mission – at scale & speed
Pioneer T&E of weapon systems built to change over time	Implement fluid & iterative T&E across the entire system lifecycle to help assure continued combat credibility as the system evolves to meet warfighter needs
Foster an agile & enduring T&E enterprise workforce	Centralize & leverage efforts to assess, curate, & engage T&E talent to quicken the pace of innovation across the T&E enterprise

What Do These Systems All Have In Common?

All lacked adequate operational testing & evaluation!

Transformation of Today's Battlefield…

SEAMLESS COLLABORATION ACROSS ALL DOMAINS

ATTRITABLE SYSTEMS AT SCALE

ENABLED BY AUTONOMY & AI

Transformation of Today's Battlefield…

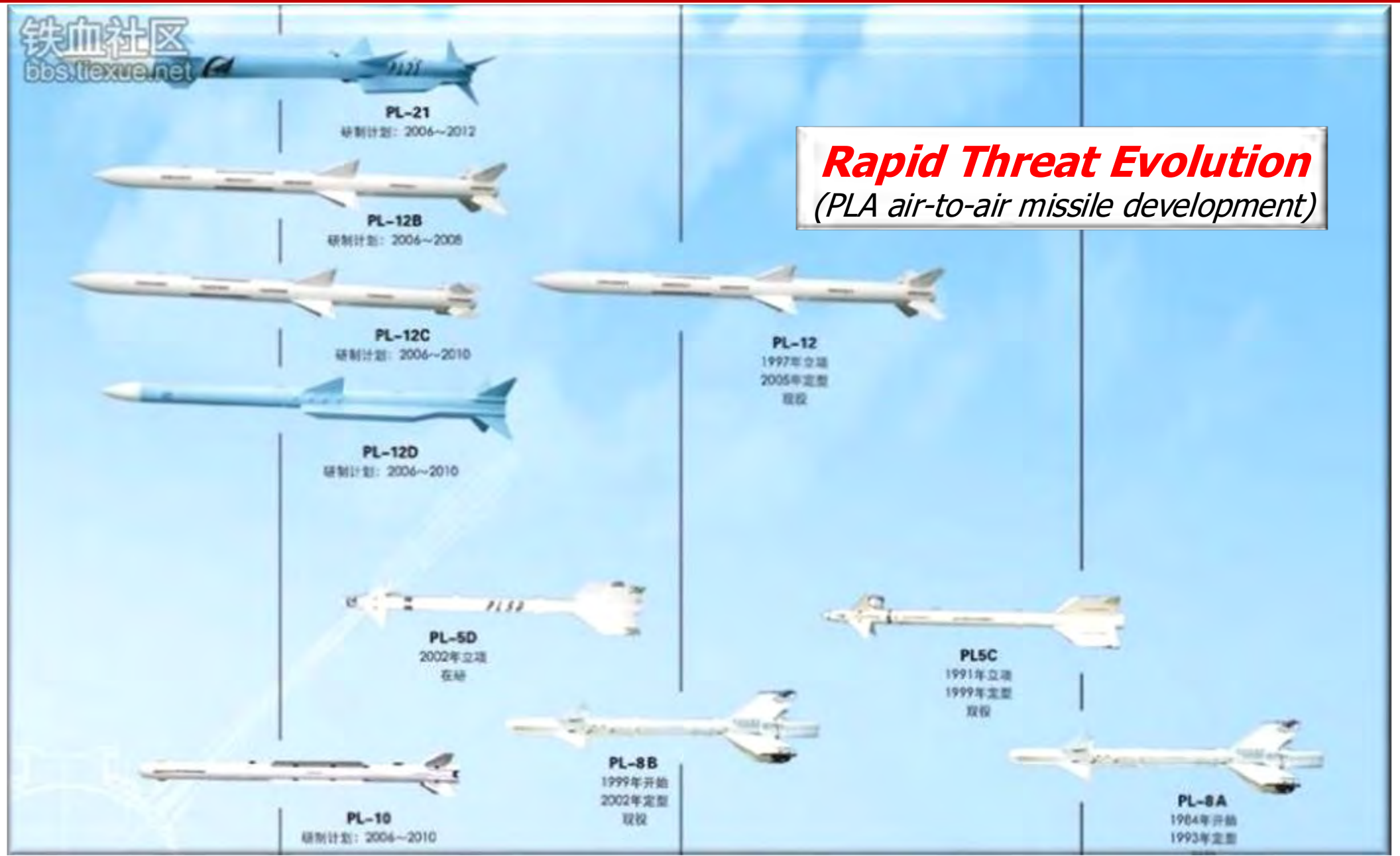
LONG-RANGE LETHALITY

ATTRITABLE SYSTEMS AT SCALE

CONTESTED ELECTROMAGNETIC SPECTRUM

ANTI-ACCESS / AREA DENIAL (A2/AD)

Transformation of Today's Battlefield…



Rapid Threat Evolution

(PLA air-to-air missile development)

DOT&E Strategic Pillars

DOT&E Strategy Implementation Plan (I-Plan)

1. Test the way we fight

- Standardize the development of a scalable and adaptive representation of the multi-domain operating environment
- Implement measures, tools, and processes to efficiently evaluate kill webs and system-of-systems performance

2. Accelerate delivery of weapons that work

- Develop and implement an enterprise-level T&E data management solution
- Integrate T&E in model-based engineering to operationalize and optimize the Shift Left approach

3. Improve the survivability of DoD in a contested environment

- Standardize and automate mission-based risk assessments
- Emphasize cyber and electromagnetic spectrum survivability
- Evaluate operational performance in a contested space environment

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

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

5. Foster an agile and enduring T&E enterprise workforce

- Identify and track T&E workforce competencies and capabilities
- Assess and address critical T&E workforce professional development needs

www.dote.osd.mil/News/News-Display/Article/3380429/dote-strategy-implementation-plan-2023

DOT&E Strategic Pillars

PILLAR 1
Test The Way We Fight

PILLAR 3
Improve the Survivability of DoD in a Contested Environment

PILLAR 4
T&E of Weapon Systems Built to Change Over Time

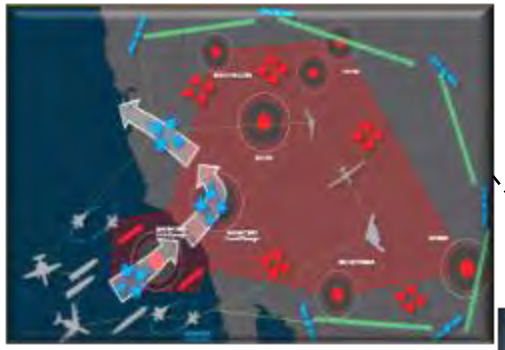
PILLAR 2
Accelerate The Delivery of Weapons That Work

PILLAR 5
Foster an Agile & Enduring T&E Enterprise Workforce

Test the Way We Fight

PILLAR 1
Test The Way We Fight

Provide tools, methods, & processes to evaluate complex missions in challenging all-domain operations



Open Air Battle Shaping (OABS)

Contested Operational Environments

Joint Simulation Environment (JSE)

Threat Radar Emulator

5th Gen Aerial Target (5GAT)

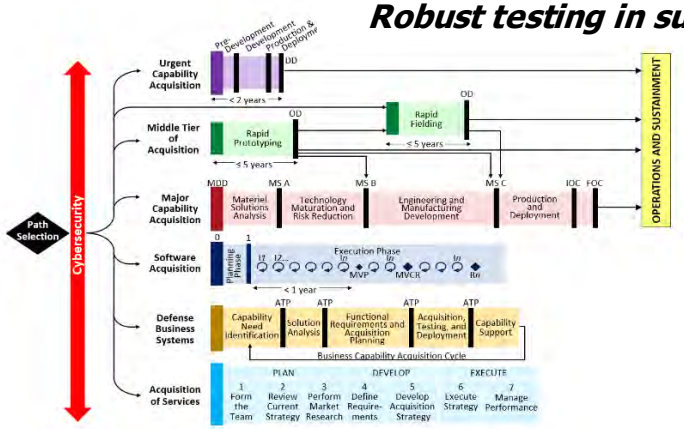
Goal: Eliminate failure on first use in combat

Accelerate the Delivery of Weapons That Work

Develop enterprise level T&E data management & analysis solutions that maximize the power of data

PILLAR 2
Accelerate The Delivery of Weapons That Work

Robust testing in support of rapid acquisition



Government cloud data access

Accessible, validated, & secure data for all stakeholders

Goal: Increase confidence that warfighters have the best available capabilities

Improve the Survivability of DoD in a Contested Environment

Satellite in an anechoic chamber

Stryker during Army Mounted Assured PNT System (MAPS) operational testing

PILLAR 3
Improve the Survivability of DoD in a Contested Environment

Enable a dynamic response to cyber & electromagnetic threats, & tests for full-spectrum survivability

Known & Expected Cyber Survivability Limits

Goal: Provide freedom of maneuver throughout the modern battlespace

T&E of Weapon Systems Built to Change Over Time

Automated evaluation of software-reliant system operational performance

PILLAR 4
T&E of Weapon Systems Built to Change Over Time

Multi-Domain / Multi-System Operations

Increase the use of digital twins & responsible AI in T&E & tools to support effective software T&E

Goal: Counter rapid threat evolution to remain current & relevant

Foster an Agile & Enduring T&E Enterprise Workforce

Automation & cyber-security skillsets

Augmented intelligence (AI+) & generative AI skillsets

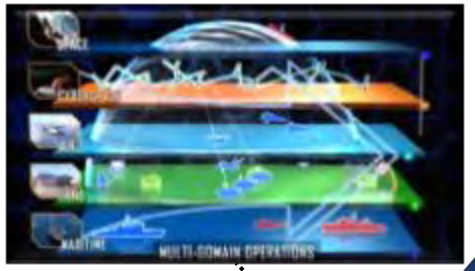
Organize, train, & equip the next-generation workforce

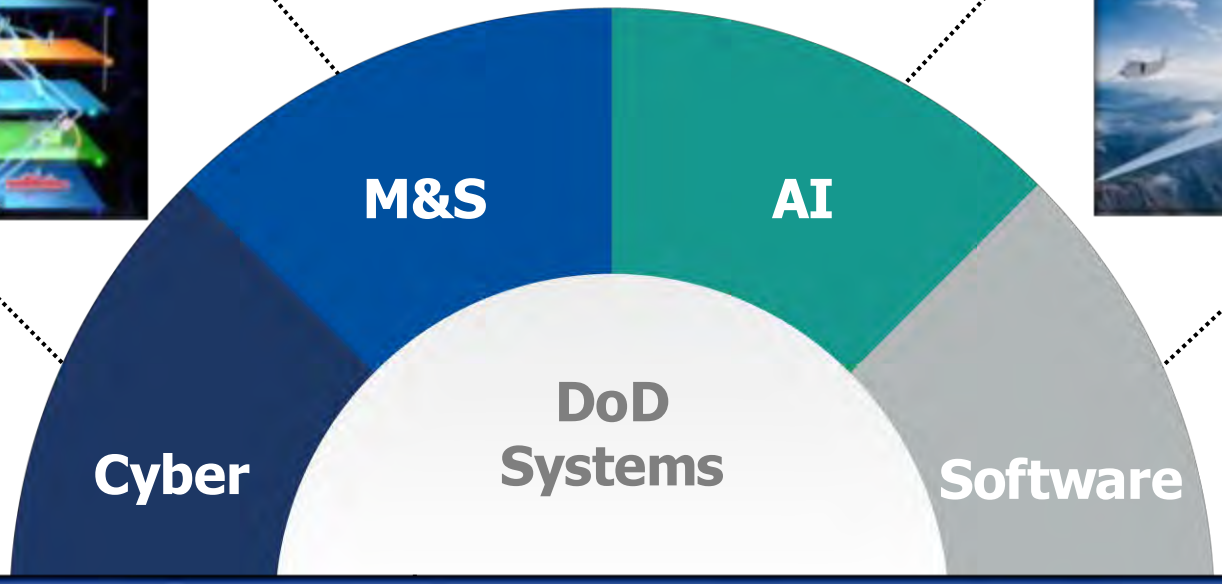
PILLAR 5
Foster an Agile & Enduring T&E Enterprise Workforce

Goal: Learn & embrace emerging technologies to leverage their benefits

Key Challenge Areas for Testing & Evaluation

Faster & more effective testing, dynamic adaptation to new scenarios & threats, & adequate testing in environments that are physically hard to replicate





The advantage in future conflicts will accrue to whichever side can fix & improve their software most rapidly & reliably

DoD Polices for OT&E & LFT&E

DoDI 5000.XF

Establishes policy, assigns responsibilities, & prescribes procedures for operational test & evaluation (OT&E) & live fire test & evaluation (LFT&E).

DoDM 5000.UX

OT&E & LFT&E input to the test & evaluation master plan (TEMP), a test & evaluation (T&E) strategy, or an equivalent artifact.

DoDM 5000.UW

Verification, validation, & accreditation (VV&A) of modeling & simulation (M&S) tools critical to meeting OT&E & LFT&E objectives.

DoDM 5000.96

OT&E & LFT&E of DoD software-intensive systems & services, & software embedded in systems & services.

DoDM 5000.UT

Realistic full spectrum survivability & full spectrum lethality testing of DoD systems & services.

DoDM 5000.UZ

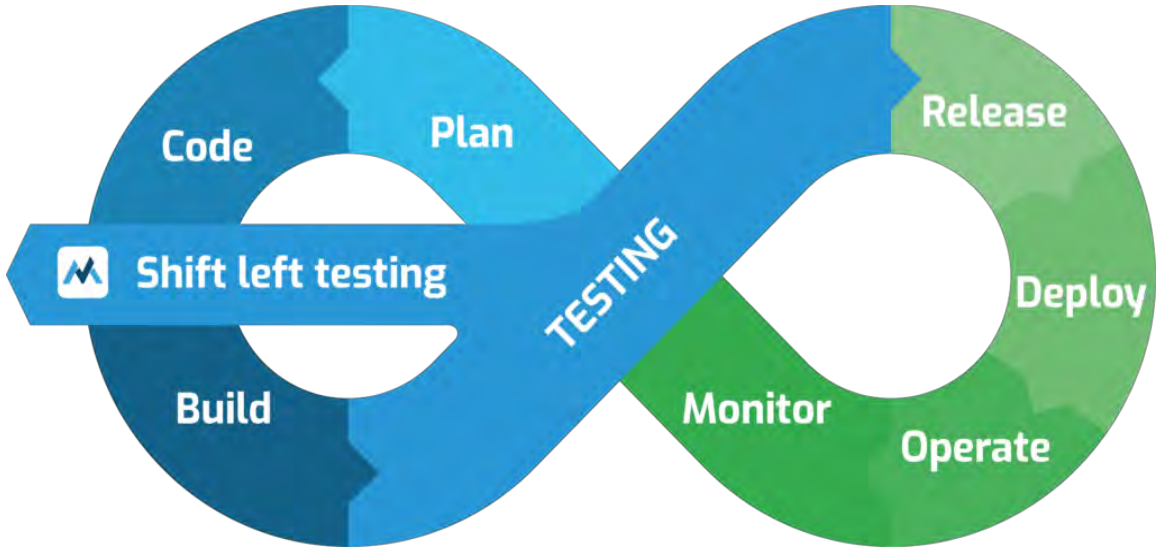
OT&E & LFT&E of artificial intelligence (AI)-enabled & autonomous systems & services.

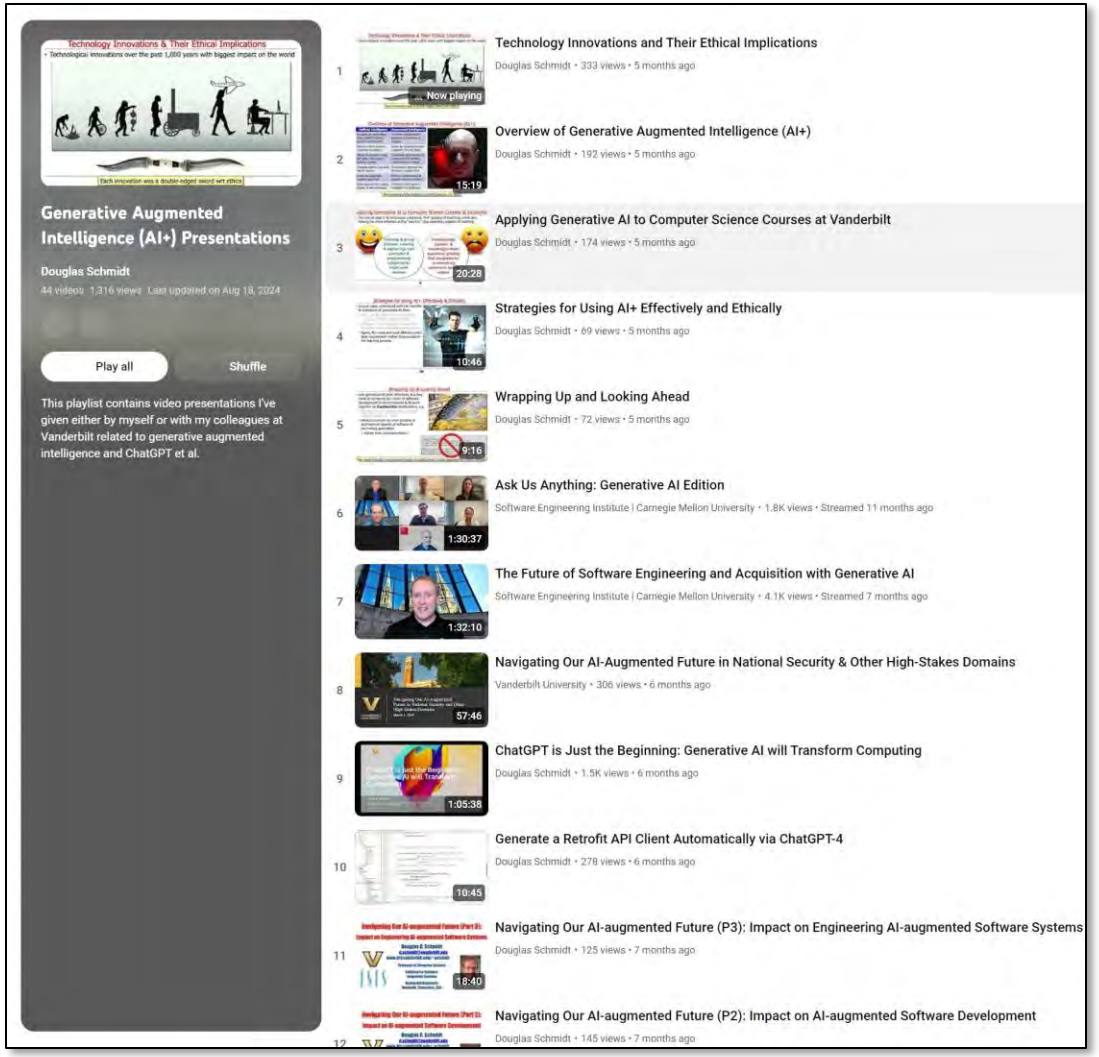
CREATE → **APPLY** → **AMPLIFY**

These policies should be released soon

Opportunities for Generative AI in T&E

- Augment various activities across the T&E enterprise





See www.youtube.com/playlist?list=PLZ9NgFYEMxp72Zo0yrTNS6utAXxYpqNGl

Opportunities for Generative AI in T&E

- Augment various activities across the T&E enterprise
  - Derive test cases from relevant design, policy, & requirement documents
    - e.g., use LLMs to analyze documents written in natural language
      - Ensure tests align w/specifications, policies, etc. from the outset

BY ORDER OF THE SECRETARY OF THE AIR FORCE

AIR FORCE MANUAL 91-119

5 JUNE 2012

Safety

SAFETY DESIGN AND EVALUATION CRITERIA FOR NUCLEAR WEAPON SYSTEMS SOFTWARE

COMPLIANCE WITH THIS PUBLICATION IS MANDATORY

ACCESSIBILITY: Publications and forms are available for downloading or ordering on the e-Publishing website at www.e-Publishing.af.mil

ChatGPT

Objectives	Using LLMs
Ensure document is clear & complete to ensure nuclear surety for key software components	Check for discrepancies • within 91-119 • between it & other documents

- **Ambiguity in Safety Certification Components:** On pages 32-33, 91-119 discusses safety certification for software components, suggesting to list them separately or combined with other safety-certified components. However, combining safety-certified components with non-safety-certified ones could complicate change tracking. Clearer guidelines are needed to avoid inconsistencies in how components are combined and tracked.

See insights.sei.cmu.edu/blog/applying-large-language-models-to-dod-software-acquisition-an-initial-experiment

Opportunities for Generative AI in T&E

- Augment various activities across the T&E enterprise
  - Derive test cases from relevant design, policy, & requirement documents
  - Use large language models (LLMs) to simulate diverse usage patterns & environments to test systems under various conditions
    - e.g., apply the *Persona* pattern

Act as a senior security engineer. You will help me investigate potential threats to my organization.

We will work together to investigate threats. I can run tools and software to gather information for us. I can cut/paste the outputs here for you to analyze.

You can ask me to do the following things:

1. Run a Linux command-line tool that I have access to and provide the output from the tool.
2. Run a Python program that you create to collect information and print it out to the terminal so that I can cut / paste it her for you to look at.
3. Write a Python script that I can run to query the NIST CVE database for known vulnerabilities related to the host, OS, services, etc. on a device and cut/paste the results for you to look at.

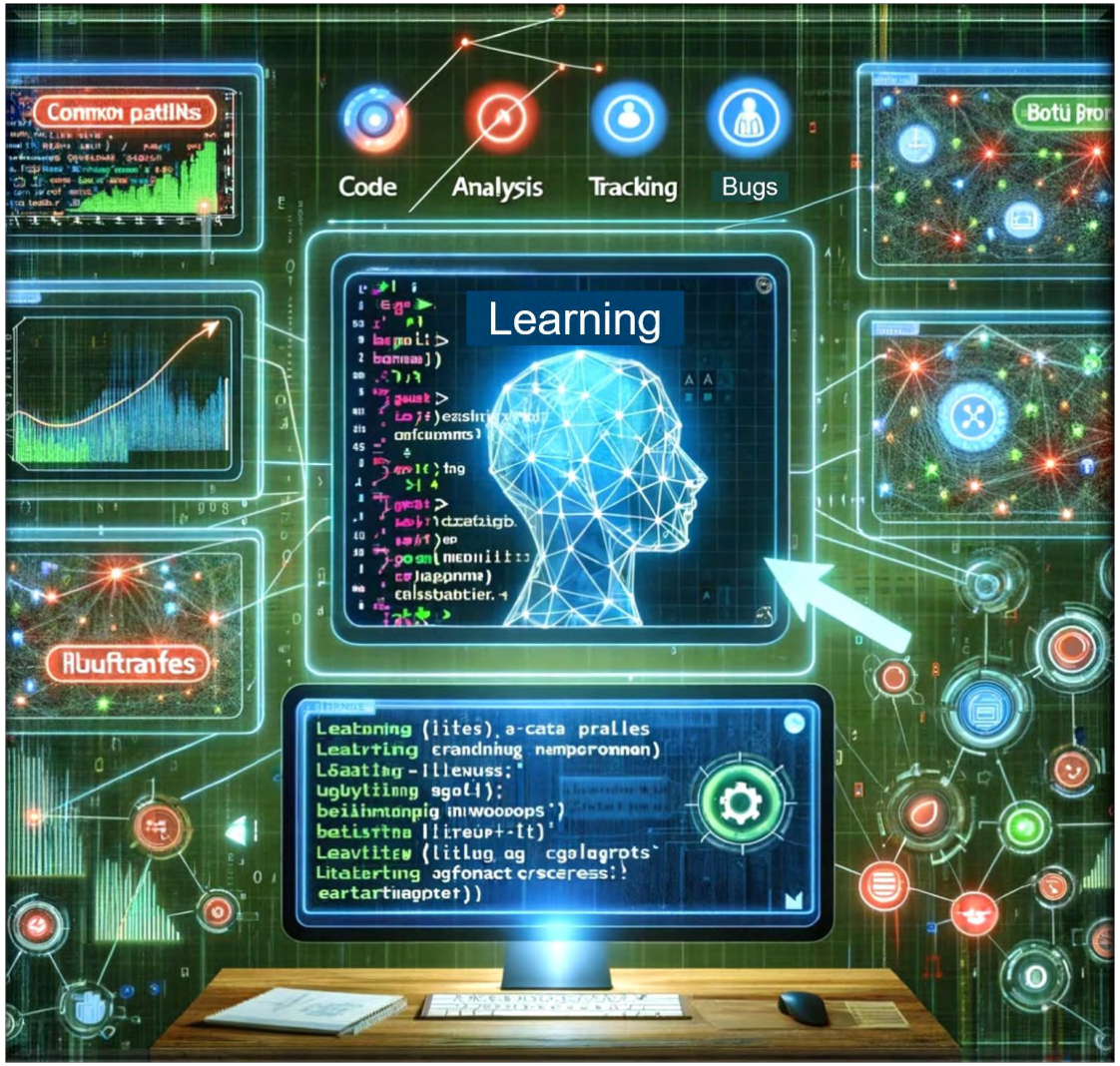
You will keep asking me to perform operations until you have enough information to recommend a plan of action. After each task you ask me to perform, remind me of what we are doing in a paragraph and then ask me for the input from the last task that you asked me to perform.

Ask me for the threat to investigate.

See www.dre.vanderbilt.edu/~schmidt/PDF/PLoP-patterns.pdf

Opportunities for Generative AI in T&E

- Augment various activities across the T&E enterprise
  - Derive test cases from relevant design, policy, & requirement documents
  - Use large language models (LLMs) to simulate diverse usage patterns & environments to test systems under various conditions
- Help testers & T&E organizations learn from prior efforts
  - Continuously improve the T&E process over time by analyzing test data to identify common pitfalls & best practices



See www.linkedin.com/pulse/5-ways-ai-disrupting-traditional-software-testing-process-sheldon

Opportunities for Generative AI in T&E

- Augment various activities across the T&E enterprise
  - Derive test cases from relevant design, policy, & requirement documents
  - Use large language models (LLMs) to simulate diverse usage patterns & environments to test systems under various conditions
  - Help testers & T&E organizations learn from prior efforts
- Identify trends in DoD acquisition processes over time
  - e.g., analyze DOT&E annual reports for the past three decades

See www.dote.osd.mil/annualreport

How to Help Deliver Weapons that Work… Faster

How to Help Deliver Weapons that Work… Faster

Enhanced Cyber Resilience

Operationally Representative Data

Digital-Physical Fusion

AI-/ML-enabled Systems & Workforce

How to Help Deliver Weapons that Work… Faster

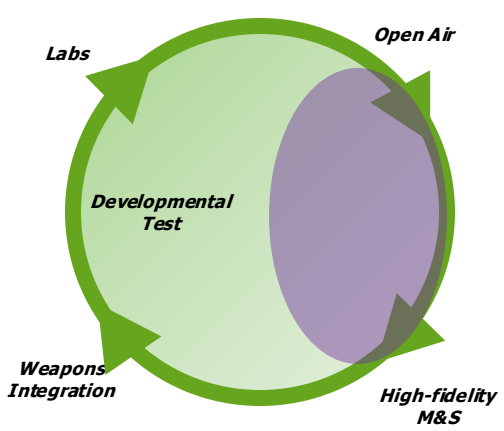
Enhanced Cyber Resilience

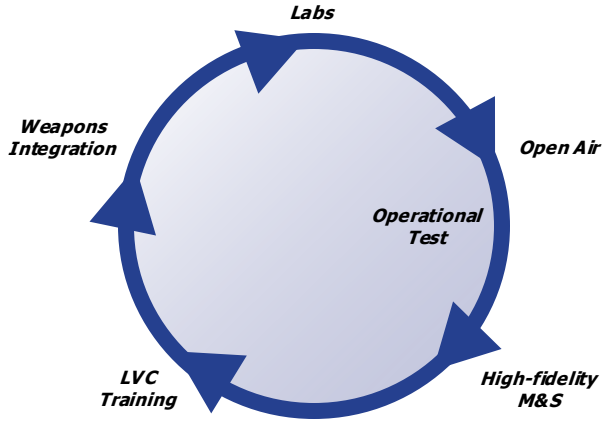
Operationally Representative Data

Digital-Physical Fusion

AI-/ML-enabled Systems & Workforce

Integrated Test





← Shift left, look right →

Questions & Answers

For more information please contact:

HON Dr. Douglas C. Schmidt
Director, Operational Test & Evaluation
Office of the Secretary of Defense
Tel. (703) 697–3655

See www.dote.osd.mil