DOT&E Background and Future Plans

HON Dr. Douglas Schmidt

Director, Operational Test & Evaluation,
Office of the Secretary of Defense

April 17, 2024

OPERATIONAL TEST AND EVALUATION

Purpose: Evaluate Operational Effectiveness, Suitability and Survivability to Defend the Homeland and Prevail in Conflict

U.S. Department of Defense Organization

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 and Engineering),
- USD (Acquisition and Sustainment),
- Director, Operational Test & Evaluation
- 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)

Services' Operational Test Agencies

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

Combatant Commands (9)

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

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 and Evaluation Command; OPTEVFOR – Operational Test and Evaluation Force; MCOTEA – Marine Corps Operational Test and Evaluation Activity; AFOTEC – Air Force Operational Test and Evaluation Center

What is the Director Operational Test & Evaluation?

Policy and Guidance

Oversight

Reporting

Congressional Tasking and Management

Strategic Initiatives

3

Education Background & Professional Experience

Education

BA & MA, Sociology

MS & PhD, Computer Science

Academics

Asst. Prof

Assoc. Prof

Cornelius Vanderbilt Professor
Associate Chair of Computer Science
Associate Provost of Research
Co-Director of the Data Science Institute

Government & Industry

Confirmed & sworn in as DOT&E

Deputy Director of Research & CTO

USAF Scientific Advisory Board

Co-Chair, Software Design & Productivity

Program Manager & Deputy Office Director

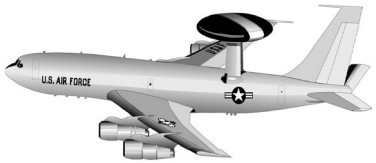
Board of Directors

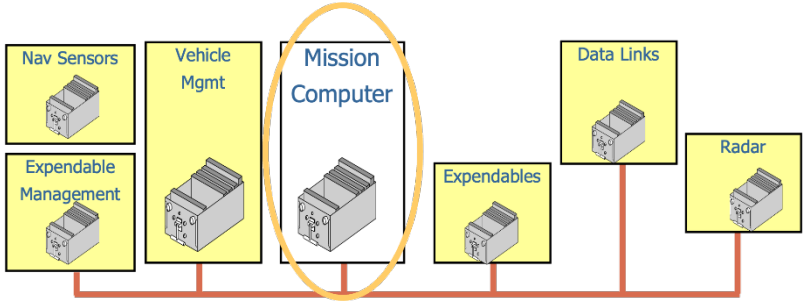
CTO

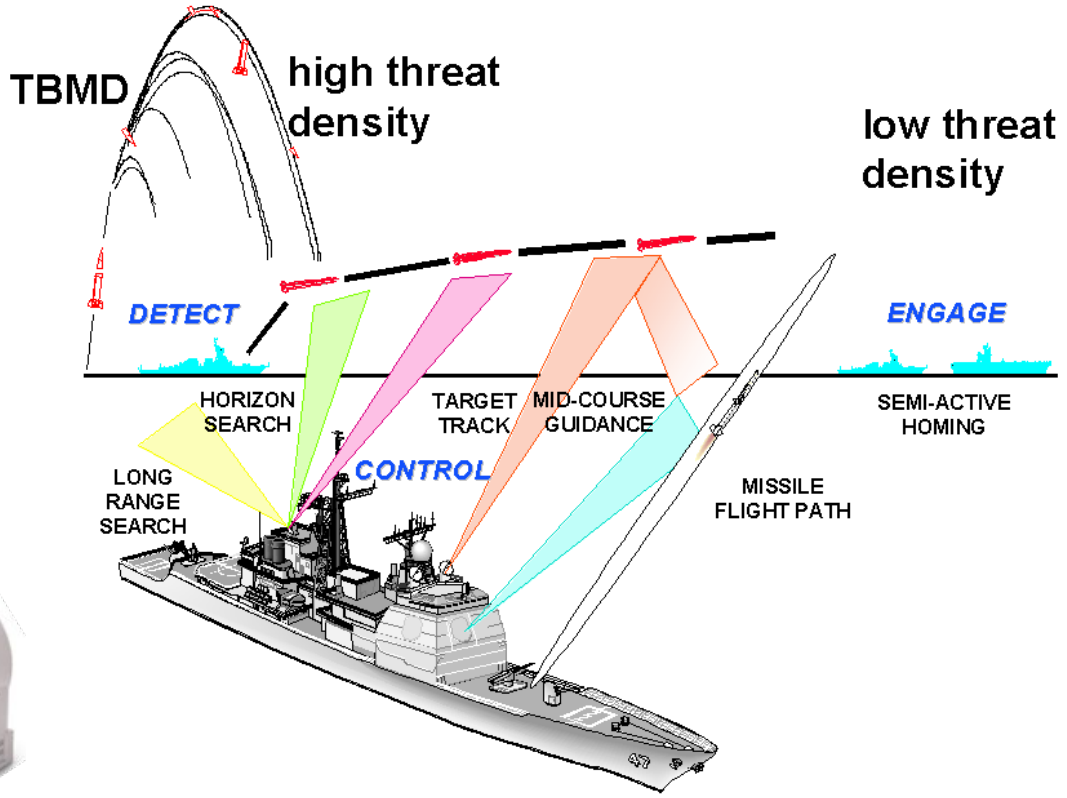
CTO

Research & Education Contributions (1990-1999)

My Focus — Software for distributed real-time & embedded (DRE) systems

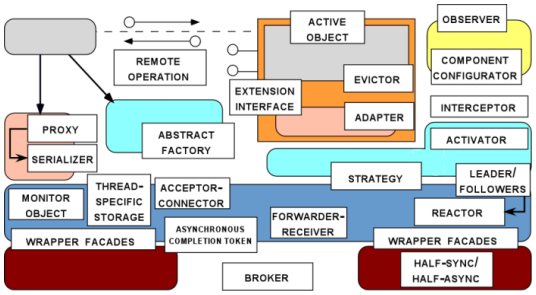






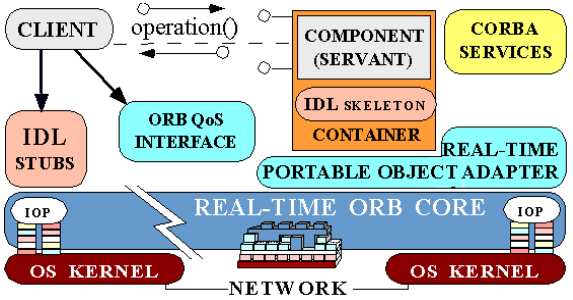
In DRE systems the “right answer” delivered too late becomes the “wrong answer”

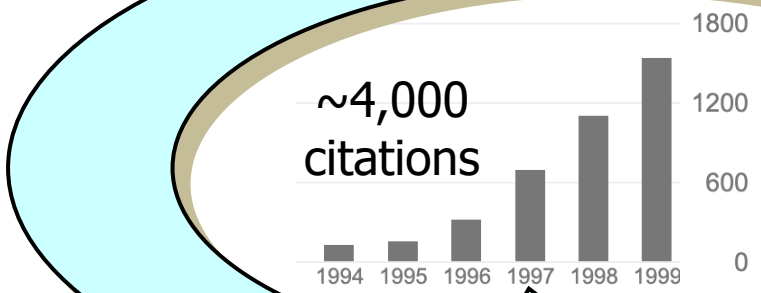
Research & Education Contributions (1990-1999)



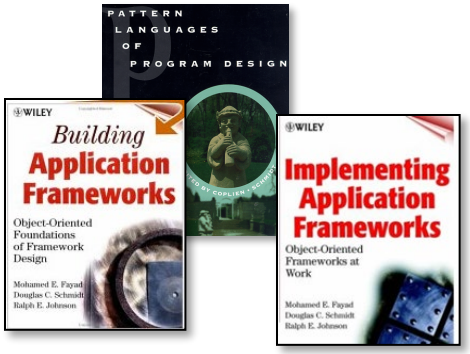
Software patterns that codify design experience

QoS-enabled middleware

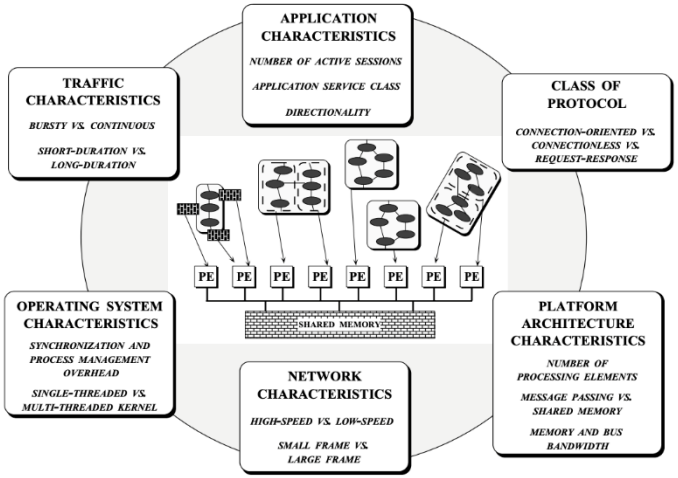




Object-oriented frameworks & patterns



Parallel protocol processing platforms



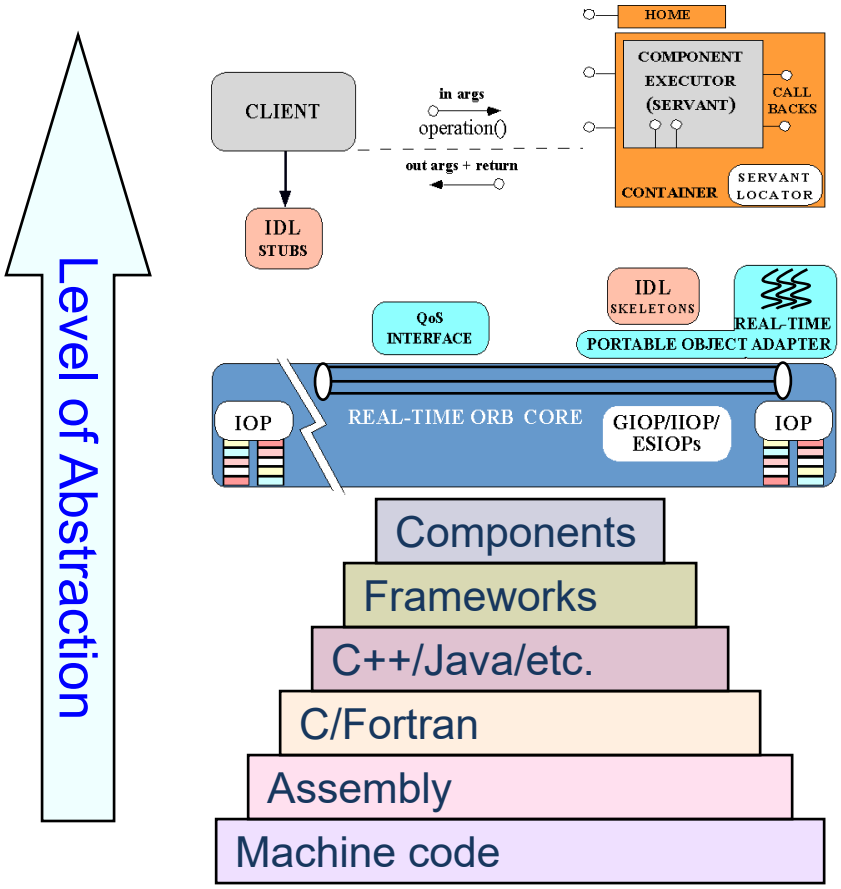
www.dre.vanderbilt.edu

Research & Education Contributions (2000-2009)

My Focus

Raising the level of abstraction for developing assured DRE & cyber-physical systems

Conventional Programming Languages & Platforms

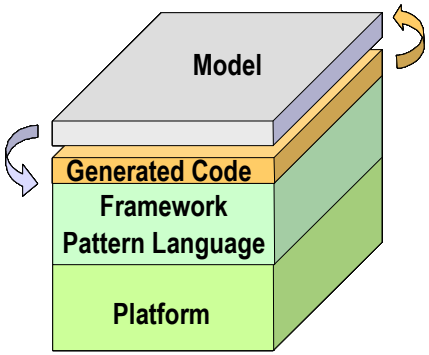


Research & Education Contributions (2000-2009)

My Focus — Raising the level of abstraction for developing assured DRE & cyber-physical systems

Conventional Programming Languages & Platforms

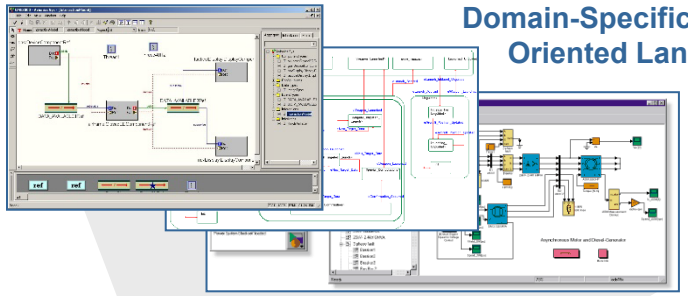
Level of Abstraction



Saturation!!!!

- Components
- Frameworks
- C++/Java/etc.
- C/Fortran
- Assembly
- Machine code

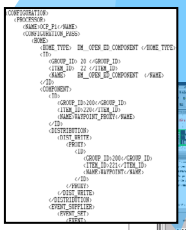
Model-Driven Engineering & Multi-Faceted Development



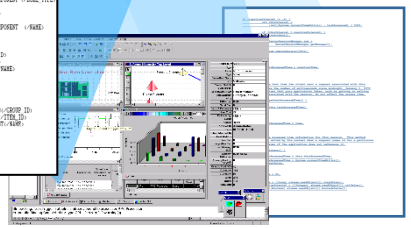
Domain-Specific & Aspect-Oriented Languages

Model-Driven Generator Technology

- Modeling of generators
- Generating generators
- Provably correct generators
- Embeddable generators



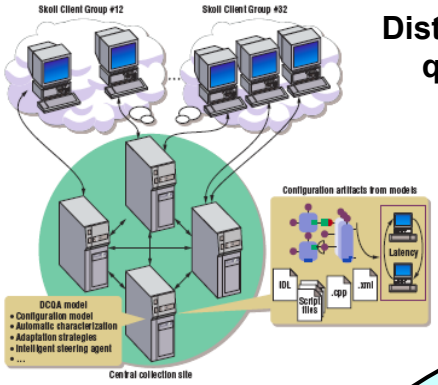
Configuration Specification



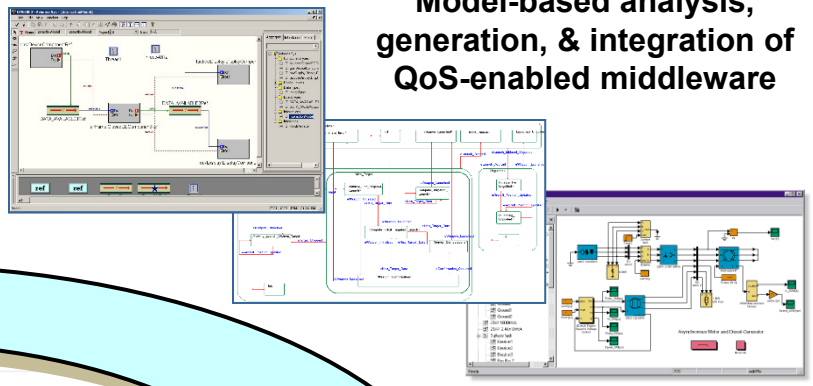
Analysis Tools

Code

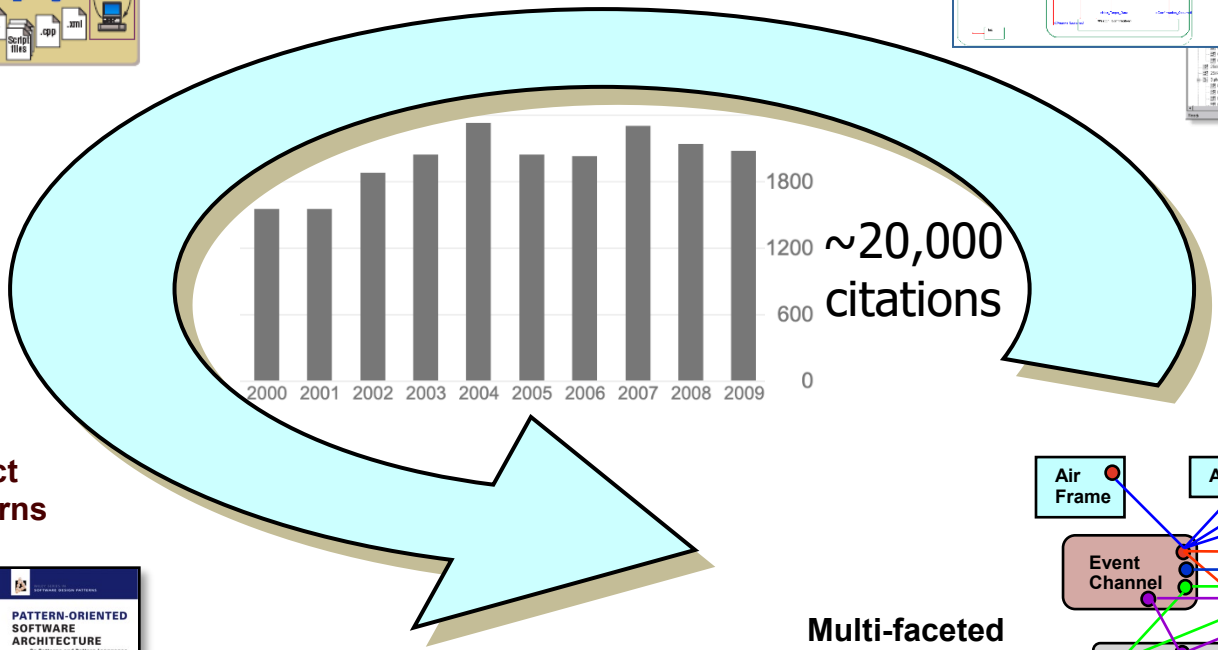
Research & Education Contributions (2000-2009)



Distributed continuous quality assurance techniques

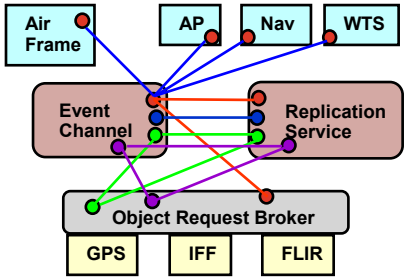


Model-based analysis, generation, & integration of QoS-enabled middleware



Distributed object computing & patterns

Multi-faceted Middleware Development



Cross-cutting Concerns

- Synchronization
- Persistence
- Memory Management
- Fault Tolerance

www.dre.vanderbilt.edu

Leadership Positions (2000-2009)

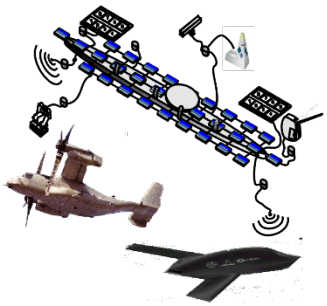
Government

Academia

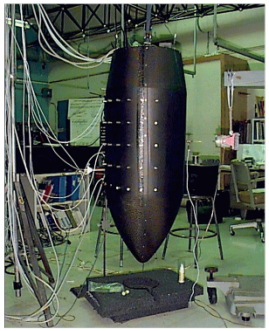
Commercial

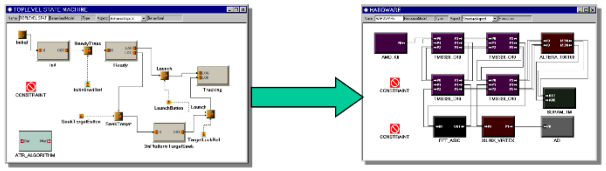
DARPA

Program Manager & Deputy Office Director, 2000-2002



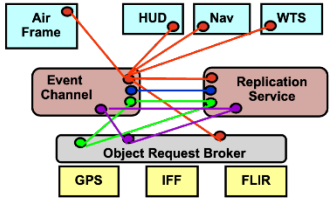
	Large Grain	Small Grain
System Technology	ARMS	ANTS
	SEC	NEST
Design Technology	MoBIES	PCES





Cross-cutting Concerns

- Synchronization
- Persistence
- Memory Management
- Fault Tolerance



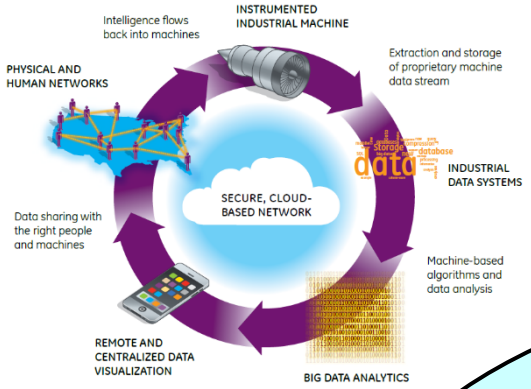
2000 2005 2010 2015 2020 2025

Research & Education Contributions (2010-2019)

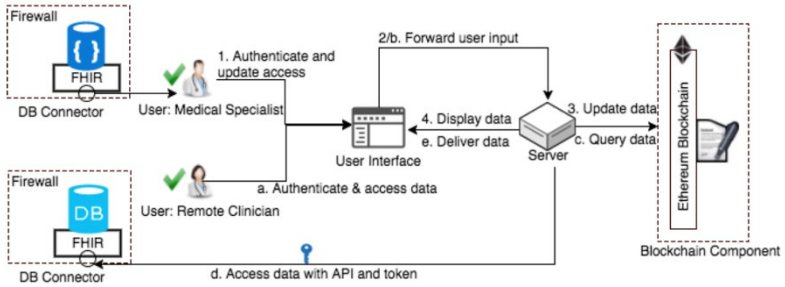
My Focus

Engineering enterprise-scale cyber-physical systems & mobile cloud computing applications

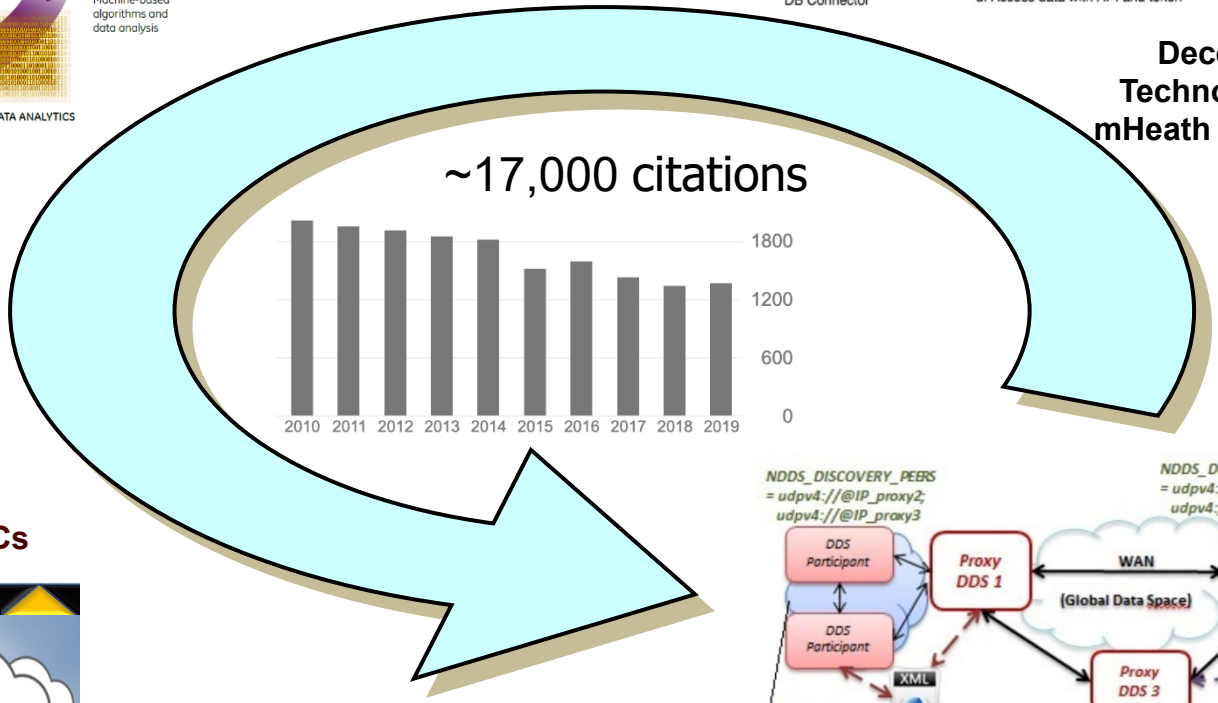
Research & Education Contributions (2010-2019)



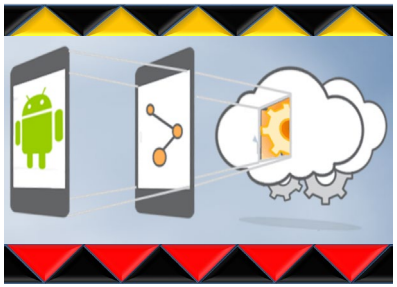
Industrial Internet



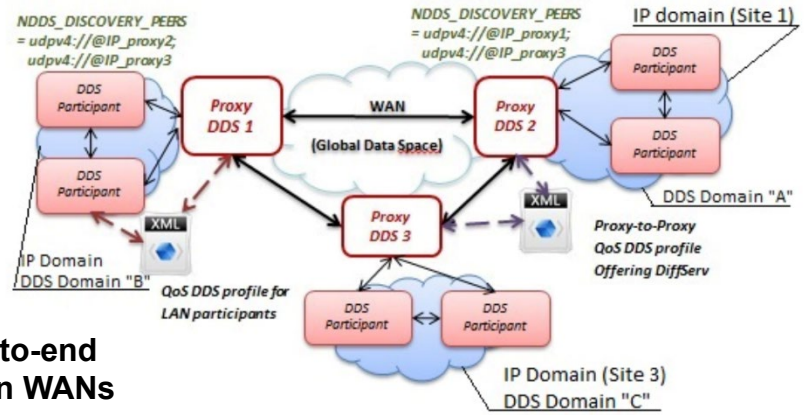
Decentralized Ledger Technologies & DApps for mHeath & Transactive Energy



Mobile Cloud Computing MOOCs



www.dre.vanderbilt.edu



End-to-end QoS in WANs

Leadership Positions (2010-2019)

Government

Academia

Commercial

Deputy Director of Research & Chief Technology Officer, 2010-2012

2000 2005 2010 2015 2020 2025

Leadership Positions (2010-2019)

Government

Academia

Commercial

Deputy Director of Research & Chief Technology Officer, 2010-2012

2000 2005 2010 2015 2020 2025

Leadership Positions (2010-2019)

Government

Academia

Commercial

Deputy Director of Research & Chief Technology Officer, 2010-2012

Advancing the Scope & Impact of SEI Research

DOUGLAS SCHMIDT (VANDERBILT UNIVERSITY)

FEBRUARY 7, 2011

When I joined the SEI last year, one of my top priorities was to advance the scope and impact of SEI R&D programs, along with increasing the visibility of the excellent work of SEI technologists who staff these programs. While the SEI is well known for its innovation and impact in several key areas, the breadth and depth of our expertise extends far beyond our most popular technologies. To increase awareness of all that we're doing, we've established a blog to highlight SEI R&D initiatives, methods, and solutions that are meeting the needs of our customers and partners in government and industry. This initial post will introduce the goals and themes of the new SEI blog.

insights.sei.cmu.edu

2000 2005 2010 2015 2020 2025

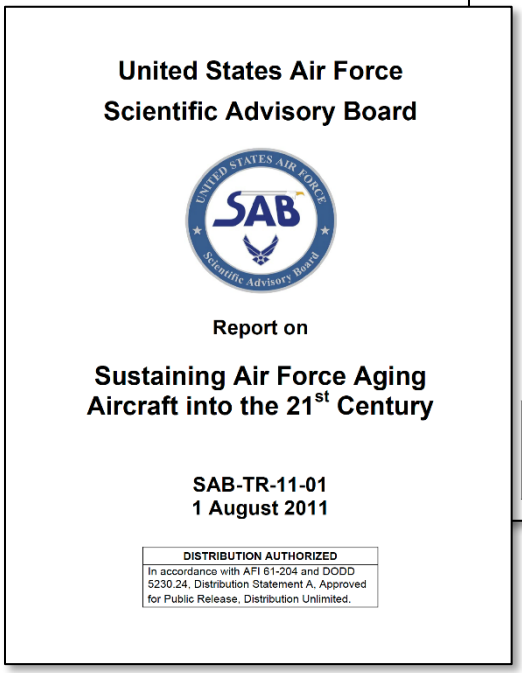
Leadership Positions (2010-2019)

Government

Academia

Commercial

Member of the Air Force Scientific Advisory Board, 2010-2014

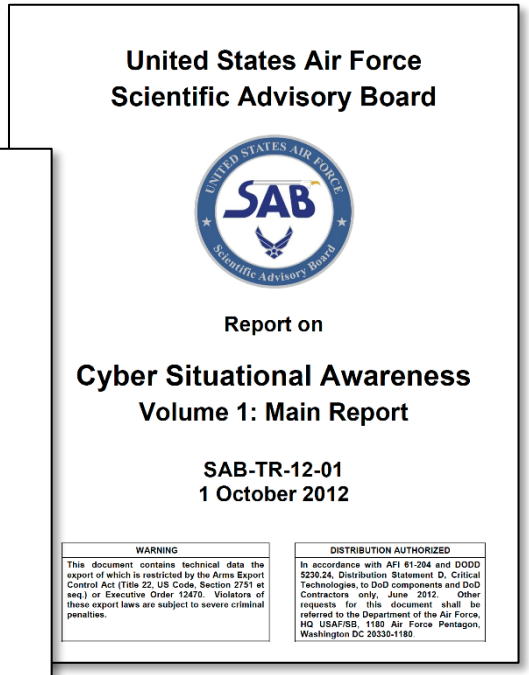

United States Air Force
Scientific Advisory Board

Report on

Sustaining Air Force Aging Aircraft into the 21st Century

SAB-TR-11-01
1 August 2011

DISTRIBUTION AUTHORIZED
In accordance with AFI 61-204 and DODD 5230.24, Distribution Statement A. Approved for Public Release, Distribution Unlimited.


United States Air Force
Scientific Advisory Board

Report on

Cyber Situational Awareness
Volume 1: Main Report

SAB-TR-12-01
1 October 2012

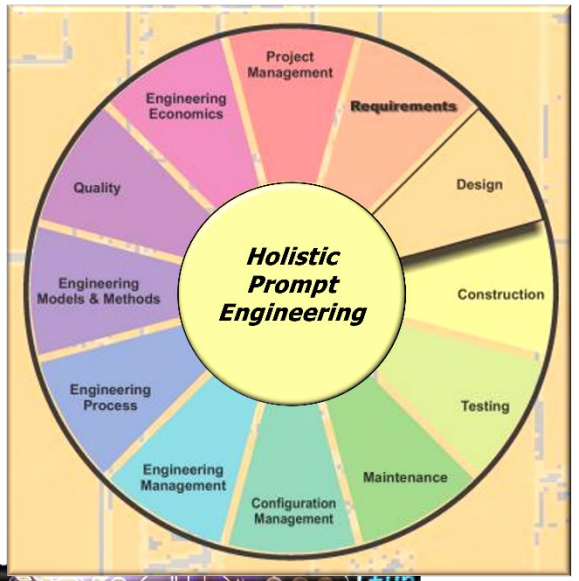
WARNING
This document contains technical data the export of which is restricted by the Arms Export Control Act (Title 22, US Code, Section 2751 et seq.) or Executive Order 12470. Violators of these export laws are subject to severe criminal penalties.

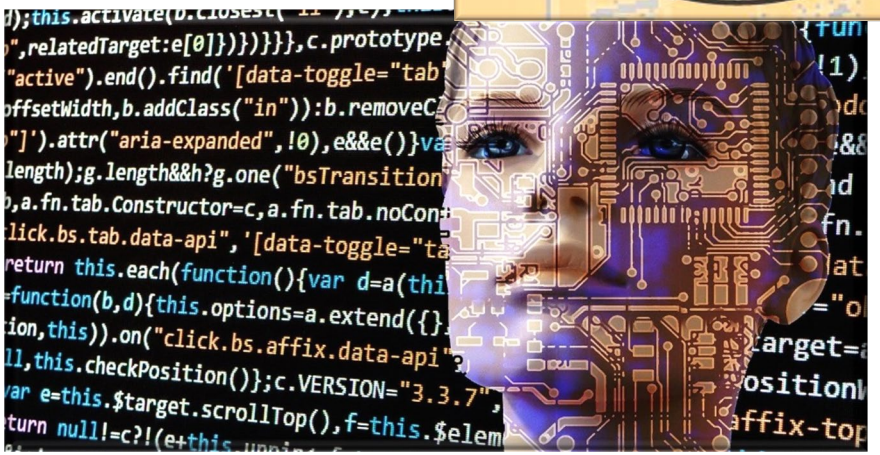
DISTRIBUTION AUTHORIZED
In accordance with AFI 61-204 and DODD 5230.24, Distribution Statement D, Critical Technologies, to DoD components and DoD Contractors only, June 2012. Other requests for this document shall be referred to the Department of the Air Force, HQ USAF/SB, 1180 Air Force Pentagon, Washington DC 20330-1180.

2000 2005 2010 2015 2020 2025

Research & Education Contributions (2020-2024)

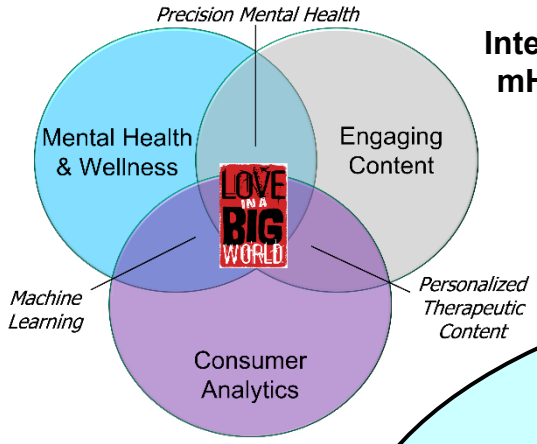
My Focus — Engineering intelligent systems at scale





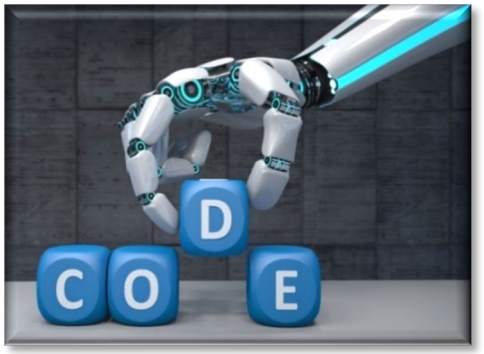
www.sei.cmu.edu/go/national-agenda

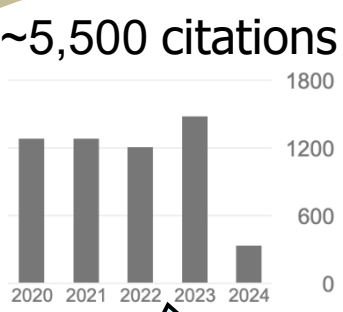
Research & Education Contributions (2020-2024)



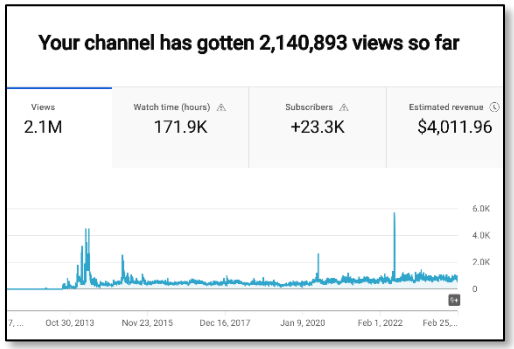
Intelligent mHealth

AI-augmented Software Engineering

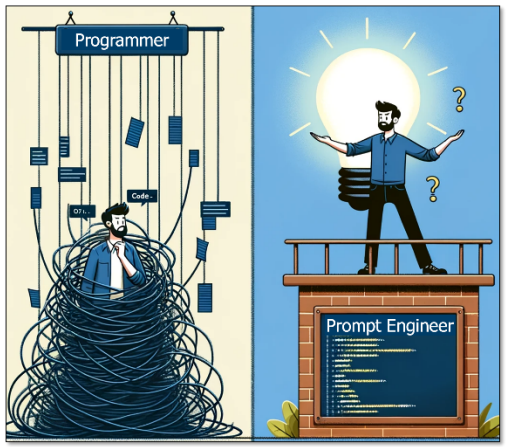




Digital Learning



www.dre.vanderbilt.edu



Prompt Patterns & Prompt Engineering

Leadership Positions (2020-2024)

Government

Academia

Commercial

Visiting Scientist, 2012-2024

2000 2005 2010 2015 2020 2025

Leadership Positions (2020-2024)

Government

Academia

Commercial

Visiting Scientist, 2012-2024

Application of Large Language Models (LLMs) in Software Engineering: Overblown Hype or Disruptive Change?

IPEK OZKAYA, ANITA CARLETON, JOHN E. ROBERT, AND DOUGLAS SCHMIDT (VANDERBILT UNIVERSITY)

OCTOBER 2, 2023

Has the day finally arrived when large language models (LLMs) turn us all into better software engineers? Or are LLMs creating more hype than functionality for software development, and, at the same time, plunging everyone into a world where it is hard to distinguish the perfectly formed, yet sometimes fake and incorrect, code generated by artificial intelligence (AI) programs from verified and well-tested systems?

LLMs and Their Potential Impact on the Future of Software Engineering

This blog post, which builds on ideas introduced in the IEEE paper *Application of Large Language Models to Software Engineering Tasks: Opportunities, Risks, and Implications* by Ipek Ozkaya, focuses on opportunities and cautions for LLMs in software development, the implications of incorporating LLMs into software-reliant systems, and the areas where more research and innovations are needed to advance their use in software engineering. The reaction of the software engineering community to the accelerated advances that LLMs have demonstrated since the final quarter of 2022 has ranged from snake oil to no help for programmers to the end of programming and computer science education as we know it to revolutionizing the software development process. As is often the case, the truth lies somewhere in the middle, including new opportunities and risks for developers using LLMs.

insights.sei.cmu.edu/blog/topics/artificial-intelligence-engineering

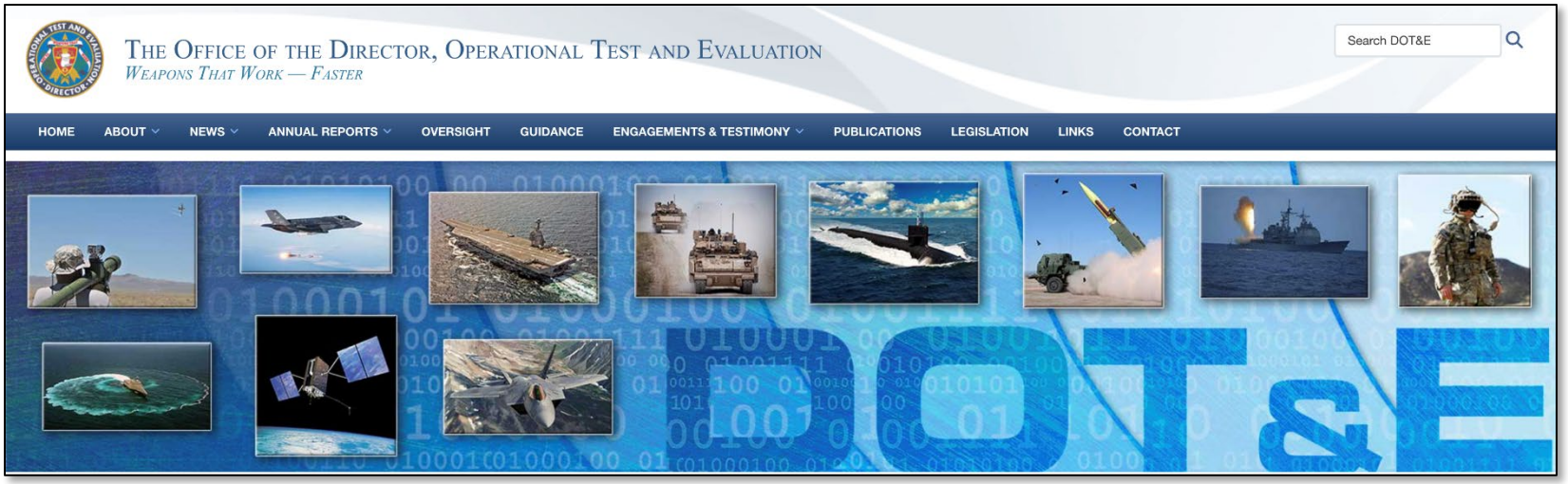
2000 2005 2010 2015 2020 2025

Leadership Positions (2020-2024)

Government

Academia

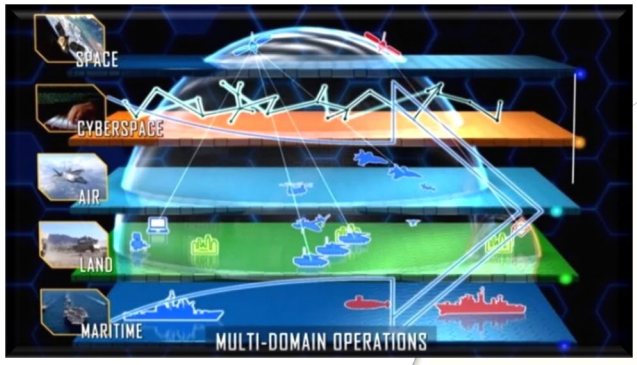
Commercial

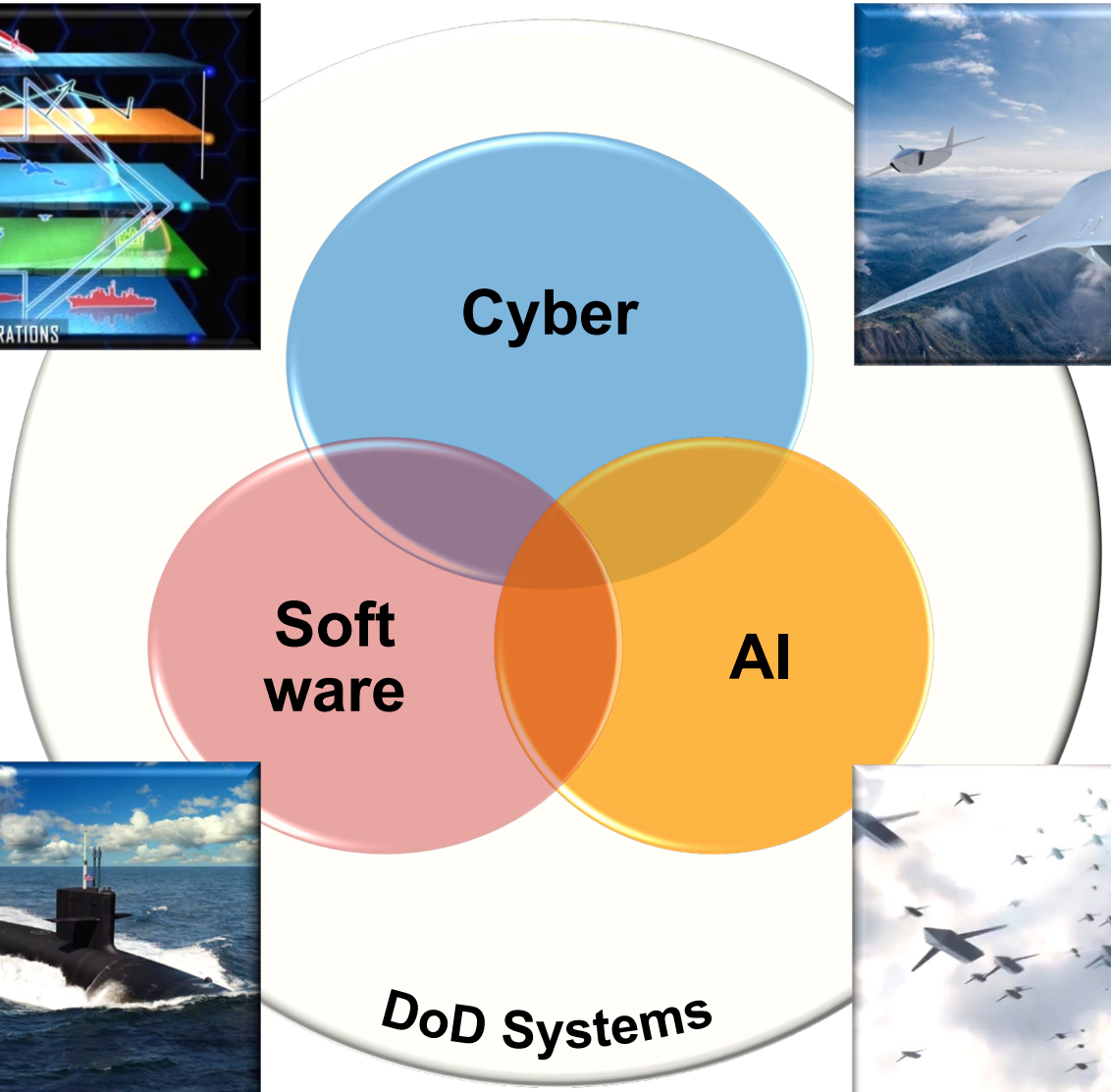


Nominated, confirmed, & sworn in as the Director of Operational Test & Evaluation, 2024-present

2000 2005 2010 2015 2020 2025

Key Challenge Areas for Testing & Evaluation





Next-steps in Rolling Out DoD Manuals for OT&E

DoDI 5000.XF	Establishes policy, assigns responsibilities, and prescribes procedures for operational test and evaluation (OT&E) and live fire test and evaluation (LFT&E) of DoD systems and services acquired via the Defense Acquisition System.
DoDM 5000.UX	Implements policy, assigns responsibilities, and provides procedures for developing OT&E and LFT&E input to the test and evaluation master plan (TEMP), a test and evaluation (T&E) strategy, or an equivalent artifact for DoD systems and services acquired via the Defense Acquisition System.
DoDM 5000.UW	Implements policy, assigns responsibility, and provides procedures for verification, validation, and accreditation (VV&A) of modeling and simulation (M&S) tools critical to meeting OT&E and LFT&E objectives of DoD systems and services acquired via the Defense Acquisition System.
DoDM 5000.96	Implements policy, assigns responsibilities, and provides procedures for OT&E and LFT&E of DoD software-intensive systems and services, and software embedded in systems and services, acquired via the Defense Acquisition System.
DoDM 5000.UT	Implements policy, assigns responsibilities, and provides procedures for realistic full spectrum survivability and full spectrum lethality testing of DoD systems and services acquired via the Defense Acquisition System.
DoDM 5000.UZ	Implements policy, assigns responsibilities, and provides procedures for OT&E and LFT&E of artificial intelligence (AI)-enabled and autonomous systems and services acquired via the Defense Acquisition System.

CREATE → **APPLY** → **AMPLIFY**

For Mapping Between DataWorks Talks & DOT&E Strategy/Implementation Plan Visit:

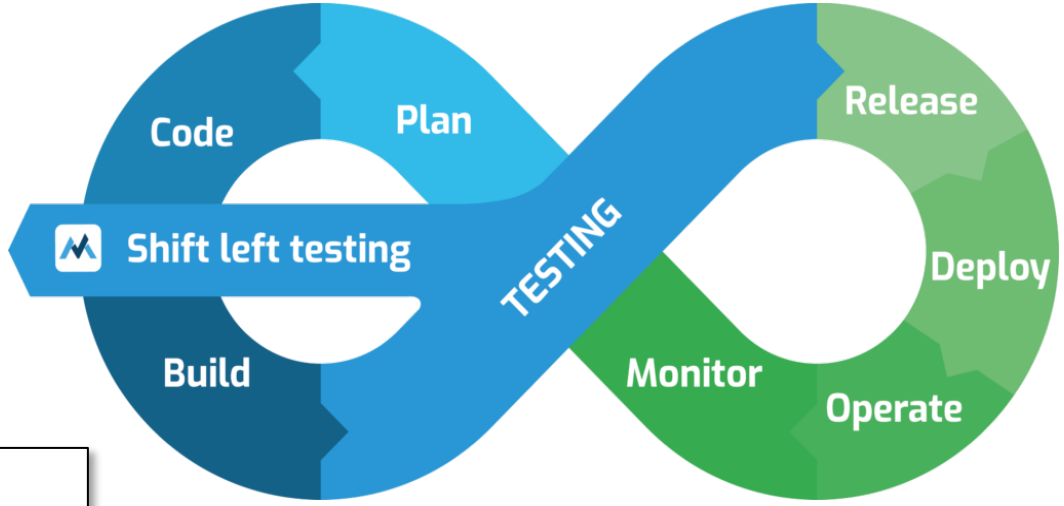
https://www.dote.osd.mil/Portals/97/pub/archive/reports/202404dote_dataworks_strategy_mapping.pdf

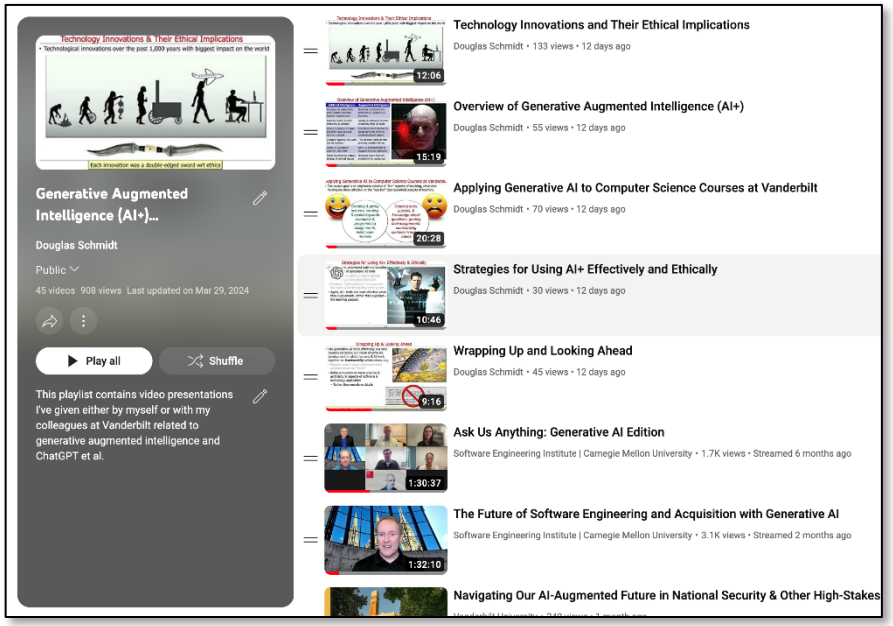
OFFICE OF PREPUBLICATION AND SECURITY REVIEW

Session Number	Session Title	Talk Title	Speaker(s)	Speaker Organization(s)	DOT&E Pillar	DOT&E Competency	Technicality
Mini-Tutorial 1	N/A	Introduction to Uncertainty Quantification	Roger Ghanem	USC	5.1	C6 - STAT; C7 - M&S	1 - Everyone
Mini-Tutorial 2	N/A	Advancing Reproducible Research: Concepts, Compliance, and Practical Applications	Boris Chernis	IDA	2.1	C4 - Info Management	1 - Everyone
Mini-Tutorial 3	N/A	Leading Change: Applying Human Centered Design Facilitation Techniques	Kelli Esser / Christina Houfek	VT	5.1	C2 - Planning; C10 - HSI	1 - Everyone
Mini-Tutorial 4	N/A	Leveraging Bayesian Methods to support Integrated Testing	Justin Krometis / Jim Ferry	VT / Metron	5.1	C6 - STAT	2 - Practitioner
Mini-Tutorial 5	N/A	The Trade-Offs in Choosing Different Apertures for the Modeling of Defense Planning	Leo Blanken / Jason Lepore	NPS / Cal Poly	5.1	C2 - Planning	2 - Practitioner
Session 1A	T&E Concepts for Modern & Emerging Technologies	Threat Integration for Full Spectrum Survivability Assessments	Russ Kupferer	DOT&E	3.2; 4.1	C2 - Planning; C3 - Policy; C5 - Cyber; C9 - EMSO; C12 - DE	1 - Everyone
Session 1A	T&E Concepts for Modern & Emerging Technologies	A Framework for OT&E of Rapidly Changing Software Systems: C3I and Business Systems	Logan Ausman	IDA	4.3	C2 - Planning; C8 - Software	1 - Everyone

Opportunities for Generative AI in T&E

- Must continue to focus on “shift left” testing intelligently

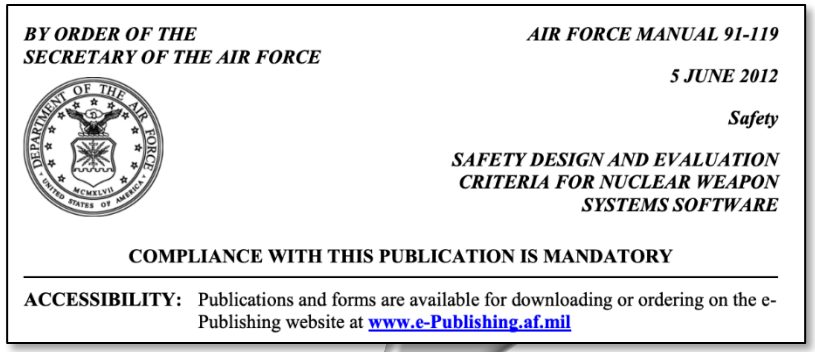




See www.youtube.com/playlist?list=PLZ9NgFYEMxp72Zo0yrTNS6utAXxYpqNGl

Opportunities for Generative AI in T&E

- Must continue to focus on "shift left" testing intelligently
  - 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 outset9



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

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

ChatGPT

- **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

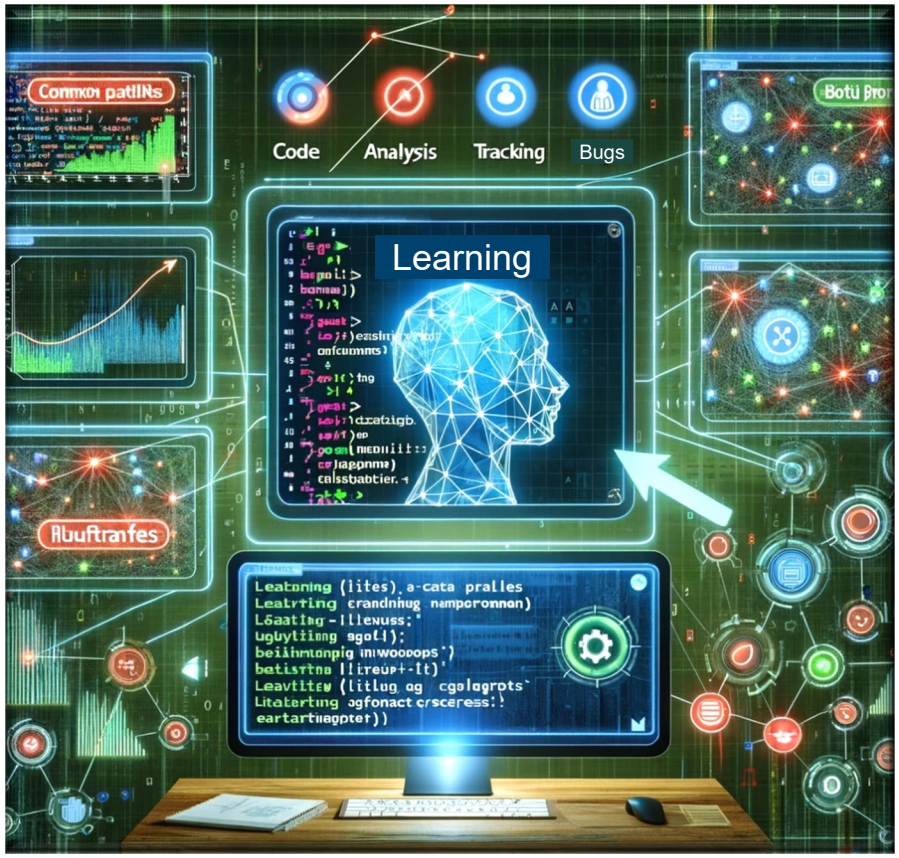
- Must continue to focus on “shift left” testing intelligently
  - Derive test cases from relevant design, policy, & requirement documents
  - Use LLMs to simulate diverse usage patterns & environments to test software under various conditions
    - e.g., apply the *Persona* pattern

> 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 arxiv.org/abs/2302.11382

Opportunities for Generative AI in T&E

- Must continue to focus on “shift left” testing intelligently
  - Derive test cases from relevant design, policy, & requirement documents
  - Use LLMs to simulate diverse usage patterns & environments to test software under various conditions
- Help testers & test organizations learn from prior projects
  - Continuously improve the testing process over time by identifying common pitfalls & best practices



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

Questions & Answers

For more information please contact:

Dr. Douglas Schmidt
Director, Operational Test & Evaluation
Office of the Secretary of Defense
Tel. (703) 697 – 3655
Email: Douglas.C.Schmidt7.civ@mail.mil