SYSTEMS ENGINEERING RESEARCH CENTER | AIRC ACQUISITION INNOVATION RESEARCH CENTER | SERC RESEARCH REVIEW 2026

SERC Research Review 2026

Dr. Paul Wach
Dr. Taylan Topcu
Mr. Tom McDermott
Dr. Joe Gregory
Dr. Alejandro Salado

WRT-2516: Systems Engineering Beyond the Horizon

Sponsor: National Nuclear Security Administration (NNSA)

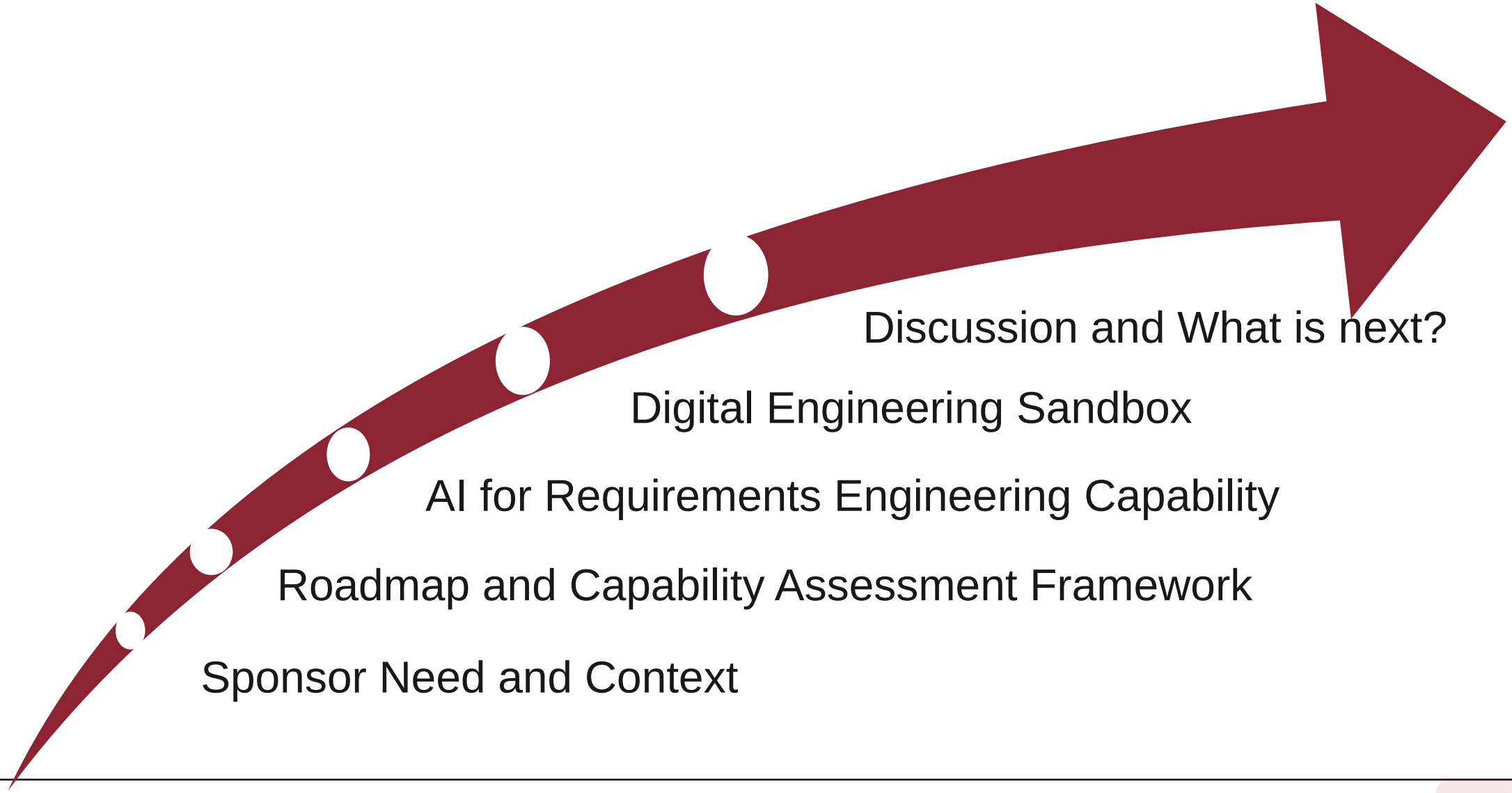


Presenter

Dr. Taylan Topcu
Grado Department of Industrial and Systems Engineering, Virginia Tech

THE UNIVERSITY OF ARIZONA | VIRGINIA TECH | STEVENS INSTITUTE OF TECHNOLOGY 1870

An Overview of This Talk



Project Introduction

Sponsor stated needs:

- *We do not want to be surprised such as we have been by digital engineering transformation…*
- *Show us what is beyond the horizon of SE and help us get ahead*
- *Better complex systems, realized faster and cheaper*



Translated to actions:

- *Project the future of SE - circa 2035*
- *Chart executable path to realize the vision*
- *Reduce risk of adoption*

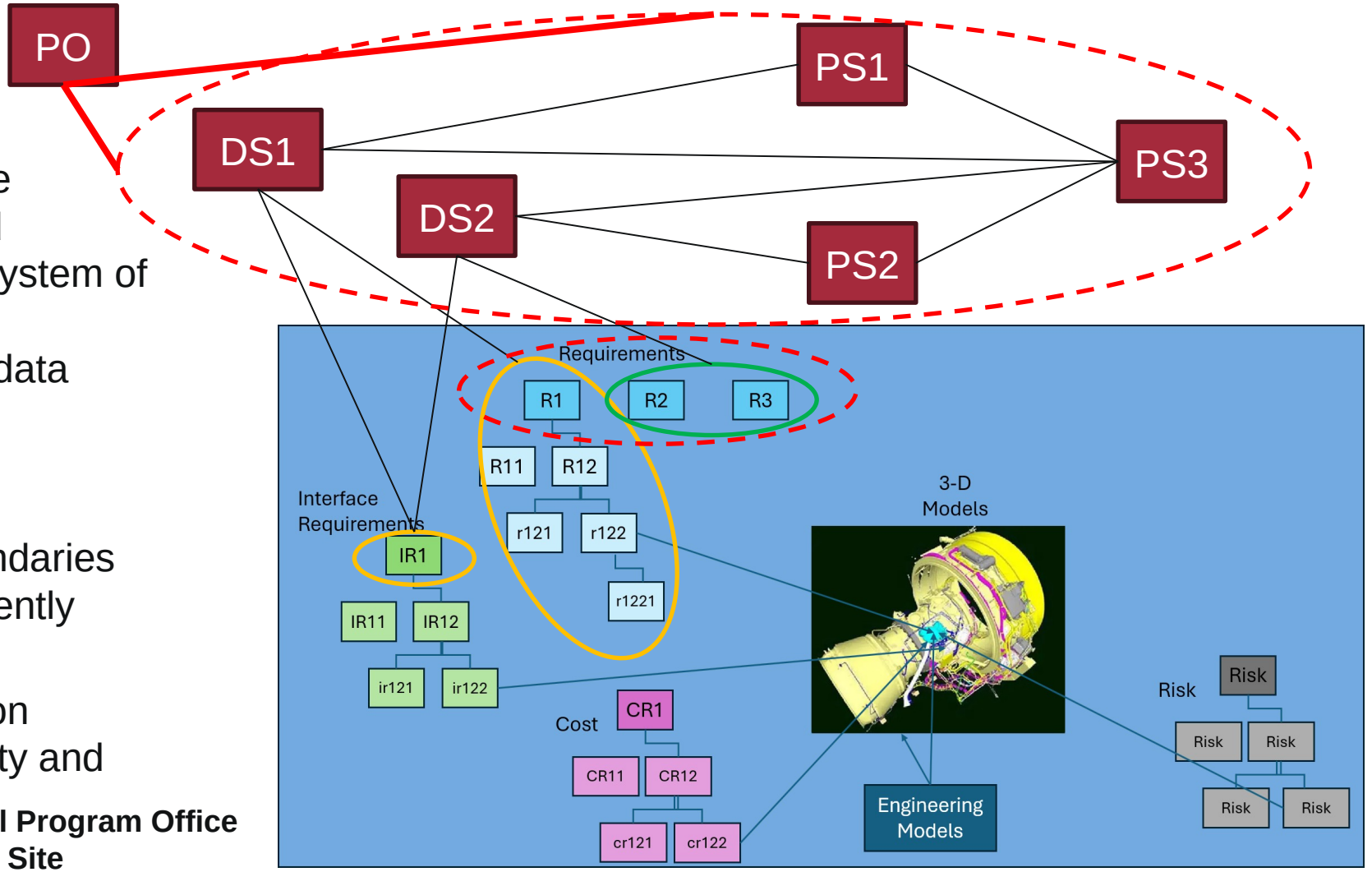
Context: NNSA Complexity and Scale

NNSA specific challenges
- Multi-organization
- Relatively flat org. structure compared to DoW pyramid
- Highly complex technical system of interest

Need connectivity of different data elements
- Data types
- Levels of system
- Across organizational boundaries
- Natively represented differently

Base year effort has focused on building towards this complexity and scale

PO – Federal Program Office
DS – Design Site
PS – Production Site



***Notional illustration of NNSA complexity**

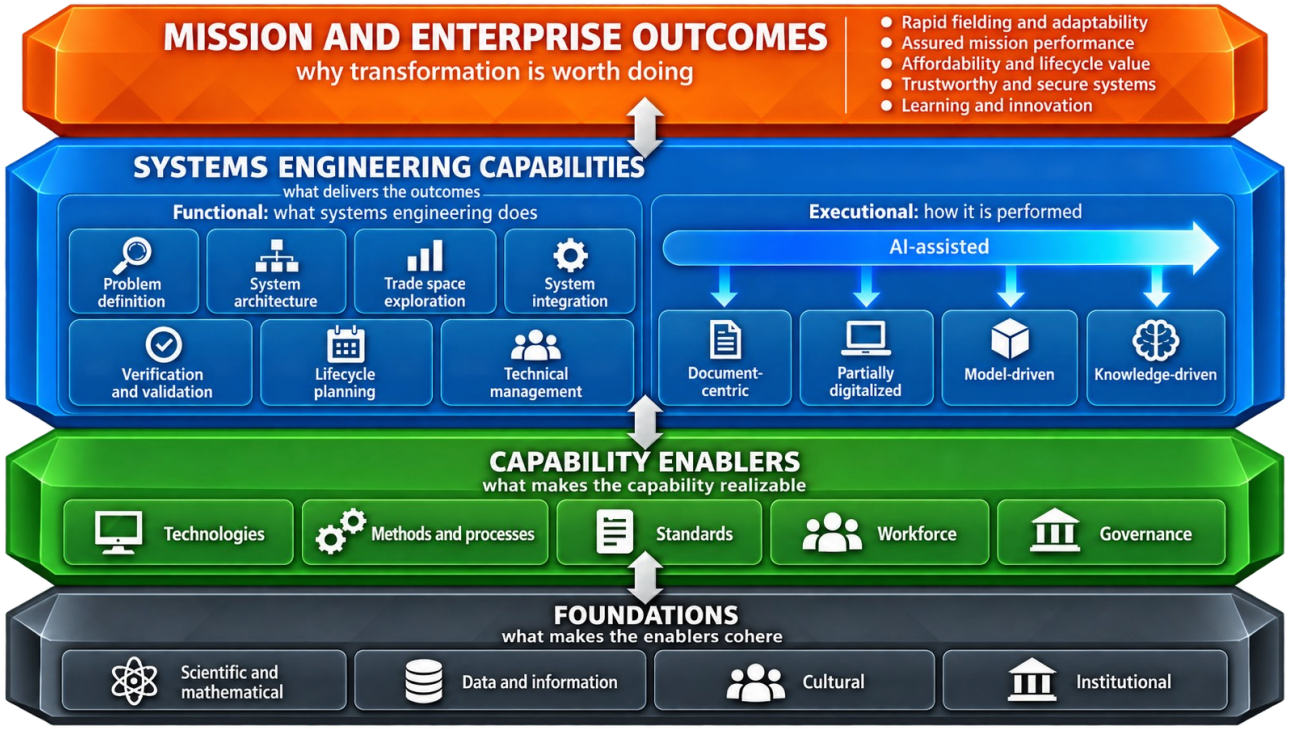
3 Main Lines of Effort

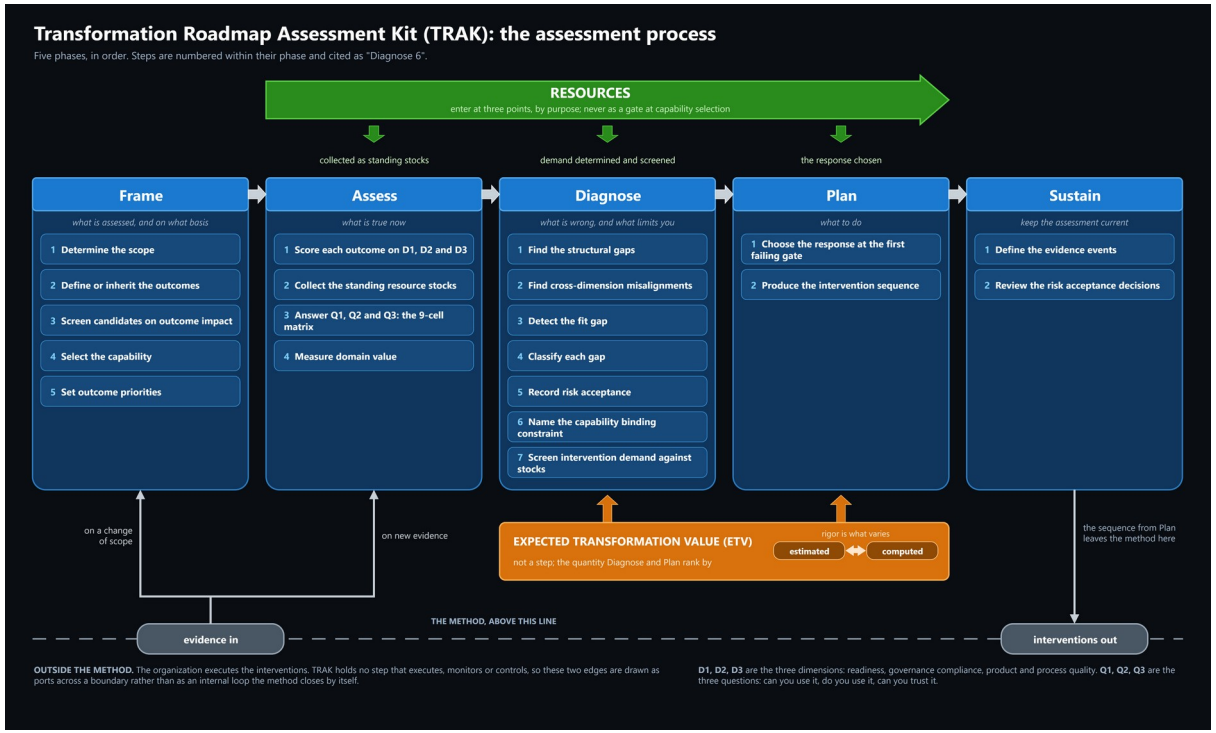
1. **Roadmap and Capability Assessment Framework**
2. **Artificial Intelligence for Requirements Engineering Capability**
3. **Digital Engineering Sandbox**

Roadmap and Capability Assessment Framework

University of Arizona and Stevens Lead

Vision, Roadmap, & Assessment Framework





Some Key Insights

- AI expected to impact all facets of systems engineering
- Rate of change is becoming continuous
- Durable substrate includes rules, ontology, ML/DL, LLMs
  - Matters for both tech and workforce development

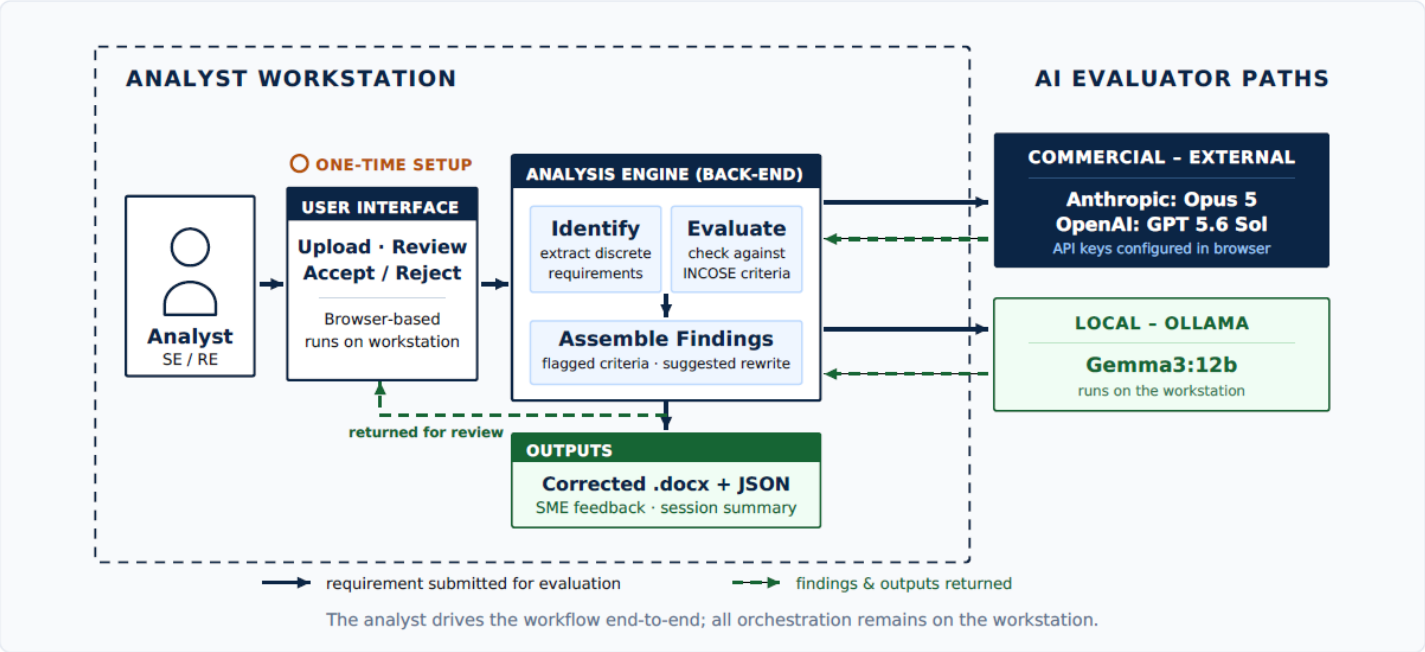
AI4RE Capability

Virginia Tech Lead

AI4RE Prototype Characteristics

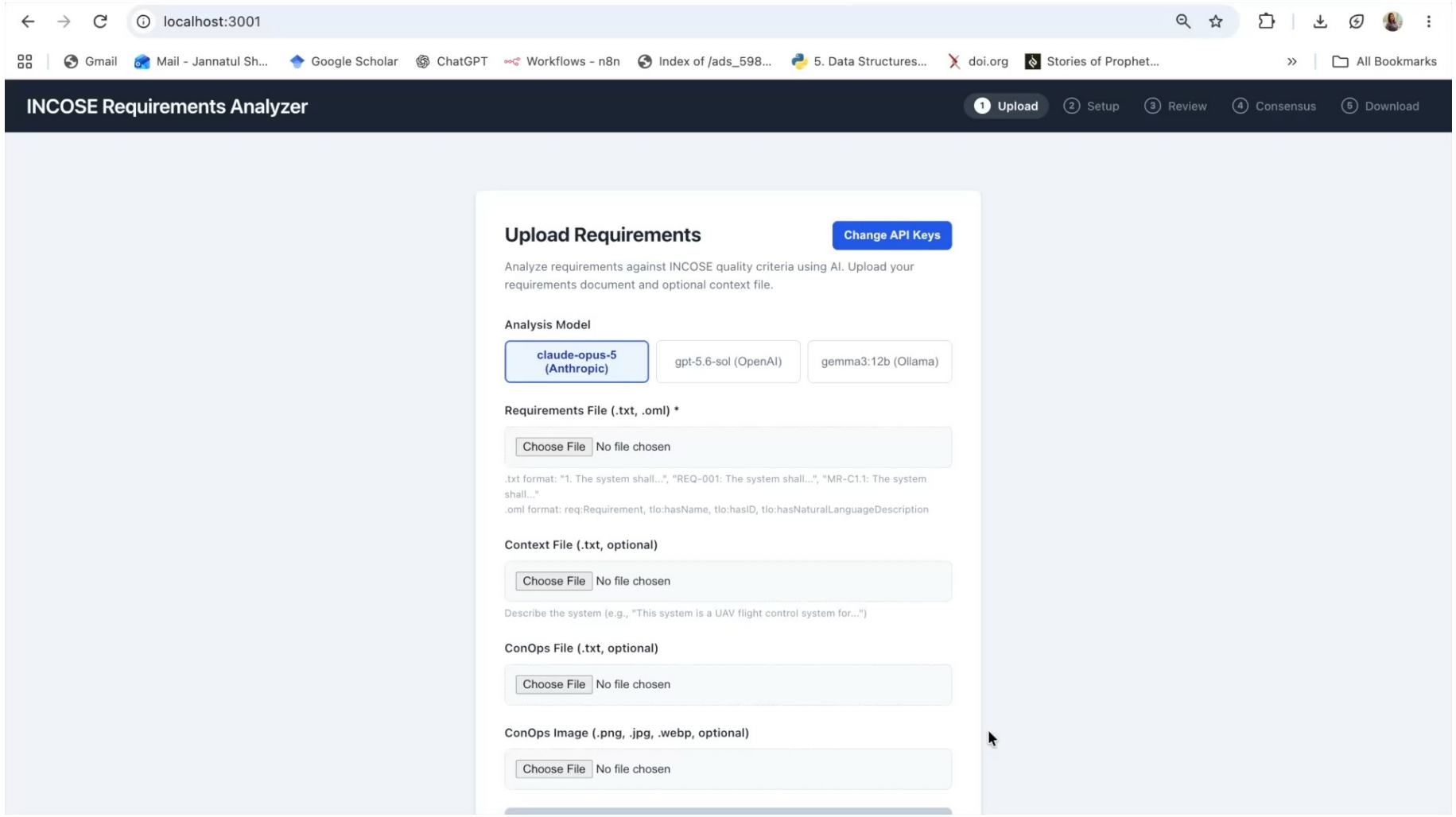
- **Flexible Design:** Designed to facilitate any foundational LLM – existing and upcoming, locally hosted or web based.
- **System and Process Agnostic:** Can analyze **any** number of requirements for **any** system or process.
  - Accepts mission context documents and uses them to judge the requirements.
- **Modular:** Will allow integration of new agentic capabilities over time.
- **Rich User Interface:** Designed to work directly with subject-matter experts, collect knowledge, and learn from these interactions over-time.
- **Long-term Human-in-the-Loop (HITL) Learning:** Stores expert feedback, analyzes conflicts and approvals, updates itself over-time.
- **Interpretable Feedback:** Provides recommendations along with justifications, enables accept, reject, and modify.

AI4RE Prototype v1

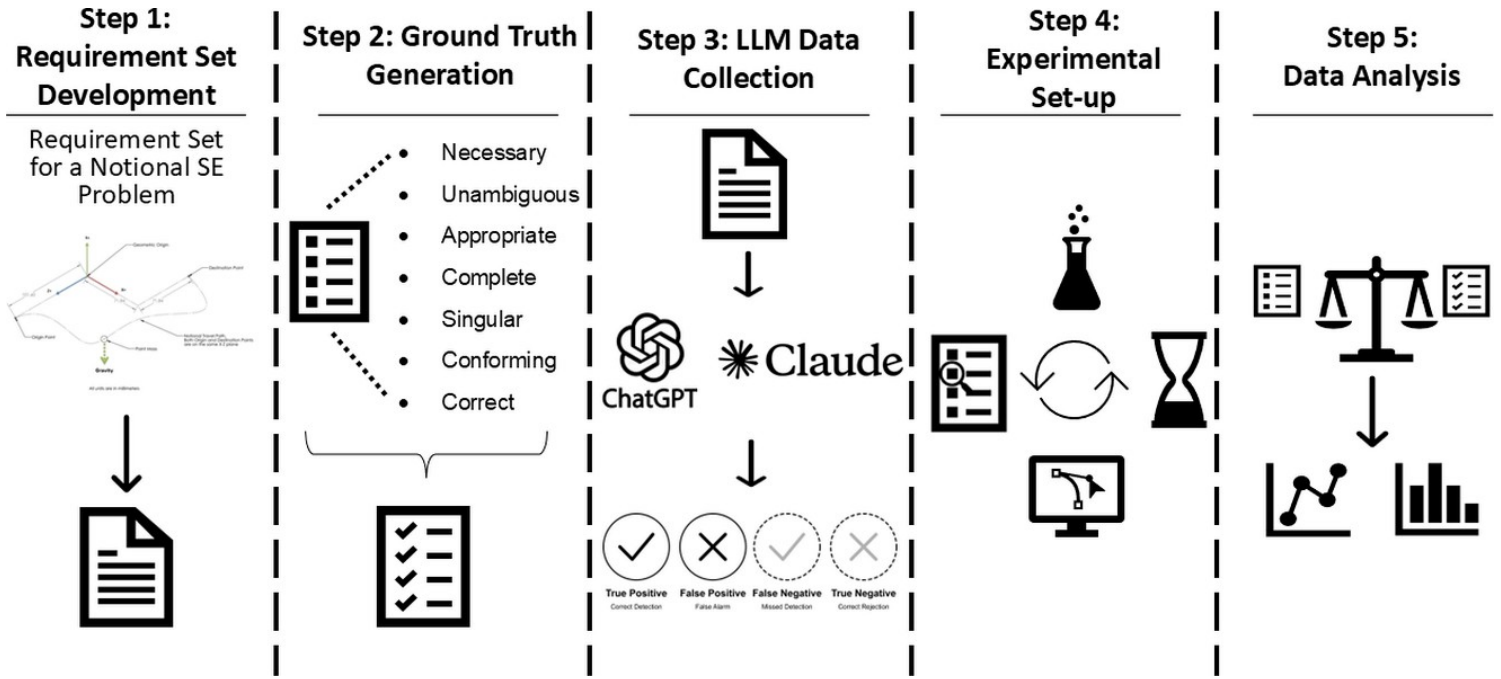


- **What is it composed of?**
  - A React based frontend and FastAPI based Backend.
- **Frontend Functionality:**
  - receives requirements, system context document (e.g., ConOps), and API key for desired LLM.
  - Communicates with back-end, provides recommendations, and captures SME judgements (accept, modify, reject).
- **Backend Functionality:**
  - SE Knowledge Stack and Analysis Engine.
  - API that facilitates LLM communication.
  - Long-term memory model and Data Export Module.

MARQ: AI for learning over-time from human-experts



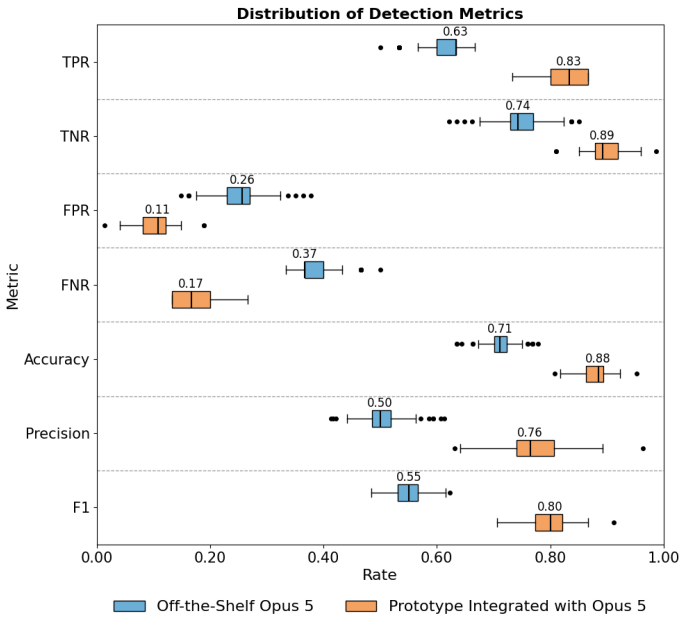
Test & Evaluation Pipeline



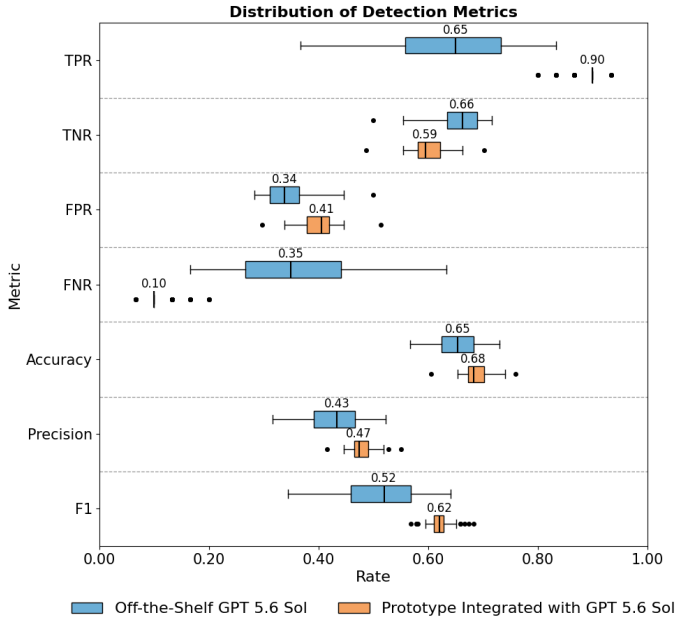
- LLMs are stochastic tools hence there is a pressing need to build trust.
- Formulated T&E Pipeline for Future Prototypes and implemented with n8n workflow.
- Performance evaluated in 7 metrics:
  - true positives, true negatives, false positives (Type 1 Error) , false negatives (Type 2 Errors), accuracy, precision, and F1 Scores.

MARQ Performance Evaluation with Various LLMs

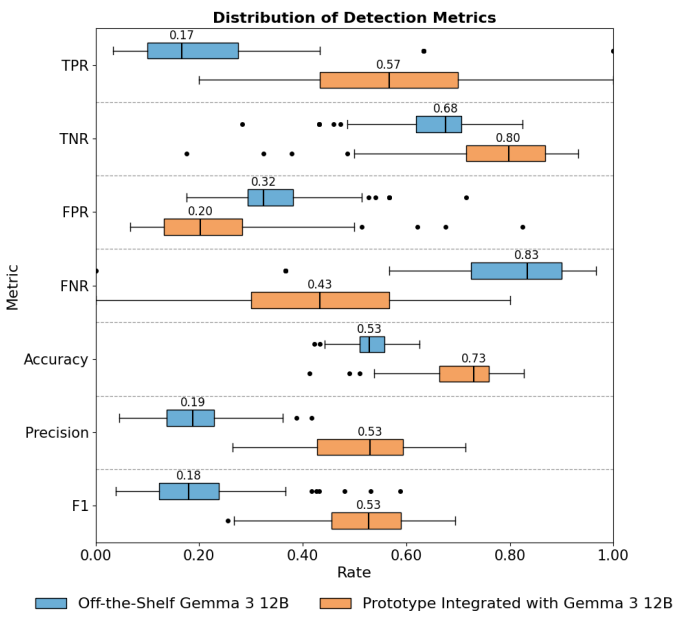
Claude Opus 5



GPT 5.6 Sol



Gemma 3 12B



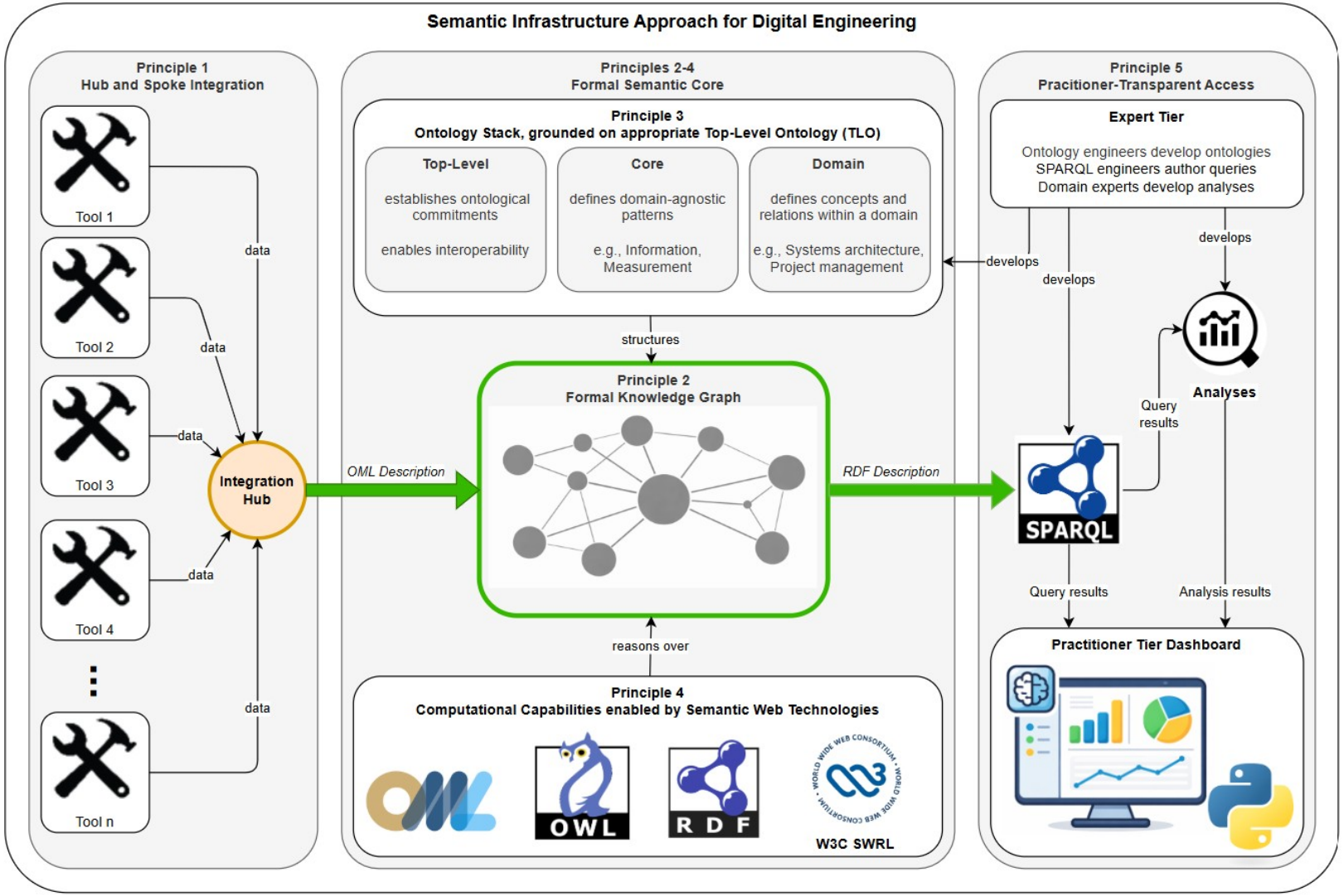
- **Findings** *(from 100 independent runs)*:
  1. **MARQ achieves median F1 scores of 0.80 for Claude Opus 5**, 0.62 for GPT-5.6 Sol, and 0.53 for Gemma 3 12B, compared with off-the-shelf baselines of 0.55, 0.52, and 0.18, respectively.
  2. **The nature of gains is model-dependent.** Opus 5 improves defect detection and false-alarm control; GPT-5.6 Sol gains detection and better consistency at the cost of more some false alarms.
  3. **Gemma 3 12B reaches the performance of off-the-shelf proprietary models**, yet remains variable.

Digital Engineering Sandbox

University of Arizona Lead

Digital Engineering (& Super Intelligence) Sandbox

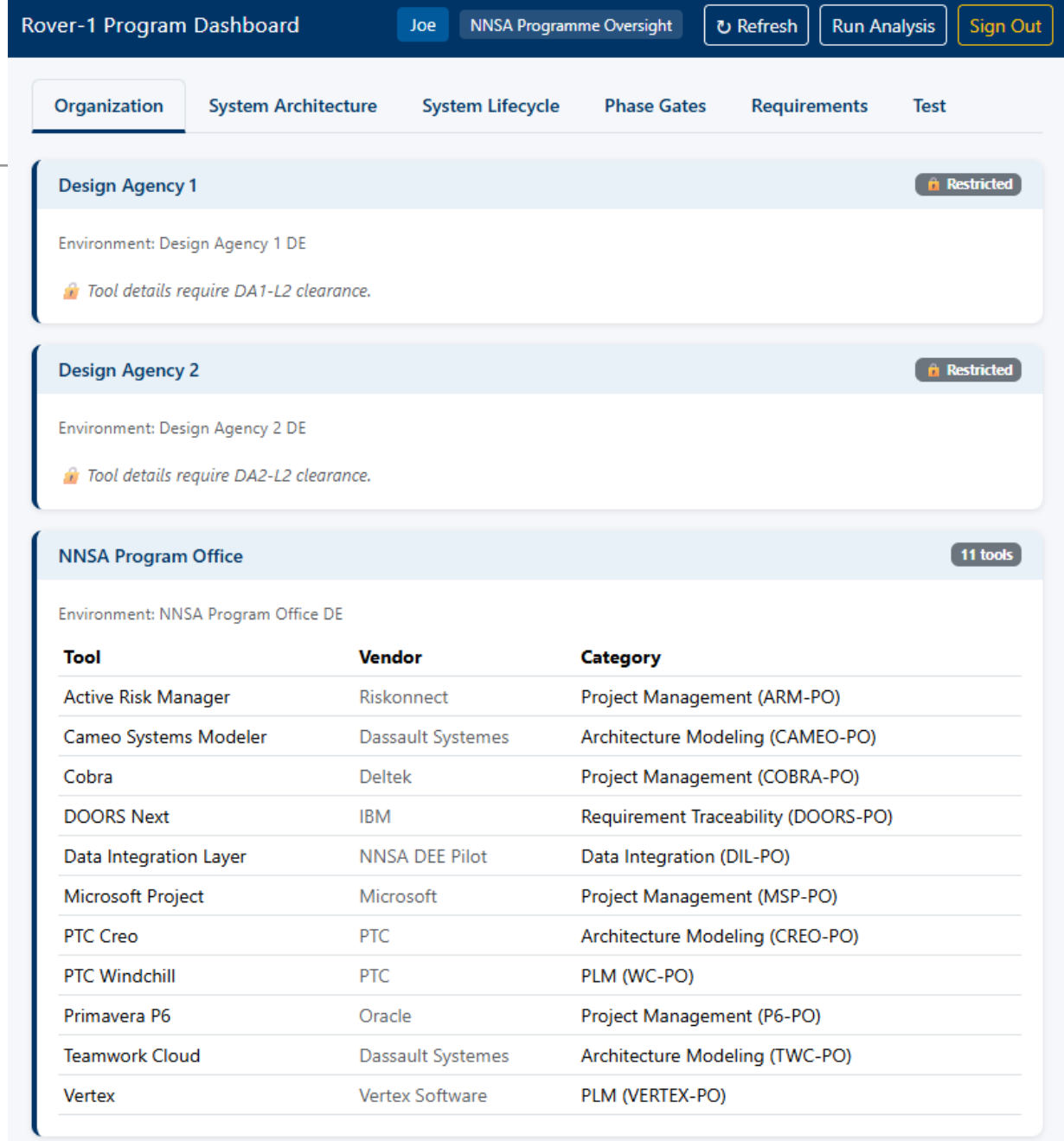
- Integration hub (left)
  - Tool/data harmonization
  - Based on UA's DE Factory
- Semantic technology (middle)
  - Ontology Development and Connected Datasets
  - Core technologies
  - Domain specific (NNSA)
- Bespoke dashboards (right)
- Work-in-progress
  - **Agentic-enhanced workflows**



DE Sandbox Status and Plan

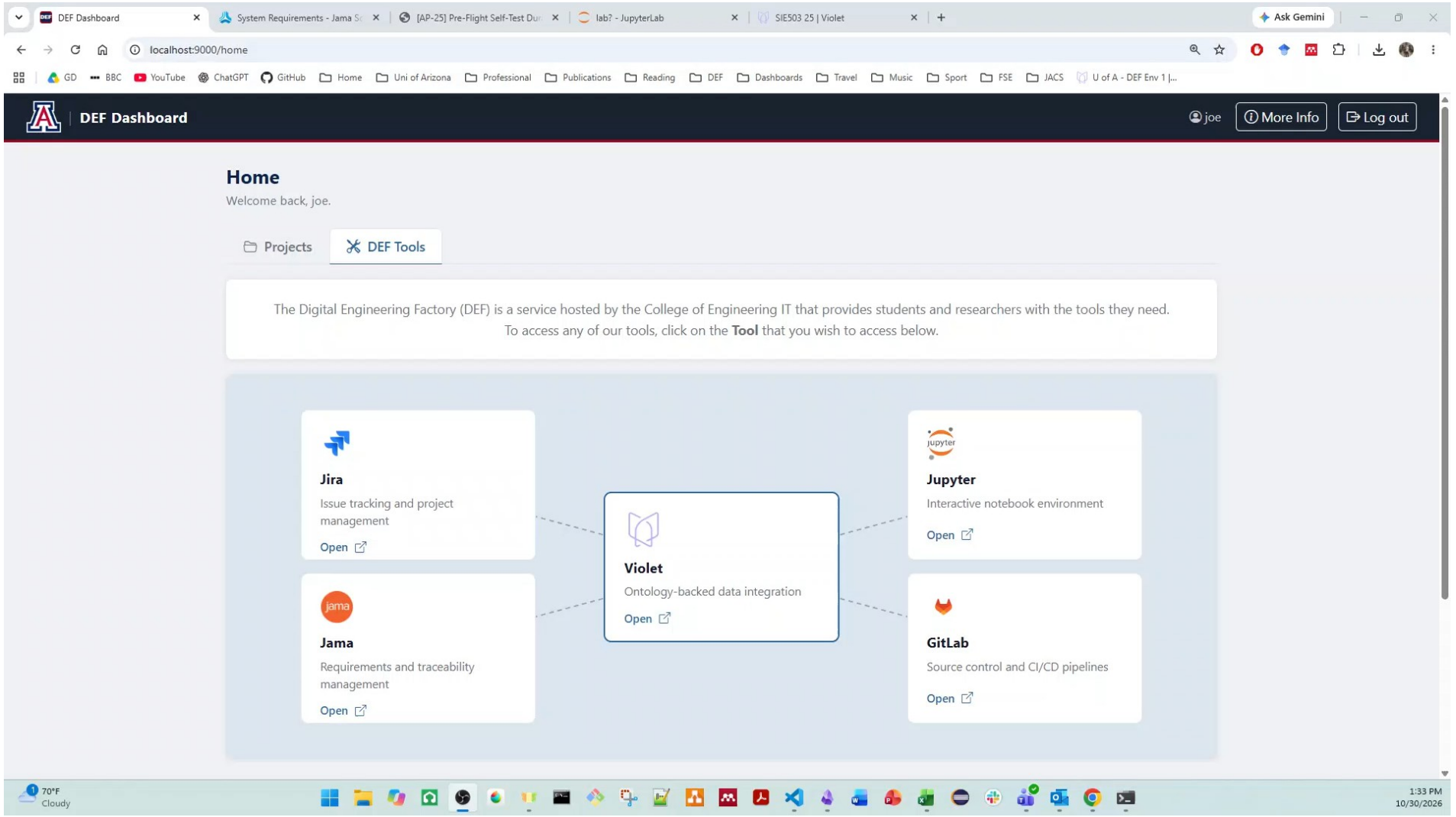
Four Threads Under Development

1. Evaluation of ***Emergent Capabilities***
   - Tool -> Digital Environment -> Program
   - Based on connectivity of tools and environments
   - How this evolves over the lifecycle
2. **Phase Gate Readiness**
   - Propagation of TBDs / TBRs across Phase X
   - Volatility of information and impact on tests, etc.
3. **Access Control**
   - People have roles within orgs – what can they see?
   - And what does this impact? Is digital thread visible?
4. **AI4RE assistant (**Embedded into workflow)
   - Users have choice of Requirements Tool



Tool	Vendor	Category
Active Risk Manager	Riskonnect	Project Management (ARM-PO)
Cameo Systems Modeler	Dassault Systemes	Architecture Modeling (CAMEO-PO)
Cobra	Deltek	Project Management (COBRA-PO)
DOORS Next	IBM	Requirement Traceability (DOORS-PO)
Data Integration Layer	NNSA DEE Pilot	Data Integration (DIL-PO)
Microsoft Project	Microsoft	Project Management (MSP-PO)
PTC Creo	PTC	Architecture Modeling (CREO-PO)
PTC Windchill	PTC	PLM (WC-PO)
Primavera P6	Oracle	Project Management (P6-PO)
Teamwork Cloud	Dassault Systemes	Architecture Modeling (TWC-PO)
Vertex	Vertex Software	PLM (VERTEX-PO)

DE Sandbox Demo



Closure



Trajectory

- **Closure**
  - DE & AI interplay create new opportunities for SE
  - This project pushes the boundary while recognizing the socio-technical nature of the challenges
- **Highlights**
  - Highest documented F1 score in requirements quality assessment and delivered prototypes
  - A rich DE sandbox to simulate organizational change processes, test new ideas, and reduce the adoption barrier
  - 4 PhD students actively working on the project
  - 3 conference articles and 1 journal article published, and 2 under review at INCOSE Systems Engineering
- **What is next?**
  - TRAK to streamline measurement
    - Of team progress, SE landscape…
  - AI4RE
    - MBSE integration, Human-AI team experiments
    - Working on Sponsor Transition
  - DE Sandbox becoming SI-enabled
    - Incorporating methods/tooling from other projects such as *agentic assistants* for **lifecycle awareness**
    - Ontology engineering, SysMLv1 &2, mission engineering, V&V and T&E, wargaming, PLM, contested supply chain management
  - Series of workshops will be conducted
  - Defense Civilian Training Corp (DCTC)
    - Transformation & Acquisition Gamification
    - Four students for AY 2026-2027

Thank you, Questions are most Welcome!

- This work was supported by: The U.S. National Nuclear Security Administration (NNSA) under SERC Grants WRT 2406 and 2516.

















Contact us -> Paul Wach: paulwach@arizona.edu & Taylan Topcu: ttopcu@vt.edu

SYSTEMS ENGINEERING RESEARCH CENTER | AIRC ACQUISITION INNOVATION RESEARCH CENTER | SERC RESEARCH REVIEW 2026

Thank you!

Email the presenter

Dr. Taylan Topcu; Email: ttopcu@vt.edu

Email the research team

Dr. Paul Wach; Email: paulwach@arizona.edu
Dr. Taylan Topcu; Email: ttopcu@vt.edu
Mr. Tom McDermott; Email: tmcdermo@stevens.edu
Dr. Joe Gregory; Email: joegregory@arizona.edu
Dr. Alejandro Salado; Email: alejandrosalado@arizona.edu

Stay connected with SERC/AIRC online