AI Challenge for Systems Engineering – A Trust-Driven Architecture for Multi-Agent System Design

David Du
Ph.D. student, Department of Industrial and Systems Engineering, Virginia Tech

Anirudh More
Ph.D. candidate, Department of Industrial and Systems Engineering, Virginia Tech

12/10/24

VIRGINIA TECH

Outline

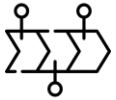
Problem Description

Objectives

Architecture

Optimization Algorithm

Trust



Next Steps

Problem Description

The Rapid Safe Passage (RSP) mission involves…

- Complex, mine-laden terrain navigation with a UGV as the primary platform
- Multiple agents: UAV for scanning, AI for rapid mine detection, human operators for accurate verification
- Uncertain environmental conditions that degrade AI reliability and complicate decision-making
- Discrepancies between predictions and reality erode operator trust, affecting system performance
- Ensuring trust in system performance is as critical as reducing uncertainty

Objectives

Minimize Traversal Time

- Formulate UGV navigation as an optimization problem
- Dynamically plan routes leveraging differing detection capabilities

Build Human Trust:

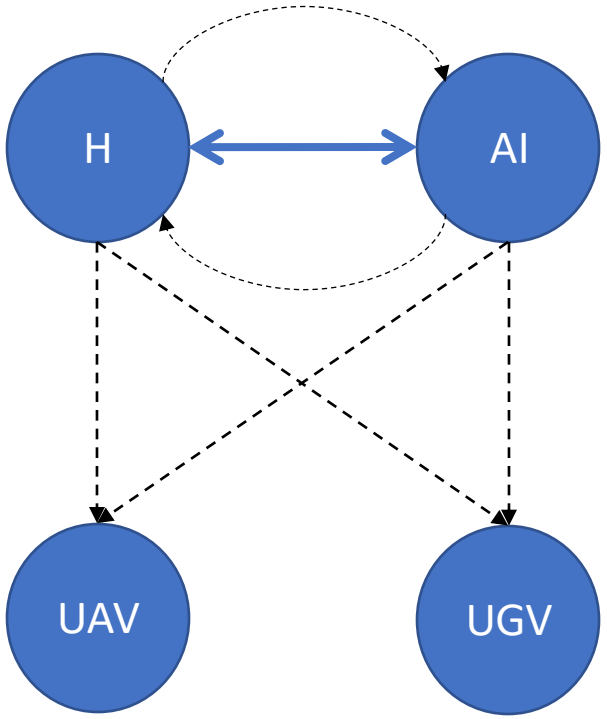
- Consider trust as a core design factor, not just a byproduct of speed
- Ensure humans remain assured of system performance

Architecture

System architecture includes:

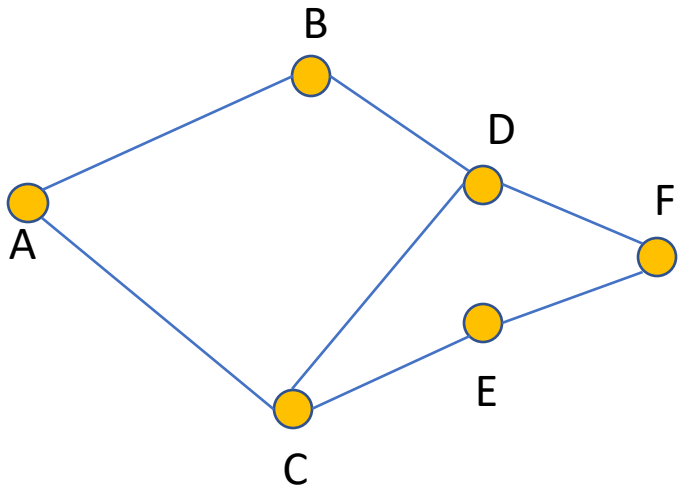
- Human: Predicts mine or clear
- AI: Predicts mine or clear
- UAV: Collects video
- UGV: Mine-clearing ground robot

Proposed architecture: Role allocation between Human and AI



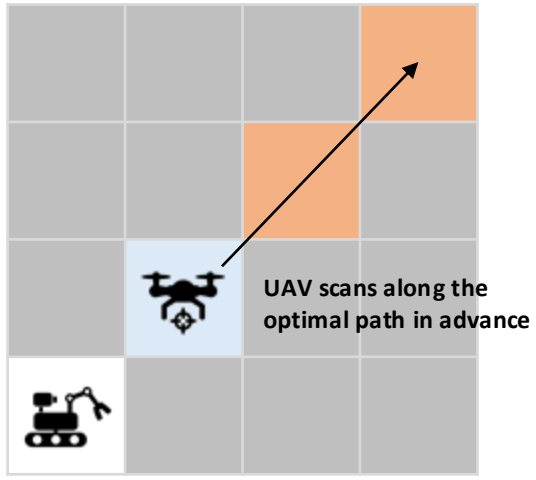
Optimization Initialization

- Map is an undirected cyclic graph G=(V,E)
  - V: Set of passage points (node)
    - v ∈ V represents a decision point
  - E: Set of links (edges/paths) connecting the points
    - e ∈ E represents a potential path between nodes

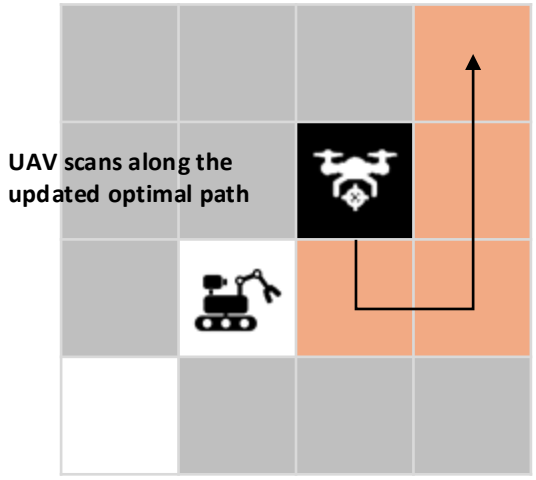


Optimization Algorithm

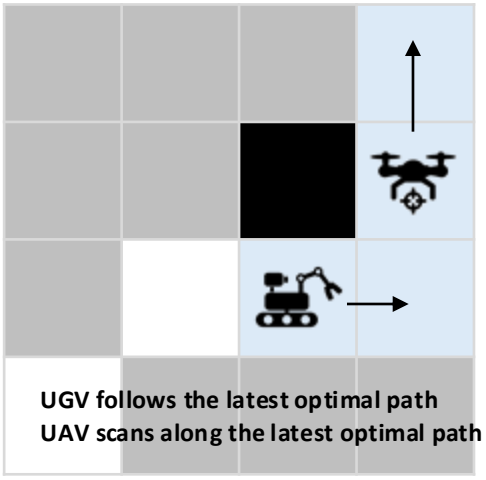
1. Generate an initial optimal path: scan and identify the path with UAV

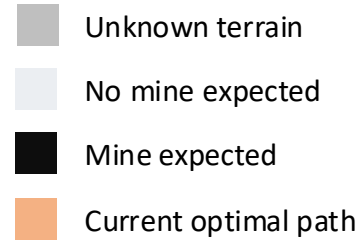


2. Gather information and update the cost → update the optimal path if needed



3. Follow the updated path, scan and identify the path



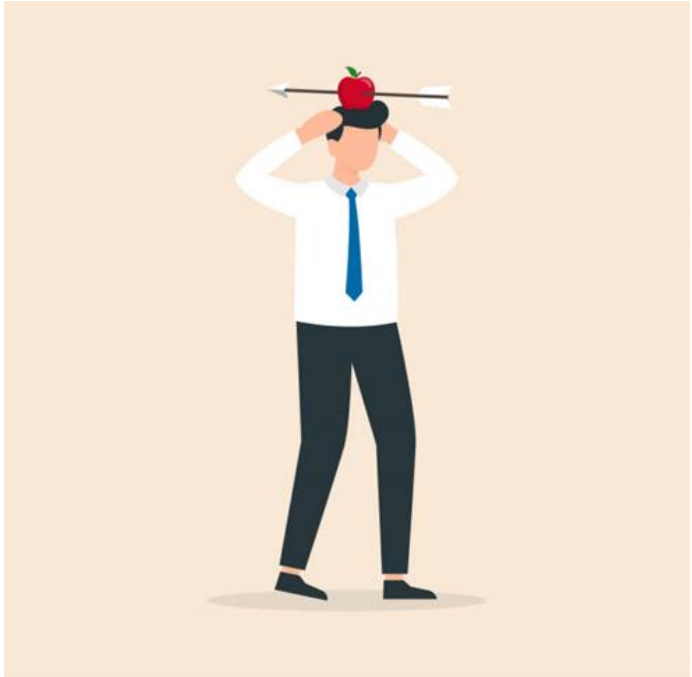


VIRGINIA TECH

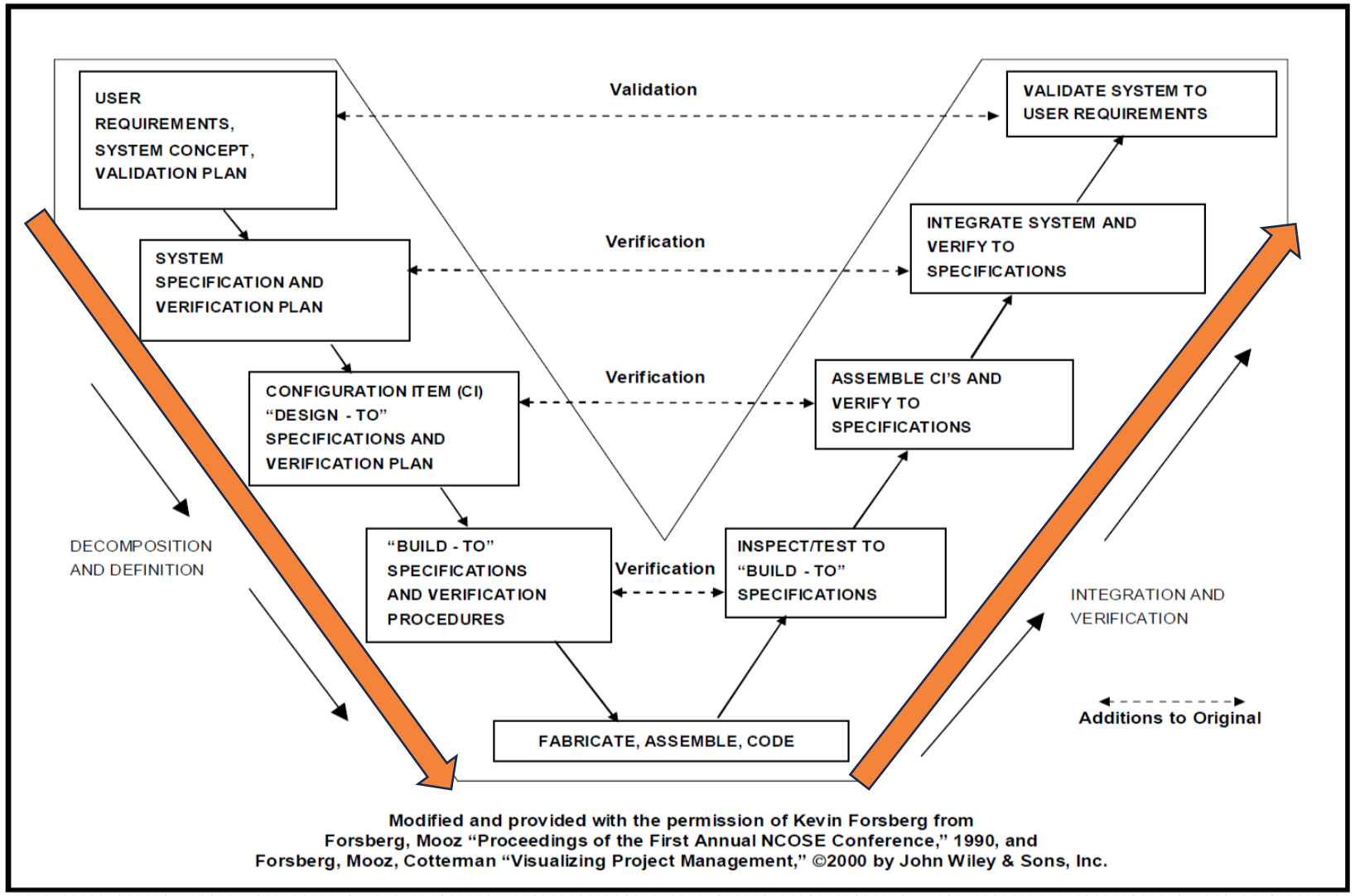
Trust

Trust in AI-Enabled Systems…

- Influences how operators use, question, or override AI outputs
- Balances efficiency gains with risk of catastrophic errors if misplaced
- Not static—evolves as system, interfaces, and operator understanding improve
- Critical in uncertain, high-stakes conditions (e.g., mine-laden terrain)



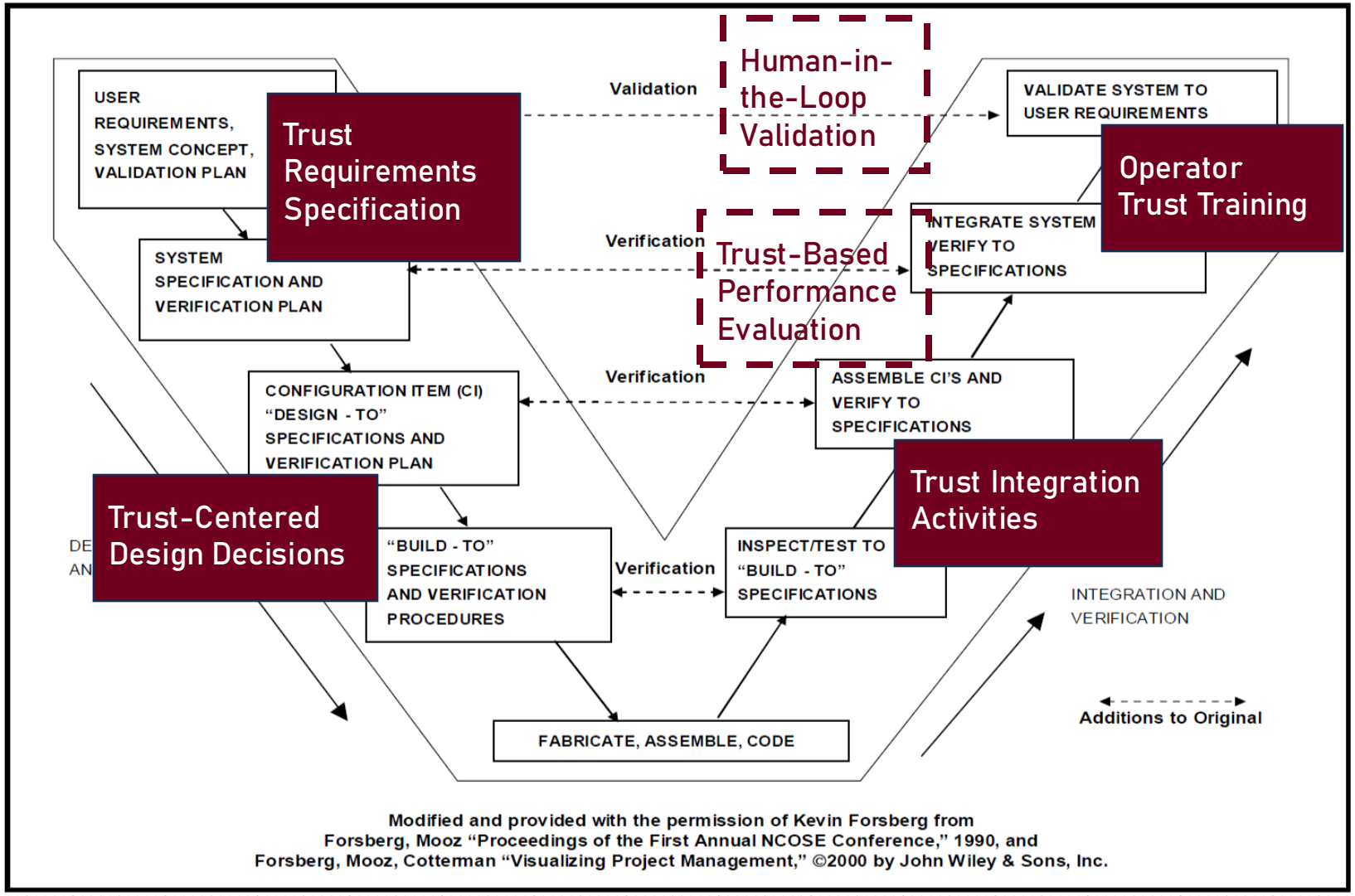
DoD V-Model



Modified and provided with the permission of Kevin Forsberg from
Forsberg, Mooz "Proceedings of the First Annual NCOSE Conference," 1990, and
Forsberg, Mooz, Cotterman "Visualizing Project Management," ©2000 by John Wiley & Sons, Inc.

VIRGINIA TECH

Extended V-Model – HSI & Trust



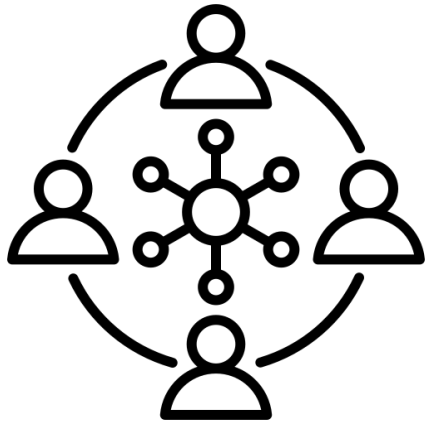
Modified and provided with the permission of Kevin Forsberg from
Forsberg, Mooz "Proceedings of the First Annual NCOSE Conference," 1990, and
Forsberg, Mooz, Cotterman "Visualizing Project Management," ©2000 by John Wiley & Sons, Inc.

VIRGINIA TECH

Activities & Artifacts to Build Trust

1. Trust Requirements Specification:
- Acceptable AI misclassification rates
- Operator confidence thresholds
- Explainability criteria

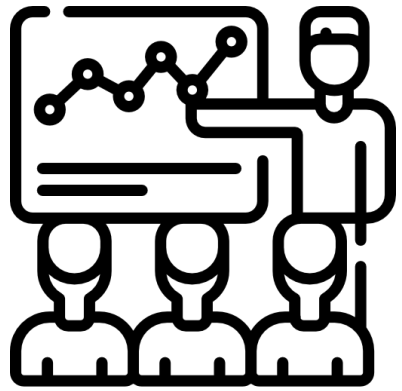
2. Participatory Design Workshops:
- Identify user expectations for transparency, explainability, and confidence-building features



Activities & Artifacts to Build Trust

3. XAI-Enhanced User Interfaces:

- Visualize confidence levels and explain AI recommendations
- Highlight environmental factors affecting AI accuracy



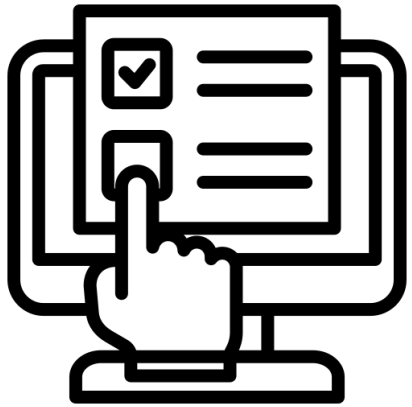
4. Scenario-Based Training and Simulations:

- Train operators under uncertain conditions (e.g., lighting, terrain variability)
- Help operators develop accurate mental models of AI performance

Activities & Artifacts to Build Trust

5. Human-in-the-Loop Verification and Validation:

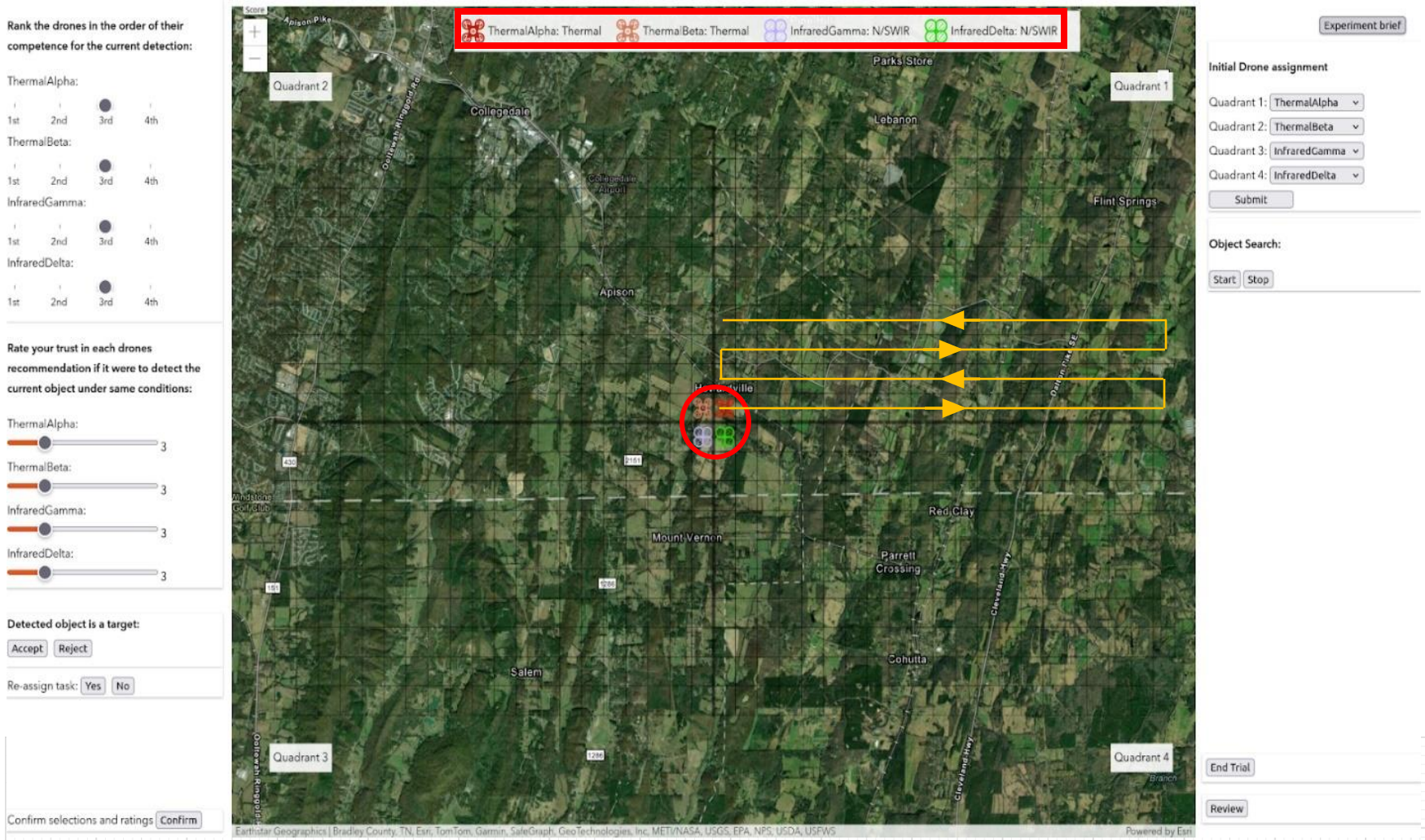
- Involve operators in testing AI's mine detection performance
- Quantify trust evolution to refine system design and training



6. Continuous Feedback and Trust Monitoring:

- Use tools like questionnaires, physiological measures, and interaction logs

Platform for trust development and measurement



Platform for trust development and measurement – Custom-built UI

VIRGINIA TECH

Next steps

- Update to the Platform to study trust development:
  - Various types of UxVs
  - Various vehicle movement patterns/routes
  - How the allocation of affects trust dynamics
- Investigate parallel modes of task allocation
  - The AI and human operator focus on different terrain segments
- Redundant parallel modes of task allocation.
  - The human and AI might examine the same areas

Thank you!

David Du
Department of Industrial & Systems Engineering, Virginia Tech
daviddu@vt.edu

Anirudh More
Department of Industrial & Systems Engineering, Virginia Tech
anirudh1@vt.edu

VIRGINIA TECH