

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

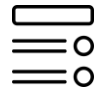
Ph.D. candidate, Department of Industrial and Systems Engineering, Virginia Tech

12/10/24

Outline



Problem Description



Objectives



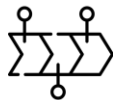
Architecture



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:

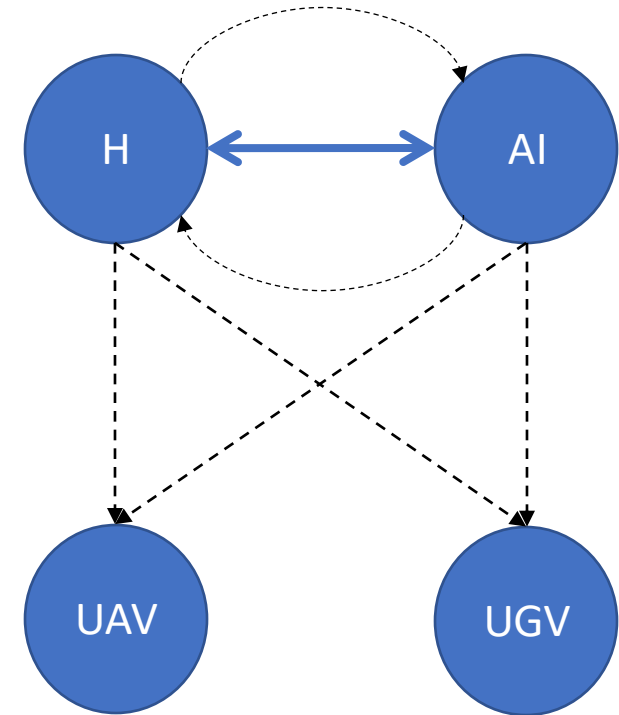
- Consider trust as a core design factor, not just a byproduct of speed
- Ensure humans remain assured of system performance

Architecture

System architecture includes:

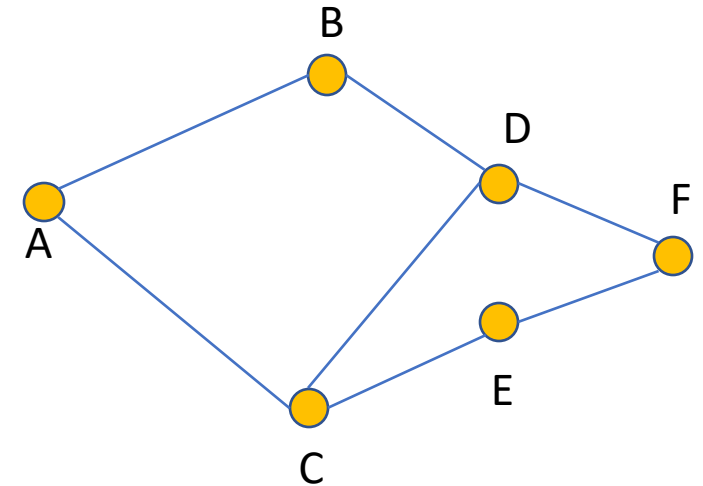
- Human: Predicts mine or clear
- AI: Predicts mine or clear
- UAV: Collects video
- UGV: Mine-clearing ground robot

Proposed architecture: Role allocation between Human and AI



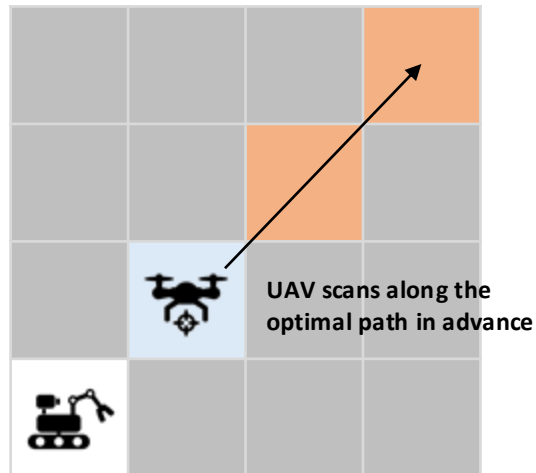
Optimization Initialization

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

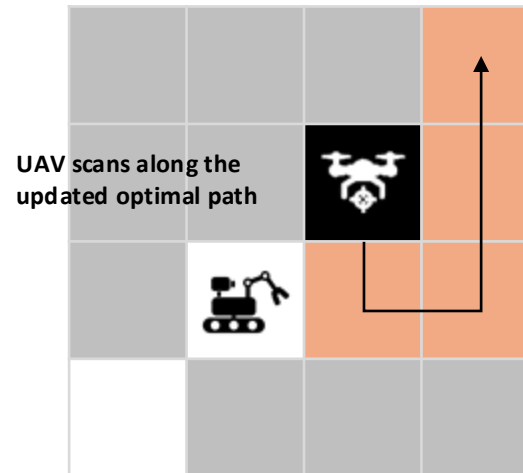


Optimization Algorithm

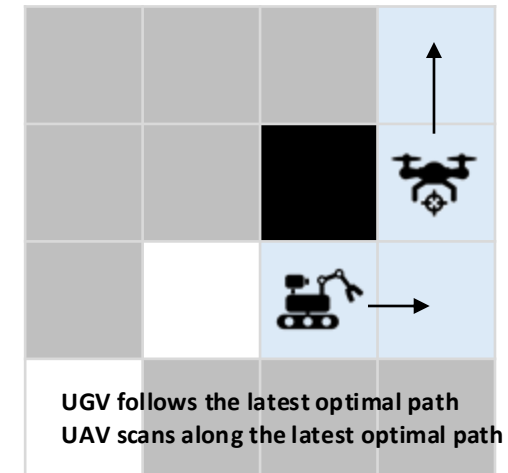
1. Generate an initial optimal path: scan and identify the path with UAV



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



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

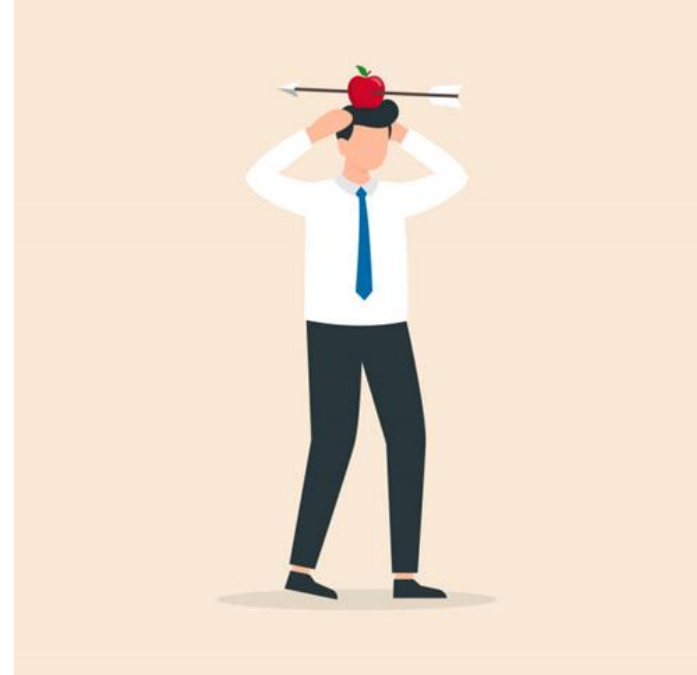


- Unknown terrain
- No mine expected
- Mine expected
- Current optimal path

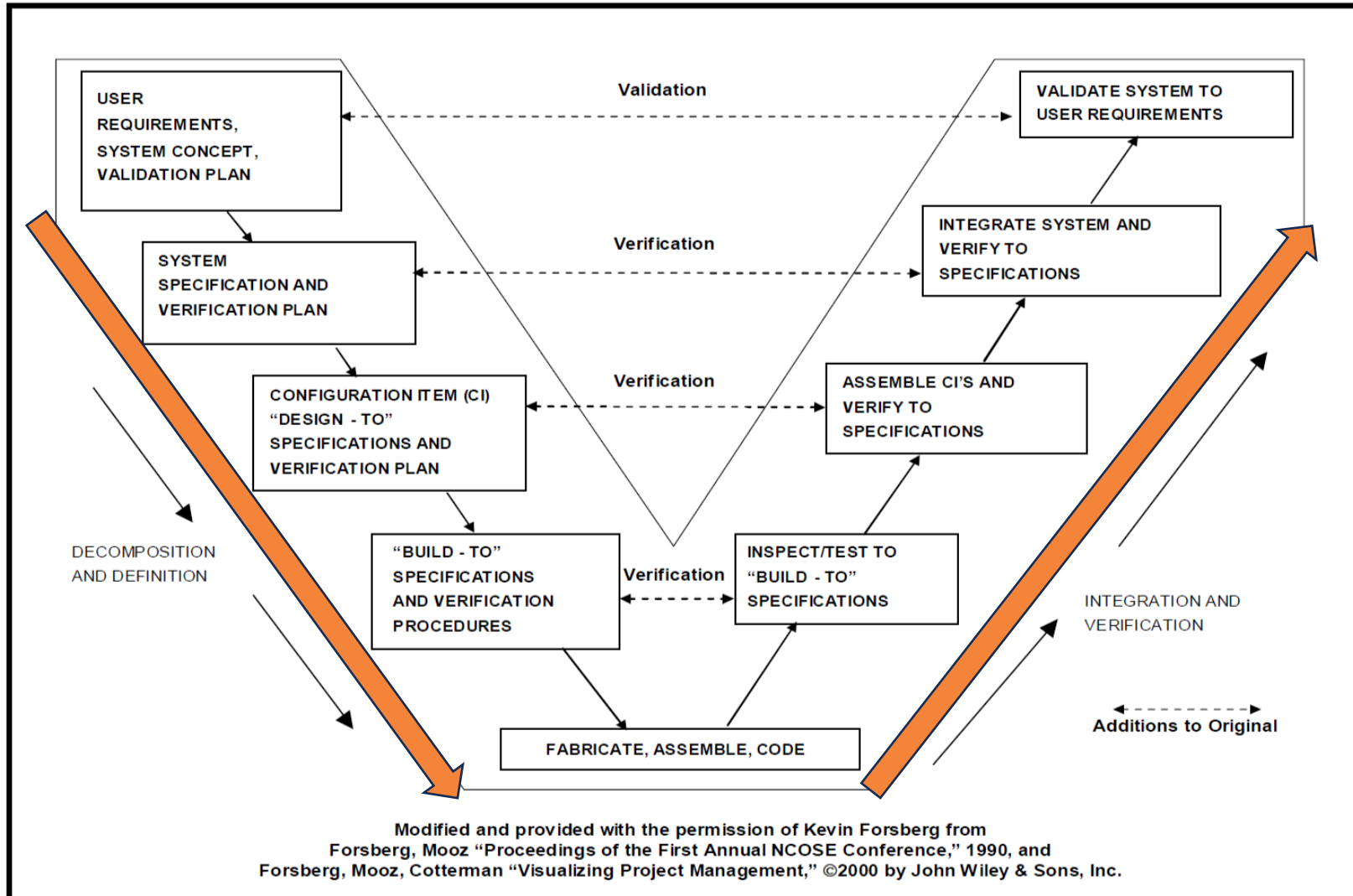
Trust

Trust in AI-Enabled Systems...

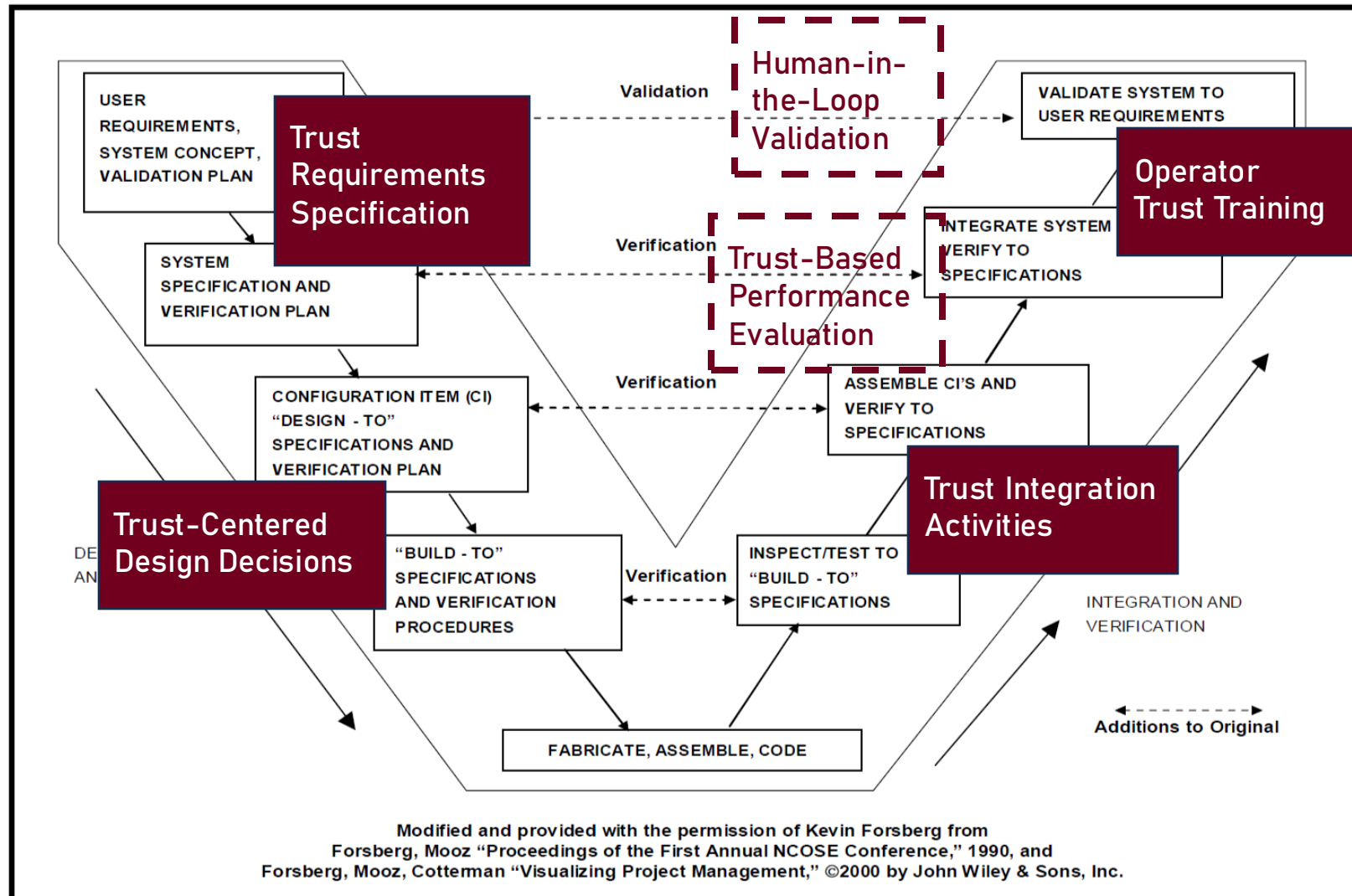
- Influences how operators use, question, or override AI outputs
- Balances efficiency gains with risk of catastrophic errors if misplaced
- Not static—evolves as system, interfaces, and operator understanding improve
- Critical in uncertain, high-stakes conditions (e.g., mine-laden terrain)



DoD V-Model



Extended V-Model - HSI & Trust



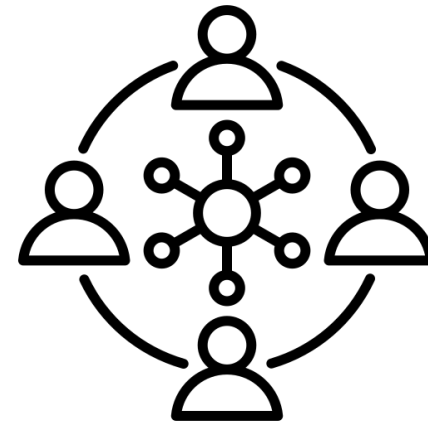
Activities & Artifacts to Build Trust

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



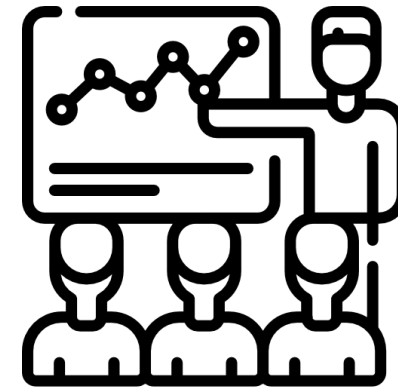
Activities & Artifacts to Build Trust

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



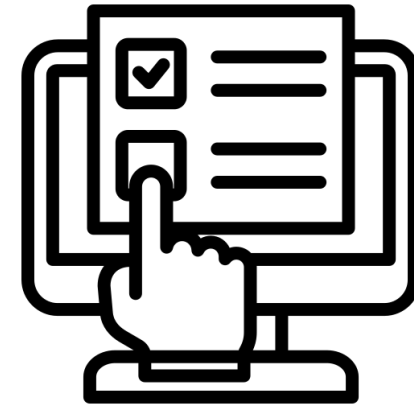
Activities & Artifacts to Build Trust

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

Rank the drones in the order of their competence for the current detection:

ThermalAlpha:
1st 2nd 3rd 4th

ThermalBeta:
1st 2nd 3rd 4th

InfraredGamma:
1st 2nd 3rd 4th

InfraredDelta:
1st 2nd 3rd 4th

Rate your trust in each drones recommendation if it were to detect the current object under same conditions:

ThermalAlpha:
3

ThermalBeta:
3

InfraredGamma:
3

InfraredDelta:
3

Detected object is a target:

Re-assign task:

Confirm selections and ratings

Score

ThermalAlpha: Thermal ThermalBeta: Thermal InfraredGamma: N/SWIR InfraredDelta: N/SWIR

Quadrant 2

Quadrant 1

Quadrant 3

Quadrant 4

Map showing a grid of quadrants (Quadrant 1, 2, 3, 4) over a satellite image of a rural area. A red circle highlights a detected object in the center of the grid. Yellow arrows point from the object towards the right side of the map.

Experiment brief

Initial Drone assignment

Quadrant 1: ThermalAlpha

Quadrant 2: ThermalBeta

Quadrant 3: InfraredGamma

Quadrant 4: InfraredDelta

Object Search:

Platform for trust development and measurement – Custom-built UI

Rank the drones in the order of their competence for the current detection:

ThermalAlpha:

1st 2nd 3rd 4th

ThermalBeta:

1st 2nd 3rd 4th

InfraredGamma:

1st 2nd 3rd 4th

InfraredDelta:

1st 2nd 3rd 4th

Rate your trust in each drones recommendation if it were to detect the

ThermalAlpha:

ThermalBeta:

InfraredGamma:

InfraredDelta:

Detected object is a target:

Accept Reject

Re-assign task: Yes No

Confirm selections and ratings Confirm

Score

ThermalAlpha: Thermal ThermalBeta: Thermal InfraredGamma: N/SWIR InfraredDelta: N/SWIR

Quadrant 2

Quadrant 1

Quadrant 3

Quadrant 4

Object detected at -85.01, 35.00

Detecting drone	InfraredGamma
Object	Non-living distractor
Lighting condition	Low Lighting
Object Detection Reliability	80%

Experiment brief

Current Drone Assignment

quadrant1: ThermalAlpha
quadrant2: ThermalBeta
quadrant3: InfraredGamma
quadrant4: InfraredDelta

Object Search:

Stop

End Trial

Review

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