Trusted AI Challenge
Phase I: Architectures and Approach

SERC Trusted Artificial Intelligence Systems Engineering Challenge, Summer 2024

Aug 9, 2024

Athul C D, Zichong Yang, Bradley Feng, Ian Walter, Yupeng Zhou, Jitesh H. Panchal

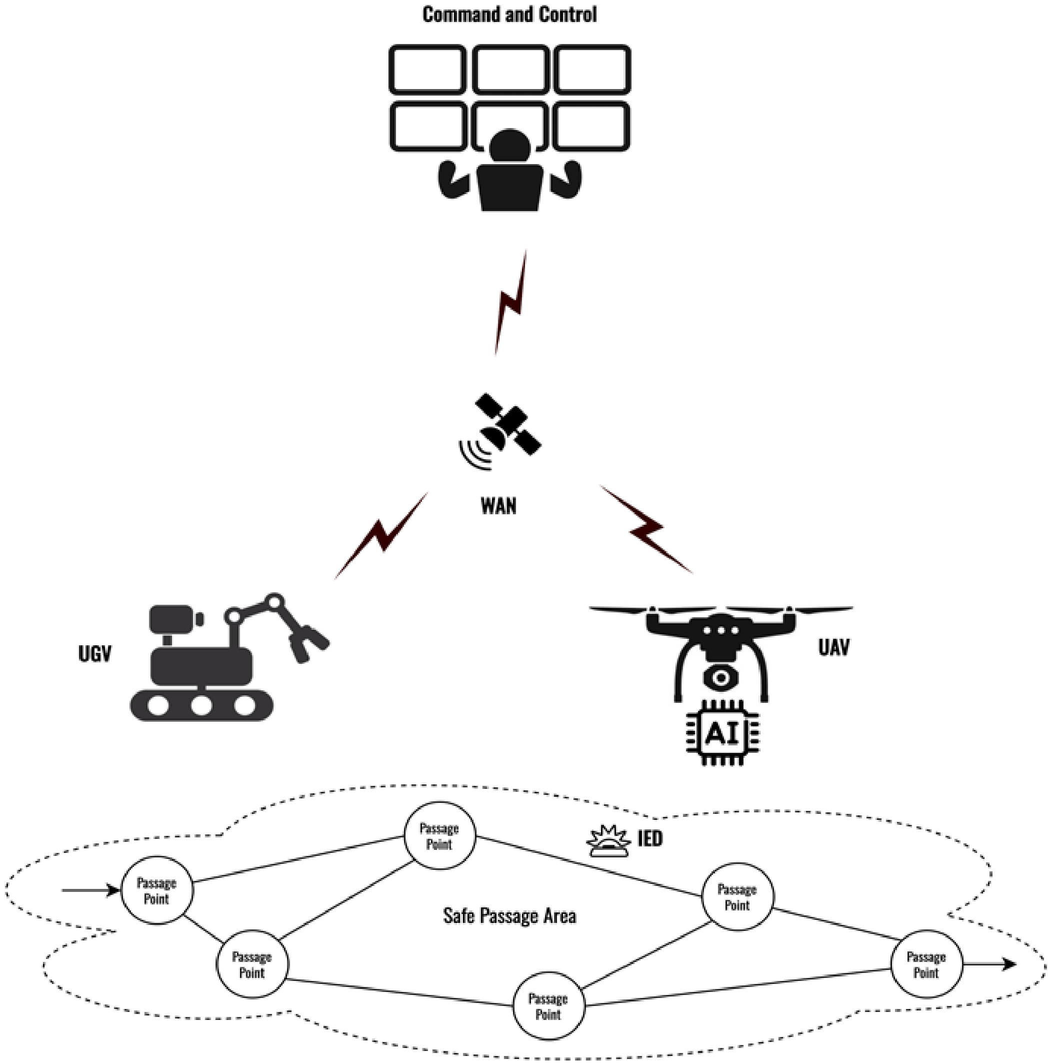
School of Mechanical Engineering, Purdue University

Design Engineering Laboratory @ Purdue
http://engineering.purdue.edu/DELP

Problem Statement

Objectives:

- Ensure safe passage from point A to point B along a defined network
- Develop the notion of trust between Artificial intelligence (AI) models and human operators
- Develop architectures to use human operators in conjunction with AI
- Identify the roles humans and AI should play in different architectures
- Compare different architectures across measures of performance



Key Assumptions (1)

- UGV
  - Can only receive instructions; no built-in processing or feedback mechanisms to send data to the Command and Control Center (C2)
  - No lethality/failure; UGV can clear the mines every time

- UAV
  - Limited computational and communication capabilities; processing power onboard is weaker than C2 and can only run a lower fidelity version of the models, can’t transfer all the high-resolution sensor data instantaneously (depends on the bandwidth)
  - On top of receiving instructions from C2, can share raw data collected from sensors (Video, images, LIDAR, etc.) and/or processed output (probability of encountering a mine)

Key Assumptions (2)

- Dynamics of the environment
  - Initially, the conditions will be assumed static, i.e., the UAV needs only one pass to know the state of the network, and no changes are accounted for during operations. Later, this can be relaxed to account for changes in the network that happen during operation (fire breaking out, change in weather, etc.)

- Command and Control Center
  - The control center has enough human personnel available at all times to make decisions and analyze the inputs

- Coordination
  - Current strategy is for a single UAV and UGV. Later, the architectures can be expanded to a swarm of UAVs and UGVs exploring the network in tandem

Characterizing Architectures

- We describe the architecture as the decision-making architecture for the task of navigating the network
- Each architecture is characterized by the flow of information and control instructions between the different stakeholders – who makes what decision? In which sequence?
- Communication between agents are of two types: information or control input

Images Captured

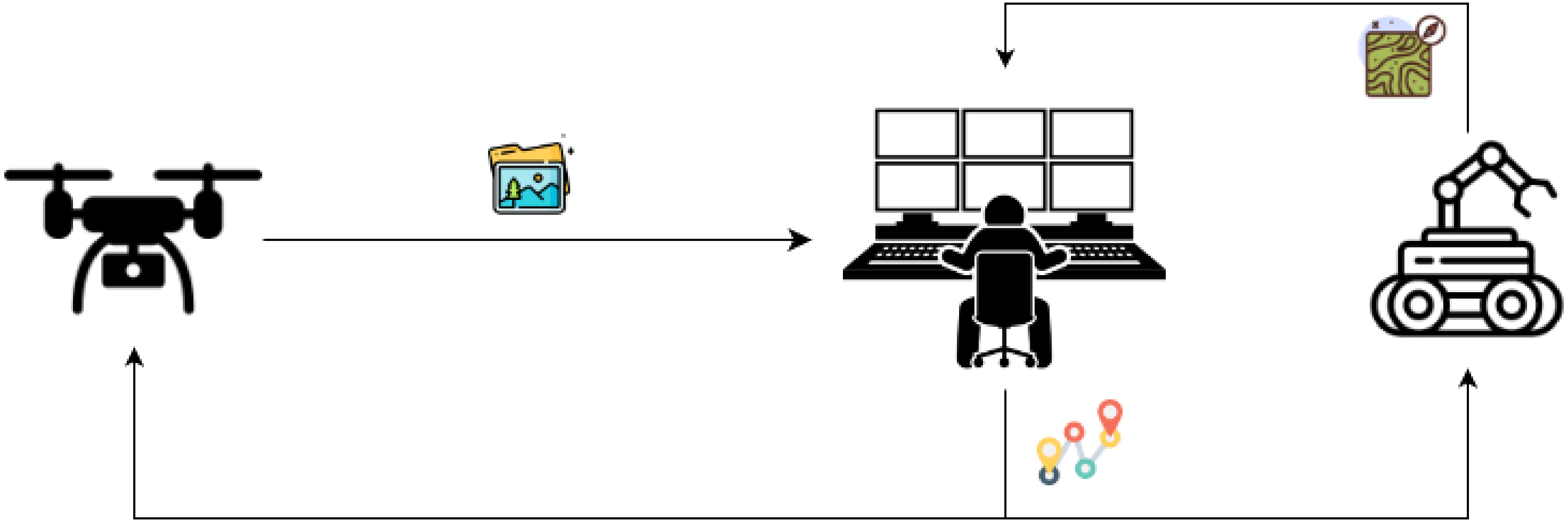
Updated Route

AI Confidence

Actual Terrain Info

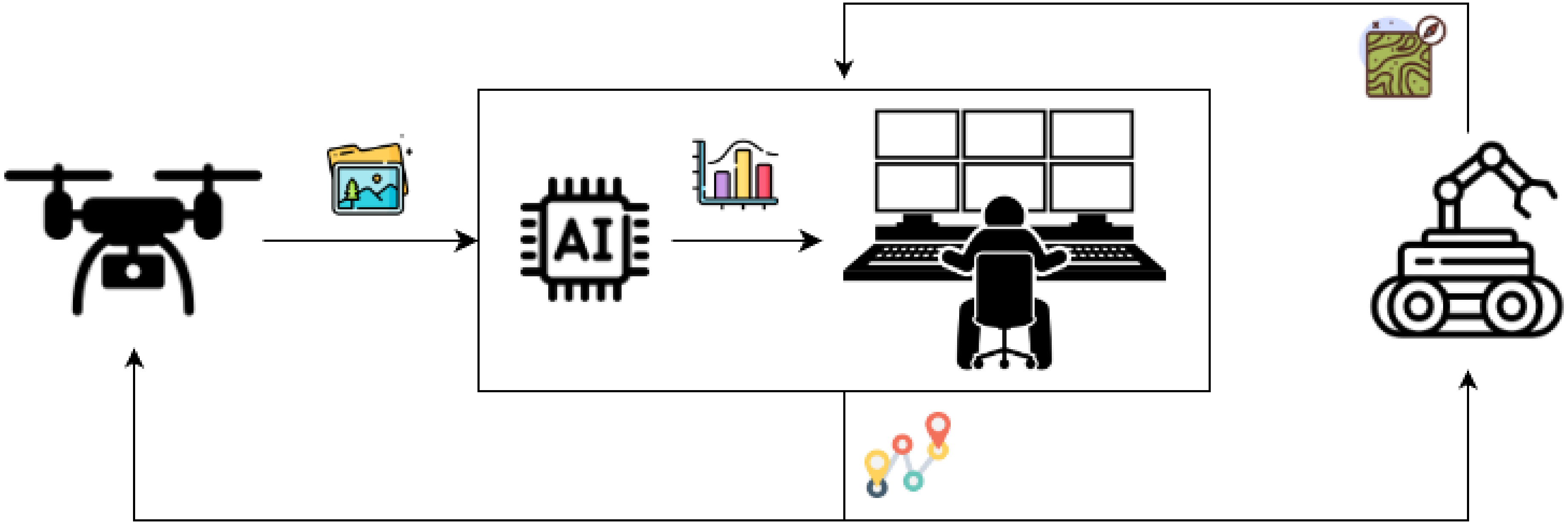
Architecture 1 – Centralized, Human

- The UAV has no intelligence. It goes to the blocks in the route received and send captured images to C2.
- C2 Human is completely responsible for analyzing images and updating the route to avoid as many mines as possible.
- The UGV has no intelligence. It proceeds following the latest route it received and report actual information of the current terrain to C2.
- Human reviewer makes the key decisions: Route for UGV and UAV by choosing the next node.



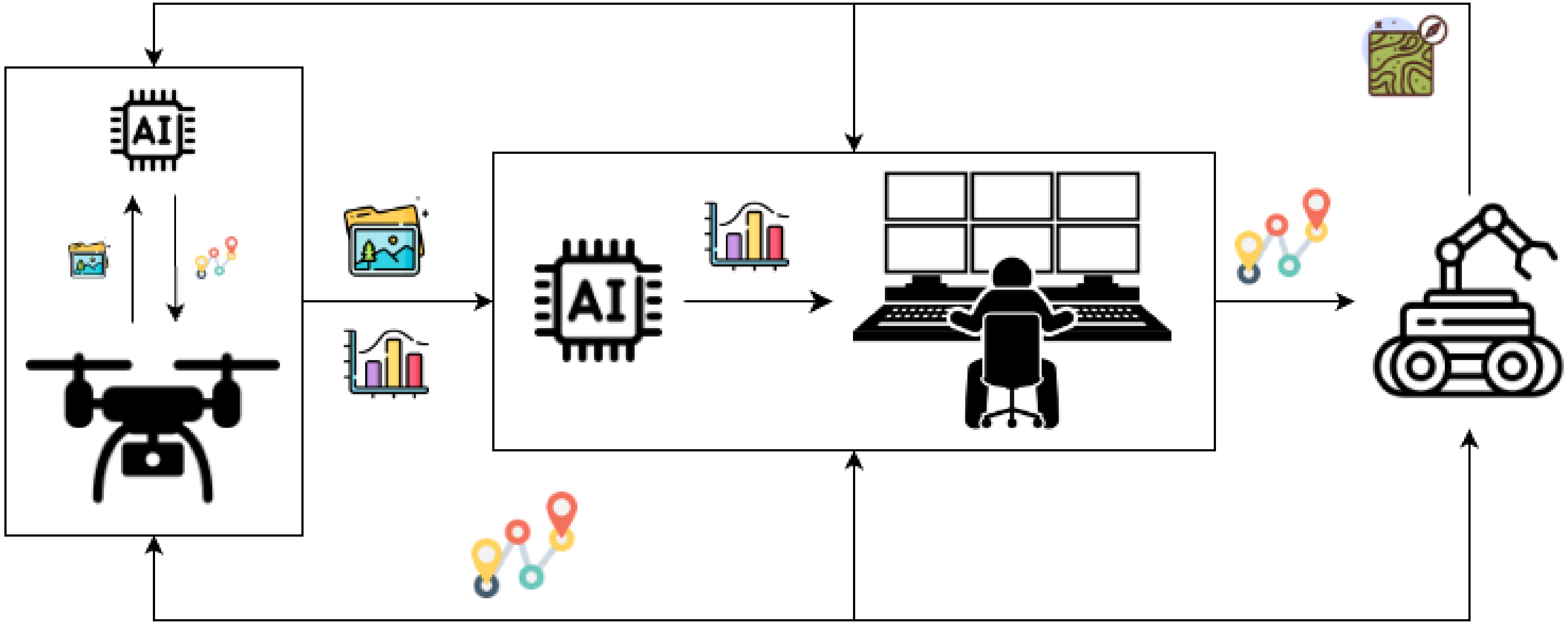
Architecture 2 – Centralized, AI

- New Entity Involved
  - Command and Control Center AI (C2 AI): C2 AI and C2 Human will act as one, sharing the channel of receiving and sending information.
- Changes:
  - C2 AI will first receive the image captured and decide whether a human reviewer should get involved.
  - The decision is based on multiple factors like weather, terrain type, lighting condition, and its confidence on the detection results.
  - The ground truth information from UGV will be added to history data for future decision making.



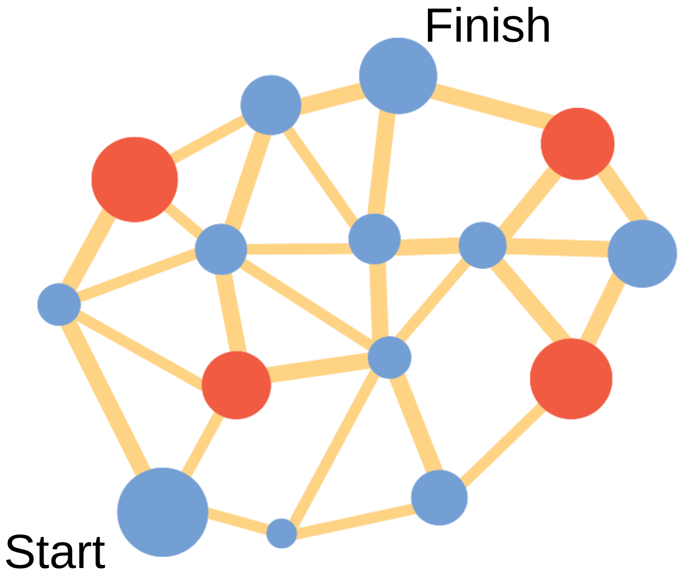
Architecture 3: Decentralized

- Updated Entity:
  - Instead of only sending pictures to C2, UAV has a smaller on-board AI to help make decisions.
  - C2 help making decision only when AI on UAV has high uncertainty.



How do we compare the architectures?

- Metrics for comparison:
  - Expected time to go from start to finish
  - Suitability across different network types
  - Hardware requirements and limitations
  - Lethality



Blue – explored by UAV
Red – unexplored by UAV
Width – proportional to the time

Variables Required – Simulation Arguments

In order to run the simulation, we need to know:

- A map with start and end goal, terrain information and ground truth mine location.
- A simulated human review and AI detection results and on all the blocks in the map.
- Planning algorithms required to be tested.

Row Index

Column Index	1	2	3	4	5	6	7	8	9	10	
1							🟥				
2		🟥		🟥							
3					x	x			🟥		
4	🟥			x		🟥	x				
5		🟥		x	🟥		🟥	x	x	x	**goal**
6			🟥	x		🟥				🟥	
7				x			🟥				
8			x	🟥	🟥	🟥			🟥		
9		x	🟥								
10		x			🟥	🟥		🟥		🟥	
		start									

x	planned route
🟥	mine location

Expected output map from the simulation, where x denotes the route planned by C2

Variables Required – Time Calculation

- From the given information, we've already know:
  - Time of decision making
    - ○ $t_{ai\_decision} = 1\ min$ - time for AI model to determine whether there is a mine using the image UAV captured
    - ○ $t_{human\_decision} = 30\ min$ - time for human to determine whether there is a mine using the image UAV captured
  - Time of moving physical entities
    - ○ $t_{uav} = 1\ min$ –time for UAV to capture and transmit imagery to Command Control
    - ○ $t_{ugv\_block\_no\_mine} = 20\ min$ – time for UGV to traverse the terrain if mine not encountered
    - ○ $t_{ugv\_clear\_mine} = 40\ min$ – time for UGV to clear a mine
    - ○ $t_{ugv\_block\_with\_mine} = t_{ugv\_clear\_mine} + t_{ugv\_block\_no\_mine} = 60\ min$ - time for UGV to traverse a block with mine
- To calculate the total time, we also need to know:
  - Total time used to find the initial block UGV to move to T_{prep} (affected by UAV capture time and total decision time)
  - number of terrains UGV traversed with mine, N_{mine}, and number of terrains UGV traversed without mine, N_{no_mine}
- The total time can then be calculated as:
  - $T = T_{prep} + N_{mine}\ \times\ t_{ugv\_block\_with\_mine} + N_{no\_mine}\ \times\ t_{ugv\_block\_no\_mine}$

Phases II & III

Phase II

- Phase II will software implementation of methods and simulation of the architectures proposed in Phase I.
- Measures of performance identified in Phase I will be computed.
- Architectures will be compared based on the performance.
- Account for changes in the operational environment during the task.
- Account for limited bandwidth and errors/loss of information during communication.

Phase III

- Phase III will involve testing the architecture on previously unseen network structures.
- New mission scenarios will be used to evaluate the different architectures.
- Consider Lethality while making decisions.
- Expand strategy to leverage multiple UAVs/UGV.
- Plan human movement along with UGVs.

Possible Extension: Physical Implementation

- Plan:
  - Set up a test network structure with obstacles simulating mines
  - Use UAVs to scan the network, with view obfuscated in parts
  - Benchmark the architectures in terms of expected time, computational power, bandwidth required

- Hardware
  - Turtle Bots
  - Drones
- Leverage existing facilities at Purdue for test and evaluation
  - Purdue UAS Research and Test Facility (PURT)

Acknowledgements

SERC WRT-1085: Trusted Artificial Intelligence (AI) Systems Engineering (SE) Challenge: Seed Funding

Prime contract number: *HQ003419D0003, DO HQ003423F0495*

Subcontract Number 2103596-05

SYSTEMS ENGINEERING RESEARCH CENTER