Trusted AI Challenge
Phase II: Architectures and Approach

Athul C. Dharmarajan, Vikranth Gadi, Zichong Yang, Bradley Feng, Jitesh H. Panchal

School of Mechanical Engineering, Purdue University

Supported by: SERC WRT-1085: Trusted Artificial Intelligence (AI) Systems Engineering (SE) Challenge: Seed Funding, Prime contract number: HQ003419D0003, DO HQ003423F0495, Subcontract Number 2103596-05

Abstract

This whitepaper discusses the work done towards phase II of the Systems Engineering Research Center (SERC) Trusted Artificial Intelligence Systems Engineering Challenge. We describe our interpretation of the problem, methodology proposed to explore and solve the problem, as well as experiments and results. We first characterize the key assumptions required to propose a decision system for the task of navigating from the start to the end node of the network based on our observations on the given information. Then, we propose three different architectures A1-A3 for the decision system and metrics to compare the performance of proposed architectures. Finally, we conduct the experiments by applying our proposed architectures on 10,000 random missions and find that architecture A1 is effective when AI is costly, architecture A2 is effective when AI is cheap and architecture A3 performs the best across a range of parameters and can balance accuracy with cost.

1 Problem Statement

The primary objective of the task is to ensure safe passage from the start point to the endpoint along a defined network. An overview of the problem is shown in Figure *1*. The task involves identifying the path taken along the network. The network has links that might contain a mine/IED (Improvised explosive device). There is a central command and control center (C2) responsible for the operation along with a UAV (Unmanned Aerial Vehicle) and UGV (Unmanned Ground Vehicle), which travel along the network. UGV can remove a mine upon encounter and the UAV can predict the likelihood of encountering a mine along a path using images fed to an AI-based model. C2 also has a human reviewer who can review the images and identify the likelihood of encountering a mine. We need to develop architectures of the decision system that integrates human operators with AI systems, ensuring that both entities can work together seamlessly. Another goal is to foster ‘trust’ between AI models and human operators by developing the notion of ‘trust’ between them. This involves identifying the specific roles that humans and AI should play within different architectural frameworks to optimize performance and identifying when a decision should be deferred to human reviewers. Finally, comparing various architectures across different performance metrics will help determine the most effective configurations for achieving these objectives.

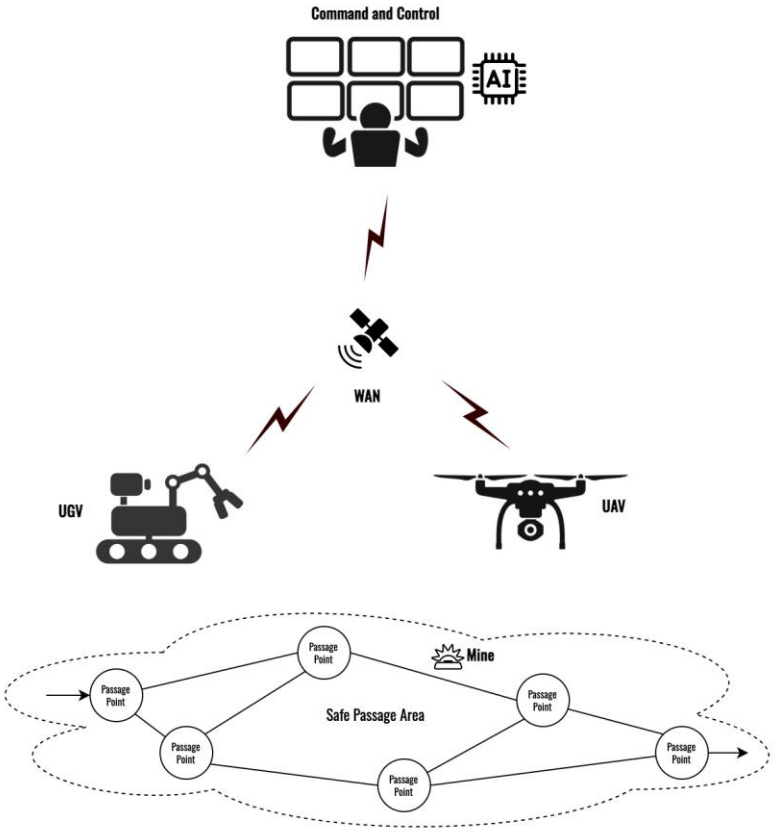

Figure 1: Overview of the problem

Based on the given problem, we made the following assumptions for each component are as follows to help framing our methodology:

- UGV
  - 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
  - UAV has 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, the UAV can share raw data collected from sensors (Video, images, LIDAR, etc.) and/or processed output (probability of encountering a mine)
- 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 in Phase III, 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 always enough human personnel available to make decisions and analyze the inputs

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

2 Methodology

2.1 Agent Behaviors

To implement the given problem in the simulation environment, there are still a few undefined behaviors for each agent not related to the variations in architectures this whitepaper discussed. Therefore, we design the behavior of each agent as follows, based on the information provided and key assumptions in *Section 1*.

- **UAV Scanning:** The UAV is only responsible for obtaining images from each nodes of the terrain. In our methods, all UAV agents will follow a path planned by A* algorithm. Each time the UAV traverses to a new node, it will send the images to the C2 and wait until the results. If a mine is detected, the C2 will plan a new path, by treating the node with a mine as an obstacle. The UAV will then retreat to the previous node and follow the new path.
- **Human and C2:** The human, along with the Command and Control Center, will be responsible for centralized communication, which includes receiving images from UAV, sending images to AI, sending planned path to UAV and UGV, as well as getting the current status of UAV and UGV. The human will also be responsible for reviewing the images if the algorithm of the architecture requires the images of the node to be reviewed manually.
- **AI Model:** The AI model is only responsible for reviewing the images and determining whether there is a mine at the given node or not. The results of the AI predictions are in the form of confidence percentage, which will be used to determine whether additional human review is required.
- **UGV:** UGV starts moving once a safe passage has been identified after scanning using UAV. There can still be mines as both human and AI are not 100% accurate. In that case, the UGV can clear the mine and continue moving along the path.

2.2 Decision-Making Architectures

We describe the architecture of the decision-making system that determines the key decisions involved in the operation during a mission. An example of the decision can be which path to take at each step, depending on the likelihood of encountering a mine/IED. The architecture will define the flow of information, who is responsible for making key decisions, and how the decisions will be deferred to human reviewers due to lack of information.

2.2.1 Architecture A1: Centralized with Human Reviewer

In Architecture A1 (Figure *2*), the Human reviewer is responsible for all the key decisions. The human reviewer decides the path for the UAV to scan based on the images fed by the UAV. The reviewer also determines the path for the UGV to navigate based on the information from the UAV. Being a central architecture, this architecture depends on having enough bandwidth and low latency to ensure smooth communication between agents and the C2. This architecture makes minimal use of artificial intelligence-based models, indicating a low amount of trust. Human review can be slow and comes at the cost of time.

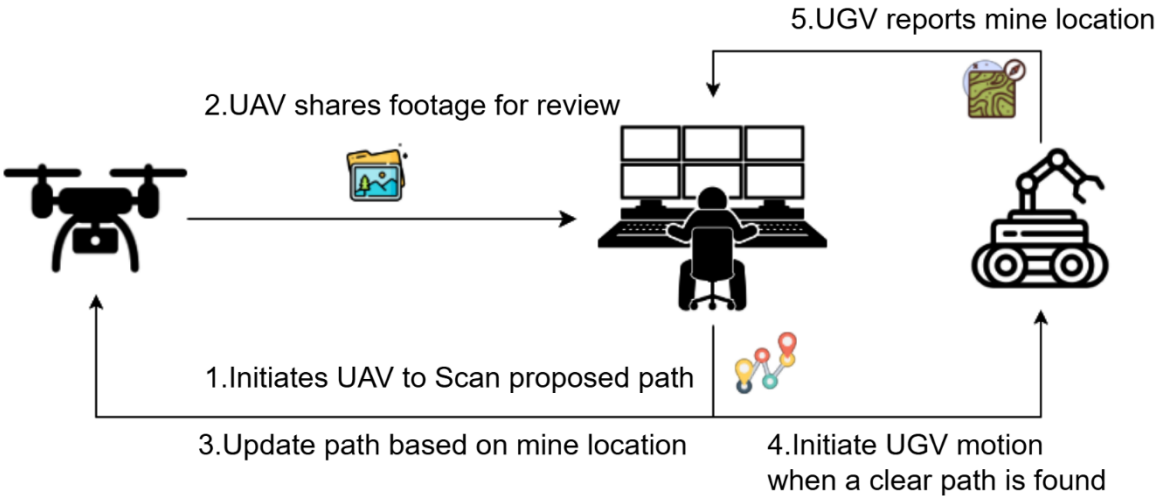

Figure 2. Architecture A1: Information Flow

2.2.2 Architecture A2: Centralized with AI

Compared to A1, AI assistance is introduced to Command and Control in Architecture A2 (Figure *3*). As using AI to detect mines is significantly faster, C2 can make faster decisions to return to the updated route and direct the UAV to capture pictures along the route faster, increasing overall efficiency. The faster processing time comes at the cost of accuracy as AI is less accurate than human on average. This architecture represents the highest level of trust with human reviewer comfortable with delegating all the tasks to an AI.

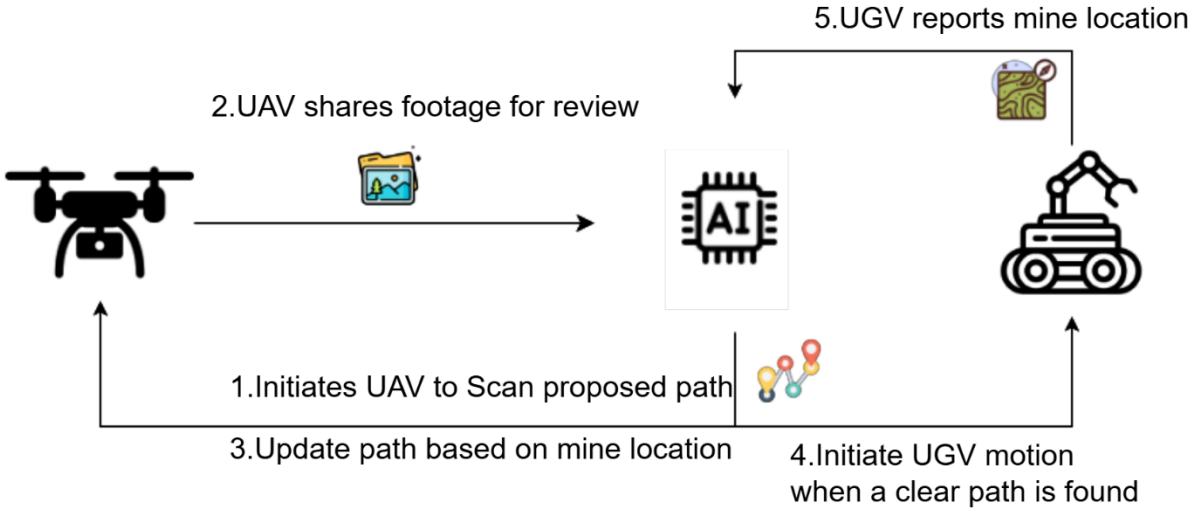

Figure 3. Architecture A2: Information Flow

Architecture A3: Centralized, Human-AI Teaming

Figure *4*). In this architecture, multi-arm bandit agents are responsible for determining whether to use a human reviewer or AI to review the footage. The multi-arm bandits make their decision based on the terrain, weather, cost and any information available which creates a discrepancy in the performance and cost of the AI and human reviewer.

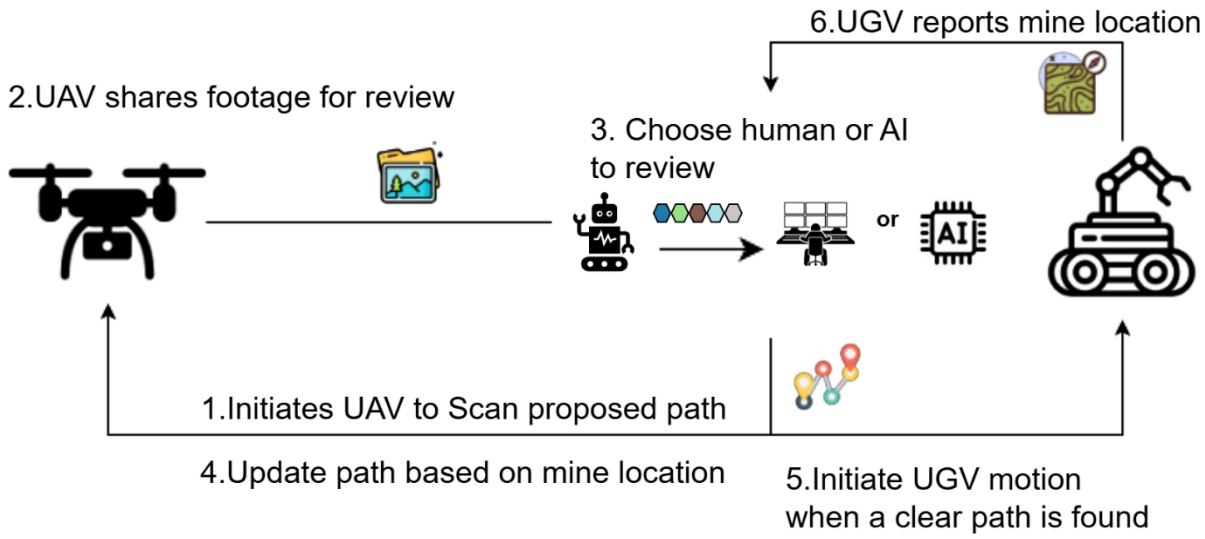

Figure 4. Architecture A3: Information Flow

3 Experiments

3.1 Experiment Setup

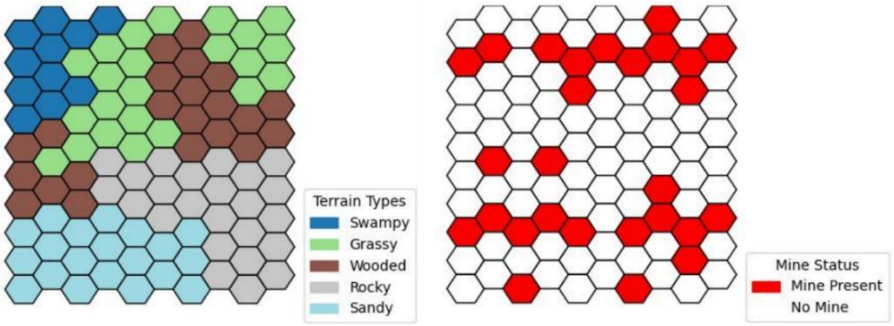

Figure 5. Given environment with terrains and mine locations

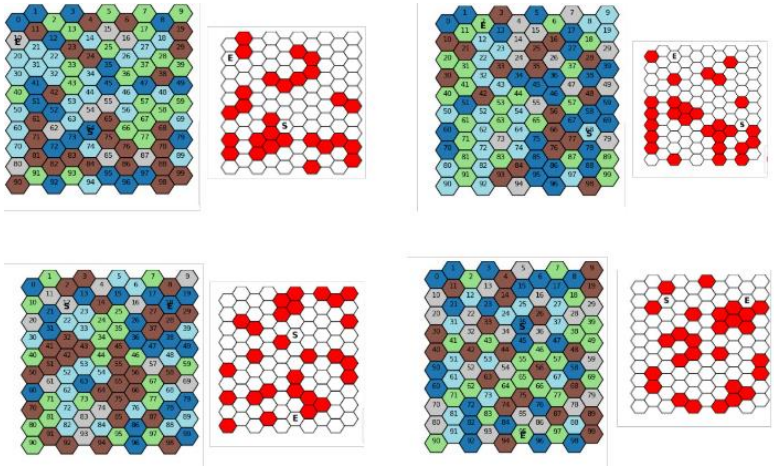

Figure 6. Example environments generated with randomized terrains, mine locations and start and end positions.

To make comparisons between different architectures, we generate environments randomly with different distributions of mines and terrain each time as shown in *Figure 6*. The accuracies are simulated for various terrain types and are drawn from beta distributions with fixed parameters. This helps us benchmark our approaches across a range of mission scenarios. The solving process is then simulated under different architectures to estimate the total mission cost. We also studied the impact of varying different parameters on the relative performance of different architectures.

3.2 Comparision of A1 and A2

We simulate A1 and A2 for the given costs (UAV movement: 1, AI query: 5, Human query: 30, UGV movement: 20, UGV clearance: 60). We also simulate for a higher AI cost (UAV movement: 1, AI query: 15, Human query: 30, UGV movement: 20, UGV clearance: 60). We observe that at a low cost for AI, the architecture A2 is better than A1. When the cost of AI is higher, the architecture A1 is better than A2 as shown in *Figure 7*.

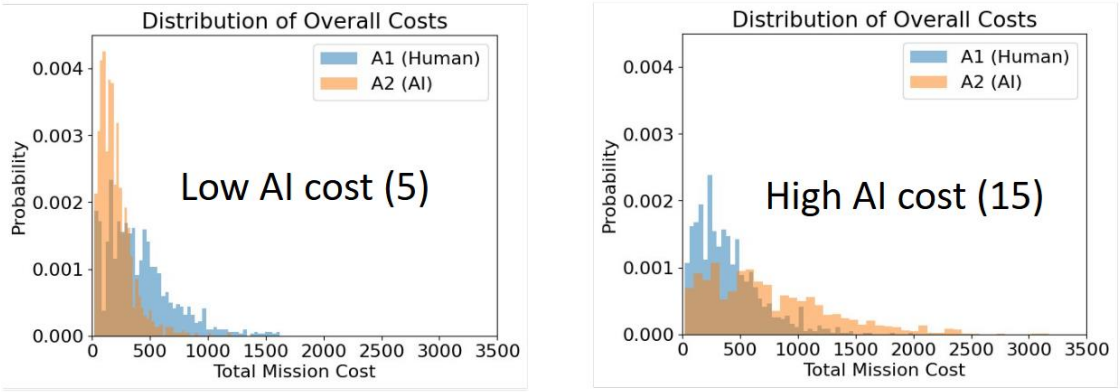

Figure 7. Results on missions with randomly initialized terrain and mine locations

3.2.1 Results on Reinforcement Learning for Architecture 3

We initialize five multi armed bandits corresponding to the five terrains. The MAB agents are defined as follows:

1. **Action Space:**
   Two arms represent the decision options:
   i. Arm 1: Delegate the task to the AI model.
   ii. Arm 2: Delegate the task to the human reviewer.
2. **Reward Mechanism:**
   The reward is calculated as a weighted sum of **accuracy** and **cost**:
   $Reward = (1-\lambda) \times Accuracy \times (-Normalized\ Cost)$
   where the parameter λ determines the priority between accuracy (for e.g.: $\lambda=0$) and cost-efficiency (for e.g.: $\lambda=1$).
3. **Reinforcement Learning Policy:**
   The MAB agent employs the **Upper Confidence Bound (UCB)** algorithm to explore and exploit its decision-making options. Initially, the agent assigns equal probabilities to both arms, ensuring unbiased exploration. Over successive episodes, the agent updates its policy based on received rewards, converging to an optimal decision strategy.

The MAB agent is initialized with a random policy and trained over 1000 episodes (trials), where terrain types (rocky, swampy, wooded, grassy, sandy) are randomized. For each task, the terrain type is identified. Based on the MAB policy, the agent selects an arm (human or AI). The agent observes the resulting reward and updates its policy accordingly. We describe the reinforcement learning results for simulations with reward function having high priority to accuracy ($\lambda=0$). After approximately 400 episodes, the agent's policy stabilizes, effectively allocating tasks to maximize performance as shown in Figure 8.

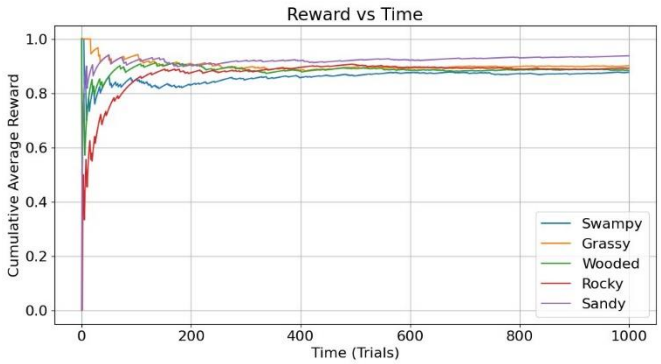

Figure 7. The policy of the MAB agents converges after 400 episodes

The probabilities shown in Figure 9 represent the likelihood of the MAB agent selecting AI or human reviewers for task execution based on terrain type. It is observed that AI is preferred in sandy, whereas the remaining terrains prefer Human reviewers.

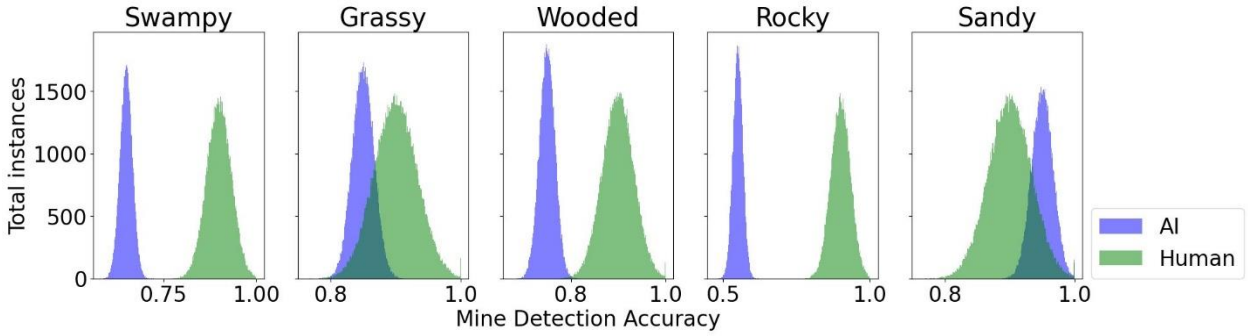

Figure 9. Rewards received by the MAB agents after using its arms

3.3 Heuristics for Human-AI teaming

We simulate 10,000 environments with 1000 episodes with four different reward functions (lambda values correspond to 0, 0.33, 0.67 and 1, with varying emphasis on cost and accuracy). Figure 10 shows how the preference between humans and AI changes for different terrains based on the tradeoff between cost and accuracy. If the priority is only accuracy, the human reviewer is overwhelmingly preferred. On the other hand, if priority is only cost, AI is overwhelmingly preferred. When we prioritize both cost and accuracy, the preference is dependent on the terrain.

Prioritize Accuracy ← → Prioritize Cost

	λ = 0	λ = 0.33	λ = 0.67	λ = 1
Rocky	Human	Human	Human	AI
Swampy	Human	Human	AI	AI
Wooded	Human	AI	AI	AI
Grassy	Human	AI	AI	AI
Sandy	AI	AI	AI	AI

Figure 10. Rewards received by the MAB agents after using its arms

4 Next Steps

In Phase III, the focus will be on expanding the scope of architecture explored and simulations performed in Phase II. As shown in *Section 2.1*, the behaviors of the agents defined are not optimal and can still be improved. Furthermore, some actions of different agents can be performed in parallel, which can further reduce the overall time required. In Phase III, the architecture will be tested on previously unseen network structures to assess its robustness and adaptability. This phase will involve using new mission scenarios to evaluate the performance of the different architectures under varied conditions. Another aspect of this phase will be considering lethality in decision-making processes to ensure the effectiveness and safety of operations. Additionally, the strategy can be expanded to leverage the capabilities of multiple UAVs and UGVs, enhancing the overall operational efficiency. Human movement can also be planned in conjunction with UGVs, leveraging seamless coordination and integration between human operators and autonomous systems. Additionally, it is crucial to account for any changes in the operational environment that may occur during the task. This includes considering factors such as limited bandwidth and potential errors or loss of information during communication, which could impact the overall performance and reliability of the system.