Project Information
Project Type: Team
Student Name(s): Jaydine Stiles, Joseph Ko
Mentor Name: Dr. Yugyung Lee
Research Title: Quantum-Enhanced Knowledge Graphs and Optimization for Trustworthy Multi-Agent Coordination in Digital Twin Robotics

Problem Statement, Hypothesis, and Research Questions:

Problem Statement:

As robotic systems become central to high-stakes operations—such as autonomous delivery, search and rescue, smart healthcare, and industrial automation—the need for secure, trustworthy, and scalable multi-agent coordination becomes critical. These systems often rely on digital twins: real-time virtual replicas of physical robots, environments, and tasks that support simulation, planning, and monitoring.

However, ensuring reliable multi-robot and human-robot collaboration in dynamic, uncertain environments remains a major challenge. This is further complicated by increasing risks around cybersecurity, data integrity, and trust erosion—where digital twins may become desynchronized or manipulated by adversarial attacks (e.g., sensor spoofing or communication interference), compromising decision-making and safety.
This research investigates how Quantum AI, particularly quantum-enhanced graph-based optimization, can be used to:

1. Improve task allocation, path planning, and load balancing in real time across robot teams and human collaborators.
2. Enable trust-aware coordination by detecting anomalies and optimizing decisions using secure graph structures within a digital twin framework.

We will construct dynamic, multi-layered knowledge graphs that represent physical robots, digital twins, environmental objects, task flows, and human operators—allowing for both operational efficiency and the detection of inconsistencies or adversarial manipulation.

Hypothesis:

Integrating Quantum AI with graph-based optimization in digital twin systems will improve multi-agent coordination, enable real-time anomaly detection, and enhance trust in robot decision-making.

Research Questions:

1. Optimization & Performance

- How can quantum-enhanced graph-based optimization improve real-time task allocation, path planning, and load balancing across robot teams and human operators?

2. Security & Trust

- How effective are dynamic graph structures in detecting desynchronization or adversarial attacks (e.g., spoofed sensor data) in a digital twin framework?
- Can trust-aware coordination models adaptively reassign tasks or isolate compromised agents in real time?

3. Digital Twin Integrity & Synchronization

- How can we design secure, real-time synchronization mechanisms between physical agents and their digital twins that prevent or recover from adversarial desynchronization while minimizing system downtime and information loss?

Literature Review

1. Modeling and Evaluating Trust Dynamics in Multi-Human Multi-Robot Task Allocation
    - Authors: Ike Obi, Ruiqi Wang, Wonse Jo, Byung-Cheol Min
    - Publication Year: 2025
    - Summary: This paper introduces the Expectation Confirmation Trust (ECT) model to analyze trust dynamics in multi-human, multi-robot (MH-MR) teams. The ECT model is grounded in the idea that trust plays a critical role in communications between MH-MR teams. These communications include things like effective coordination, adaptive decision making, and overall team performance in complex environments; the study demonstrates that incorporating trust models into task allocation algorithms improves task success rates and reduces completion times across various team configurations. The authors evaluated ECT against five other existing models and found that the ECT model is consistently better than the other models in performance. However, the study was limited to simulations and does not test with any real world deployment. Our work will extend on this study to test how well the ECT model does when tested with real deployment, not just simulations.
    - Link: arXiv:2409.16009arxiv.org+2arxiv.org+2arxiv.org+2

2. **TIP: A Trust Inference and Propagation Model in Multi-Human Multi-Robot Teams**

   - Authors: Guo, Yang, & Shi
   - Publication Year: 2023
   - Summary: Summary: This paper introduces the Trust Inference and Propagation (TIP) model, which accounts for both direct and indirect human experiences when modeling trust in human-robot teams. Direct experience refers to firsthand interactions, while indirect experience involves observing or learning about others' interactions with robots. TIP defines trust as a weighted combination of these two types of experiences and uses a gradient-based optimization method to fit personalized trust curves for each human-robot pair. The model was validated through a 15-session user study involving 30 participants performing a drone-based search task. Results show that TIP significantly reduces trust prediction error (RMSE ≈ 0.06–0.08) compared to models relying only on direct experience (RMSE ≈ 0.08–0.11). The authors also provide a theoretical guarantee that the inferred trust values will converge over repeated interactions. Limitations include limited scalability (tested only on two humans and two drones) and reliance on a single modality (task outcomes only), excluding potentially informative data like latency or communication behaviors.
   - Relevance to Our Project:
     - Digital-Twin Integration: TIP's trust update mechanism can be embedded into our digital twin knowledge graph to continuously track evolving trust in real-time.
     - Quantum-Enhanced Detection: TIP trust scores can feed into quantum-enhanced graph-based anomaly detection to flag potential manipulations (e.g., spoofed outcomes).
     - Adversarial Robustness: By simulating cyber-physical attacks in our digital twin system, we can observe TIP's behavior under adversarial conditions and adapt it to resist manipulation (e.g., through selective trust decay).

3. **Attacking Digital Twins of Robotic Systems to Compromise Security and Safety**

   - Authors: Carr, Wang, Wang & Han
   - Publication Year: 2022
   - Summary: This paper investigates the security vulnerabilities of digital twin systems (DTS) in robotic platforms that rely on the Robot Operating System (ROS). The authors demonstrate how *Person-in-the-Middle* (PitM) attacks can compromise the communication channel between a physical robot and its digital twin, leading to unsafe or malicious behaviors. Two case studies are presented: a TurtleBot 3 operating in Gazebo and a Universal Robot 10 in Unreal Engine. In both scenarios,

attackers inject falsified velocity or joint-trajectory messages, effectively altering robot behavior without detection. The paper emphasizes that DTS architectures often inherit the security flaws of ROS 1, including the lack of built-in encryption and authentication. While the authors suggest mitigations such as topic encryption and migrating to ROS 2, they do not provide implementation or empirical evaluation of these defenses. Furthermore, the work does not incorporate anomaly detection or trust mechanisms to enable system-level resilience. Limitations include the exclusive focus on ROS 1 (excluding more secure middleware like ROS 2) and a lack of implemented or tested countermeasures.

- Relevance to Our Project:
  - We can embed TIP-style trust scores into the DTS so that sudden drops in trust (or unexpected message flows) automatically trigger alerts or fail-safe behaviors.
  - We can simulate these PitM scenarios in our digital-twin framework, measure resilience metrics under attack, and evaluate specific countermeasures rather than only proposing them.

4. Exploring the Synergy of Human-Robot Teaming, Digital Twins, and Machine Learning
   - Authors: Evan Langas, Muhammad Zafar, Filippo Sanfilippo
   - Publication Year: 2025
   - Summary: This paper discusses the integration of human-robot teaming with digital twin technology and machine learning. It highlights how these technologies can enhance perception, prediction, and decision-making in human-centric industrial systems. The methodology consists of a progressive framework: they start with exploring human-robot interaction (HRI), to human-robot collaboration (HRC), to physical HRC (pHRC), to true human-robot teaming (HRT). By integrating real time sensing, simulation, and intelligent decision-making, digital twins and machine learning can enhance robot perception, safety, and adaptability in dynamic industrial environments. They also discuss the ethical considerations of giving a robot autonomy, privacy, and how the designs have to be human-centric to make it more ethical. Some of the limitations include the fact that this paper is mostly conceptual; there is no real empirical data, no tests/experiments really conducted. Our work will extend on this by attempting to enhance digital twin technology using machine learning using some of the conceptual ideas they have came up with for enhancements.
   - Link: https://arxiv.org/abs/2410.18195
5. NeuronsGym: A Hybrid Framework and Benchmark for Robot Navigation With Sim2Real Policy Learning
   - Authors: Haoran Li, Guangzheng Hu, Mingjun Ma, Yaran Chen, Dongbin Zhao
   - Publication Year: 2024

- Summary: This paper presents NeuronsGym, a hybrid framework and benchmark for studying Sim2Real policy learning in robot navigation. It combines a Unity3D-based simulator with a real-world testbed using RoboMaster EP robots to enable efficient training and direct policy deployment. The simulator models robot dynamics and sensor behavior (LiDAR, odometer, camera) with realistic noise and anomalies to better match physical environments. A novel metric, Safety-Weighted Path Length (SFPL), is introduced to evaluate navigation safety by penalizing collisions, improving upon traditional metrics like SPL.NeuronsGym supports tunable motion and sensor parameters to study the Sim2Real gap under varying conditions. Experiments show that domain randomization methods like UDR and SimOpt improve real-world performance, especially under the PointGoal navigation task. LiDAR fidelity has a greater impact on navigation transfer than dynamics modeling, and increased robot speed worsens Sim2Real performance. The simulator is highly efficient, supporting fast, headless-mode training for large-scale experiments. While the framework emphasizes physical safety during navigation, it does not explore ethical issues like robot autonomy or privacy. A key limitation is the use of fixed environments. Our work will use this to help us know how to simulate an environment in Unity, and have the robot path around objects, and we can add onto this by adding dynamic pathing aspects using the quantum AI to allow for our robot to be able to move around dynamic objects (for more adaptability in real world applications).
- Link: https://ieeexplore.ieee.org/abstract/document/10750009

6. Learning Attribute Attention and Retrospect Location for Instance Object Navigation
   - Authors: Yanwei Zheng, Yaling Li, Changrui Li, Taiqi Zhang, Yifei Zou, Dongxiao Yu
   - Publication Year: 2025
   - Summary: This paper presents a novel cascade architecture to improve instance-level object navigation (ION), where agents must locate specific objects using fine-grained attributes like color, material, and reference cues. Current ION approaches often struggle with attribute confusion and weak memory of explored areas. To address these limitations, the authors introduce two main components: the Object-Attribute Attention Graph (OAAG) and the Objective Retrospect and Location Module (ORLM). OAAG enhances object discrimination through two sub-graphs: the Object-Aware Graph (OAG), which dynamically learns relationships among observed objects, and the Attribute-Attention Graph (AAG), which uses attention to focus on key distinguishing attributes. This helps the agent identify specific instances, even within the same category. ORLM improves memory and spatial reasoning through a Back-tracker, which retains temporal and spatial object memory, and a Locator, which maps where targets were seen during exploration. Together, OAAG and ORLM

provide stronger perception and memory capabilities.Integrated with an A3C reinforcement learning framework and tested in the AI2-THOR simulator, the model achieves state-of-the-art performance on Instance-Localization, Instance-Navigation, and Category-Localization tasks. The approach nearly doubles success rates over baseline methods. This will be useful to our project because we can use their improved ideas of ION to help us navigate dynamic simulated environments.

- Link: https://dl.acm.org/doi/10.1145/3706423

7. Personalized Instance-based Navigation Toward User-Specific Objects in Realistic Environments
   - Authors: Luca Barsellotti, Roberto Bigazzi, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara
   - Publication Year: 2024
   - Summary: This paper introduces Personalized Instance-based Navigation (PIN), a task where agents must locate a specific object instance (e.g., a child's favorite teddy bear) among similar items, without relying on contextual cues. Unlike traditional navigation tasks, PIN emphasizes fine-grained object recognition and personalized retrieval in complex, realistic environments.To support this, the authors present PInNED, a dataset built from Habitat-Matterport3D scenes augmented with photo-realistic 3D objects from Objaverse-XL. Each episode provides the agent with reference images on neutral backgrounds and textual descriptions, without environmental context. Objects are procedurally placed on furniture like beds and tables, and same-category distractors are included to increase difficulty. The dataset includes 338 unique instances across 18 categories, with over 865k training and 1.2k validation episodes. Target objects used in validation are unseen during training, requiring zero-shot generalization. PIN differs from existing tasks by focusing on movable, injected objects and emphasizing instance-level recognition rather than category-based search. Experiments show that modular navigation approaches outperform end-to-end models, which struggle with distinguishing similar objects and recovering from mistakes. Despite improvements, the task remains challenging, especially when distractors are present. This benchmark establishes a foundation for future work on personalized embodied navigation.This will help us with our project because we can use their ideas of how to identify specific objects, make it better (maybe through the use of upgraded knowledge graphs and quantum AI) and then apply it to our robot.
   - Link: https://arxiv.org/abs/2410.18195

High Level Proposed Research:

AI Models:

- Graph Neural Networks (GNNs) for modeling complex relationships between robots, environments, and tasks within dynamic knowledge graphs.
- Quantum Approximate Optimization Algorithm (QAOA) for real-time task allocation and trust-aware path planning.
- TIP and ECT trust updates embedded in a live knowledge graph (robots, tasks, sensors, humans).
- Autoencoders/anomaly detection models to identify compromised data flows or adversarial manipulation.

Tools and Libraries:

- Qiskit for QAOA and other quantum needs.
- PyTorch for developing and training classical machine learning models like GNNs or autoencoders.
- Hugging Face Transformers for integrating LLM -based reasoning into agent decisions or knowledge graph summarization.
- NetworkX/DGL for graph-based modeling of robots, environments, and trust relationships.
- Digital Twin Platforms (simulation, Unity3D specifically) to represent the physical-virtual interaction part of this project.

Cybersecurity Context:

Since we are focusing on implementing a digital twin for our project, these are the main cybersecurity concerns and focuses:

- Anomaly detection for identifying inconsistent or malicious behavior in sensor data or coordination plans.
- Trust modeling among multi-agent systems to simulate and adapt to potential deception or interference.
- Resilience analysis under adversarial scenarios (e.g., communication interference, spoofed commands) to test robustness of coordination protocols.

Experimental Design:

*Note: All models, methods, and resources outlined in this design are subject to refinement as the project evolves and based on feedback from our mentor and experimental findings.

1: Simulation and Baseline Reproduction

- Goal: Replicate existing state-of-the-art (SOTA) models in simulation for object segmentation, navigation, and adversarial testing.

- Tasks:
    - Deploy Segment Anything (SAM) model to create visual masks in Unity-based RoboTHOR environments.
    - Integrate navigation baselines (e.g., A3C, PPO) for instance-based navigation using AI2-THOR/NeuronsGym.
    - Simulate failure episodes (e.g., path obstruction, repetitive motion loops) and log vision + action data.
- Tools:
    - Unity3D + AI2-THOR / NeuronsGym
    - PyTorch + SAM (ViT-H) in Colab
    - Stable Baselines3
- Evaluation:
    - Compare instance-navigation accuracy before and after segmentation
    - Visual validation of segmentation quality
    - Time-to-target and success rate metrics under default conditions

2: Adversarial Failure Injection and Trust Modeling

- Goal: Test how trust-aware models detect and adapt to compromised episodes in a digital twin environment.
- Tasks:
    - Modify RoboTHOR to simulate spoofed vision or sensor data (e.g., wrong position, blacked-out image, swapped object label).
    - Inject Expectation Confirmation Trust (ECT) and TIP trust scoring into agent-environment graphs.

    - Label and store failure episodes in JSON logs.
    - Analyze changes in trust scores during adversarial conditions.
- Tools:
    - AI2-THOR / Unity3D for digital twin
    - Custom wrapper to manipulate metadata (e.g., spoofed agent position)
    - TIP/ECT logic implemented using PyTorch or NetworkX
- Evaluation:
    - Trust decay rate vs. error severity
    - System reaction (e.g., isolation of bad agents, task reallocation)
    - Precision/recall of anomaly detection vs. baseline

3: Quantum-Enhanced Graph Optimization

- Goal: Test if quantum-enhanced graph optimization improves real-time coordination and decision-making.
- Tasks:
    - Construct knowledge graphs with nodes for agents, tasks, environment states, trust values.
    - Apply Quantum Approximate Optimization Algorithm (QAOA) for path planning, load balancing, or task assignment.
    - Compare against classical GNN-based solutions.
- Tools:
    - Qiskit for QAOA

    - NetworkX/DGL for graph construction
    - PyTorch for classical baselines
- Evaluation:
    - Task completion time and accuracy
    - Anomaly recovery time
    - Comparison of classical vs. quantum performance under identical graph structures

4: Cross-System Integration and Evaluation

- Goal: Evaluate how all modules interact in a closed-loop simulation.
- Tasks:
    - Integrate segmentation, trust scoring, anomaly injection, and QAOA-based planning in a single episode.
    - Visualize system decisions, trust flow, and agent paths over time.
    - Track digital twin synchronization status and anomaly recovery.
- Tools:
    - Streamlit for interactive interface (optional)
    - Matplotlib + PIL for visual logs
    - JSON logs and knowledge graphs for data storage
- Evaluation:
    - System-wide trust dynamics
    - Digital twin synchronization accuracy
    - Recovery rate from spoofing or logic inconsistencies

Reproduce One or More SOTA Baselines:

- Model Name: Automatic Mask Generator (based on Segment Anything)
- Source: Barsellotti, Luca, et al. “Personalized Instance-Based Navigation toward User-Specific Objects in Realistic Environments.” Advances in Neural Information Processing Systems, vol. 37, 16 Dec. 2024, pp. 11228–11250, https://arxiv.org/pdf/2410.18195.. Accessed 15 June 2025.
- Implementation Process:
    - Forked the Official Github repository, and used the Segment Anything model in Google Colab.
    - Loaded the pretrained Segment Anything model (ViT-H variant).
    - Used the SamAutomaticMaskGenerator to generate masks from input images.
    - No major code changes were made to the original repository aside from adapting it to a test image for demonstration.
- Environment:
    - Python 3.9
    - PyTorch
    - Torchvision
    - Segment-anything
- Evaluation:
    - Successfully reproduced the functionality of the SOTA model on a test image.
    - The model generated masks accurately and rapidly, validating its zero-shot capability to segment novel objects without task-specific training. However, the “happy medium” of masks is unknown, as too many masks creates too many objects and does not accurately represent the environment, and too few masks has a similar problem; the only difference is that it captures too few objects, not too many.
- Our own approach:
    - As of now, we have only reproduced the baseline SOTA model. While we haven’t built our own segmentation model yet, this reproduction sets a strong and reliable starting point to improving mask generation.
    - Our next steps would be to figure out if we can find out the “happy medium” of mask generation, and maybe find a way to use AI/Quantum AI to improve the accuracy of the mask generation to more accurately represent the objects in the room.