Multi-Robot Learning from Human Feedback

Connor Mattson

Kalhert School of Computing
University of Utah,
Salt Lake City, USA
c.mattson@utah.edu

Abstract

Learning from human feedback enables AI systems and robots to learn policies that align with human intent. While existing work has primarily examined learning from demonstrations, corrections, and preferences in single-agent settings, these ideas have yet to be fully extended to multi-agent domains—where cooperation, decentralization, and non-stationary dynamics demand new methods. In this thesis summary, I highlight my current work and outline future directions for multi-robot learning from human feedback, offering deployment strategies that align supervisor intent with robot teams in the real world.

Introduction

As robot systems become more integrated into human life, the complexity of their tasks—and the cooperation required to complete them—will continue to increase. While some tasks can be adequately performed by a single robot, multi-robot systems extend robotics into domains that demand teamwork, collaboration, and heterogeneous embodiments to solve complex problems like autonomous transportation, warehouse navigation, and swarm surveillance. Prior work has studied cooperation in these systems through the direct optimization of reward or cost functions, but real-world scenarios are often difficult to formalize, leading to deployments that fail to capture a human's true intent. My research seeks to answer the following research question: **How can we leverage human feedback to align multi-robot behavior with human intent?** My thesis focuses on two directions for answering this question:

- **Aim 1:** Leverage human data and representation learning to learn similarity measures for team behavior.
- **Aim 2:** Design novel methods that enable a single human supervisor to teach cooperative tasks to robot teams.

The following sections highlight both my existing contributions and my plans for advancing these aims, as summarized in Figure 1. By developing strategies for human-centered multi-agent learning, *my work focuses on reducing the expertise and effort required to train and fine-tune multi-robot systems in the real world*.

Aim 1: Behavior Representation Learning

A key challenge in training multi-agent policies lies in the increased complexity of state spaces and the representation of cooperative behavior. Traditionally, feature representations in multi-agent systems are either hand-crafted (Brown et al. 2018) or constructed by concatenating the raw observations of all agents into a single vector (Albrecht, Christianos, and Schäfer 2024), which scales poorly as the number of agents N increases. Learning behavior representations that map similar behaviors into the same region of a latent space provides a powerful tool for comparing multi-agent behaviors. Such learned spaces can be leveraged to formulate reward functions that more accurately capture a team's progress toward goal states, or as a search domain for exploring novel behavioral strategies.

Completed Work My existing work demonstrates that humans can collect counterfactual triplet examples of robot swarm behaviors to train a similarity embedding space. Using this space, I show that evolutionary optimization can effectively explore the learned behavioral space of multi-agent controllers, producing unique behaviors that may hold practical value for human engineers (Mattson and Brown 2023). Building on this idea, my subsequent work incorporates human-gating into the evolutionary process, allowing the reported behaviors to better align with a human's conception of novelty (Mattson, Clark, and Brown 2023), and explores methods for directly deploying these controllers on mobile robots in the real world (Mattson et al. 2025). In parallel work, I demonstrate that representation learning from human preferences can be used to formulate reward functions in single-agent tasks that zero-shot transfer to previously unseen robot embodiments. This represents a step towards **generalizable policy learning**, where *multiple candidate robots with differing embodiments* can execute tasks using a single learned reward function (Mattson, Aribandi, and Brown 2024).

Aim 2: Learning Cooperative Policies from Human Feedback

While optimization and multi-agent reinforcement learning can accomplish significant tasks, imitation learning and human feedback have proven particularly effective in producing policies aligned with human intent when a well-defined

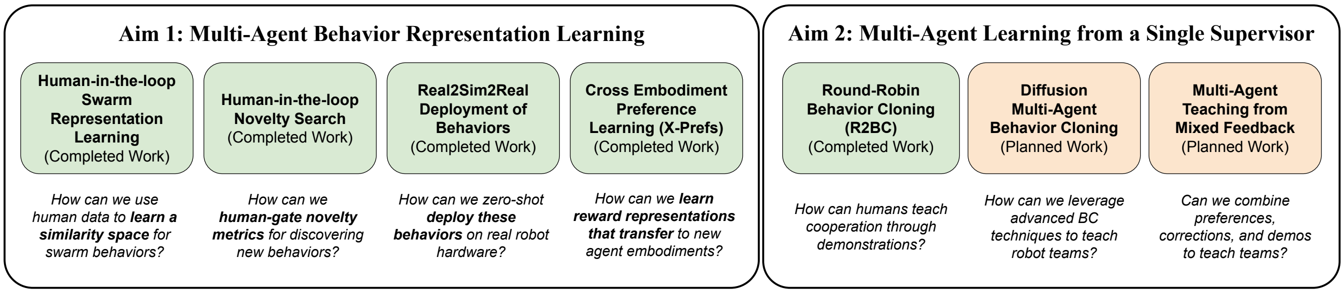

Figure 1: Research Aims and Contributions

reward function is unavailable. Even in single-robot settings, my PhD work demonstrates the difficulty that many humans have in teleoperation robots efficiently (Belsare et al. 2025). To provide demonstrations to robot teams, advanced teleoperation setups or multiple human demonstrators are often required (Turco et al. 2024). Without such infrastructure or additional operators, it is intuitively difficult for a single human to teach a team of N robots. My second research aim focuses on developing new algorithms and realistic teaching paradigms that enable a single human supervisor to team robot teams in cooperative tasks.

Completed Work In my recent submission to ICRA 2026, I introduced Round-Robin Behavior Cloning (R2BC), a method that allows a single human operator to iteratively provide demonstrations to individual agents in an online imitation learning framework (Mattson et al. 2026). This approach is the **first to ease the burden on human demonstrators** by relaxing the assumption that one person can simultaneously show all agents how to act in real time. *R2BC outperforms several single-human baselines by more than 3x*, enabling cooperative training and deployment for robot teams without requiring a fully specified reward function or multiple human demonstrators.

Proposed Work While R2BC provides a promising paradigm for multi-agent teaching, it has limitations: it relies on consistent unimodal demonstrations and remains vulnerable to covariate shift. To address these challenges, I plan to extend round-robin imitation learning with advanced behavior cloning strategies, such as diffusion policies. In parallel, I will study *how to effectively combine multiple forms of feedback into human-centered learning paradigms* for multi-agent systems, drawing from and expanding upon prior work in single-agent domains (Mehta and Losey 2024). At least one of these studies will include IRB-approved human trials, where I will evaluate both quantitative and qualitative improvements across diverse scenarios in which a human is tasked with teaching a robot team to complete a task.

Expected Progress Before AAAI'26 By January 2026, I expect to have preliminary findings from both proposed works. These results will discussed at the consortium and will highlight the unique challenges of teaching multi-robot teams to work together. *I believe these results will spark discussion within the AAAI community and promote future directions for human-centered multi-agent learning.*

References

Albrecht, S. V.; Christianos, F.; and Schäfer, L. 2024. *Multi-agent reinforcement learning: Foundations and modern approaches*. MIT Press.

Belsare, A.; Karimi, Z.; Mattson, C.; and Brown, D. S. 2025. Toward Zero-Shot User Intent Recognition in Shared Autonomy. In *Conference on Human-Robot Interaction*.

Brown, D. S.; Turner, R.; Hennigh, O.; and Loscalzo, S. 2018. Discovery and exploration of novel swarm behaviors given limited robot capabilities. In *Distributed Autonomous Robotic Systems*. Springer.

Mattson, C.; Aribandi, A.; and Brown, D. S. 2024. Representation alignment from human feedback for cross-embodiment reward learning from mixed-quality demonstrations. *Reinforcement Learning Conference (RLC)*.

Mattson, C.; and Brown, D. S. 2023. Leveraging human feedback to evolve and discover novel emergent behaviors in robot swarms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 56–64.

Mattson, C.; Clark, J. C.; and Brown, D. S. 2023. Exploring behavior discovery methods for heterogeneous swarms of limited-capability robots. In *International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*.

Mattson, C.; Raveendra, V.; Novoseller, E.; Waytowich, N.; Lawhern, V.; and Brown, D. S. 2026. R2BC: Multi-Agent Imitation Learning from Single-Agent Demonstrations. In *International Conference on Robotics Automation (ICRA)*, (In Review).

Mattson, C.; Raveendra, V.; Vega, R.; Nowzari, C.; Drew, D. S.; and Brown, D. S. 2025. Discovery and Deployment of Emergent Robot Swarm Behaviors via Representation Learning and Real2Sim2Real Transfer. *International Conference on Autonomous Agents and Multiagent Systems*.

Mehta, S. A.; and Losey, D. P. 2024. Unified Learning from Demonstrations, Corrections, and Preferences during Physical Human-Robot Interaction. arXiv:2207.03395.

Turco, E.; Castellani, C.; Bo, V.; Pacchierotti, C.; Prattichizzo, D.; and Baldi, T. L. 2024. Reducing cognitive load in teleoperating swarms of robots through a data-driven shared control approach. In *2024 International Conference on Intelligent Robots and Systems (IROS)*.