AI-Assisted Coordination of Human Teams in Situated Tasks

Sangwon Seo, Zhiqin Qian, Vaibhav Unhelkar
Rice University
Houston, TX 77005, USA
{sangwon.seo, bill.qian, vaibhav.unhelkar}@rice.edu

Abstract

Effective teamwork is crucial in high-stakes domains, yet it is highly challenging to achieve. Team members often must make decisions with limited information and under constraints on communication and time. Recognizing both the value of human coaches as well as the challenges of integrating them into practical settings, we envision *AI-based coaching agents* to enhance team coordination and performance. This extended abstract introduces *AI Coaches* and *Coordinators*, highlights key research questions from both human and AI perspectives that must be addressed to realize them, and summarizes our recent work in developing algorithms and systems to bring AI Coaches and Coordinators to fruition.

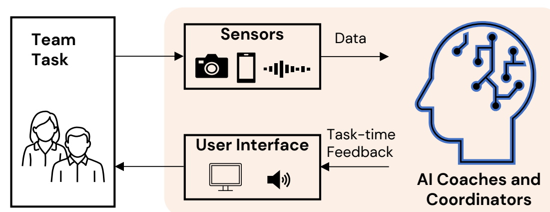

Figure 1: Schematic of the envisioned AI-based Coaching agents and Coordinators: agents that monitor a team during task execution and deliver interventions to enhance team coordination, training, and performance.

Introduction

Effective human teamwork is critical in many high-stakes domains, such as healthcare and disaster response. Tasks in these fields often demand goal-oriented collaboration within application-specific, real-world environments. Poor coordination among team members can result in suboptimal task performance and, in the worst cases, catastrophic failures (Seo et al. 2021). At the same time, optimizing teamwork is highly challenging. Team members often have limited observability of the task environment and may hold different mental models regarding their roles and plans. This issue is particularly pronounced in high-stakes domains, where time constraints restrict opportunities for communication.

In certain contexts, such as sports, teams rely on expert coaches to improve coordination and performance. These coaches monitor team execution in real-time and provide interventions, often in the form of verbal feedback. However, resource constraints make it difficult to include specialized coaches in most high-stakes teaming scenarios. Recognizing both the value of human coaches and the challenges of integrating them into practical settings, we envision *AI-based coaching agents* and *coordinators* (henceforth, *AI Coach*) to enhance team coordination and performance. Our research is building the computational foundation of AI Coaches (Seo and Unhelkar 2022; Seo, Han, and Unhelkar 2023; Seo and Unhelkar 2025). Concurrently, in collaboration with domain experts, we are actively exploring their potential to enhance surgical teamwork (Seo et al. 2021, 2025).

Similar to human coaches, an AI Coach monitors teams as they perform situated tasks and provides real-time interventions to improve teamwork. As illustrated in Figure 1, sensors are used to monitor both the team and the environment in which the situated task takes place. When the AI Coach detects imperfect coordination, it delivers interventions through its user interface to enhance teamwork. At its core, the AI Coach is driven by a suite of machine learning and decision-making algorithms that translate sensor-derived observations into interventions that enhance team coordination. In this abstract, we outline the key requirements of the computational core and summarize our progress in developing algorithms to fulfill them.

Scope Teamwork is influenced by various factors, and defining what constitutes effective teamwork can vary depending on the context and perspective. Hence, constructing the computational core of an AI Coach is a complex research endeavor, involving a vast design space and challenging AI problems. To arrive at a practical solution, our work has focused on two (of many) axes of teamwork: shared mental models and team performance. Shared mental models are a critical construct of teamwork that enables implicit coordination in a team (Bisbey, Traylor, and Salas 2021). Team performance is defined as the team's achievement of shared task objectives. Real-world scenarios often witness poor *team performance* stemming from a lack of *shared mental models*, reaffirming the utility of an AI Coach that can help establish shared mental models.

Copyright © 2025, Rice University. All rights reserved. Extended abstract presented at the AAAI Spring Symposium on the Current and Future Varieties of Human-AI Collaboration 2025.

Imitation Learning of Team Behavior

To develop computational approaches for enhancing teamwork, a mathematical model of teamwork that captures team members' behavior driven by their mental models is essential. Specifically, while predicting team behavior, this model should consider not only observable task contexts but also each member's unobservable plans and intents. Multi-agent imitation learning (MAIL) offers a promising avenue to develop such models using data of past task executions. More concretely, conventional imitation learning designs an agent behavior as a policy $\pi(a|o)$, where a_i is an action and o is an observation, and seeks to learn it from demonstrations. However, traditional MAIL algorithms neither explicitly model nor learn the influence of team members' mental models or intents on their decision-making.

Consequently, learning the required teamwork models necessitates specialized MAIL algorithms that explicitly model team members' intents. To derive such algorithms, we explicitly represent the i-th team member's policy as $\pi_i(a_i|o, x_i)$, where x_i denotes an intent. Also, since the team members may update their intent during the task, we model its temporal dynamics as $\zeta_i(x_i|o, x_i^-)$, where x_i^- is the previous intent. Consequently, imitation learning of team behavior requires learning both π and ζ from demonstrations in collaborative settings. Since annotations of the intents are highly challenging to collect, we assume only partial amount of them are optionally available.

To address this partially observable MAIL problem, we have developed Bayesian Team Imitation Learner (BTIL) and Deep Team Imitation Learner (DTIL) (Seo and Unhelkar 2022, 2025). Built upon mean-field variational inference, BTIL achieves sample- and label-efficient learning of teaming behaviors. However, BTIL struggles to scale up to domains with complex or continuous state spaces as it is based on a Bayesian approach. DTIL, on the other hand, utilizes distribution-matching-based online imitation learning and function approximation (specifically, neural networks) to learn team behavior models from demonstrations (Seo and Unhelkar 2024). DTIL effectively learns team member policies even for complex multi-agent scenarios (such as those with high-dimensional states) through factored optimizations.

Automated Task-Time Team Intervention

Building on the teamwork models described above, the AI Coach requires reasoning algorithms to assess team alignment during task execution and prescribe interventions to enhance performance. To address this need, we have developed TIC: an algorithm that automatically generates interventions based on its sensor-derived observations and a learned model of teamwork (Seo, Han, and Unhelkar 2023). TIC specifically targets imperfect coordination resulting from a lack of shared mental models. Accordingly, its interventions are limited to offering suggestions that help align team members' intents.

TIC consists of two key components: an intent inference module and an intervention strategy module. The intent inference module employs a filtering algorithm that accounts for the possibility that team members may not always accept the AI Coach's suggestions. The intervention strategy module optimizes intervention decisions by balancing their costs and benefits. Rather than relying on rule-based methods, TIC algorithmically generates interventions by evaluating intent compatibility through the expected cumulative return, computed using task objectives and the teamwork model. Importantly, the cost of interventions must be considered, as excessive interventions may disrupt task execution and introduce unintended side effects.

In recent work, we developed a proof-of-concept AI Coach, Socratic, which integrates TIC with an interactive user interface (Seo et al. 2025). Human subject studies show that Socratic significantly improves team performance with minimal interventions. Participants also found Socratic helpful and trustworthy, reinforcing its potential for adoption.

Future Directions

Alongside these promising results highlighting the potential of AI Coaches and Coordinators, our findings also motivate research directions for developing more capable ones.

First, while BTIL and DTIL introduce novel team modeling capabilities, both are limited to shared mental models represented as a scalar intent and require the set of possible intents to be predefined. Future methods that leverage high-dimensional representations of mental models and adapt them dynamically during learning are highly relevant.

Second, current interventions are restricted to suggestions that align team members' mental models. Expanding the types interventions to include more varied and generalized recommendations would lead to more capable AI Coaches. Two particularly relevant directions are: (1) incorporating explainable AI to generate more interpretable suggestions – e.g., not only suggesting a more optimal shared plan, but also the reason behind its selection; and (2) leveraging novel methods for reward design to refine the AI Coach's intervention strategy to specify – both which plans the team should undertake and how they should execute these plans.

Finally, given that AI Coaches and Coordinators can make errors, a critical area for further research is ensuring their safe and responsible deployment. This includes studying how trust in AI Coaches and Coordinators can be effectively built, calibrated, and maintained to foster successful human-AI collaboration.

Acknowledgments

This extended abstract summarizes our prior and ongoing research on AI-assisted human teamwork. For more details, we direct readers to the papers referenced on the following page. We gratefully acknowledge our collaborators who have been instrumental in defining the concept of an AI Coach, especially Marco Zenati, Roger Dias, Julie Shah, and Eduardo Salas. Together, we are demonstrating the utility of this *Human-AI Collaboration*-focused research in the practical contexts of team coordination and clinical training. This research is supported in part by the National Science Foundation (NSF award #2205454), the Army Research Office, and Rice University.

References

Bisbey, T.; Traylor, A.; and Salas, E. 2021. Transforming teams of experts into expert teams: eight principles of expert team performance. *Journal of expertise*, 4(2).

Seo, S.; Han, B.; Harari, R.; Dias, R.; Zenati, M.; Salas, E.; and Unhelkar, V. 2025. Socratic: Enhancing Human Teamwork via AI-enabled Coaching. In *AAMAS*.

Seo, S.; Han, B.; and Unhelkar, V. 2023. Automated Task-Time Interventions to Improve Teamwork using Imitation Learning.

Seo, S.; Kennedy-Metz, L.; Zenati, M.; Shah, J.; Dias, R.; and Unhelkar, V. 2021. Towards an AI coach to infer team mental model alignment in healthcare. In *2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 39–44. IEEE.

Seo, S.; and Unhelkar, V. 2022. Semi-Supervised Imitation Learning of Team Policies from Suboptimal Demonstrations.

Seo, S.; and Unhelkar, V. 2024. IDIL: Imitation Learning of Intent-Driven Expert Behavior. In *AAMAS*, 1673–1682.

Seo, S.; and Unhelkar, V. 2025. Hierarchical Imitation Learning of Team Behavior from Heterogeneous Demonstrations. In *AAMAS*.