University of Pittsburgh | School of Computing and Information

Proposal Defense

Doctor of Philosophy in Intelligent Systems

"**A Computational Framework for Efficient Human-agent Teamwork**" by **Huao Li**

Date: Nov. 22, 2024
Time: 12 – 2 p.m.
Place: Room 301, Information Sciences Building, 135 N Bellefield Ave., Pittsburgh, PA 15213

Committee:

- Michael Lewis, Professor, School of Computing and Information, University of Pittsburgh
- Na Du, Assistant Professor, School of Computing and Information, University of Pittsburgh
- Erin Walker, Associate Professor, School of Computing and Information; Learning Research and Development Center, University of Pittsburgh
- Katia Sycara, Research Professor, Robotics Institute, Carnegie Mellon University

Abstract:
This thesis proposal addresses critical challenges in developing intelligent agents that can collaborate effectively with human teammates. As AI systems become increasingly prevalent in collaborative environments, there is a pressing need for agents that can understand, adapt to, and communicate naturally with humans. To meet this challenge, this research explores the development of a computational framework aimed at enhancing human-agent teamwork. The framework consists of three core components: computational Theory of Mind (ToM) modeling, mutual adaptation within human-agent teams, and interpretable communication for collaborative interaction.

A key element of the proposal is the development of a neural network-based Theory of Mind model, designed to infer human mental states, beliefs, and intentions in team tasks. This model allows agents to reason about their human teammates' knowledge and perspectives, facilitating more effective collaboration. The research also investigates an adaptive agent architecture, enabling agents to recognize human strategies and adjust their own behavior in real time during collaborative tasks. Preliminary results demonstrate novel methods for generating synthetic human data through embodied large language model (LLM) agents and aligning multi-agent reinforcement learning (MARL) communication with human language, promoting more natural interactions in human-agent teams. Through experiments conducted in simulated teamwork environments, the thesis shows how these computational approaches can improve coordination, adaptation, and communication in human-agent collaboration.

In the proposed work, we will focus on integrating computational Theory of Mind models to guide agents in learning adaptive behaviors and communication strategies within human-agent teams. The ultimate goal is to develop socially intelligent agents that can engage in natural and intuitive teamwork with humans.