COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

DISSERTATION DEFENSE

Paul Bara

Theory of Mind in Collaborative Tasks between Embodied Agents

Monday, November 28, 2022
11:00 – 1:00 pm
Hybrid – Zoom
BBB 3901

ABSTRACT: An ideal integration of autonomous agents in a human world implies that they are able to collaborate on human terms. In particular, theory of mind plays an important role in maintaining common ground during human collaboration and communication. To enable the theory of mind modeling in situated interactions, we introduce a fine-grained dataset of collaborative tasks performed by pairs of human subjects in the 3D virtual blocks world of Minecraft. It provides information that captures partners' beliefs of the world and of each other as an interaction unfolds, bringing abundant opportunities to study human collaborative behaviors in situated language communication.

Collaborative tasks often begin with partial task knowledge and incomplete plans from each partner. To complete these tasks, partners need to engage in situated communication with their partners and coordinate their partial plans towards a complete plan to achieve a joint task goal. While such collaboration seems effortless in a human-human team, it is extremely challenging for human-AI collaboration. To address this limitation, we take a step towards Collaborative Plan Acquisition, where humans and agents strive to learn and communicate with each other to acquire a complete plan for joint tasks. Specifically, we formulate a novel problem that tasks agents to predict the missing task knowledge for themselves and for their partner based on rich perceptual and dialogue history. We extend a situated dialogue benchmark for symmetric collaborative tasks in a 3D blocks world and investigate computational strategies for plan acquisition.

We continue by showing that this mitigation of knowledge disparity can be effectively used to improve decision-making. Furthermore, we also show that when paired with a human subject, behavior incorporating this knowledge mitigation is preferred by the human subjects and is perceived as a more effective collaboration by gauging if the human accepts the autonomous agent as a partner and if the human feels accepted as a partner by the autonomous agent.

CHAIR: Prof. Joyce Chai