

## Teaching by Demonstration: Towards a Bayesian Approach to Explaining Robot Policies by Peizhu Qian

**Abstract:** Interactions between human and artificial agents (e.g., robots, semi-autonomous vehicles, and decision-support aids) continue to become commonplace. Artificial agents interact with their end users to provide services in homes, schools, and hospitals. They work collaboratively with human users on tasks such as disaster response, manufacturing, and transportation. While artificial agents often behave according to pre-programmed policies or objective functions, their behaviors are not always intuitive to human users. For example, a user might wonder how a Roomba (a home vacuum robot) does not miss any dirt or why it might run into a dead-end corner. To truly realize agents' benefits, ensure safety, and avoid unintended side-effects, it is essential that the end users understand the artificial agents' behavior. To facilitate this understanding, approaches that explain the behaviors of these agents will be critical. In this work, we investigate approaches to explain the policies of artificial agents to non-expert human end users through demonstrations of agents' behavior.

We propose an Agent-Teacher-Student paradigm to teach (Teacher) the policy of a robot (Agent) to a human user (Student). The agent in this paradigm behaves according to a Markovian policy, and the teacher (also a virtual agent) chooses which state-action pairs (called the teacher's instructions) from the policy to show to the student. The goal for the teacher is to select effective and informational instructions through our algorithm to maximize the student's knowledge on the agent's policy. In contrast to prior works which adopt a teacher-centric perspective, our approach adopts a student-centric perspective for teaching the agent's policy. In particular, our approach explicitly models the student's mental model and chooses instructions tailored to the student using Monte Carlo Tree Search (an anytime algorithm for decision-making under partial observability). Preliminary evaluations in synthetic grid-world tasks and with simulated users provide proof of concept for our approach. Encouraged by these results, the next steps of our ongoing research include applications to more complex tasks and evaluations with human users in the loop.