

MARCER: A Multimodal Augmented Reality System for Composing and Executing Robot Action Plans



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Bryce Ikeda*
biked@unc.edu

Maitrey Gramopadhye*
maitrey@email.unc.edu

LillyAnn Nekervis
lnekervis@unc.edu

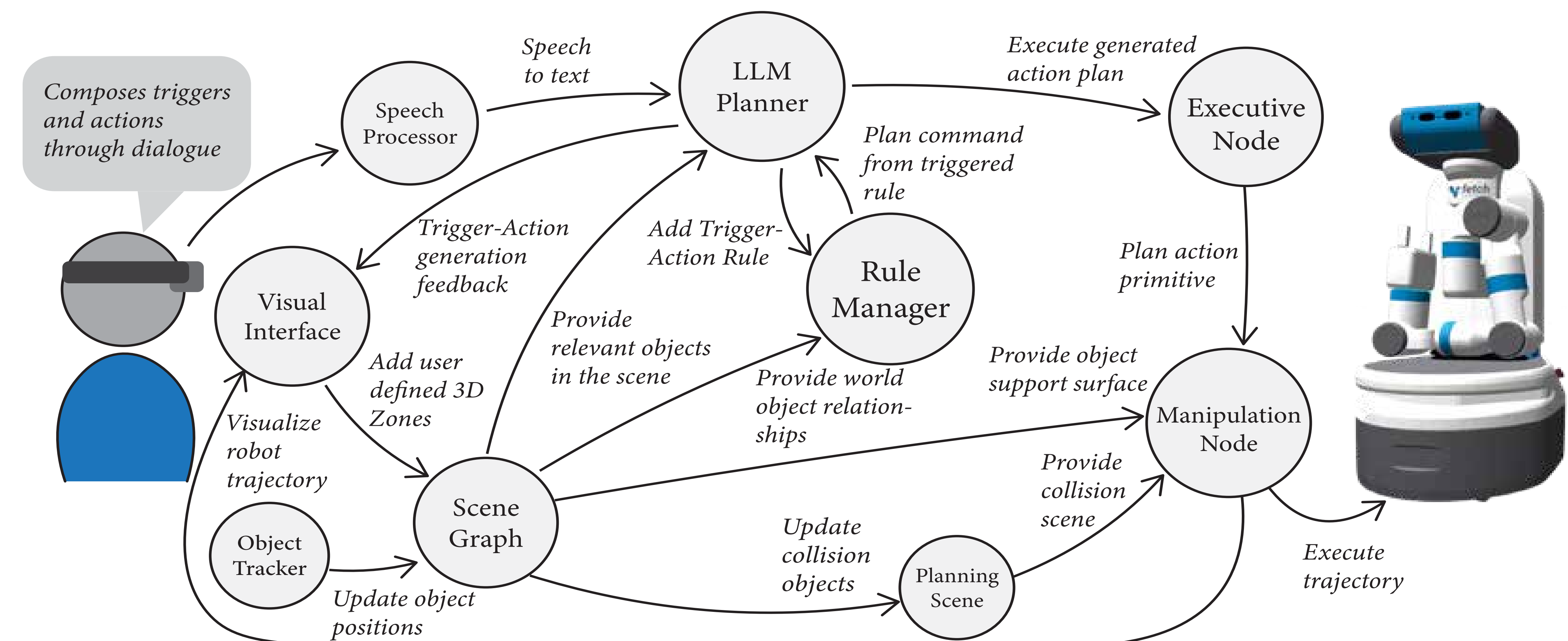
Daniel Szafir
dszafir@cs.unc.edu

Introduction

MARCER is a multimodal robot programming system that combines natural language and augmented reality for an intuitive user experience. Users program robots by articulating commands like 'Move the groceries that are on the table to the shelves.' A Large Language Model (LLM) interprets these commands and integrates environmental context to create editable triggers and Action Plans. An augmented reality (AR) interface allows users to set virtual constraints, visualize and edit trigger-action pairs and plans, and use a digital twin to preview and adjust robot activities in real time.



System Implementation



System Design

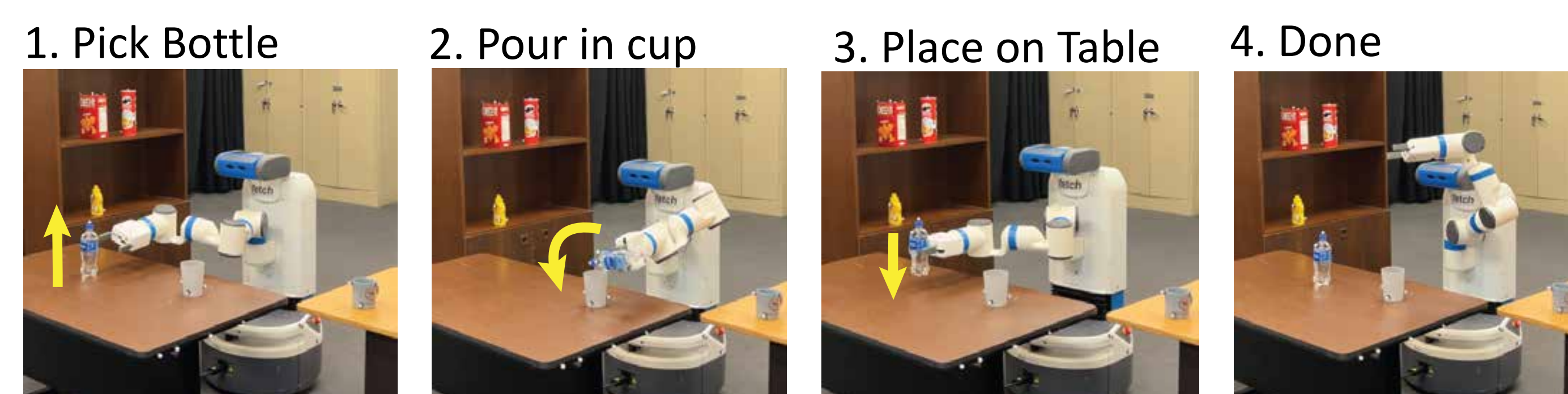
1. LLM Planner: Generates action plan from natural language input

To pour you a glass of water, I can pick the bottles of water from the table and pour it into the cup on the table. First, I will use the action "pick bottle" to grab the bottle of water from the table. Then, using the action "pour in cup," I will pour the water from the bottle into the cup on the table. This action will help quench the thirst by serving you a glass of water.

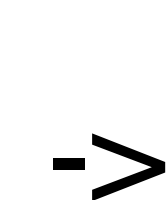
Action plan for "I'm thirsty, Can you pour me a glass of water?":

1. Pick bottle
2. Pour in cup
3. Place on table
4. Done

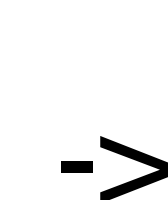
2. Translate detailed reasoning into actionable steps for robot execution.



3. Natural Language Input

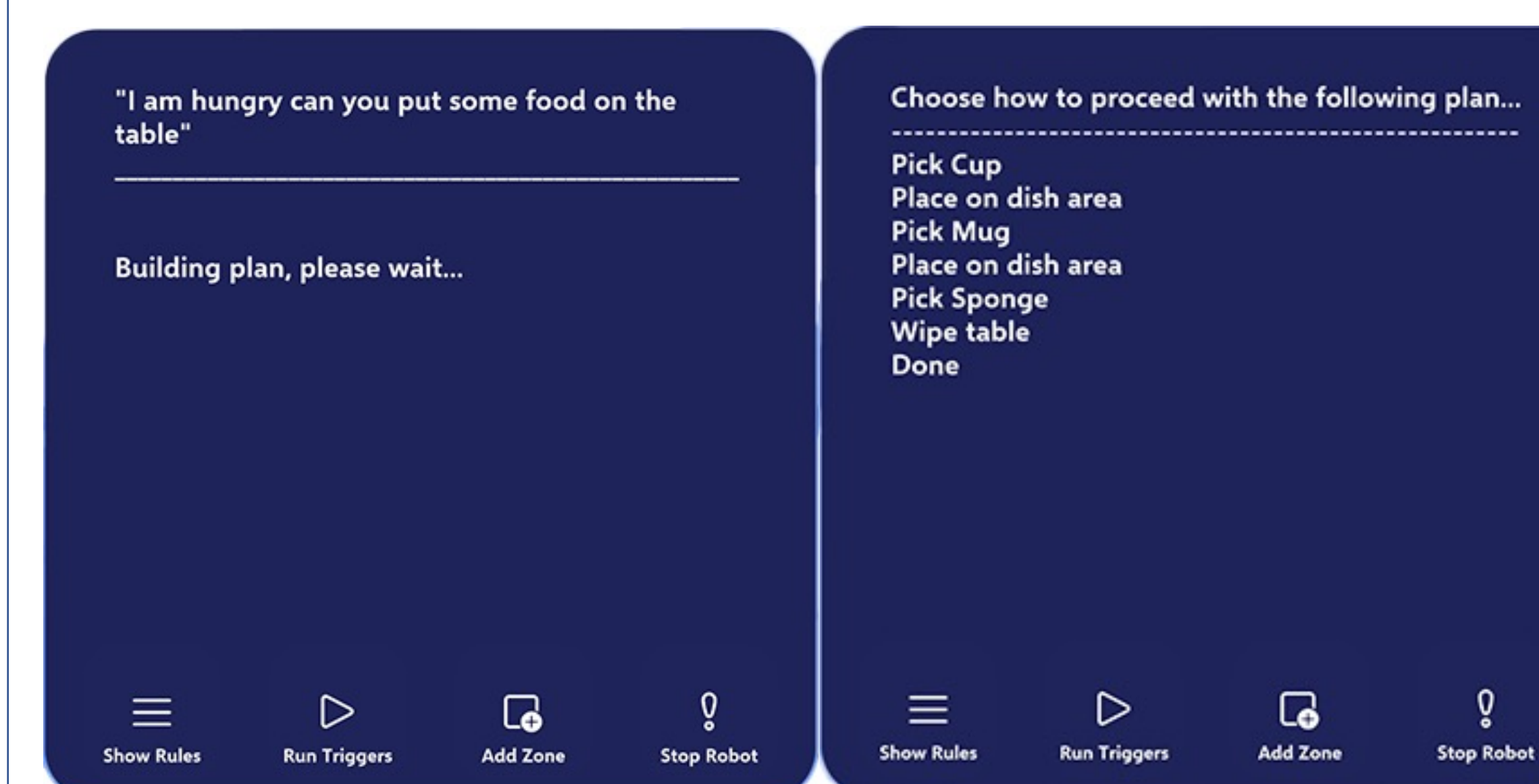


LLM generated Action plan



Execution with digital twin preview

Applications



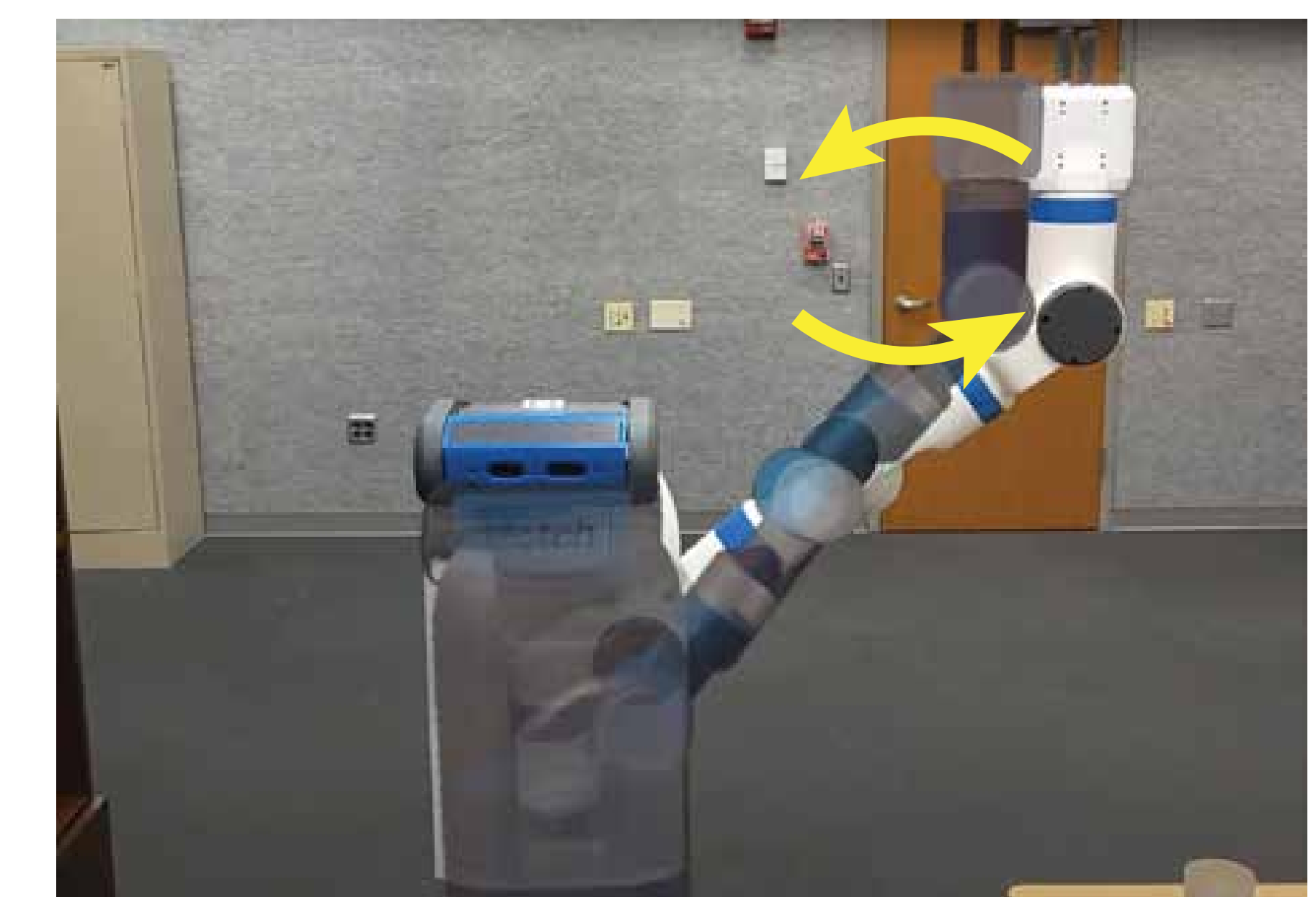
Direct Verbal Commands



Remote Interaction via World in Miniature (WiM)



Reactive Tasks and Zone Triggers



Social Interaction