

MOSAIC: A Modular System for Assistive and Interactive Cooking

Huaxiaoyue Wang*, Kushal Kedia*, Juntao Ren*, Rahma Abdullah, Atiksh Bhardwaj, Angela Chao, Kelly Y Chen, Nathaniel Chin, Prithwish Dan, Xinyi Fan, Gonzalo Gonzalez-Pumariega, Aditya Kompella, Maximus Adrian Pace, Yash Sharma, Xiangwan Sun, Neha Sunkara, Sanjiban Choudhury

Cornell University

<https://portal-cornell.github.io/MOSAIC/>



Fig. 1: **MOSAIC cooking in the kitchen.** (top) MOSAIC interacts with a user via natural language and controls a tabletop manipulator (R1) and a mobile manipulator (R2) to prepare vegetable soup with the user. (bottom) We evaluate MOSAIC on multiple recipes, involving a range of robot skills that interact with the human user and everyday objects.

Abstract—We present MOSAIC, a modular architecture for home robots to perform complex collaborative tasks, such as cooking with everyday users. MOSAIC tightly collaborates with humans, interacts with users using natural language, coordinates multiple robots, and manages an open vocabulary of everyday objects. At its core, MOSAIC employs modularity: it leverages

multiple large-scale pre-trained models for general tasks like language and image recognition, while using streamlined modules designed for task-specific control. We extensively evaluate MOSAIC on 60 end-to-end trials where two robots collaborate with a human user to cook a combination of 6 recipes. We also extensively test individual modules with 180 episodes of visuomotor picking, 60 episodes of human motion forecasting, and 46 online user evaluations of the task planner. We show that MOSAIC is able to efficiently collaborate with humans

* Denotes equal contribution.

Correspondence to: yukiwang@cs.cornell.edu