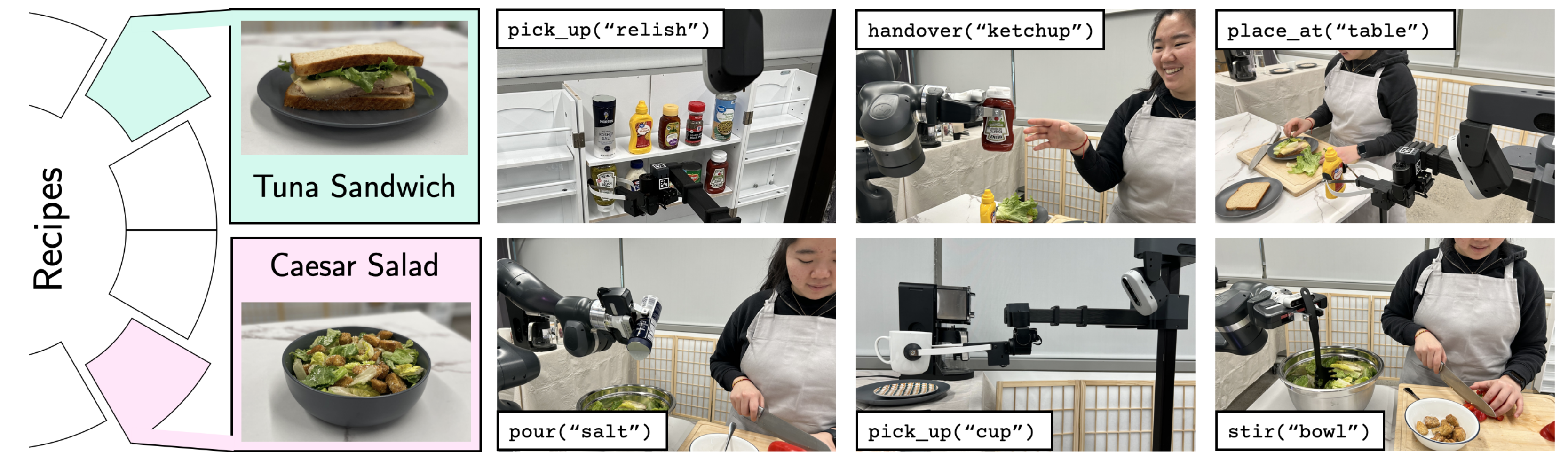# MOSAIC: Modular Foundation Models for Assistive and Interactive Cooking

Yuki 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 Pace, Yash Sharma, Xiangwan Sun, Neha Sunkara, **Sanjiban Choudhury**

Cornell Bowers C·IS Computer Science — CoRL — PORTAL — Scan me!
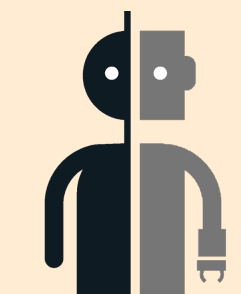
Desiderata for Collaborative Systems

1. Interact with users via natural language
2. Perform a range of skills manipulating everyday objects
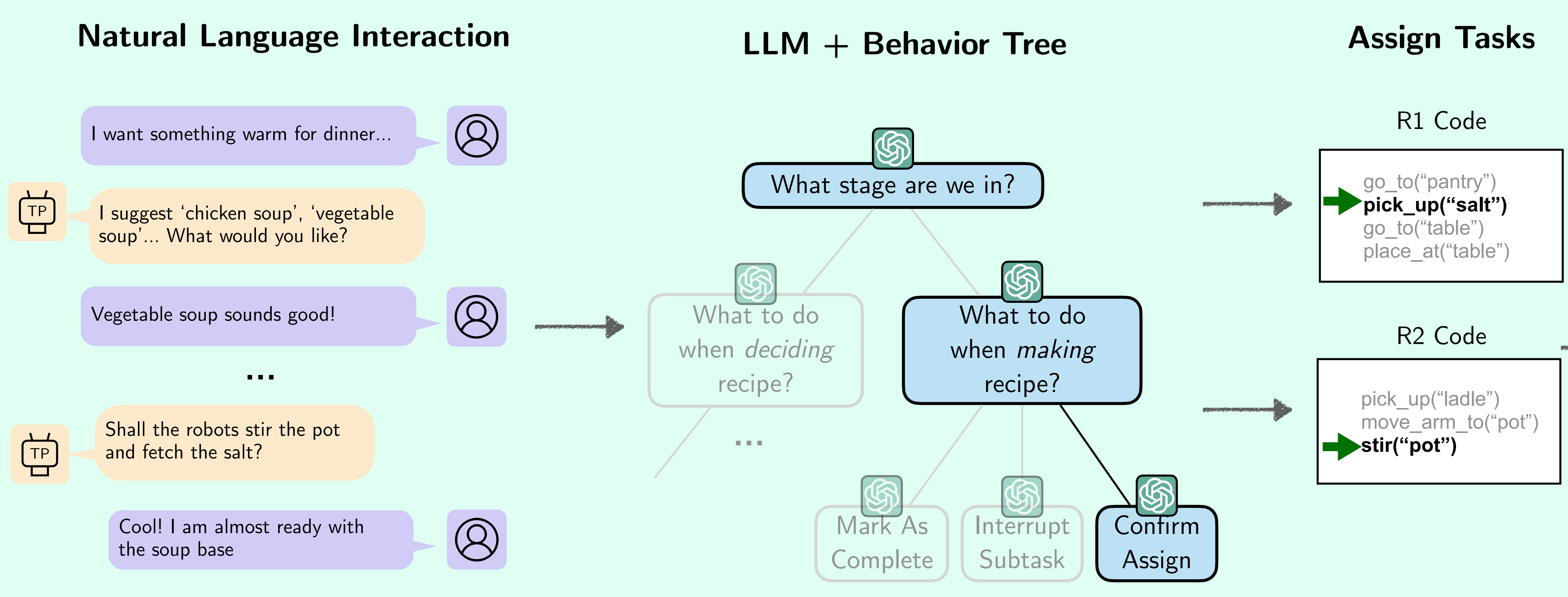3. Seamless and fluid collaboration with humans

Recipes: Tuna Sandwich, Caesar Salad

pick_up("relish") — handover("ketchup") — place_at("table")
pour("salt") — pick_up("cup") — stir("bowl")

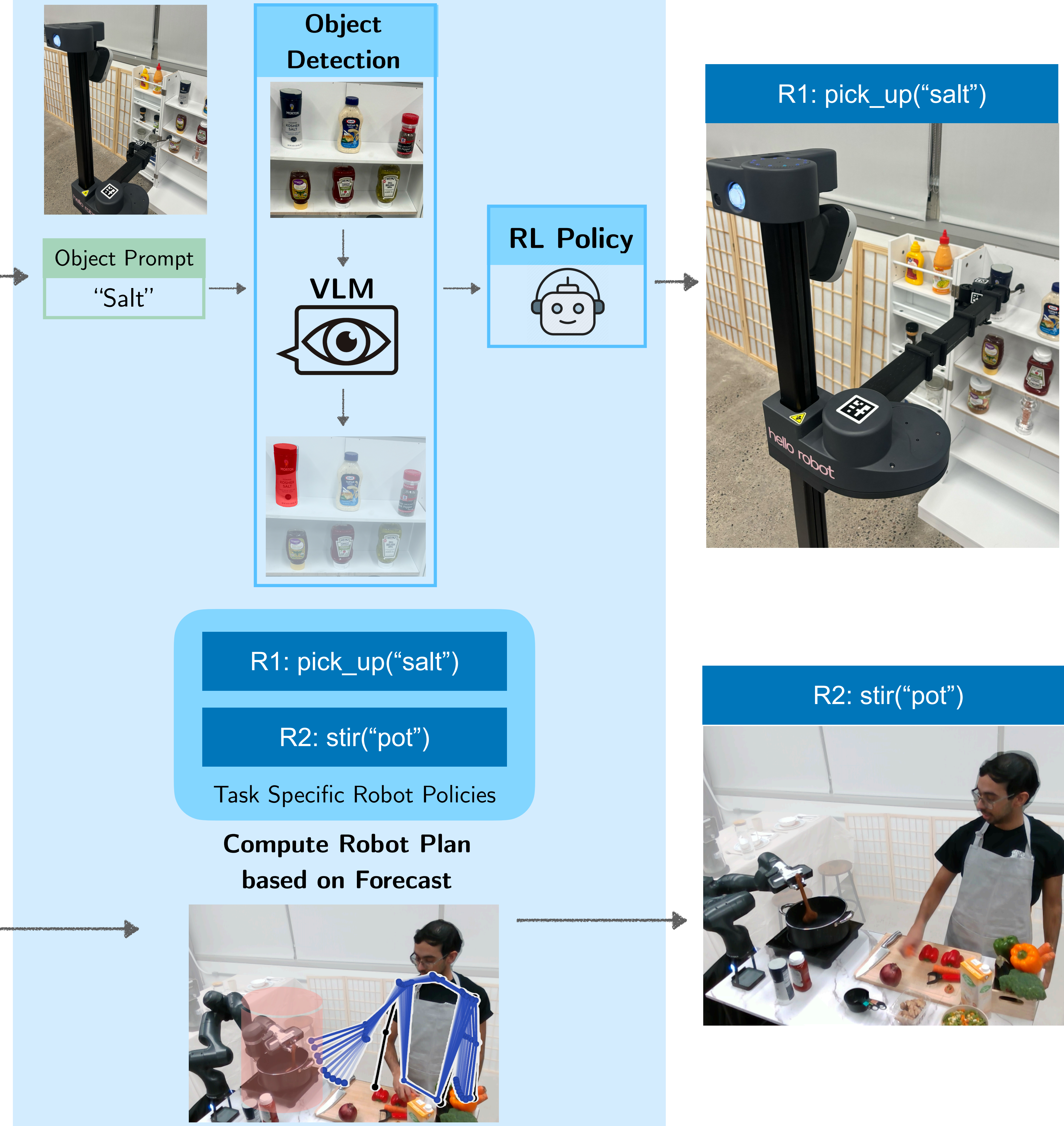## Key Design Characteristic: Employ Modularity at *Multiple* Different Levels

Large Pretrained Models for Language and Image Perception — Task Specific Modules to Compose their Outputs into Actions
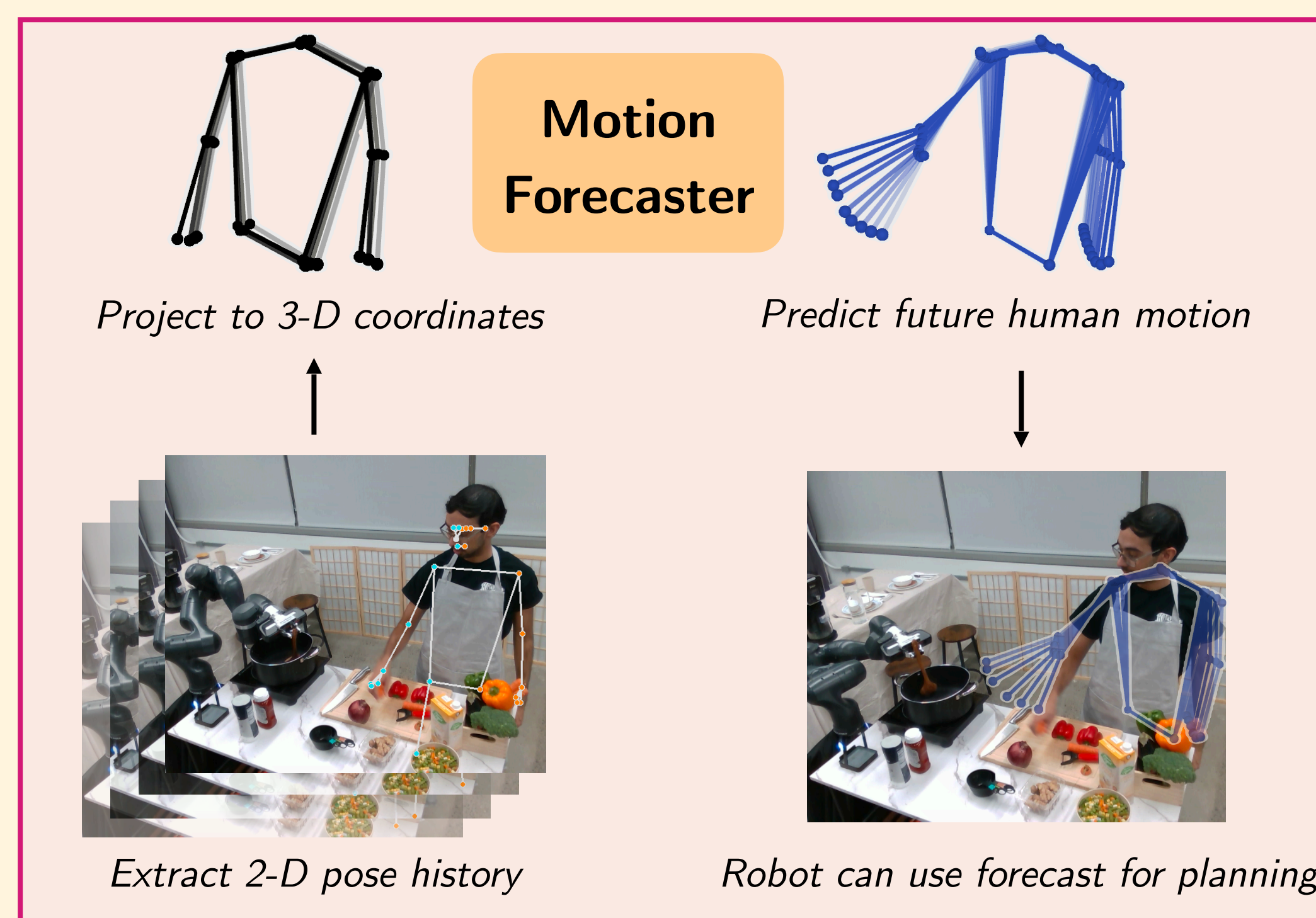
Interactive Task Planner

Natural Language Interaction

- I want something warm for dinner...
- I suggest 'chicken soup', 'vegetable soup'... What would you like?
- Vegetable soup sounds good!
- ...
- Shall the robots stir the pot and fetch the salt?
- Cool! I am almost ready with the soup base

LLM + Behavior Tree

- What stage are we in?
- What to do when *deciding* recipe?
- What to do when *making* recipe?
- Mark As Complete
- Interrupt Subtask
- Confirm Assign

Assign Tasks

R1 Code
```
go_to("pantry")
pick_up("salt")
go_to("table")
place_at("table")
```

R2 Code
```
pick_up("ladle")
move_arm_to("pot")
stir("pot")
```

Human Motion Forecasting

Close-Proximity Human-Robot Collaboration

Detect and Forecast Human Motion

Motion Forecaster

- Project to 3-D coordinates
- Predict future human motion
- Extract 2-D pose history
- Robot can use forecast for planning

Visuomotor Skills

Object Detection

Object Prompt "Salt" → VLM

RL Policy

R1: pick_up("salt")

R2: stir("pot")

R1: pick_up("salt")
R2: stir("pot")
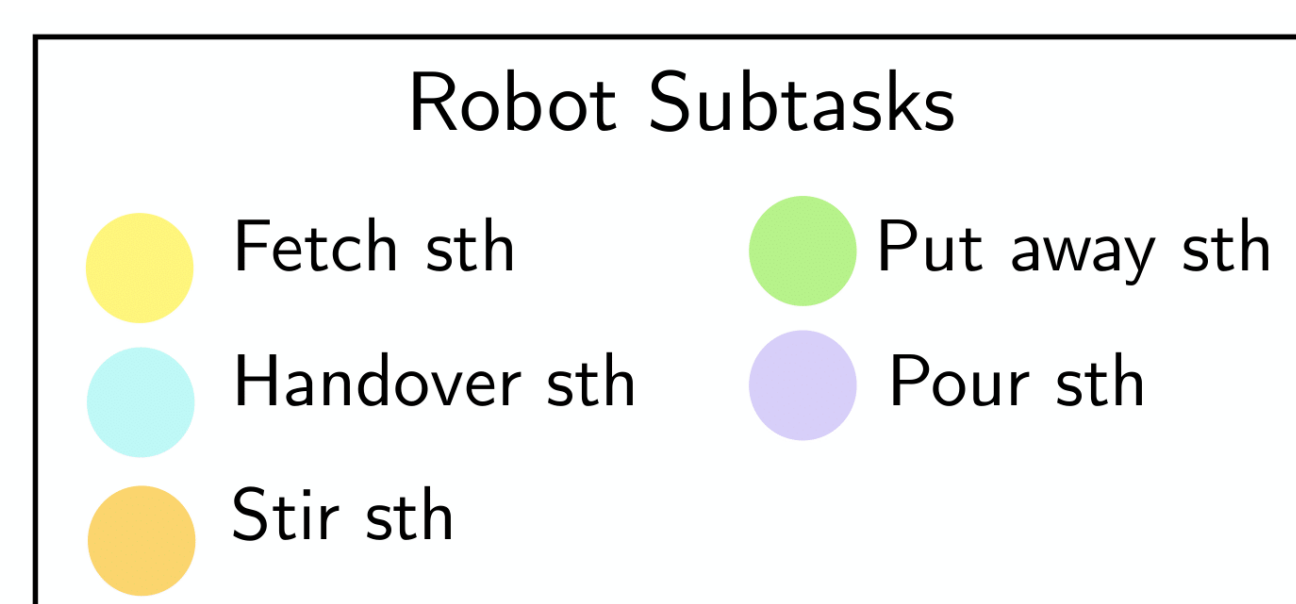**Task Specific Robot Policies**

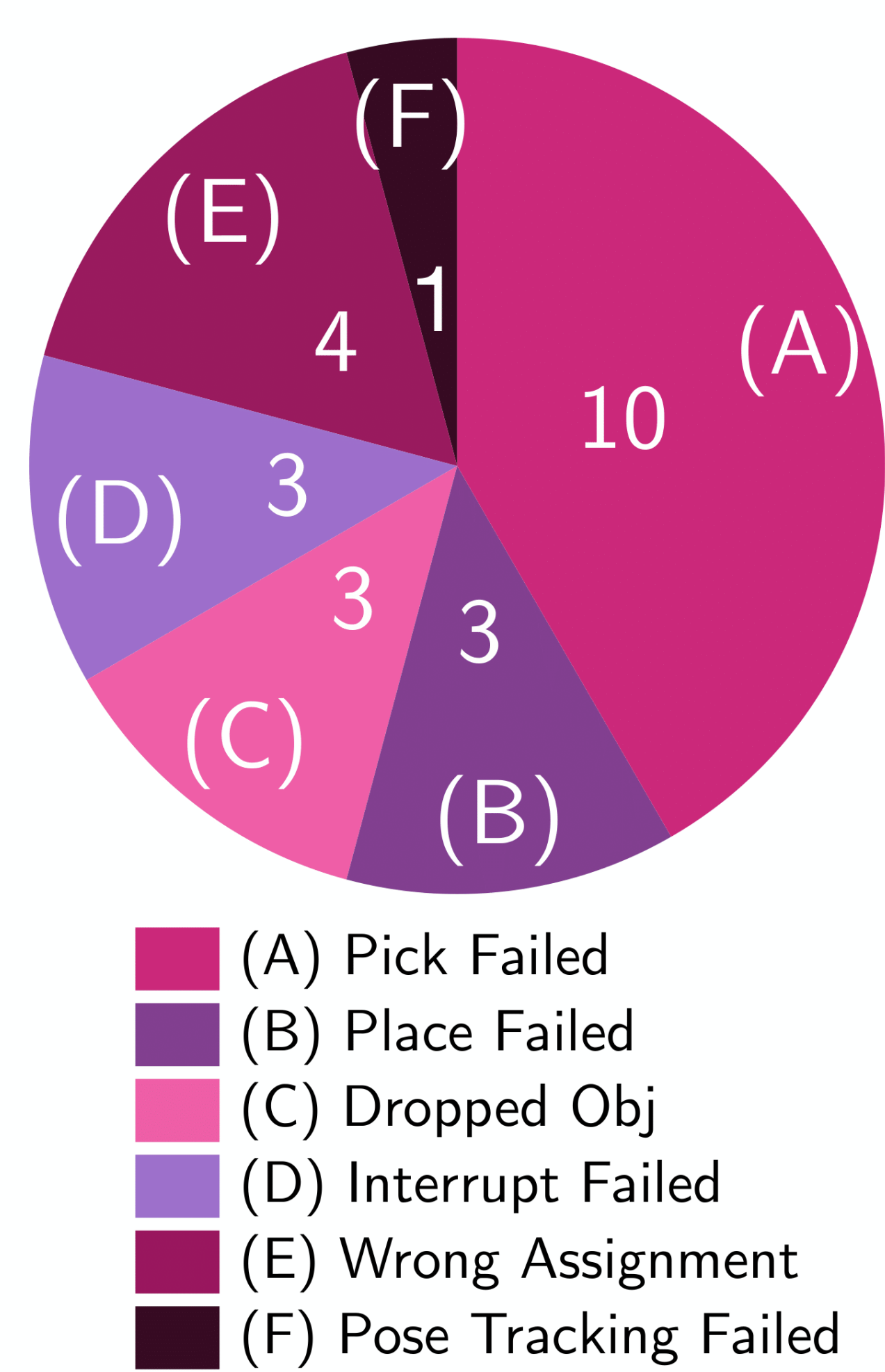**Compute Robot Plan based on Forecast**

End-to-End System Evaluation

60 end-to-end trials with a human user covering *6 different recipes*

Recipe		Typical Robot Skills Used	Success	Subtasks Comp.
Toss Salad	R1:	Go-to, Pick, Go-to, Place	8 / 10	92.5%
	R2:	Pick, Stir		
Tuna Sandwich	R1:	Go-to, Pick, Go-to, Place, Handover, Go-to, Place	8 / 10	96.0%
	R2:	Pick, Stir		
Vegetable Soup	R1:	Go-to, Pick, Go-to, Place	8 / 10	96.0%
	R2:	Pick, Handover, Pick, Stir		
Corn Soup	R1:	Go-to, Pick, Go-to, Place	6 / 10	90.0%
	R2:	Pick, Pour, Place, Pick, Stir		
Caesar Salad	R1:	Go-to, Pick, Go-to, Place	5 / 10	86.7%
	R2:	Pick, Pour, Place, Pick, Stir		
Chicken Soup	R1:	Go-to, Pick, Go-to, Place, Handover, Go-to, Place	6 / 10	91.4%
	R2:	Pick, Handover, Pick, Pour, Place, Pick, Stir		
			41 / 60	91.6%

Robot Subtasks
- Fetch sth
- Put away sth
- Handover sth
- Pour sth
- Stir sth

Failure Cases

- (A) Pick Failed — 10
- (B) Place Failed — 3
- (C) Dropped Obj — 3
- (D) Interrupt Failed — 3
- (E) Wrong Assignment — 4
- (F) Pose Tracking Failed — 1

MOSAIC enables Recoverability

- The robot will get ketchup for your meal
- Wait! Can you get the soup base instead?
- Sure! Robot will get soup base for you first.
- Thank you!

Ketchup — Soup Base
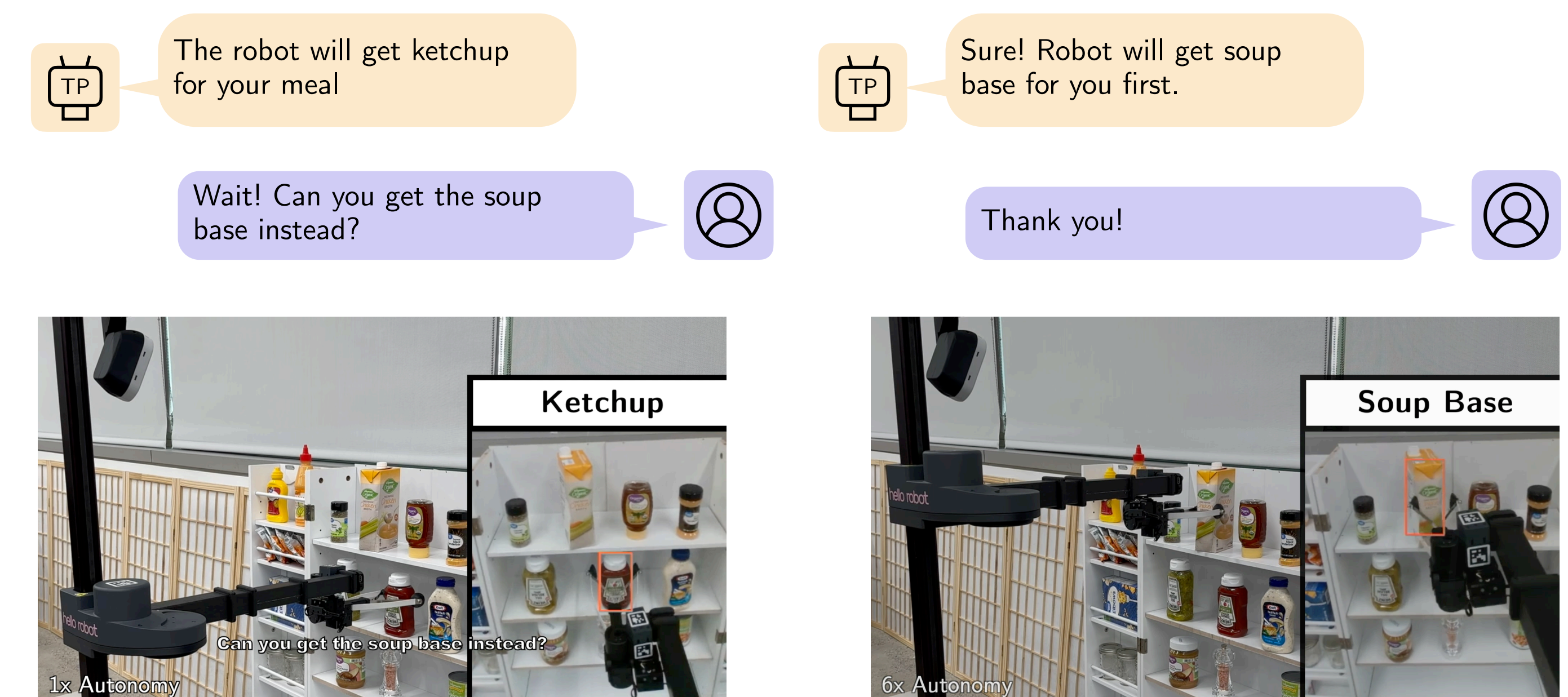
MOSAIC enables Safety

— Forecasted Pose

Safe Interaction, Quick Reaction

(Top-left scene annotations)
- I am preparing vegetables for the soup. Can you pour some salt after stirring?
- R1 Subtasks: Stir, Pour salt
- R1 Current Code: pick_up("ladle"), move_arm_to("pot"), stir("pot")
- R2 Subtasks: Fetch Salt
- R2 Current Code: go_to("pantry"), pick_up("salt"), go_to("table"), place("table")