Real-Time Anomaly Detection and Reactive Planning with Large Language Models

Rohan Sinha[1], Amine Elhafsi[1], Christopher Agia[2], Matthew Foutter[3], Edward Schmerling[4] and Marco Pavone[1,4]

Abstract—**Foundation models, e.g., large language models (LLMs), trained on internet-scale data possess zero-shot generalization capabilities that make them a promising technology towards detecting and mitigating out-of-distribution failure modes of robotic systems. Fully realizing this promise, however, poses two challenges: (i) mitigating the considerable computational expense of these models such that they may be applied online, and (ii) incorporating their judgement regarding potential anomalies into a safe control framework. In this work, we present a two-stage reasoning framework: First is a fast binary anomaly classifier that analyzes observations in an LLM embedding space, which may trigger a slower fallback selection stage that utilizes the reasoning capabilities of generative LLMs. These stages correspond to branch points in a model predictive control strategy that maintains the joint feasibility of continuing along various fallback plans to account for the slow reasoner's latency as soon as an anomaly is detected, thus ensuring safety. We show that our fast anomaly classifier outperforms autoregressive reasoning with state-of-the-art GPT models, even when instantiated with relatively small language models. This enables our runtime monitor to improve the trustworthiness of dynamic robotic systems, such as quadrotors or autonomous vehicles, under resource and time constraints. Videos illustrating our approach in both simulation and real-world experiments are available on our project page: https://sites.google.com/view/aesop-llm.**

I. Introduction

Autonomous robotic systems are rapidly advancing in capabilities, seemingly on the cusp of widespread deployment

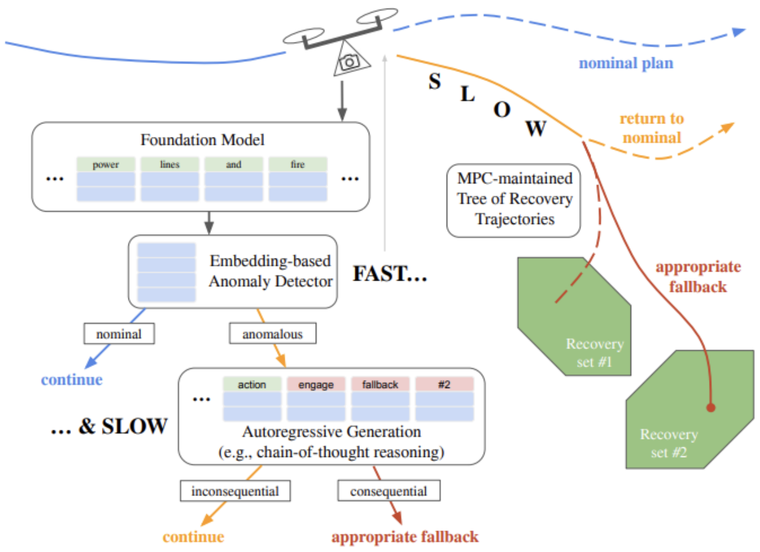
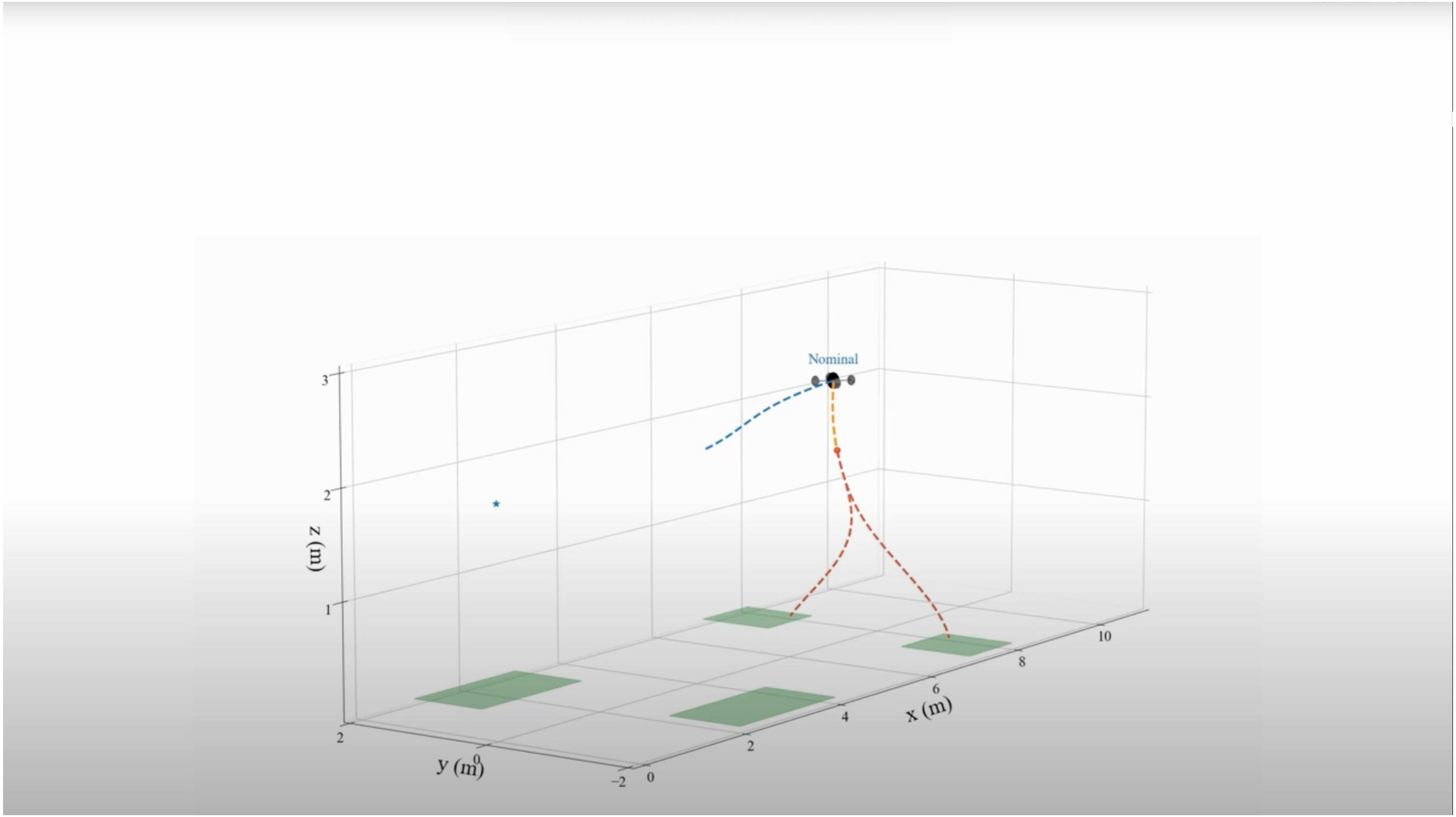


Fig. 1: We present an embedding-based runtime monitoring scheme using fast and slow language model reasoners in concert. During nominal operation, *the fast reasoner* differentiates between nominal and anomalous robot observations. If an anomaly is flagged, the system enters a fallback-safe state while *the slow reasoner* determines the anomaly's hazard. In this fallback-safe state, we guarantee access to a set of safe recovery plans (if the anomaly is consequential) and access to continued

https://www.youtube.com/watch?v=R-Fco_cevfc&t=68s



https://www.youtube.com/watch?v=GGzIgjxJuo8&t=19s

From Foresight to Forethought: VLM-In-the-Loop Policy Steering via Latent Alignment

Yilin Wu[1], Ran Tian[2], Gokul Swamy[1], Andrea Bajcsy[1]
[1]Carnegie Mellon University [2]UC Berkeley
{yilinwu, gswamy, abajcsy}@andrew.cmu.edu, rantian@berkeley.edu

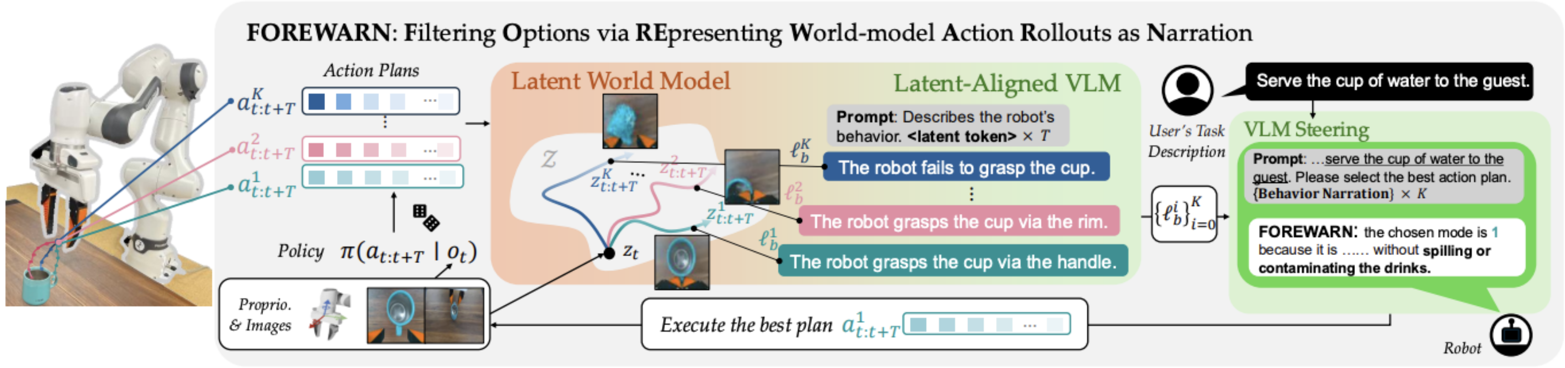
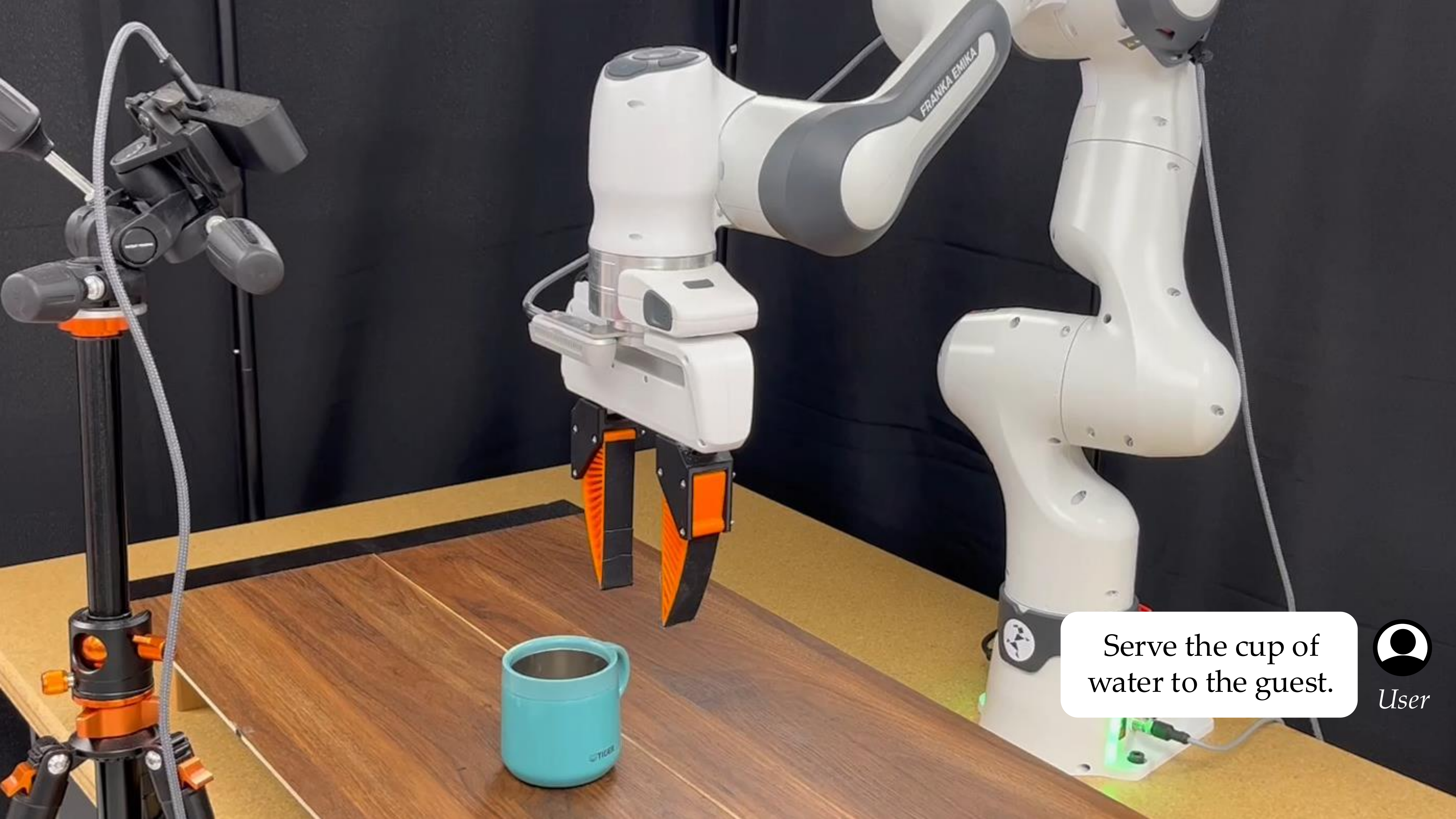


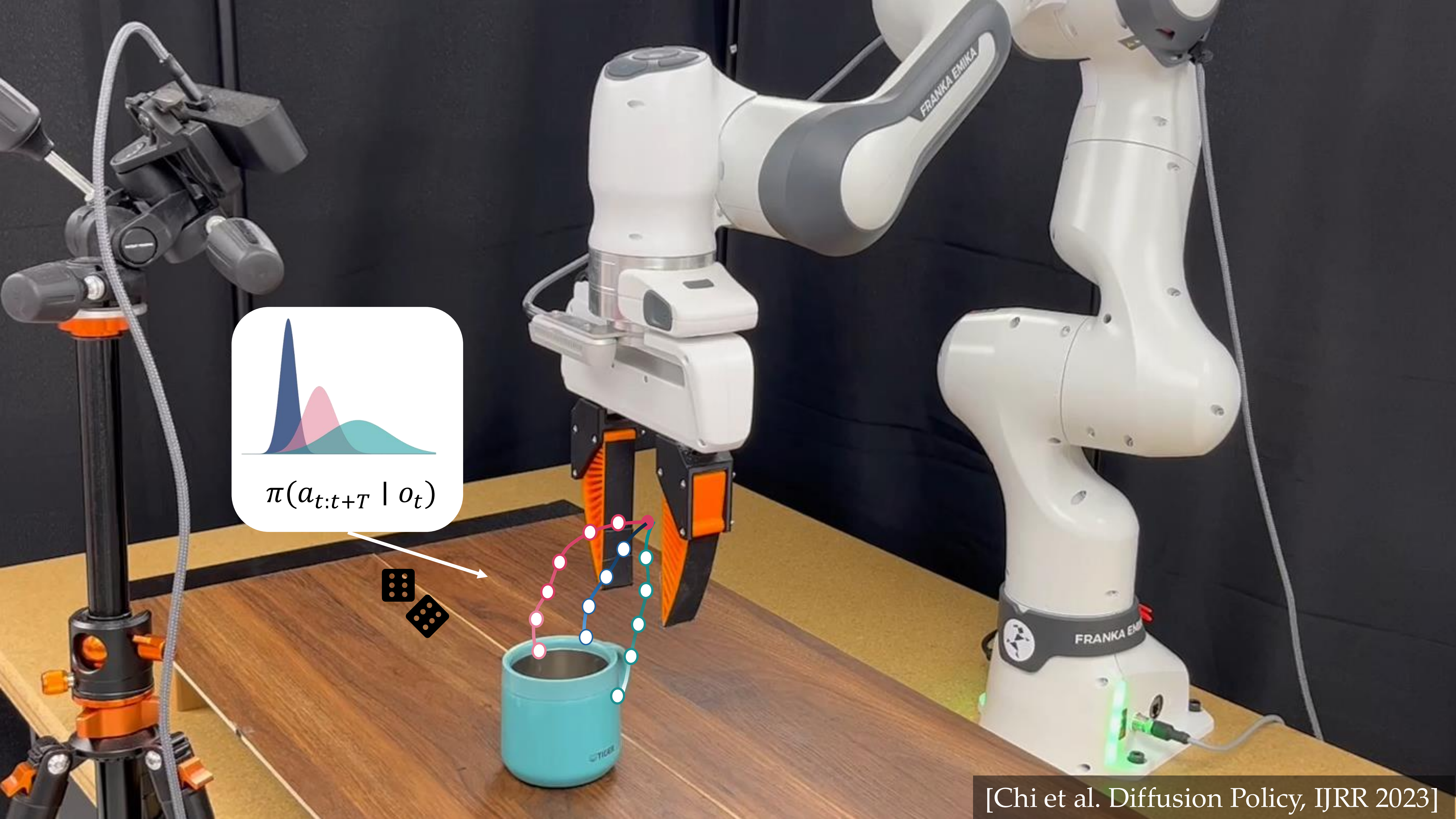
Fig. 1: We present **FOREWARN**, an VLM-in-the-loop policy steering algorithm for multi-modal generative robot policies. Our key idea is to decouple the VLM's burden of predicting action outcomes from evaluation. By predicting action outcomes with a pre-trained latent dynamics model and aligning a VLM to reason about these latent states in text, FOREWARN can select action plans at runtime that are most appropriate for new task contexts and user needs.

***Abstract*—While generative robot policies have demonstrated significant potential in learning complex, multimodal behaviors from demonstrations, they still exhibit diverse failures at deployment-time. Policy steering offers an elegant solution to reducing the chance of failure by using an external verifier to select from low-level actions proposed by an imperfect generative policy. Here, one might hope to use a Vision Language Model (VLM) as a verifier, leveraging its open-world reasoning capabilities. However, off-the-shelf VLMs struggle to understand the consequences of low-level robot actions as they are represented fundamentally differently than the text and images the VLM was trained on. In response, we propose FOREWARN, a novel**

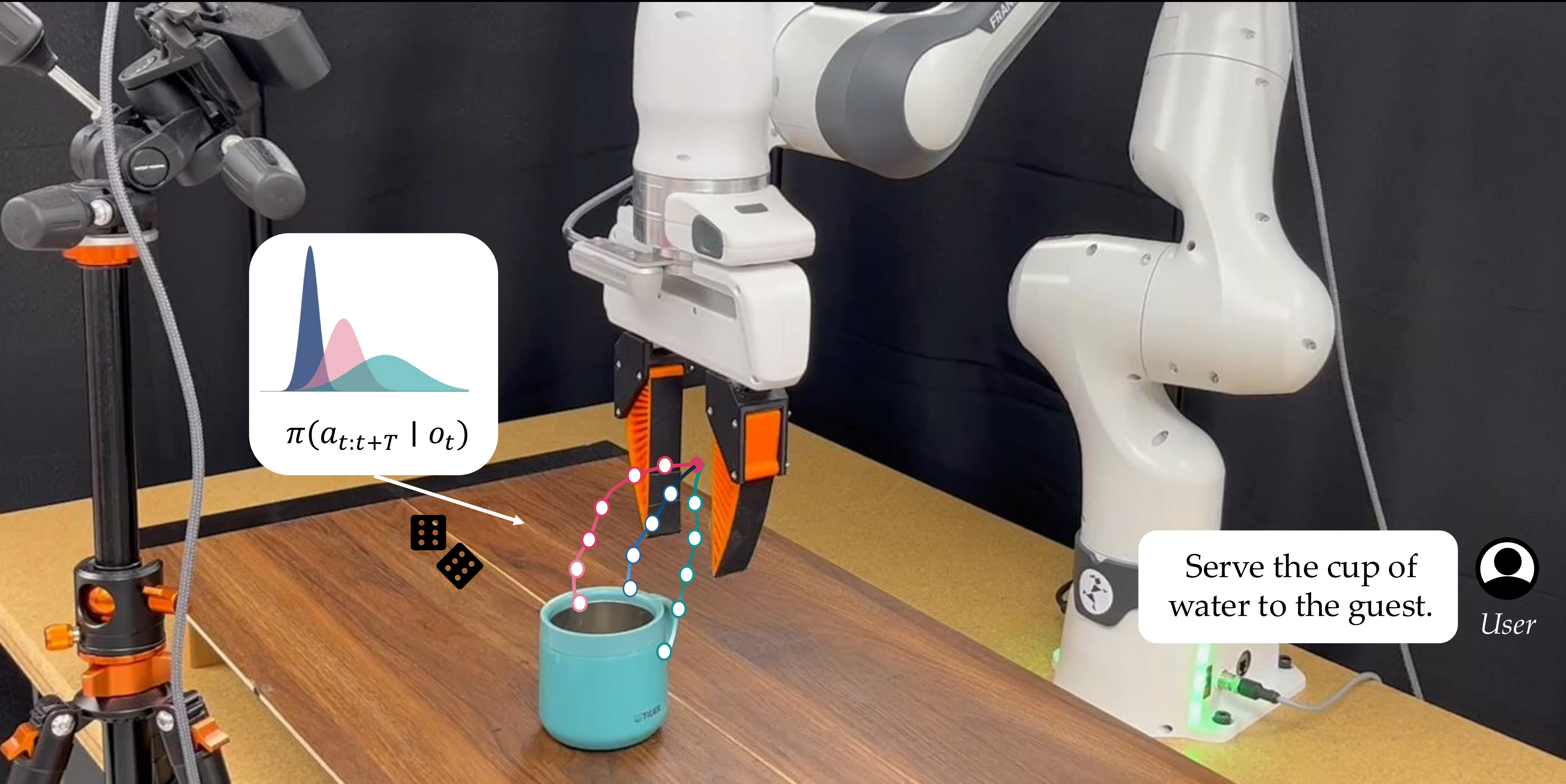
the robot in the left of Figure 1 that must pick up a mug from the table. At training time, the generative policy learns a distribution over useful interaction modes such as grasping the cup by different parts (e.g., handle, lip and interior, etc.) shown in wrist camera photo in Figure 1.

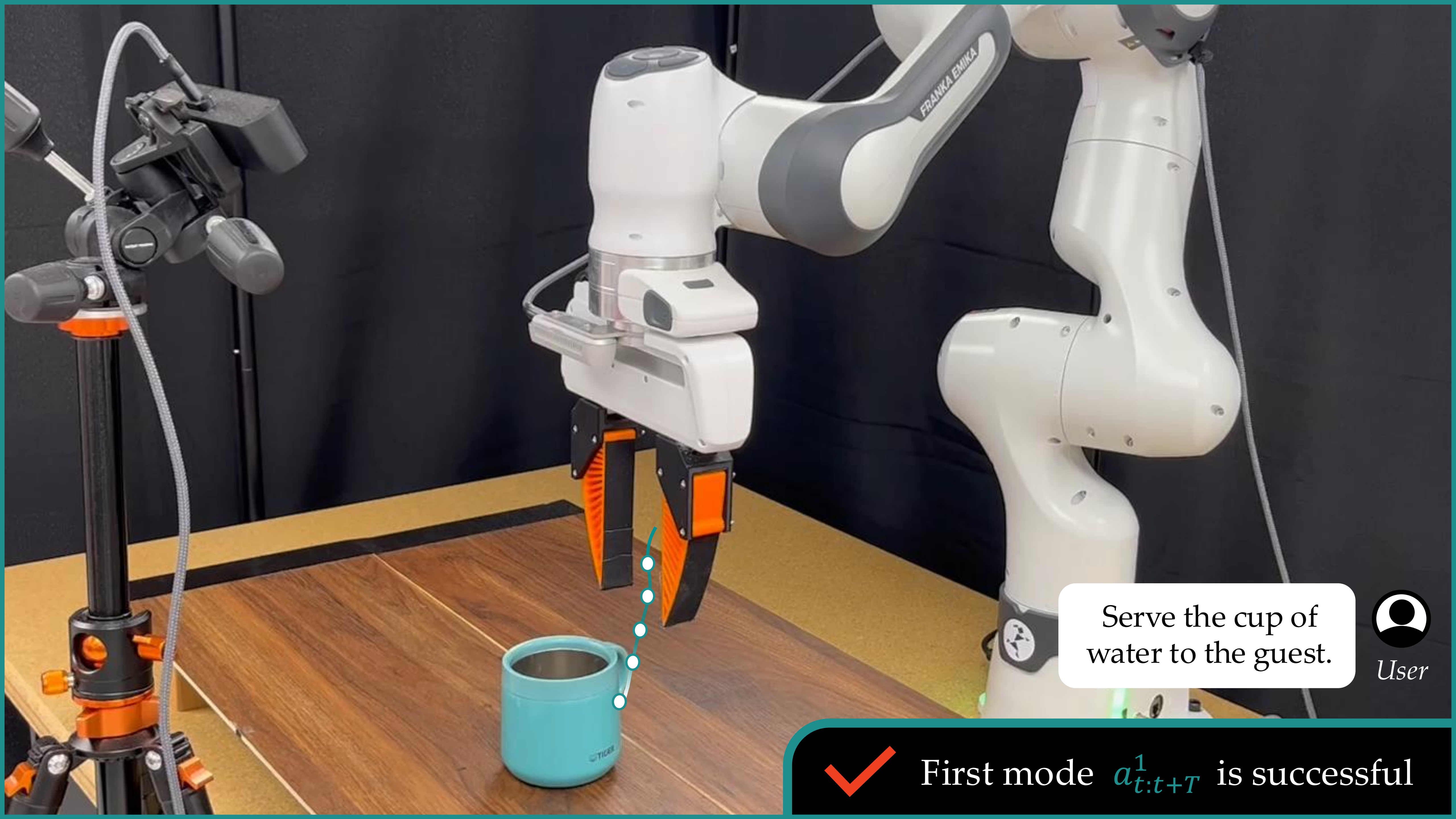
However, at runtime, the policy exhibits a range of degradations, from complete task failures (such as the robot knocking down the cup during grasping, shown in the center of Figure 1), to inappropriate behaviors that are misaligned with the deployment context or preferences of an end-user (such as


Serve the cup of water to the guest.
User
FRANKA EMIKA

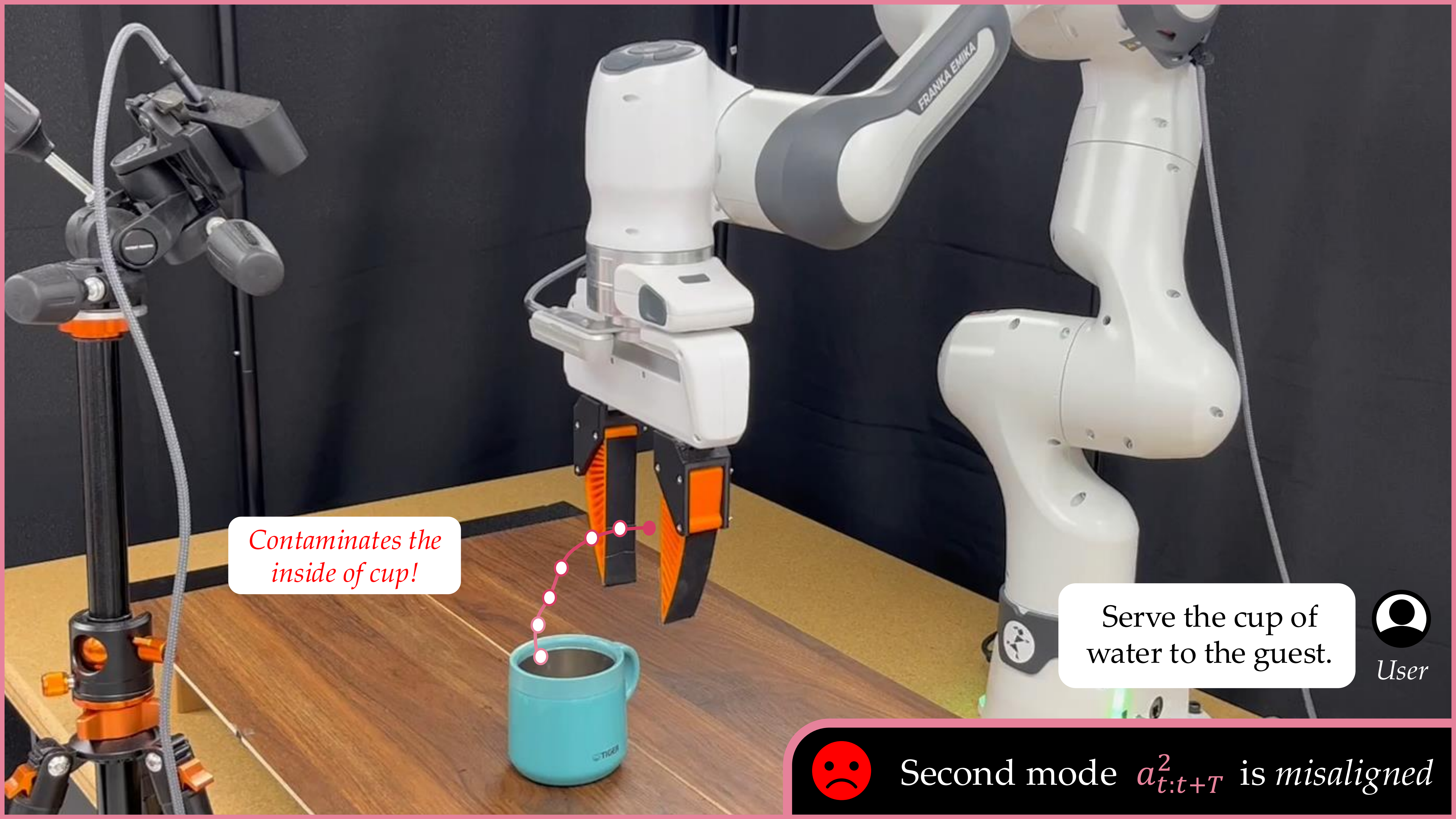

$\pi(a_{t:t+T} \mid o_t)$
FRANKA EMIKA
[Chi et al. Diffusion Policy, IJRR 2023]

But not all sampled actions result in the same task performance!

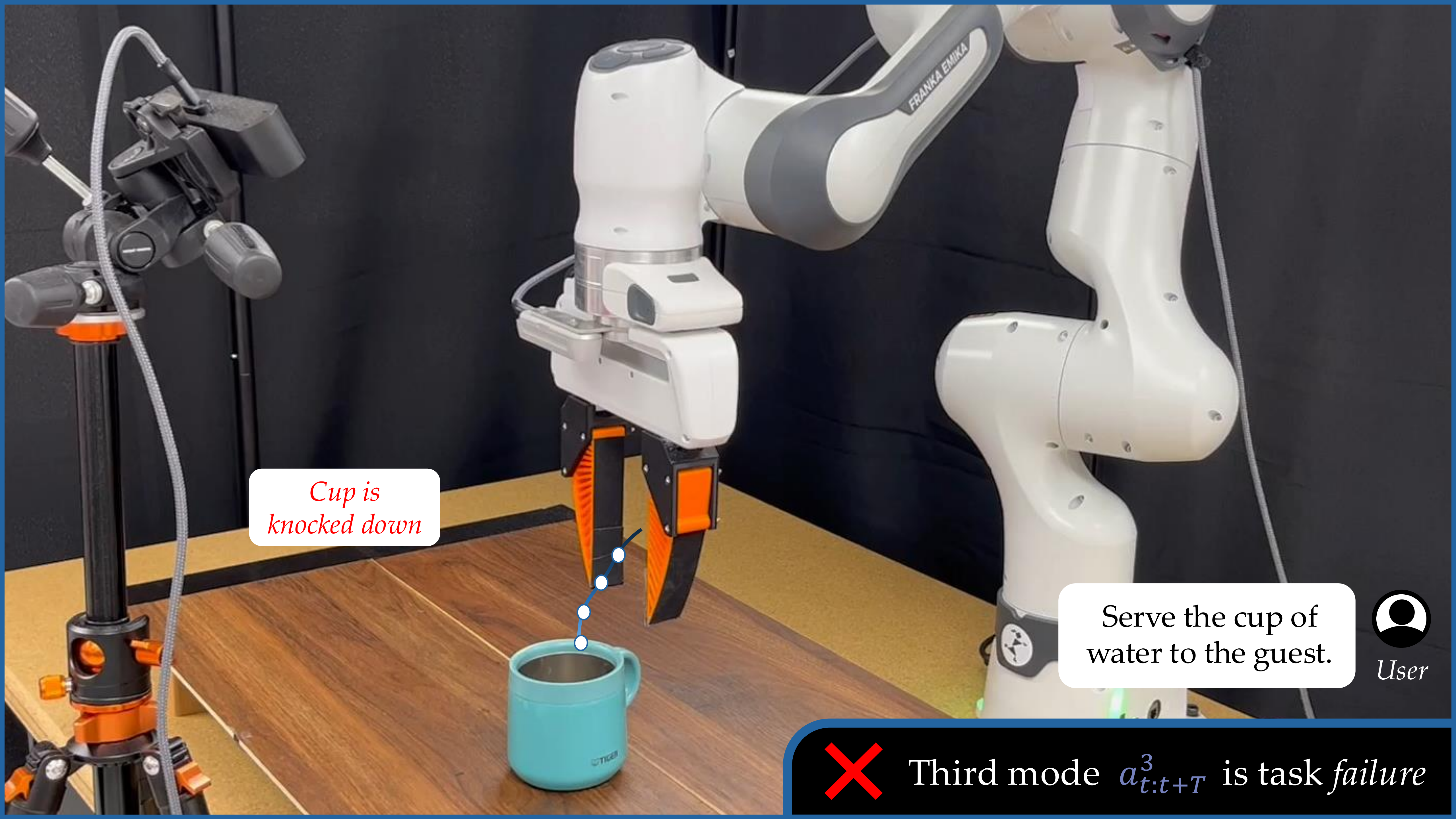




Serve the cup of water to the guest.
User
First mode $a^1_{t:t+T}$ is successful

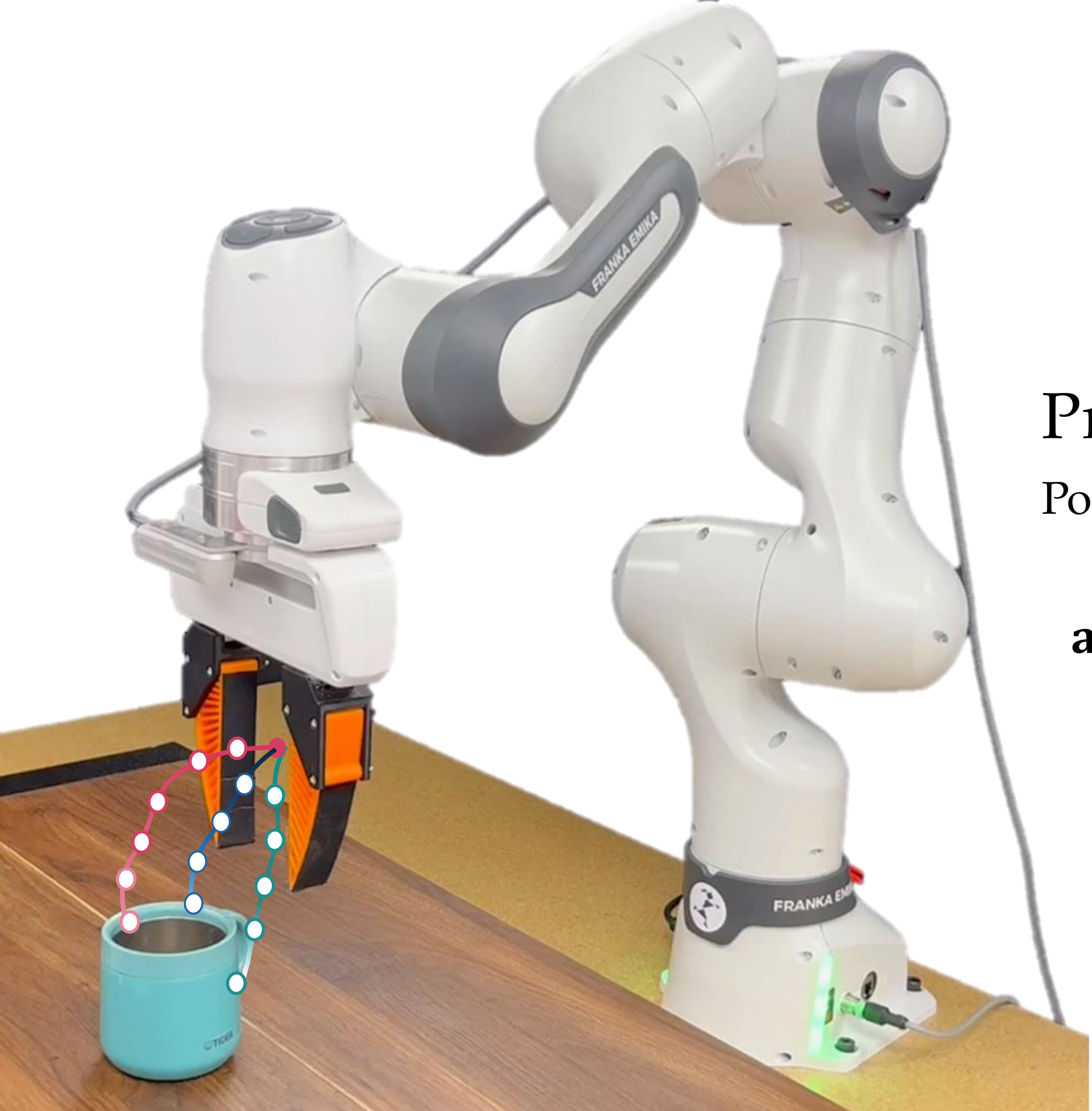


Contaminates the inside of cup!
Serve the cup of water to the guest.
User
Second mode $a^2_{t:t+T}$ is *misaligned*


Cup is knocked down
Serve the cup of water to the guest.
User
Third mode $a^3_{t:t+T}$ is task *failure*

The base policy may already contain the “right” behavior mode within its distribution….

…but we *need to verify* that the robot’s sampled action plan will lead to “good” outcomes



Problem Formulation:

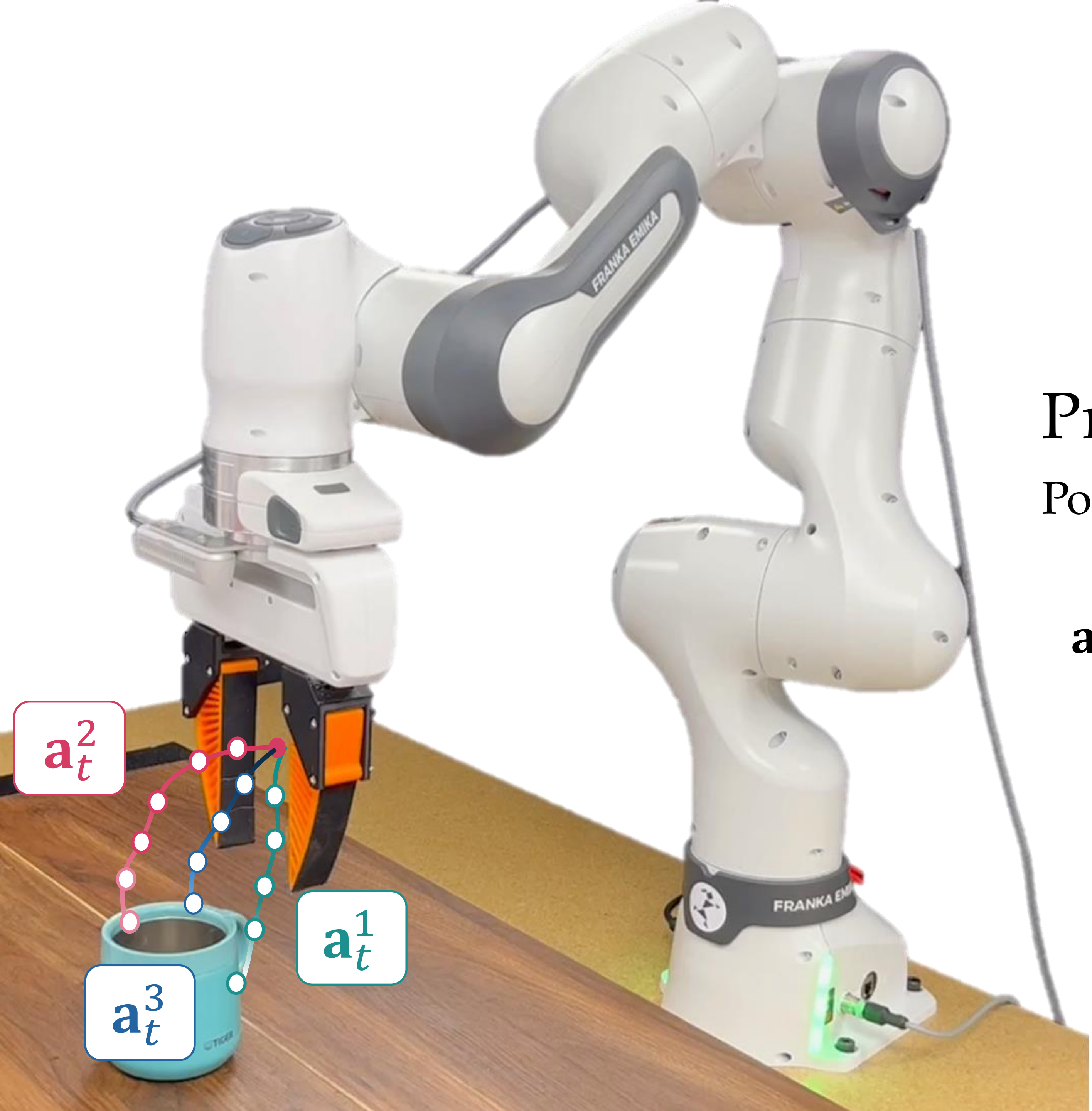
Policy Steering as Model Predictive Control

$$\mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t \in \{\mathbf{a}_t^i\}_{i=1}^{K}} \mathbb{E}_{\boldsymbol{o}_t \sim P(\cdot|o_t,\mathbf{a}_t)}\left[R(\boldsymbol{o}_t;\ell)\right]$$

$\boxed{\mathbf{a}_t^1}$

$\boxed{\mathbf{a}_t^2}$

$\boxed{\mathbf{a}_t^3}$



Problem Formulation:

Policy Steering as Model Predictive Control

$$\mathbf{a}_t^\star = \arg \max_{\mathbf{a}_t \in \{\mathbf{a}_t^i\}_{i=1}^K} \mathbb{E}_{\boldsymbol{o}_t \sim P(\cdot | o_t, \mathbf{a}_t)} [R(\boldsymbol{o}_t; \ell)]$$

outcome prediction *verification*

How do we solve this tractably?

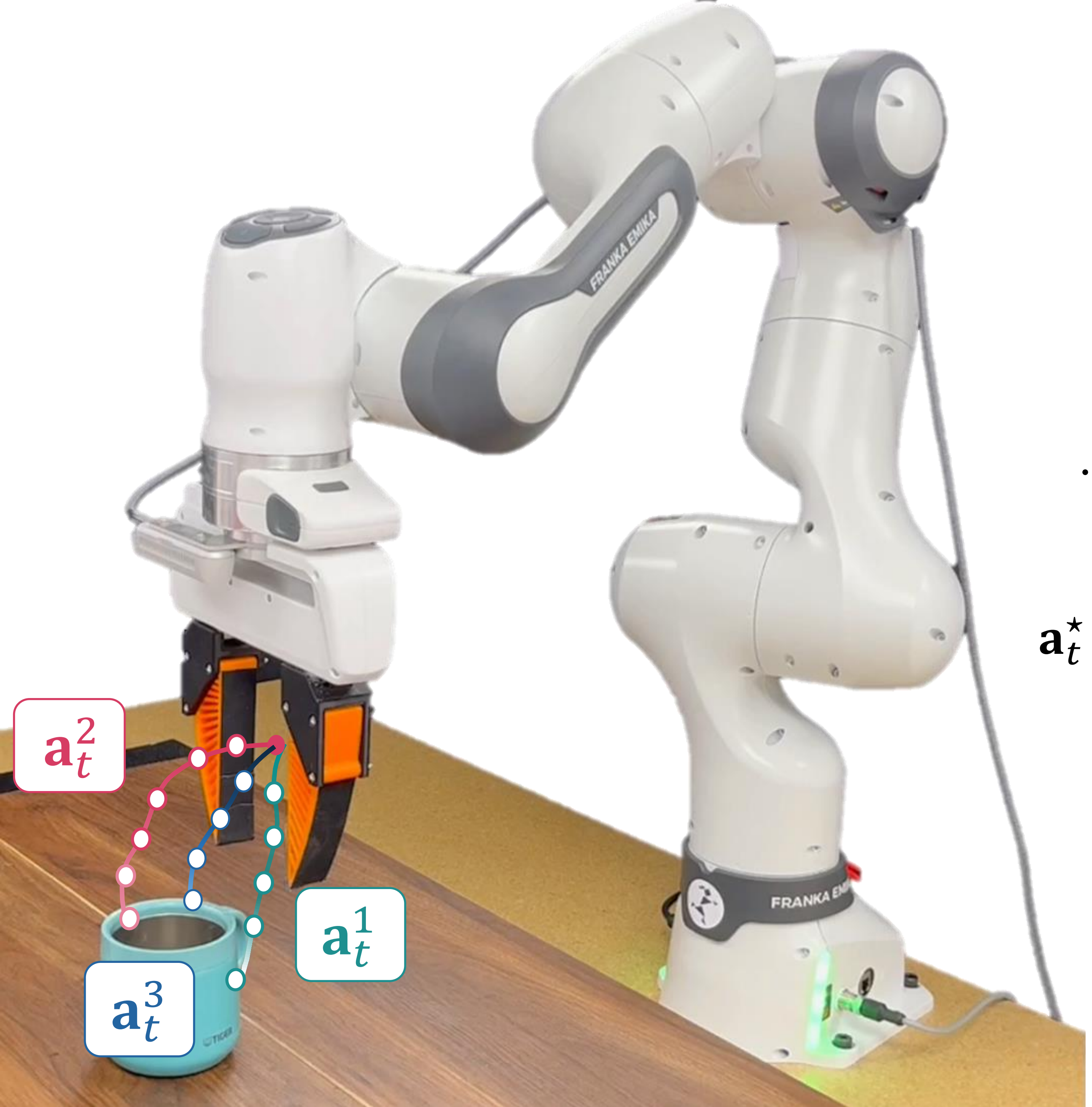
ℓ = Serve the cup of water to the guest.

Key Idea:

Reason about outcomes in a world model's latent state representation…

$$\mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t \in \{\mathbf{a}_t^i\}_{i=1}^K} \mathbb{E}_{\underbrace{\mathbf{z}_t \sim f_\phi(z_t, \mathbf{a}_t)}_{\textit{world model for outcome prediction}}} [R(\mathbf{z}_t; \ell)]$$

ℓ = Serve the cup of water to the guest.



Key Idea:

Reason about outcomes in a world model's latent state representation…

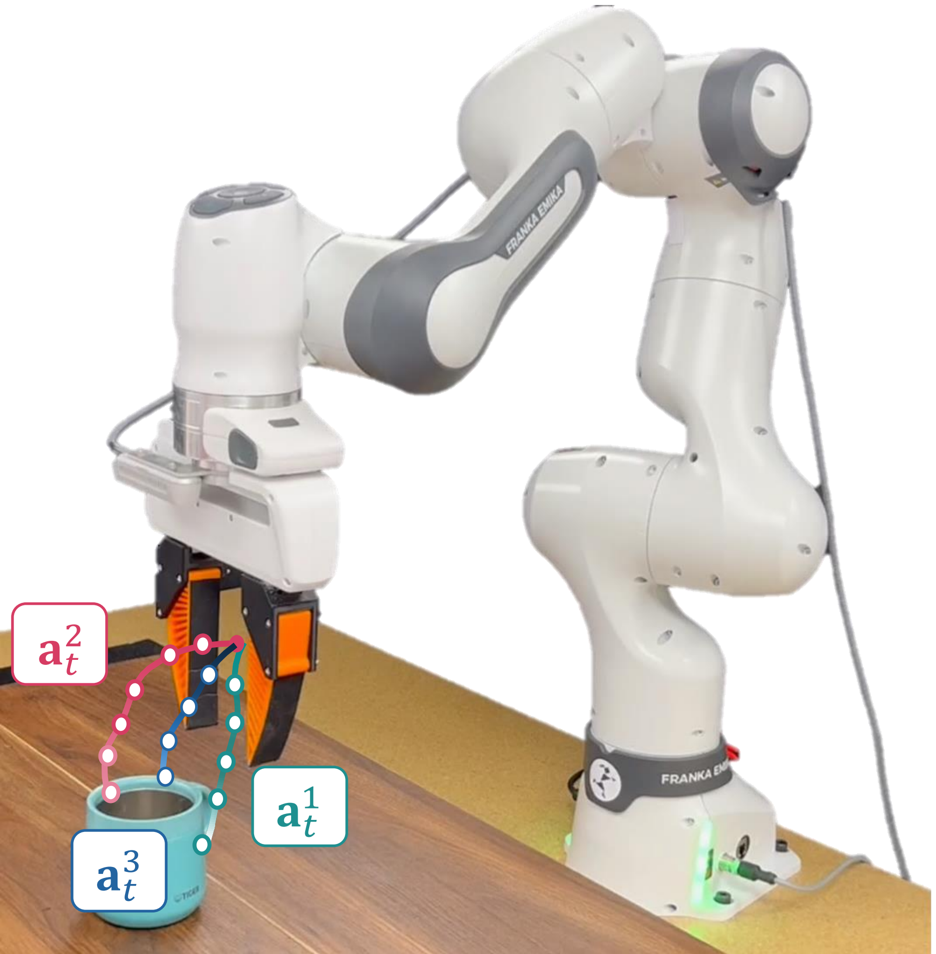
…and align a VLM to directly reason on the latent states for evaluation

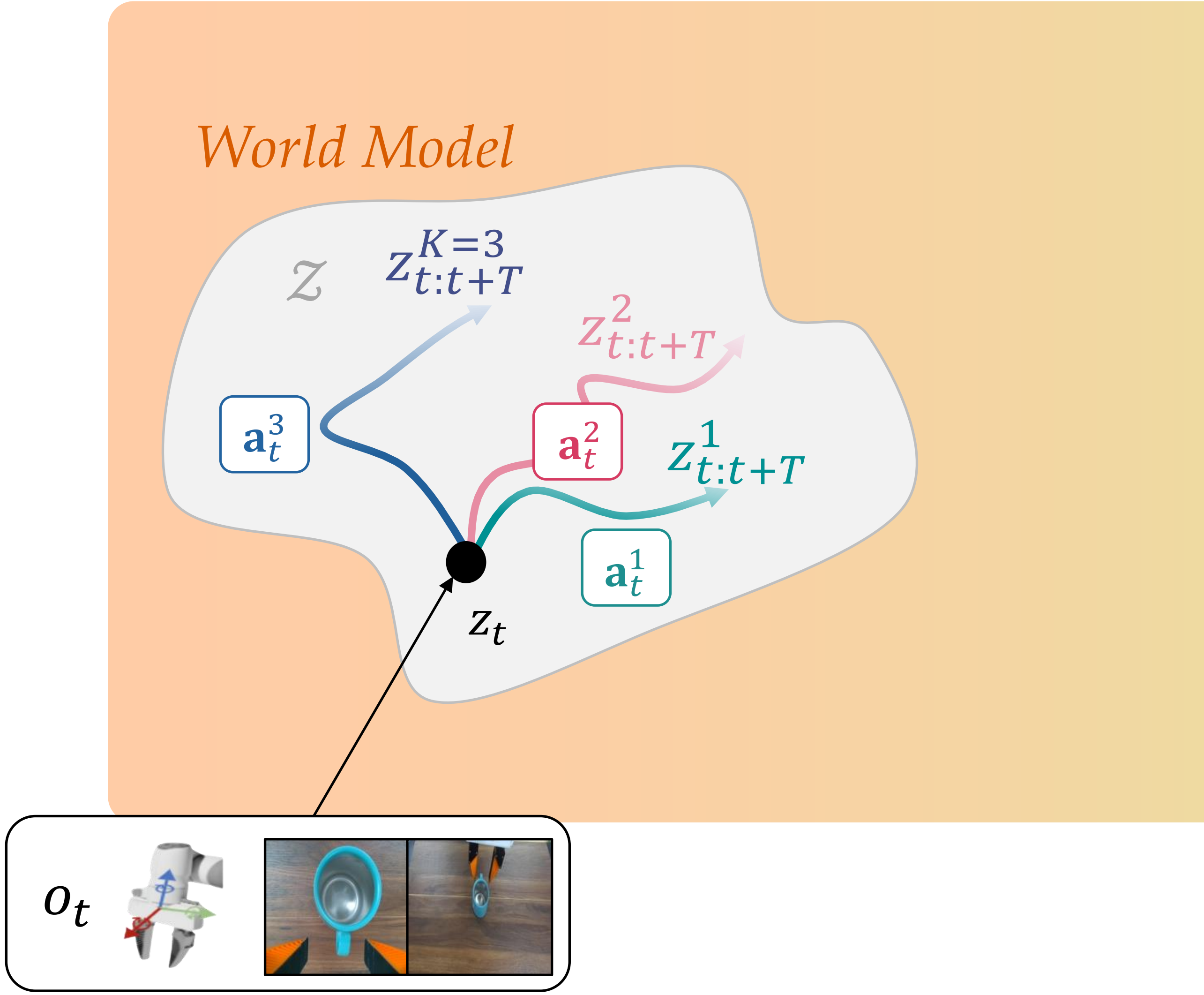
$$\mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t\in\{\mathbf{a}_t^i\}_{i=1}^K} \underbrace{\mathbb{E}_{\mathbf{z}_t\sim f_\phi(z_t,\mathbf{a}_t)}}_{\text{world model for outcome prediction}} \underbrace{\left[R_\psi^{\mathrm{VLM}}(\mathbf{z}_t;\ell)\right]}_{\text{latent-aligned VLM for verification}}$$

$\ell =$ Serve the cup of water to the guest.

VLM-in-the-Loop Policy Steering

$$\mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t\in\{\mathbf{a}_t^i\}_{i=1}^K} \mathbb{E}_{\boldsymbol{z}_t\sim f_\phi(z_t,\mathbf{a}_t)}\left[R_\psi^{\mathrm{VLM}}(\boldsymbol{z}_t;\ell)\right]$$

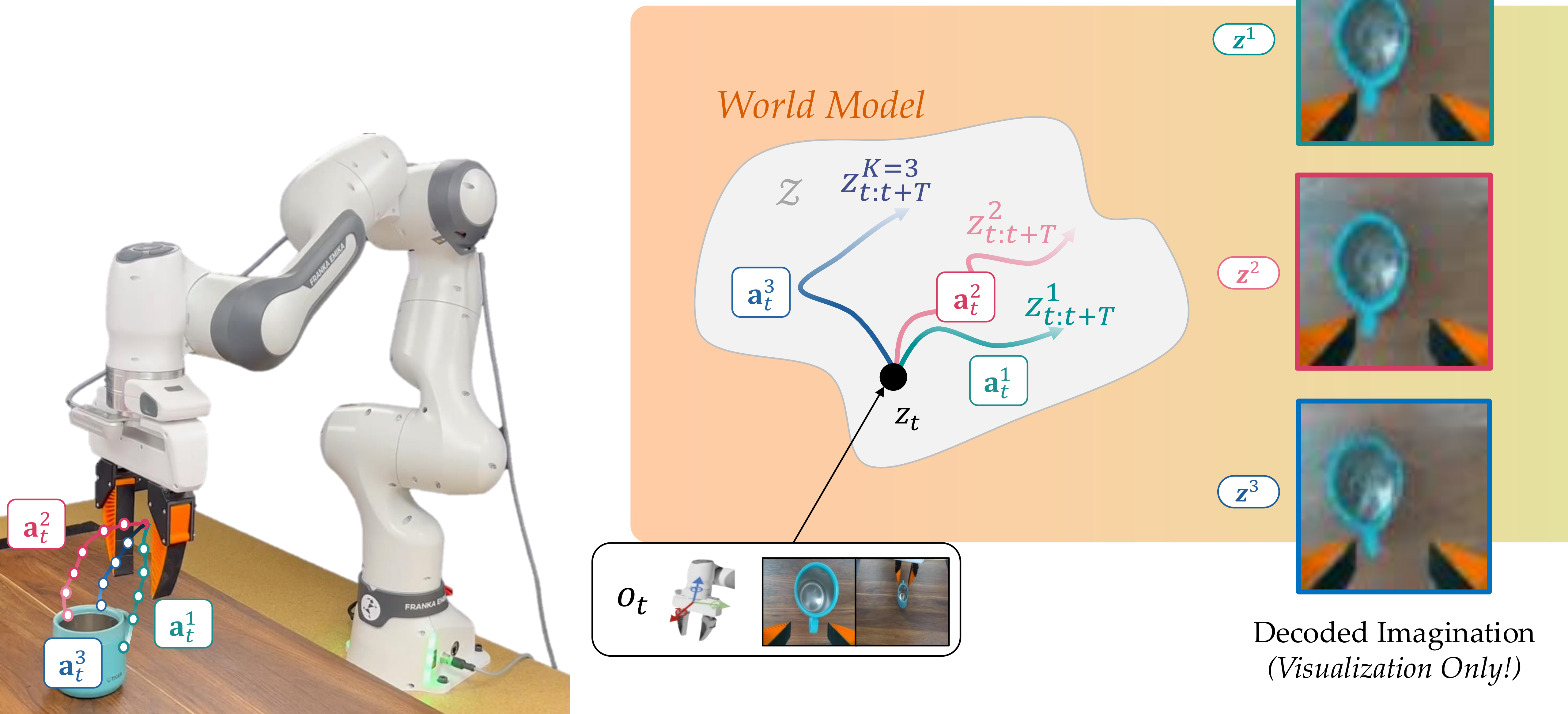




[Y. Wu, R. Tian, G. Swamy, A. Bajcsy. "VLM-in-the-loop Policy Steering via Latent Alignment." RSS, 2025.]

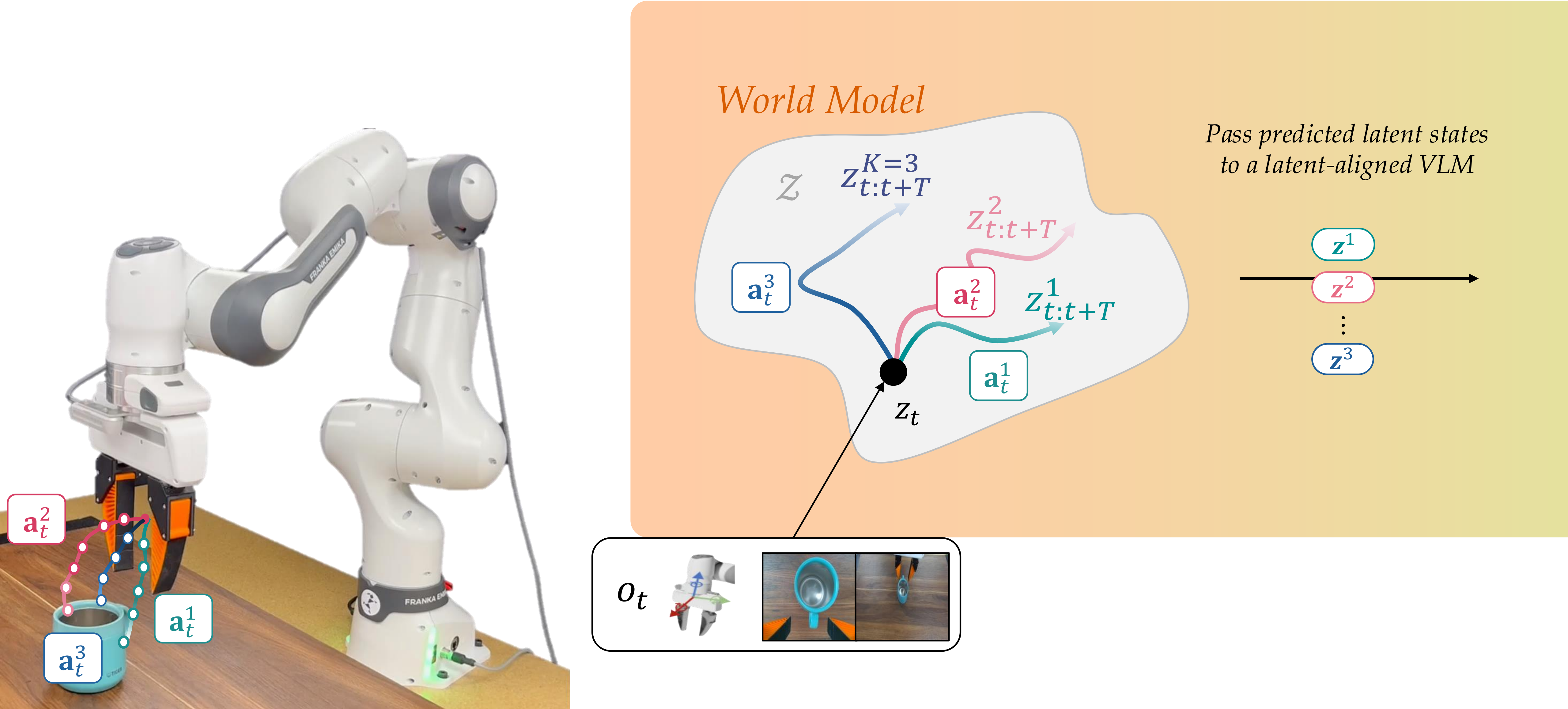
VLM-in-the-Loop Policy Steering

$$\mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t \in \{\mathbf{a}_t^i\}_{i=1}^K} \mathbb{E}_{\boldsymbol{z}_t \sim f_\phi(z_t, \mathbf{a}_t)} \left[R_\psi^{\mathrm{VLM}}(\boldsymbol{z}_t; \ell)\right]$$



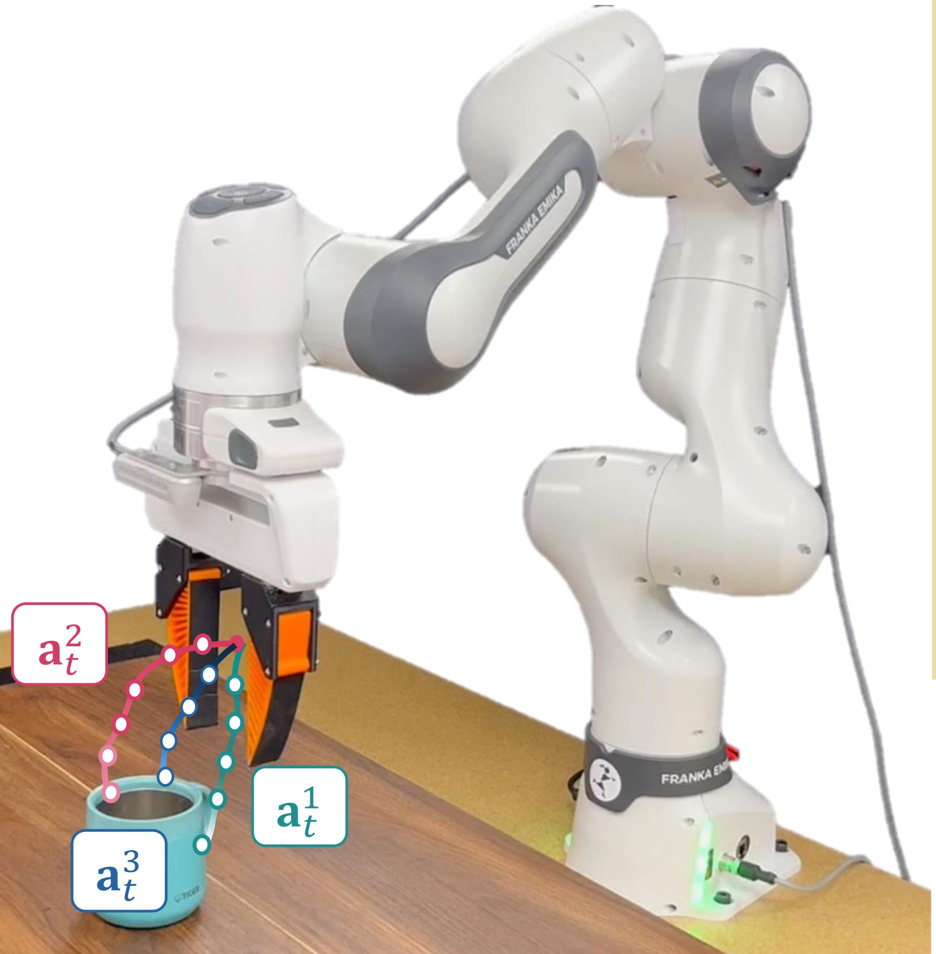
VLM-in-the-Loop Policy Steering

$$\mathbf{a}_t^\star = \arg \max_{\mathbf{a}_t \in \{\mathbf{a}_t^i\}_{i=1}^K} \mathbb{E}_{\mathbf{z}_t \sim f_\phi(z_t, \mathbf{a}_t)} \left[R_\psi^{\text{VLM}}(\mathbf{z}_t; \ell) \right]$$

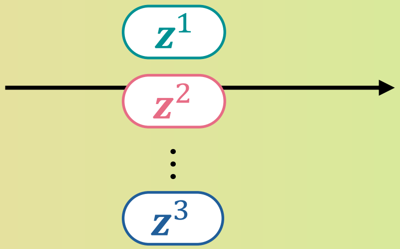


VLM-in-the-Loop Policy Steering

$$\mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t \in \{\mathbf{a}_t^i\}_{i=1}^K} \mathbb{E}_{\mathbf{z}_t \sim f_\phi(z_t, \mathbf{a}_t)} \left[R_\psi^{\mathrm{VLM}}(\boldsymbol{z}_t; \ell) \right]$$

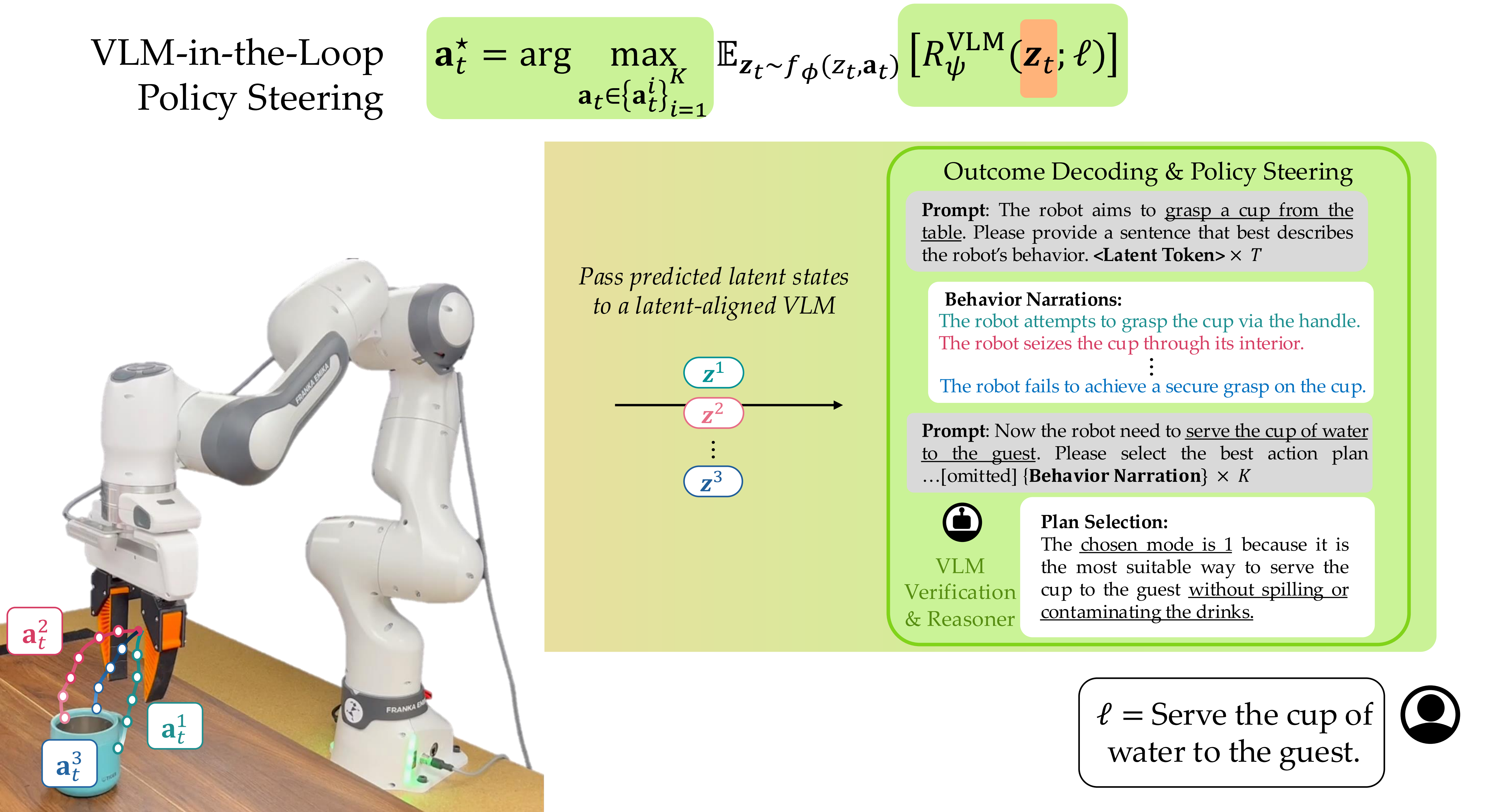


Pass predicted latent states to a latent-aligned VLM

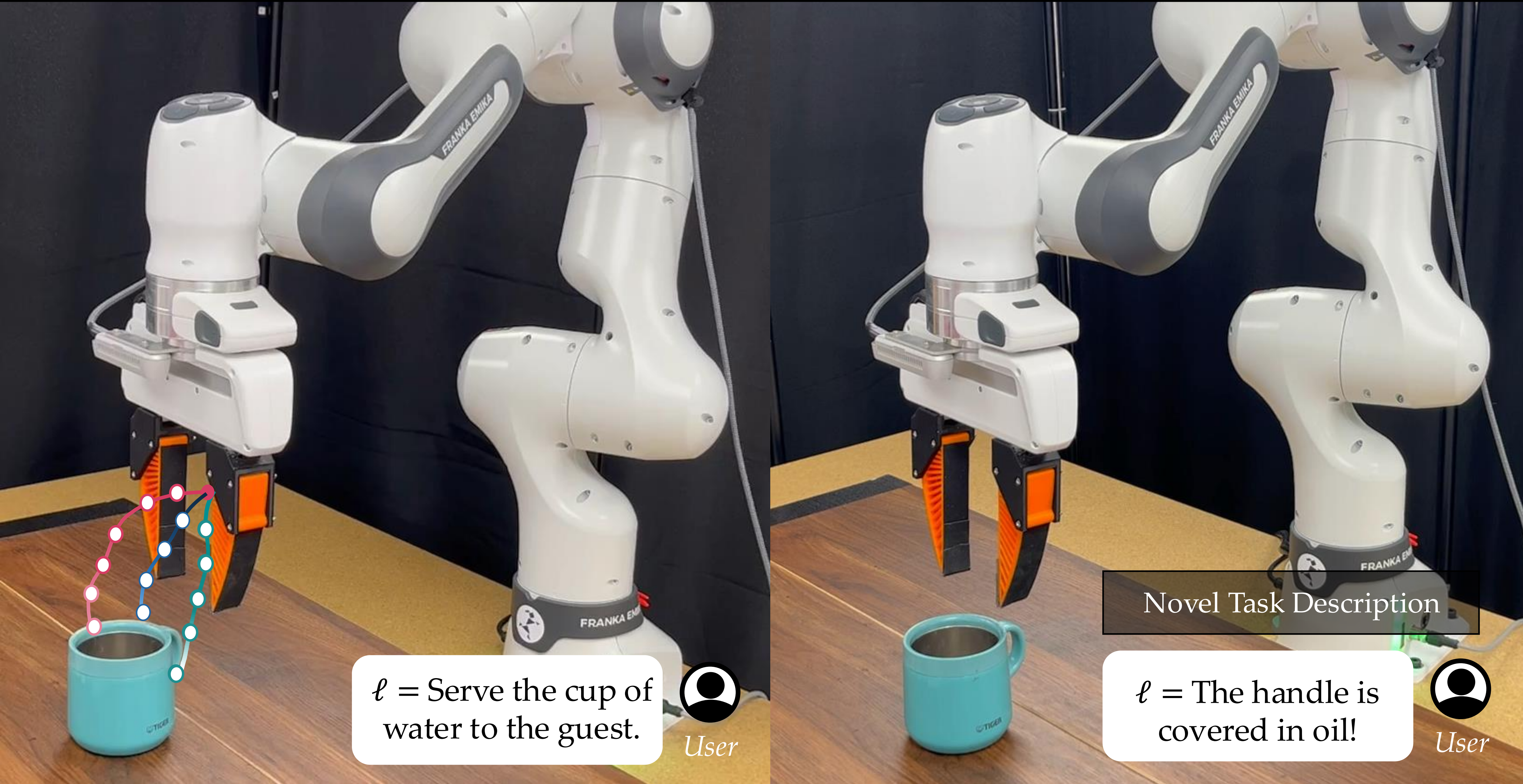


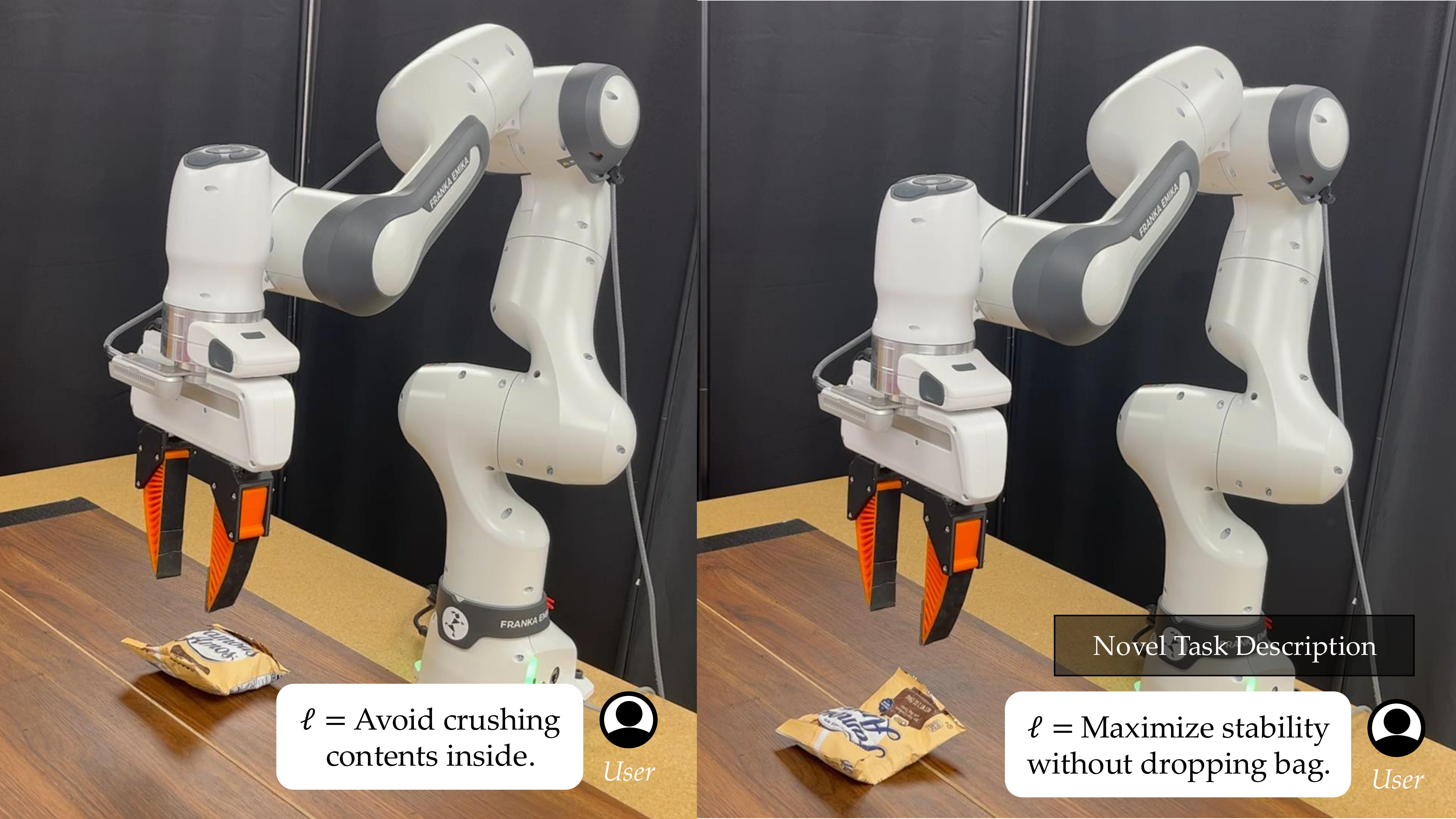
*We pose a latent-text alignment problem, **fine-tuning** this model via a **visual Q&A task***

Intuition: *VLM <u>describes in text</u> what is going on in the latent state*

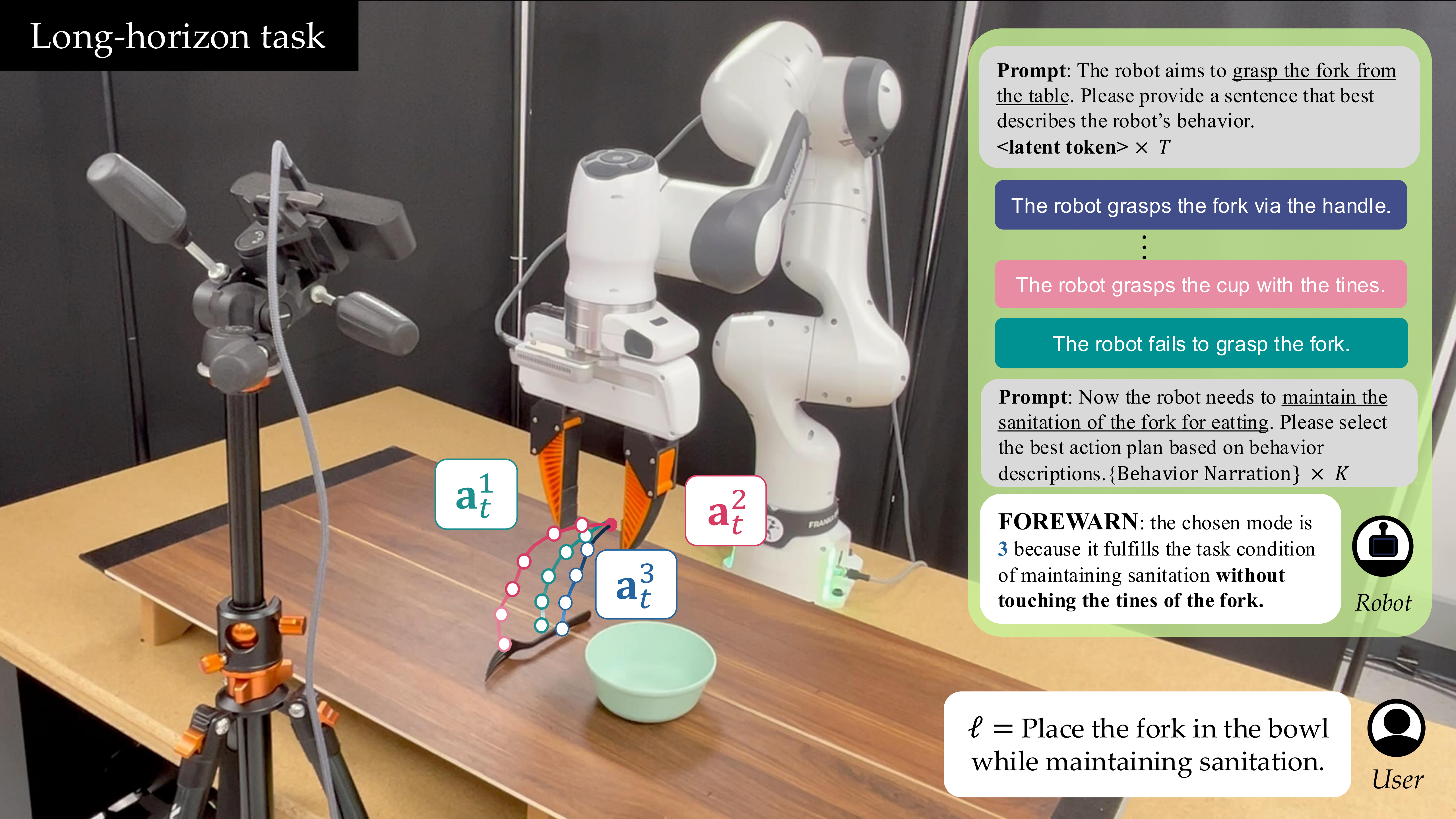

VLM-in-the-Loop Policy Steering
$\mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t \in \{\mathbf{a}_t^i\}_{i=1}^K} \mathbb{E}_{z_t \sim f_\phi(z_t, \mathbf{a}_t)} \left[R_\psi^{\mathrm{VLM}}(z_t; \ell) \right]$
Pass predicted latent states to a latent-aligned VLM
z^1
z^2
z^3
Outcome Decoding & Policy Steering
Prompt: The robot aims to grasp a cup from the table. Please provide a sentence that best describes the robot's behavior. <Latent Token> × T
Behavior Narrations:
The robot attempts to grasp the cup via the handle.
The robot seizes the cup through its interior.
The robot fails to achieve a secure grasp on the cup.
Prompt: Now the robot need to serve the cup of water to the guest. Please select the best action plan …[omitted] {Behavior Narration} × K
VLM Verification & Reasoner
Plan Selection:
The chosen mode is 1 because it is the most suitable way to serve the cup to the guest without spilling or contaminating the drinks.
$\mathbf{a}_t^2$
$\mathbf{a}_t^1$
$\mathbf{a}_t^3$
FRANKA EMIKA
ℓ = Serve the cup of water to the guest.

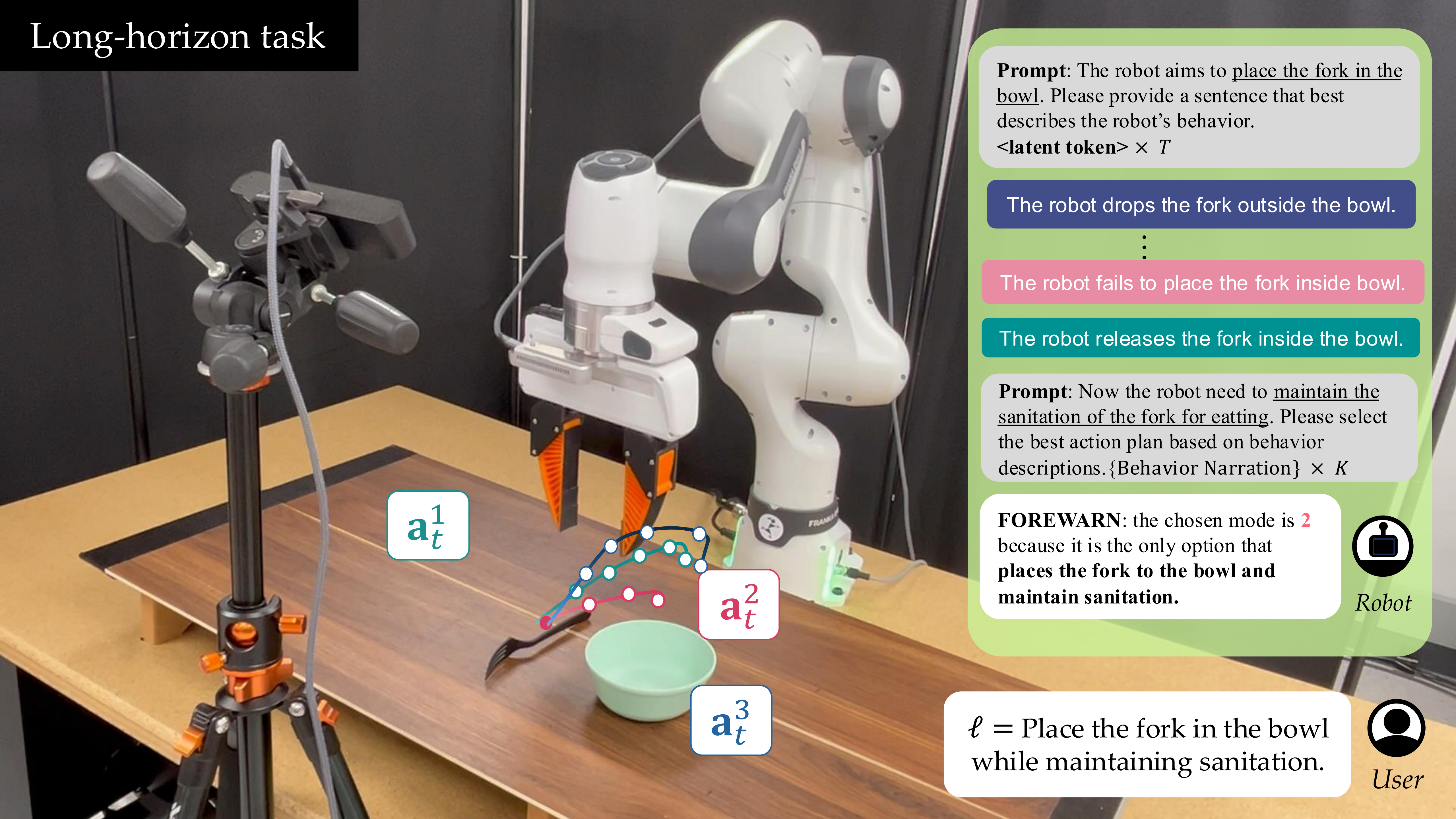
With our policy steering, correct mode is selected even with new task description





ℓ = Avoid crushing contents inside.
User
Novel Task Description
ℓ = Maximize stability without dropping bag.
User


Long-horizon task
Prompt: The robot aims to grasp the fork from the table. Please provide a sentence that best describes the robot's behavior.
<latent token> × T
The robot grasps the fork via the handle.
The robot grasps the cup with the tines.
The robot fails to grasp the fork.
Prompt: Now the robot needs to maintain the sanitation of the fork for eatting. Please select the best action plan based on behavior descriptions.{Behavior Narration} × K
FOREWARN: the chosen mode is 3 because it fulfills the task condition of maintaining sanitation **without touching the tines of the fork.**
Robot
$\mathbf{a}_t^1$
$\mathbf{a}_t^2$
$\mathbf{a}_t^3$
ℓ = Place the fork in the bowl while maintaining sanitation.
User



Long-horizon task
Prompt: The robot aims to place the fork in the bowl. Please provide a sentence that best describes the robot's behavior.
<latent token> × T
The robot drops the fork outside the bowl.
The robot fails to place the fork inside bowl.
The robot releases the fork inside the bowl.
Prompt: Now the robot need to maintain the sanitation of the fork for eatting. Please select the best action plan based on behavior descriptions.{Behavior Narration} × K
FOREWARN: the chosen mode is 2 because it is only option that places the fork to the bowl and maintain sanitation.
Robot
$\mathbf{a}_t^1$
$\mathbf{a}_t^2$
$\mathbf{a}_t^3$
ℓ = Place the fork in the bowl while maintaining sanitation.
User

Quantitative Results: Policy Steering Performance

Method	Success Rate ↑					
	Training Task Description			Novel Task Description		
	Cup	Bag	Fork	Cup	Bag	Fork
Base Policy	0.25±0.10	0.20±0.09	0.25±0.10	0.50±0.11	0.35±0.11	0.25±0.10
FOREWARN (Ours)	**0.80±0.09**	**0.70±0.10**	**0.70±0.10**	**0.80±0.09**	**0.70±0.10**	**0.65±0.11**
VLM-Act	0.45±0.11	0.25±0.10	0.20±0.09	0.30±0.10	0.50±0.11	0.25±0.10

20 trials on hardware

Directly fine-tune the original Llama model to take as input $(\mathbf{a}_t^i, o_t)$ and predict behavior narrations <u>without utilizing a world model.</u>

[Y. Wu, R. Tian, G. Swamy, A. Bajcsy. "VLM-in-the-loop Policy Steering via Latent Alignment." arXiv, 2025.]

Qualitative Results: Behavior Narration

Ground-truth observations

Cup Task

Prompt: The robot aims to grasp a cup from the table. Please provide a sentence that best describes the robot's behavior.

Ours: The robot works on seizing the cup through its interior.

VLM-Act: The robot attempts to grab the mug by its handle.

Bag Task

Prompt: The robot aims to grasp a bag of chips from the table. Please provide a sentence that best describes the robot's behavior.

Ours: The robot grips the chip bag directly in the middle.

VLM-Act: The robot holds the chip bag by the corner.

VLM-Act *struggles to capture accurate motion details and thus provides no useful signal for VLM to steer the policy.*

[Y. Wu, R. Tian, G. Swamy, A. Bajcsy. "VLM-in-the-loop Policy Steering via Latent Alignment." arXiv, 2025.]