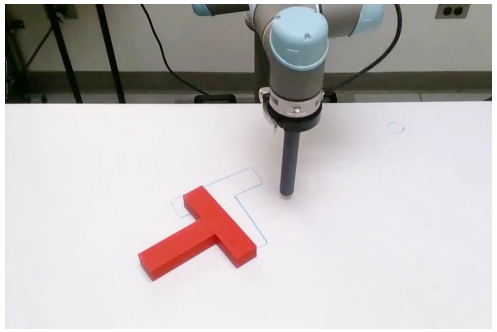
VIKI-R: Coordinating Embodied Multi-Agent Cooperation via Reinforcement Learning

Li Kang
Shanghai Jiaotong University
kangli02@sjtu.edu.cn

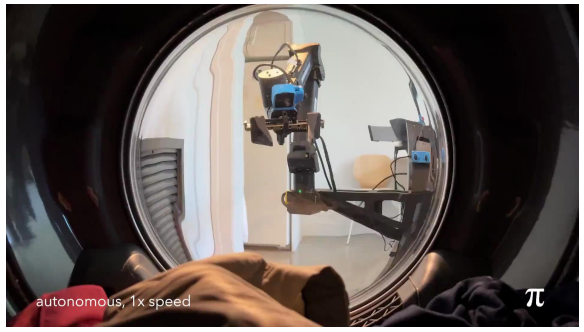
Project Page: https://faceong.github.io/VIKI-R/

Why Multi-Agent Systems Matter

Single agent can perceive, plan, and act—but their ability limited.



Diffusion Policy [1]

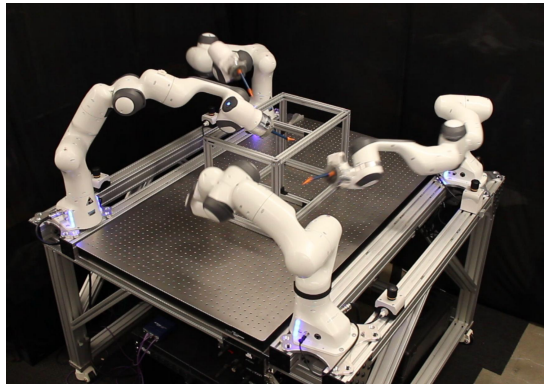


π0 [2]

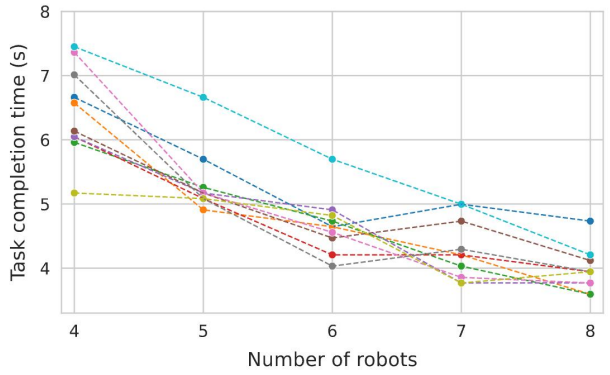
Cross-species collaboration is the future of AGI.

Human-Robot Interaction

Multi-agent systems are the natural scaling law of embodiment.



RoboBallet[3]



Efficiency & Robustness & Generalization

Simulation-Native ⟷ Real2Sim2Real

↘ ↙

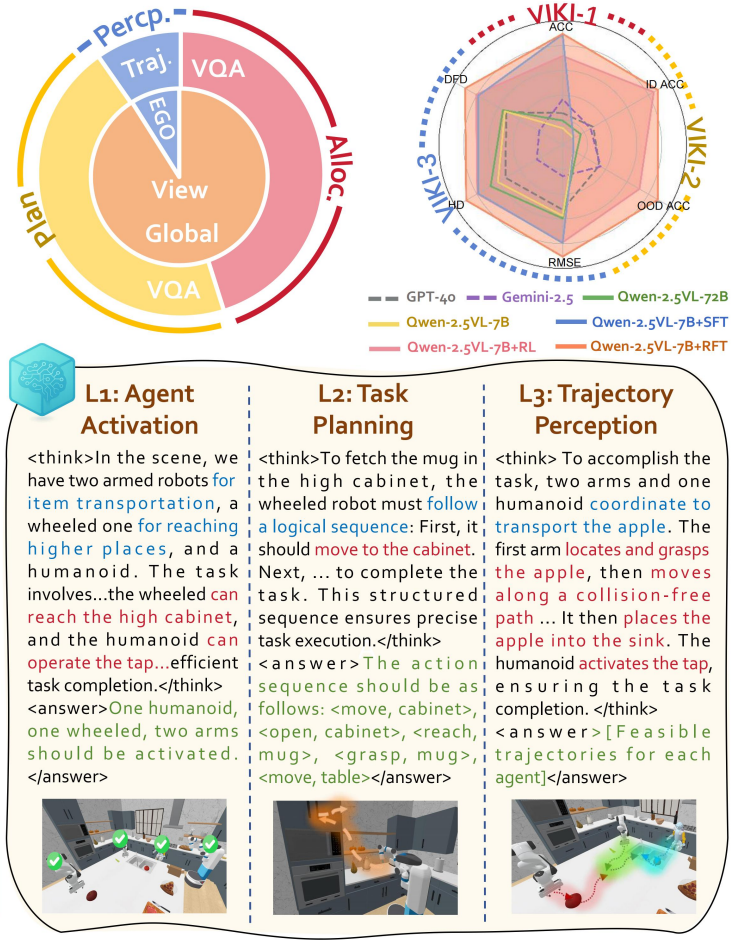
Compositional Intelligence

[1] C. Chi et al., Diffusion Policy: Visuomotor Policy Learning via Action Diffusion, Robotics: Science and Systems (RSS), 2023.
[2] K. Black et al., "π^0: A Vision-Language-Action Flow Model for General Robot Control," Physical Intelligence 2024.
[3] M. Lai et al., "RoboBallet: Planning for Multi-Robot Reaching with Graph Neural Networks and Reinforcement Learning," Science Robotics, 2025.

Compositional Intelligence > Single Skill

Compositional Agents

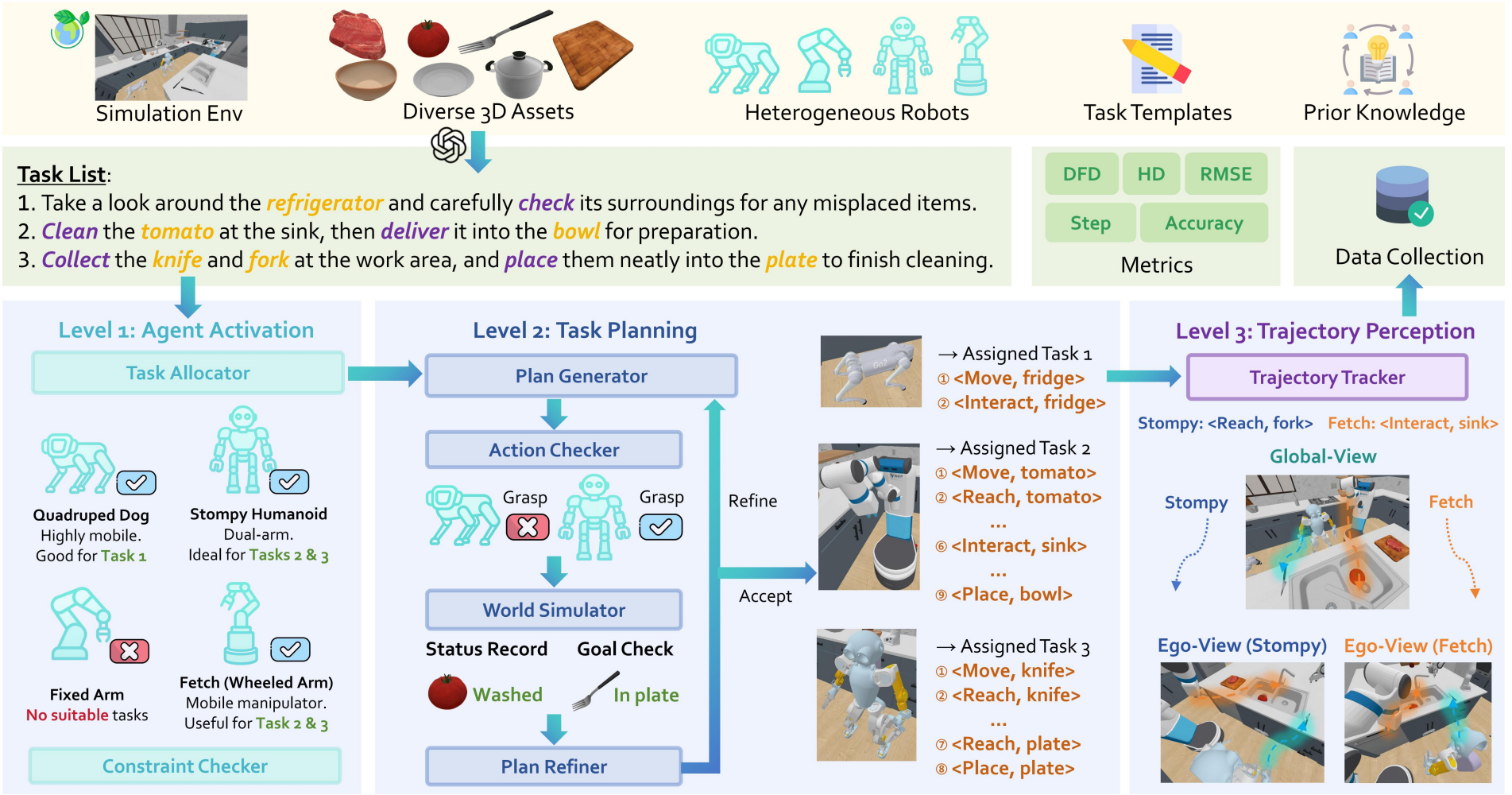
Hierarchical coordination of heterogeneous robots (activation · planning · perception)

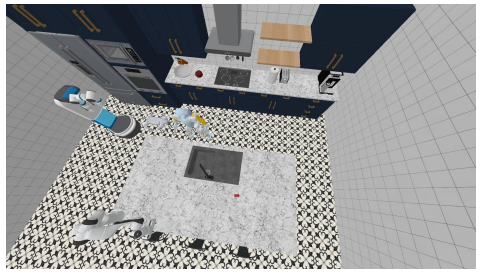


L. Kang et al., VIKI-R: Coordinating Embodied Multi-Agent Cooperation via Reinforcement Learning, NeurIPS 2025.

Simulation-Native: Scaling through Simulation

Rethinking simulation are ideal platforms for collecting large amounts of embodied data.

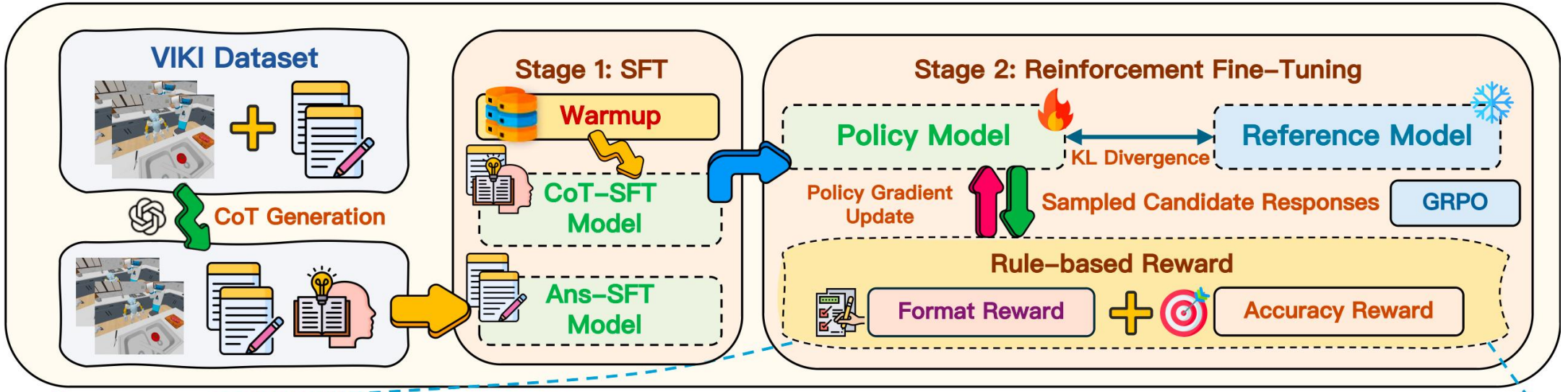


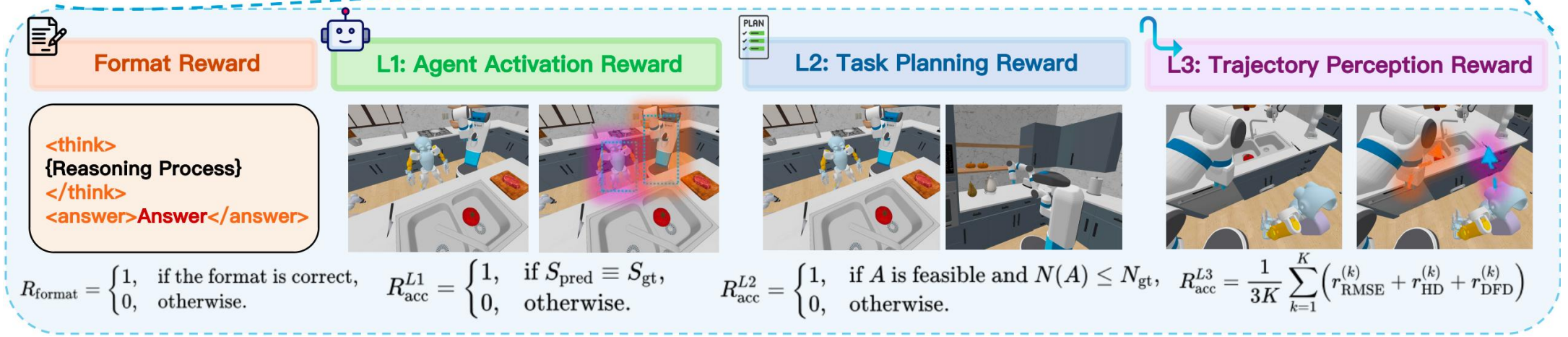


L. Kang et al., VIKI-R: Coordinating Embodied Multi-Agent Cooperation via Reinforcement Learning, NeurIPS 2025.

Learning to Reason: Two-Stage Visual Reasoning Framework

High-Quality SFT + RFT → Efficient Reasoning Models





L. Kang et al., VIKI-R: Coordinating Embodied Multi-Agent Cooperation via Reinforcement Learning, NeurIPS 2025.

Main Results

ACC_ID: Accuracy on in-domain test set. ACC_OOD: Accuracy on out-of-domain test set. Bold: best. Underline: second best.

Category	Method	VIKI-L1	VIKI-L2			VIKI-L3			
		ACC_ID ↑	ACC_ID ↑	ACC_OOD ↑	ACC_AVG ↑	RMSE ↓	HD ↓	DFD ↓	AVG ↓
Closed-Source	GPT-4o	18.40	22.56	10.02	17.50	100.80	115.34	131.05	115.73
	Claude-3.7-Sonnet	12.40	19.44	0.57	11.82	283.31	323.53	346.88	317.91
	Gemini-2.5-Flash-preview	31.40	20.00	10.51	16.17	453.89	519.14	540.80	504.61
Open-Source	Qwen2.5-VL-72B-Instruct	11.31	8.40	1.20	5.49	81.31	94.62	113.15	96.36
	Qwen2.5-VL-32B-Instruct	9.50	3.60	0.00	2.15	88.48	99.80	119.78	102.69
	Llama-3.2-11B-Vision	0.40	0.50	0.00	0.30	192.69	223.57	231.85	216.04
Qwen2.5VL-3B	Zero-Shot	1.95	0.22	0.00	0.13	96.22	114.93	130.98	114.04
	+Ans SFT	35.29	81.06	30.71	60.74	74.70	90.28	102.26	89.08
	+VIKI-R-Zero	20.40	0.00	0.00	0.00	80.36	95.36	120.27	98.66
	+VIKI-R	74.10	93.61	32.11	68.78	75.69	90.25	103.65	89.86
Qwen2.5VL-7B	Zero-Shot	4.26	0.44	0.00	0.26	81.93	103.82	112.91	99.55
	+Ans SFT	72.20	**96.89**	25.62	68.13	65.32	81.20	90.89	79.14
	+VIKI-R-Zero	**93.59**	0.17	0.00	0.10	67.42	85.30	95.32	82.68
	+VIKI-R	93.00	95.22	**33.25**	**69.25**	**64.87**	**79.23**	**89.36**	**77.82**

Additional Analysis

Observation 1

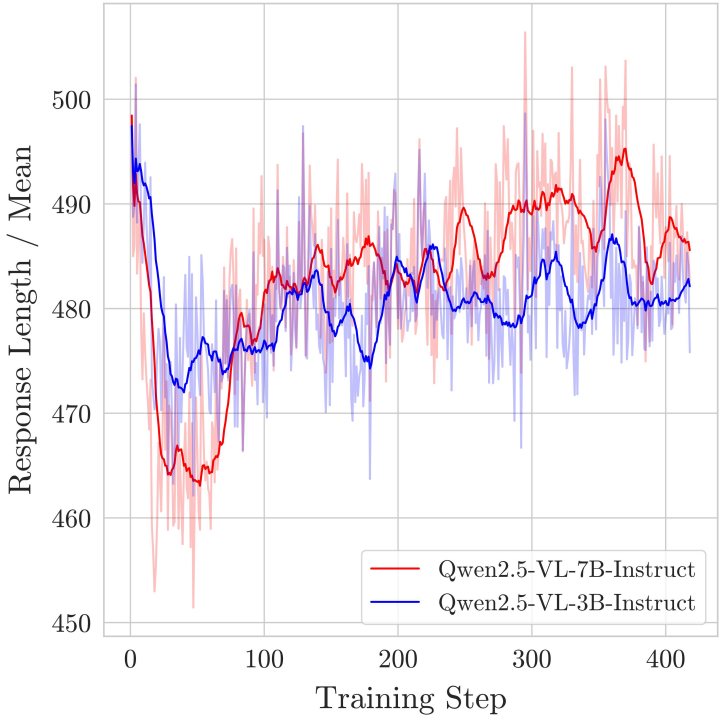
Applying the **step penalty** improves both generalization and planning efficiency, helping models produce more accurate and optimized plans.

ACC_ID: Accuracy on in-domain tasks. ACC_OOD: Accuracy on out-of-domain tasks.
Δ Steps: Difference between predicted and ground-truth plan length.

Variant	ACC_OOD ↑	ACC_ID ↑	Δ Steps ↓
VIKI-R (with step penalty)	46.8	96.0	+0.05
VIKI-R (without step penalty)	7.1	8.0	+1.97

Observation 2

The training process evolves from **format optimization** to **task reasoning**, showing a clear progression toward more accurate and detailed task execution.



Thank You — Building Together the Next Generation of Embodied System

Li Kang

Shanghai Jiaotong University

2025.11