universität freiburg

Open World Embodied Intelligence:

Learning from Perception to Action in the Wild

Abhinav Valada

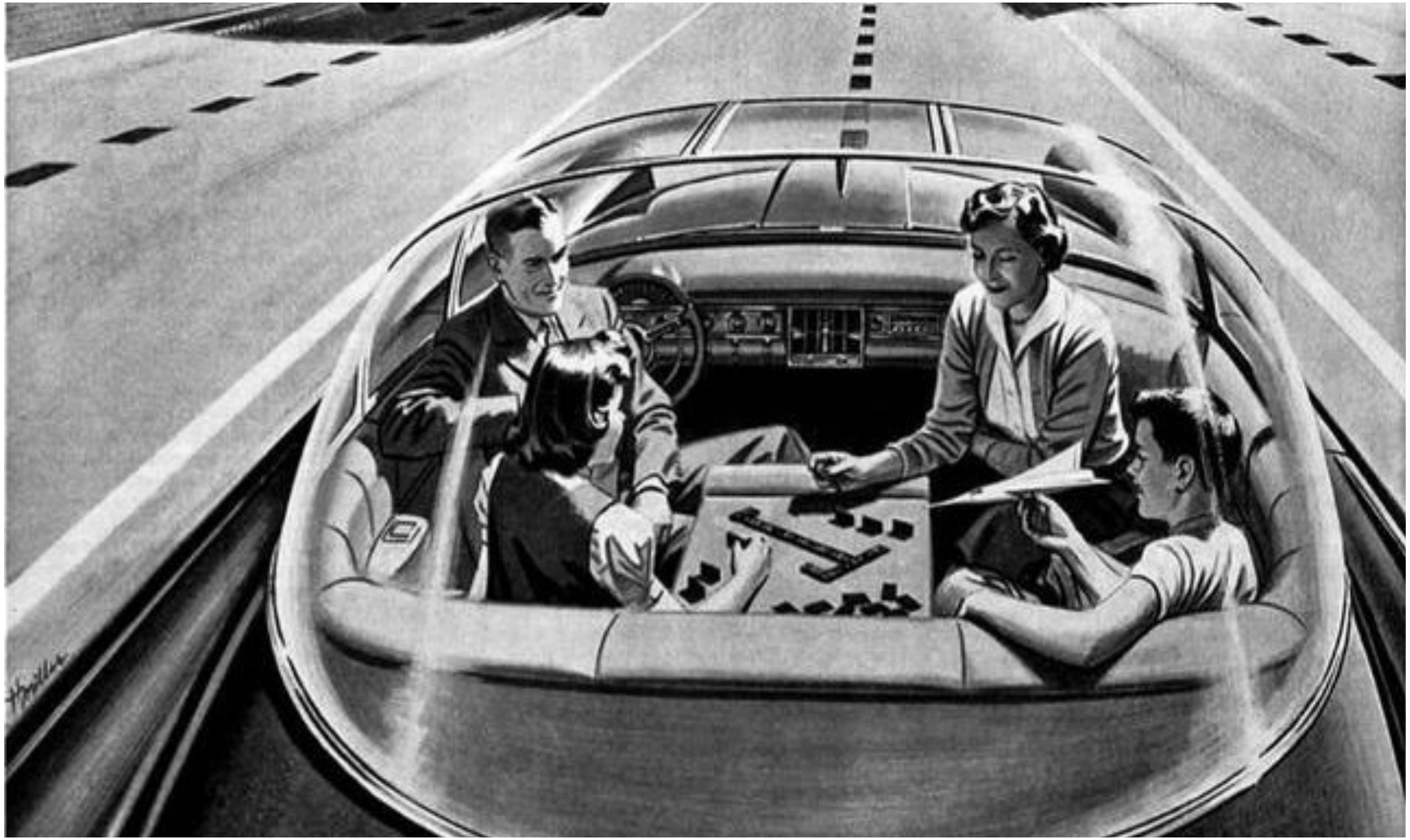
18th November 2025

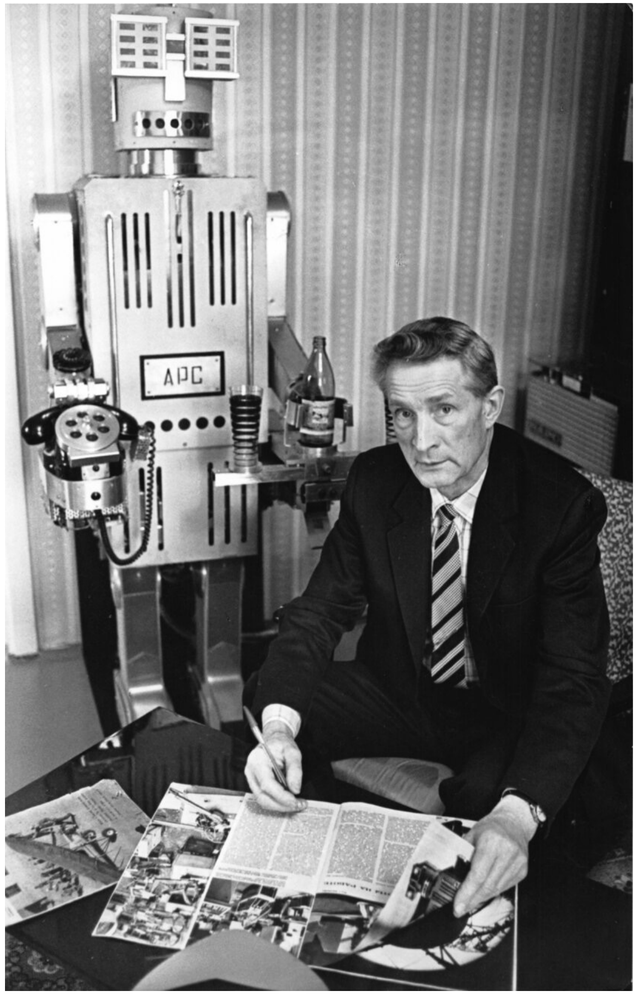
ellis UNIT FREIBURG

robot learning

Imagining Open-World Embodied Intelligence

Learn to perform complex tasks alongside people, adapt to diverse environments, objects, attributes, and skills

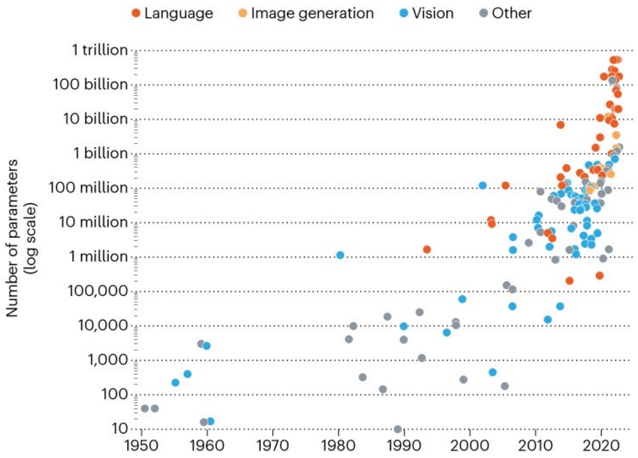


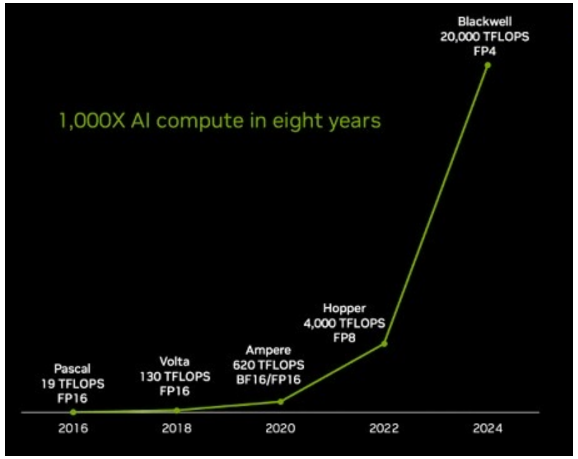


Reimagining Open-World Embodied Intelligence

A Day in the Life of AI in 2025

- Remarkable progress in capabilities of AI to generate and reason about language, images, tasks, and videos
- Primary contributors:
  - Vast amounts of suitable training data such that open-world reasoning becomes in-distribution
  - Very large models that can digest this data
  - Primarily behavior cloning for training





Robot Learning Data Pyramid

Robot Data

(24 Hours / Robot / Day)

Demonstration Interface Data

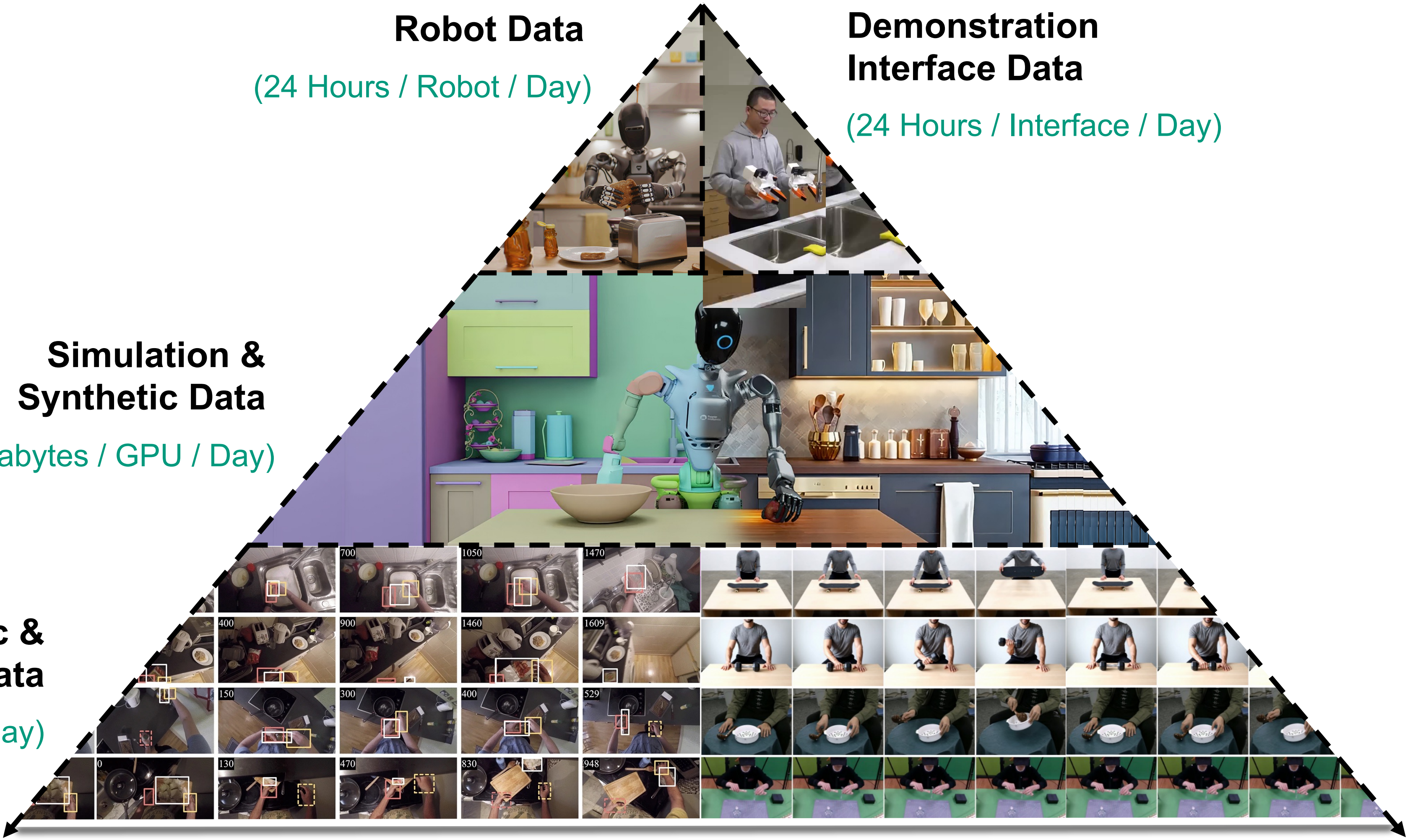
(24 Hours / Interface / Day)

Simulation & Synthetic Data

(Gigabytes / GPU / Day)

Exocentric & Egocentric Video Data

(Exabytes / Day)



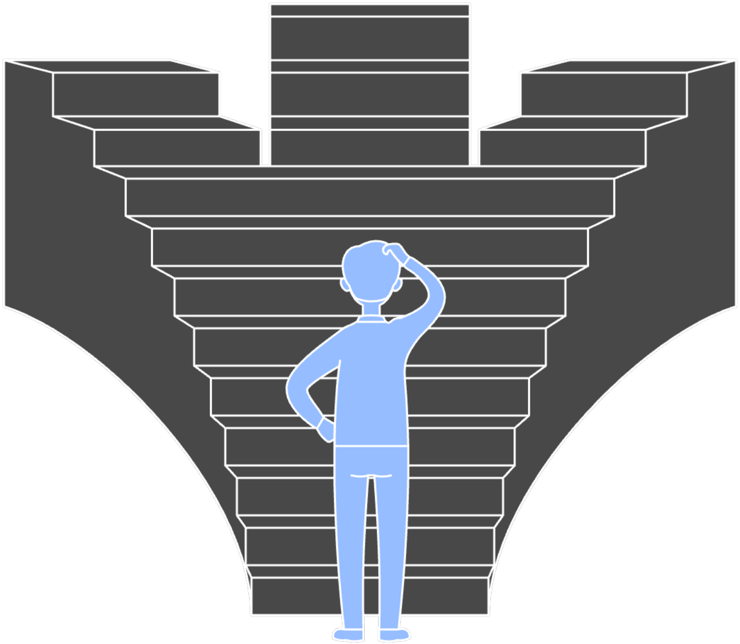
Quick Poll

Which lever matters the most for real-world generalization?

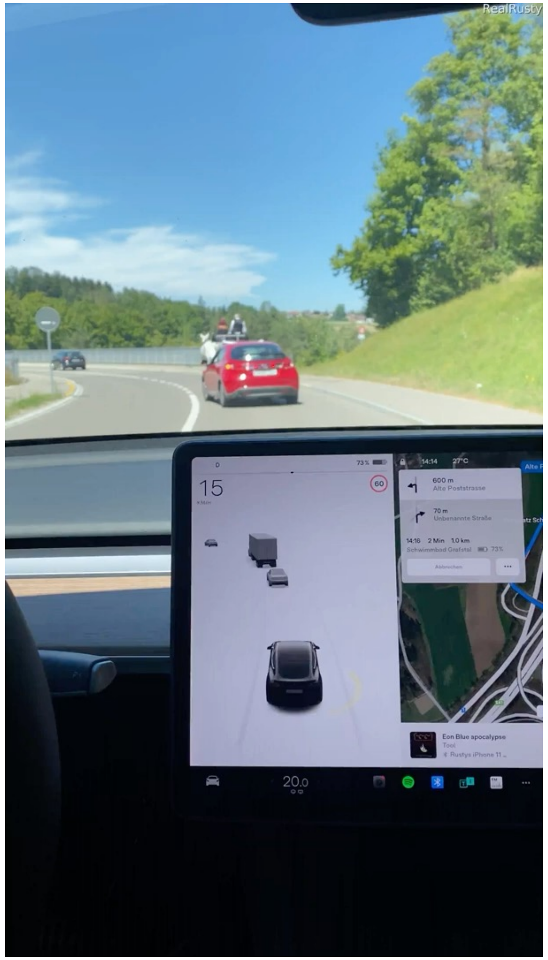
Architectures

Data

Algorithms



Physical Embodiment of Robots Breaks Assumptions

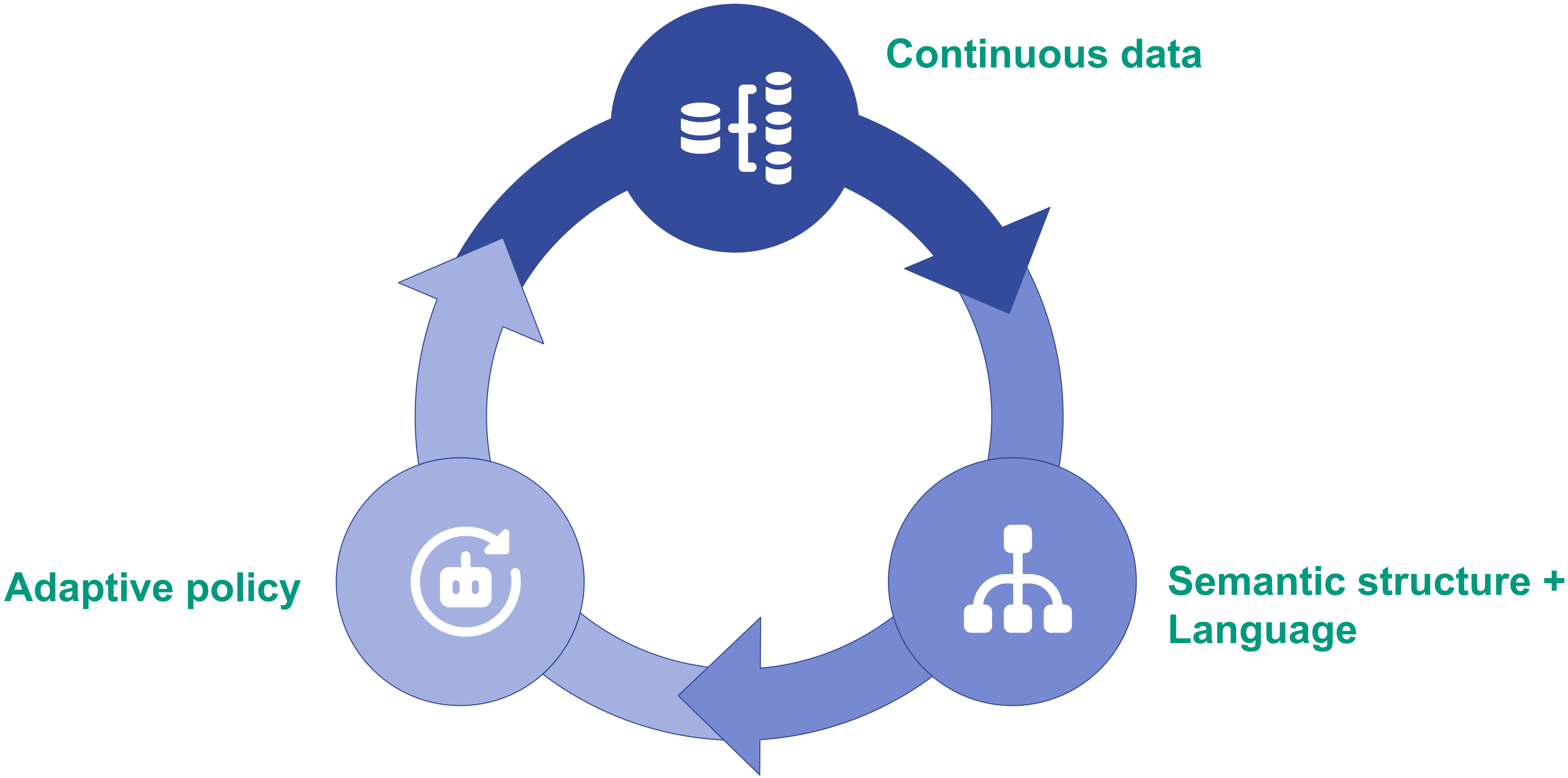


Curated data ⟶ Fixed representations ⟶ Frozen policy

How do we enable generalizable long-horizon autonomy in the open world?

Open world isn't **optional**. It's **operational**.

Towards Open-World Embodied Autonomy

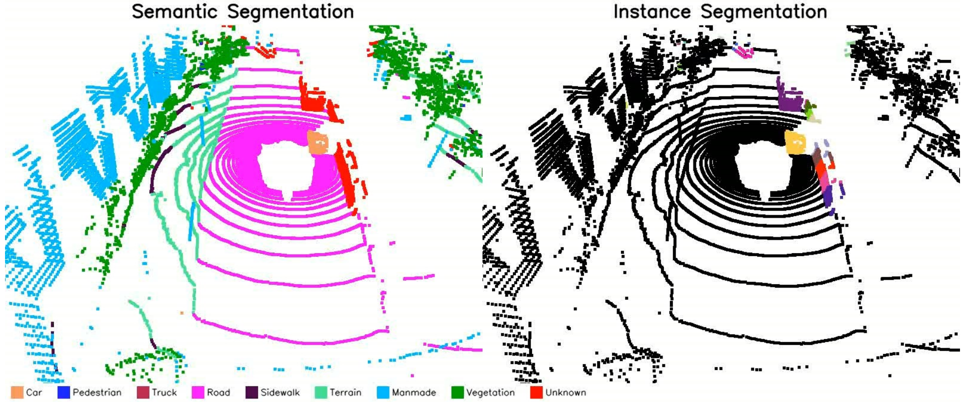


Open-World Robotic Perception

Label Efficient Panoptic Segmentation

[Hindel *et al.*, IROS 2025, Vödisch *et al.*, RA-L 2025]

Open-Set & Out-of-Distribution Segmentation



[Mohan *et al.*, IROS 2025, RA-L 2024]

Open-Vocabulary Dynamic 3D Scene Graphs

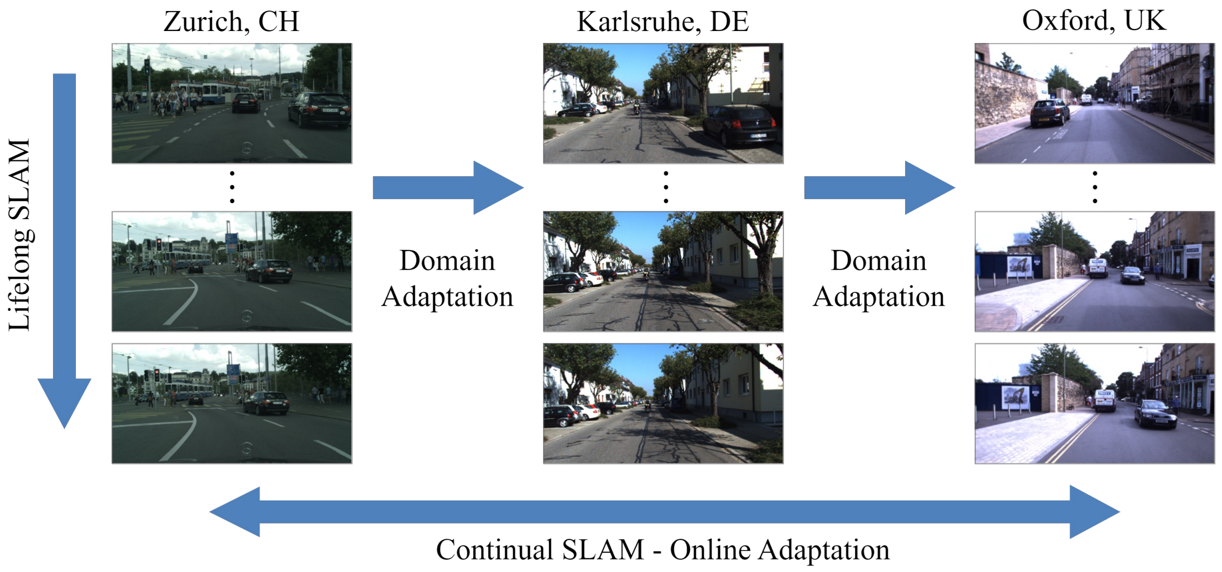
- A Environment
- B Roads & intersections
- C Static objects
- D Dynamic objects
- E Pose graph & point clouds

[Stienke *et al.*, IROS 2025]

"The Oxford Radar RobotCar Dataset", D. Barnes *et al.*, ICRA, 2020

Rethinking Lifelong SLAM

- Lifelong SLAM
  - A single environment that changes over time
- Domain adaptation
  - Directed knowledge transfer from domain A to domain B
- Continual SLAM
  - Undirected transfer between multiple domains
  - Represents deploying a robot in the open world
  - Online adaptation mitigating catastrophic forgetting



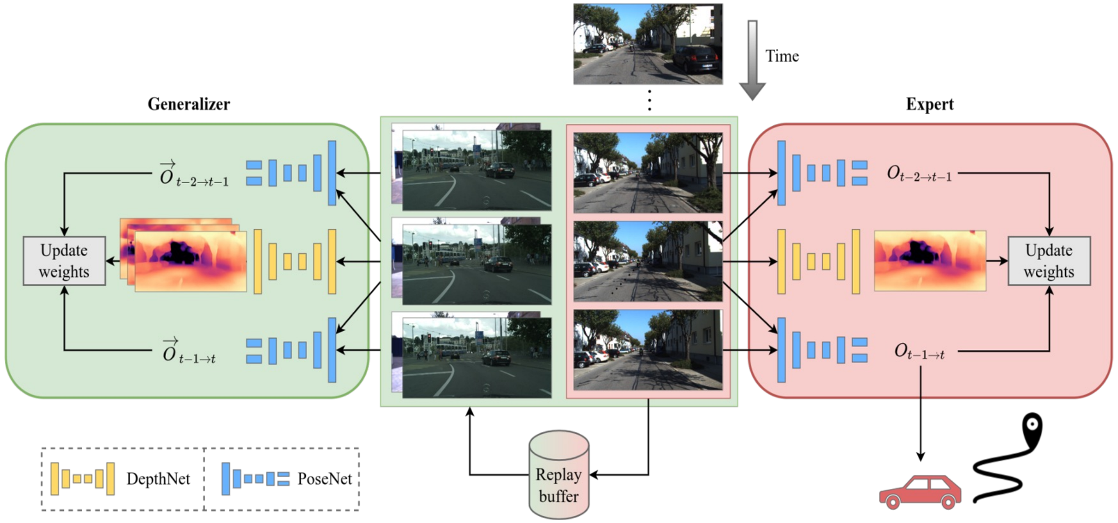
CL-SLAM: Online Continual SLAM

Continual learning setup:

1. Pretrain model on large, diverse dataset
2. Deploy robot in the open world
3. Robot adapts online and memorizes experience samples

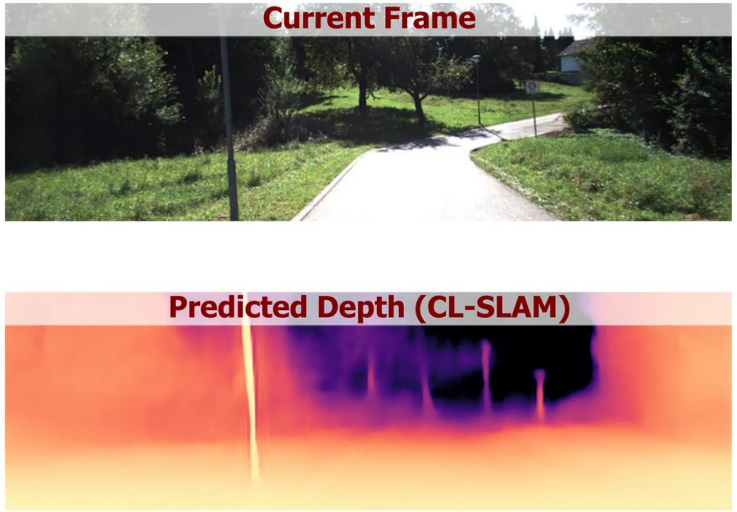
Applied strategies:

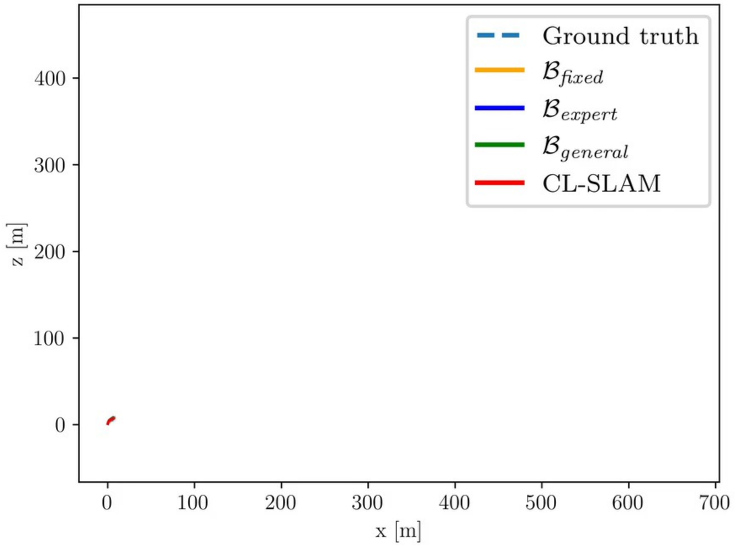
- Data rehearsal via a replay buffer
- Dual-network architecture consisting of a generalizer and an expert



[Vödisch *et al.*, ISRR 2022, CVPR 2023]

CL-SLAM Results: Memory Retention



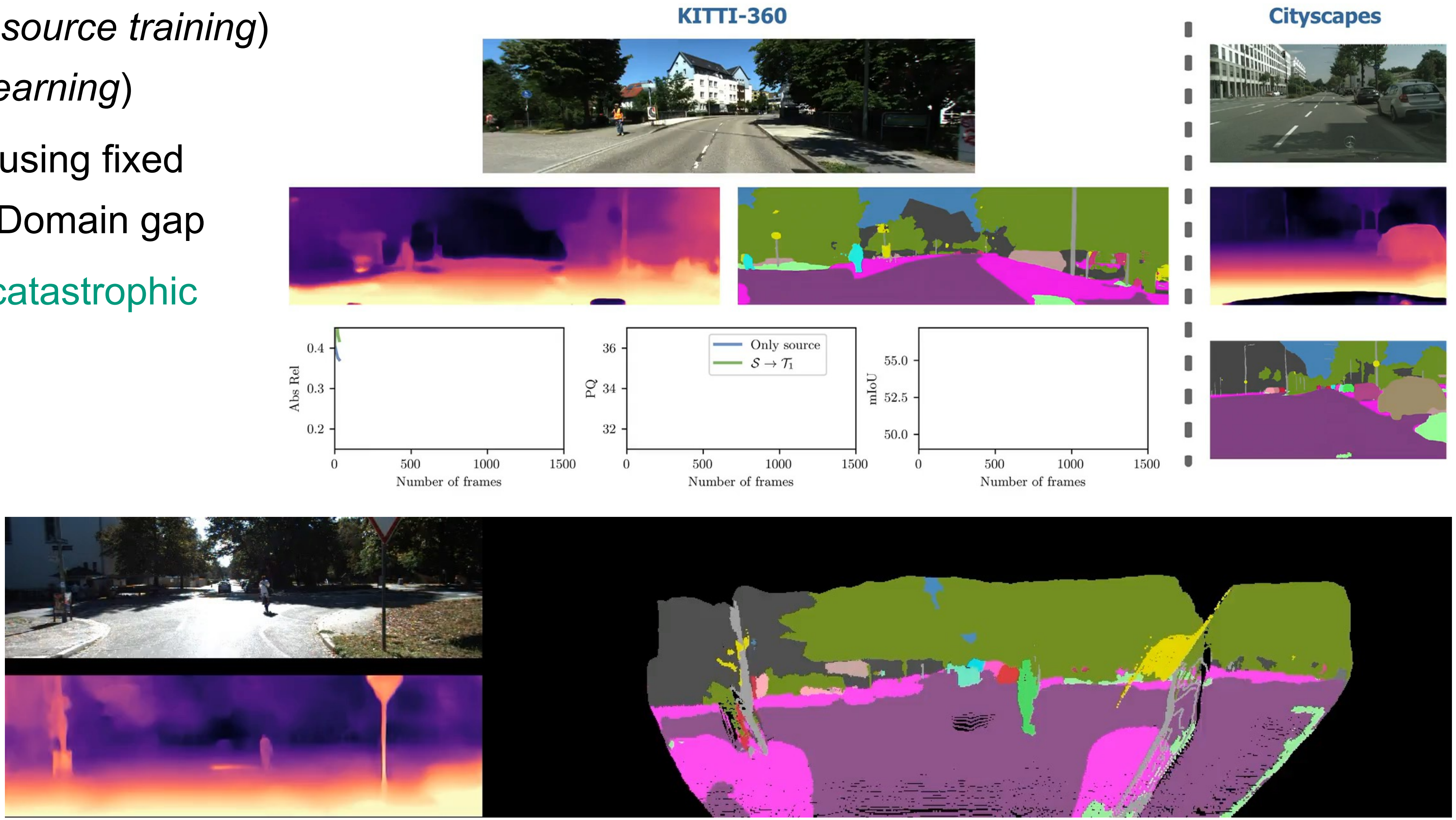


[Vödisch *et al.*, ISRR 2022, CVPR 2023]

- Deploy on a new scene from a previously seen environment
  Cityscapes → KITTI → RobotCar → KITTI
- CL-SLAM improves performance on past scenes while adapting online

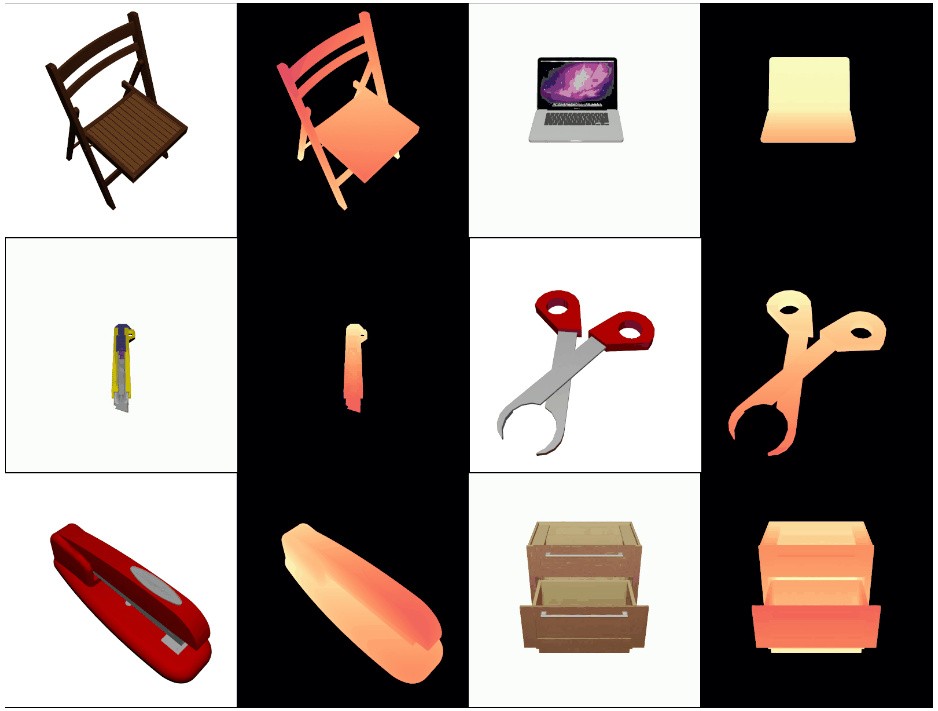
Online Continual Learning of Panoptic Depth Estimation

- Experiment: Cityscapes (*source training*) → KITTI-360 (*continual learning*)
- Better performance than using fixed weights (*only source*) → Domain gap
- Successful mitigation of catastrophic forgetting



[Vödisch *et al.*, RSS 2023]

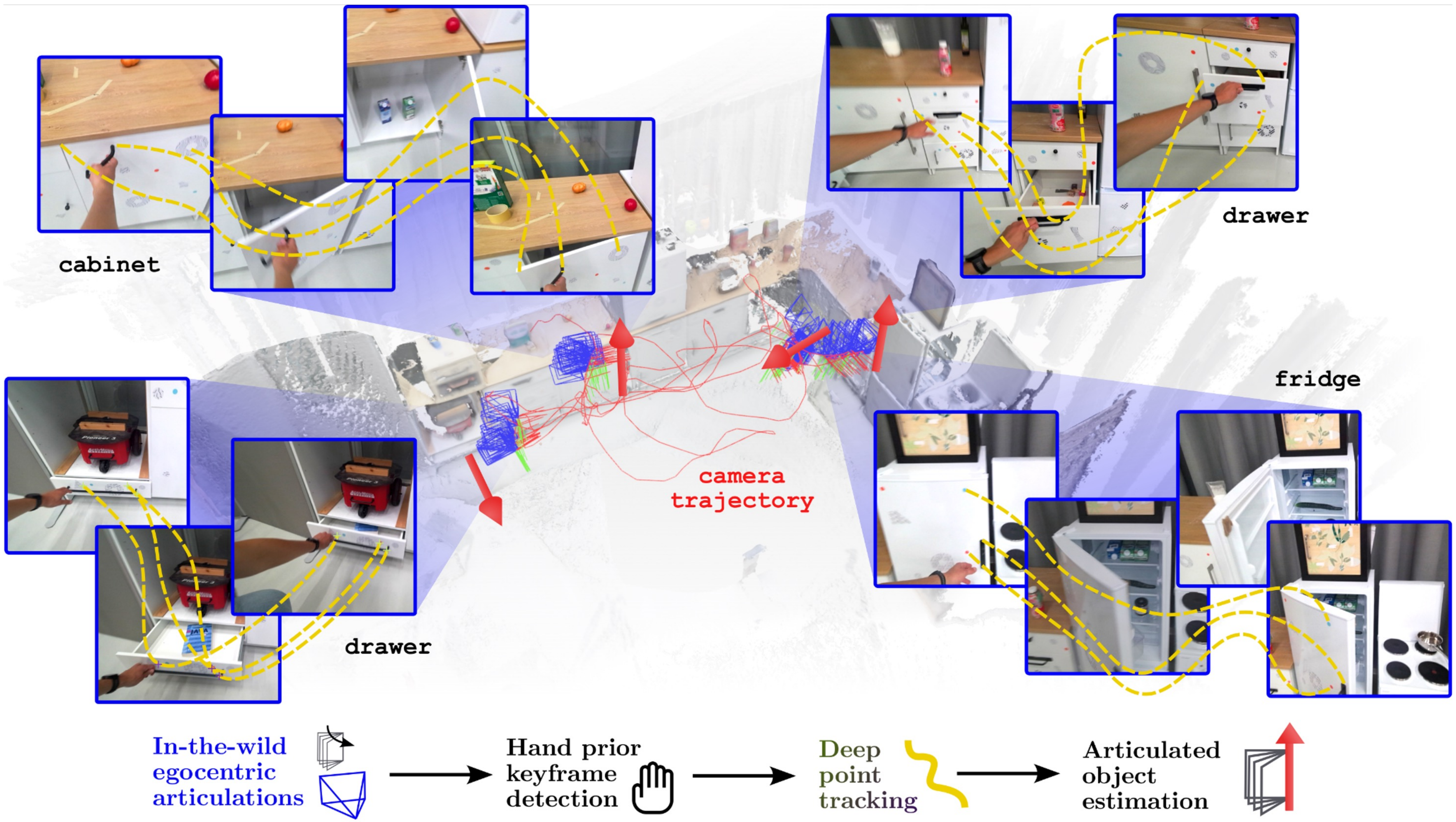
Articulation Estimation from Casual Observations



[Heppert *et al.*, CVPR 2023, Buchanan *et al.*, ICRA 2024]

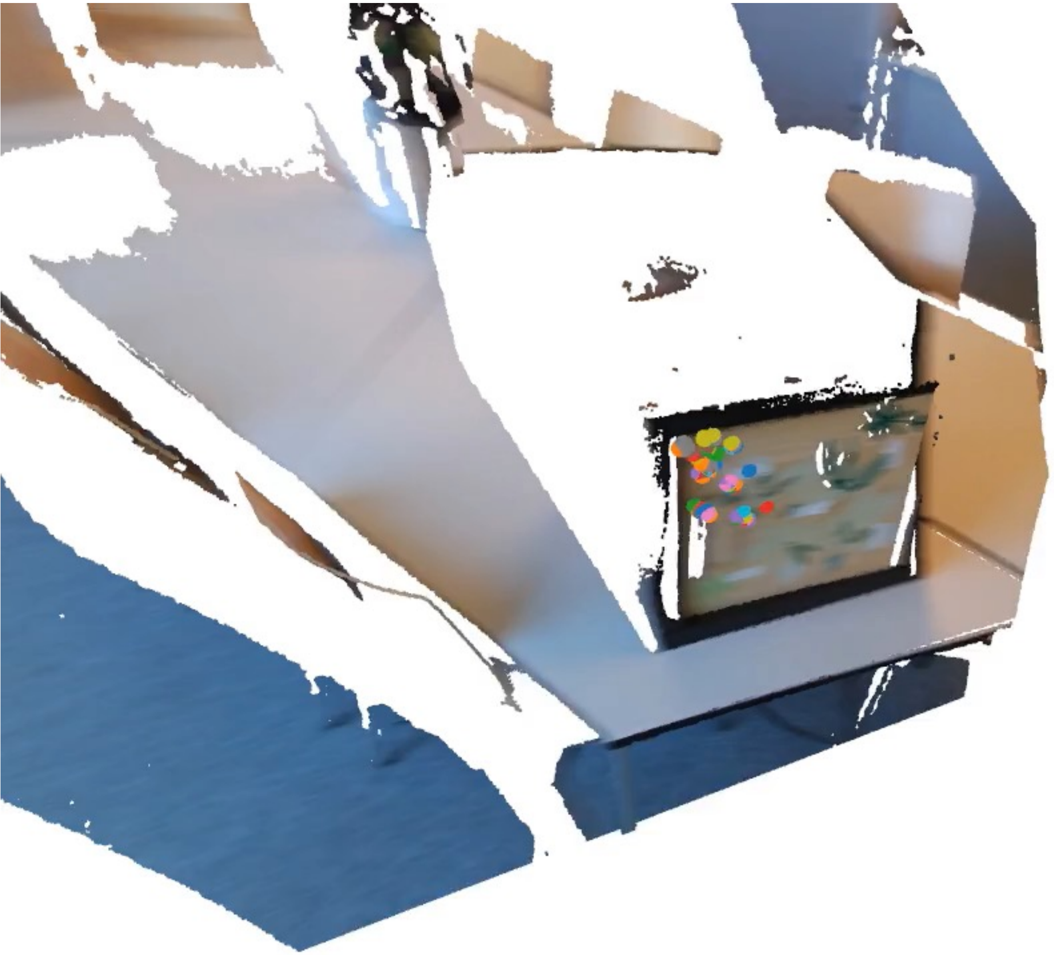
[Werby *et al.*, CoRL 2025]

ArtiPoint: Articulated Object Estimation in the Wild



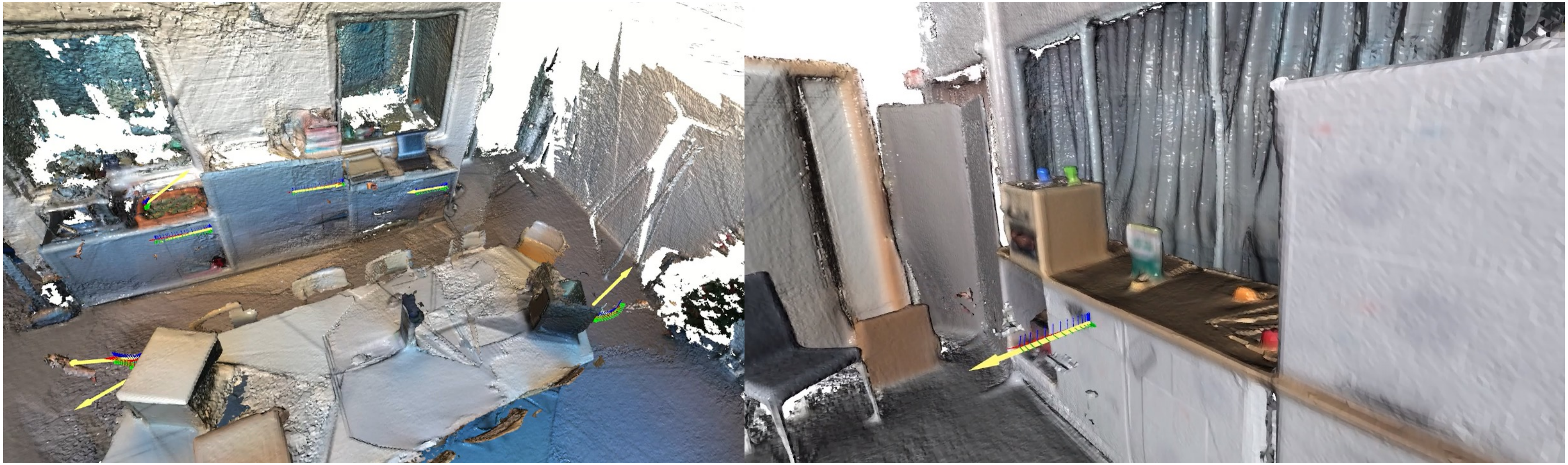
[Werby *et al.*, CoRL 2025]

Arti4D Dataset

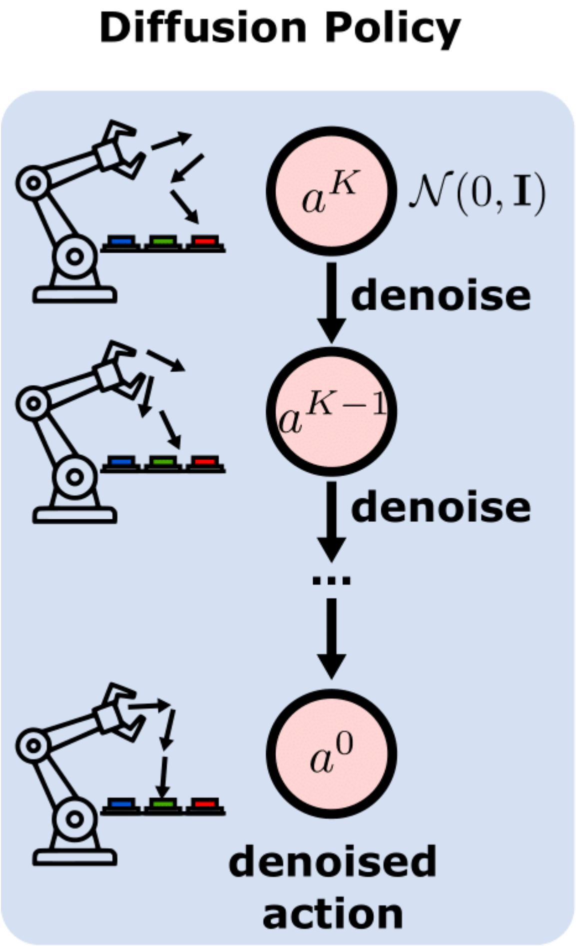


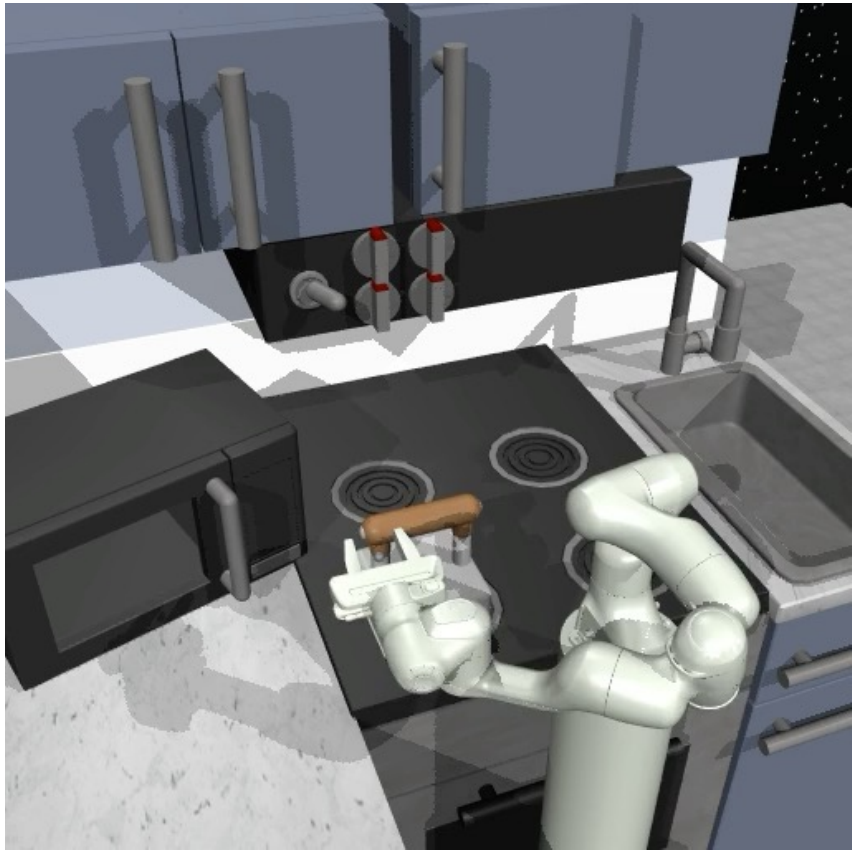
[Werby *et al.*, CoRL 2025]

Scene-Level Articulated Object Estimation on Arti4D

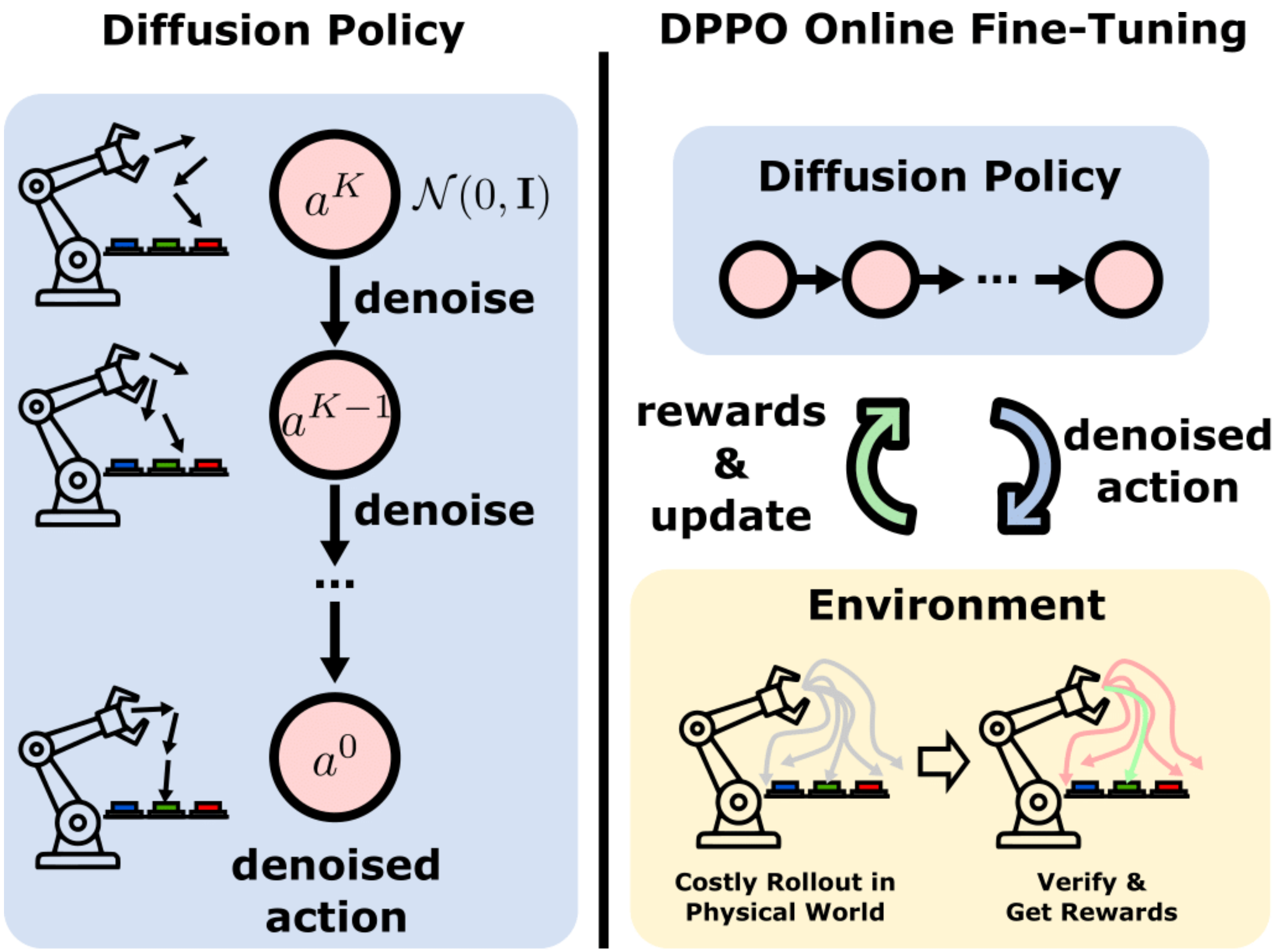


Policy Learning via Action Diffusion



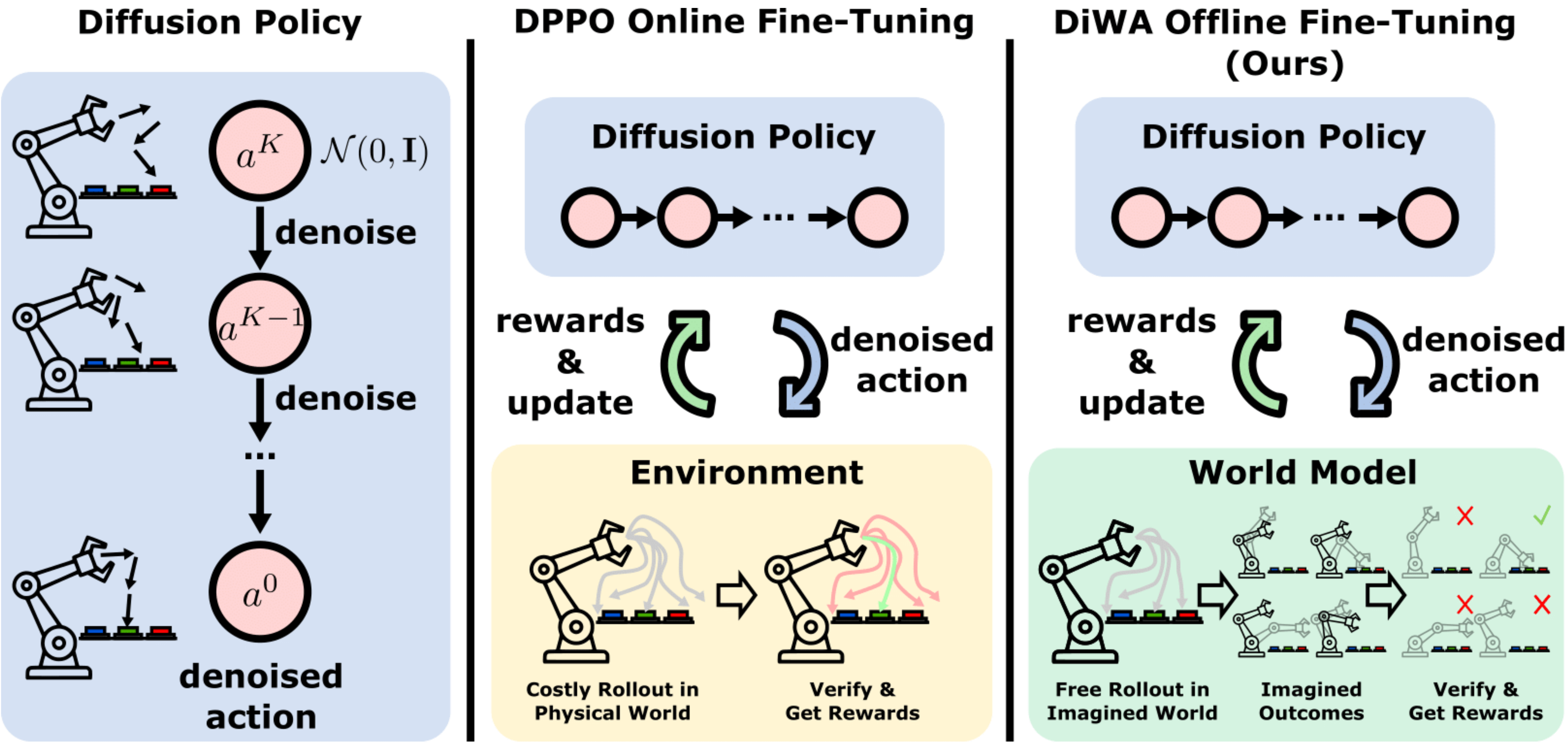


Why Online Fine-Tuning Fails



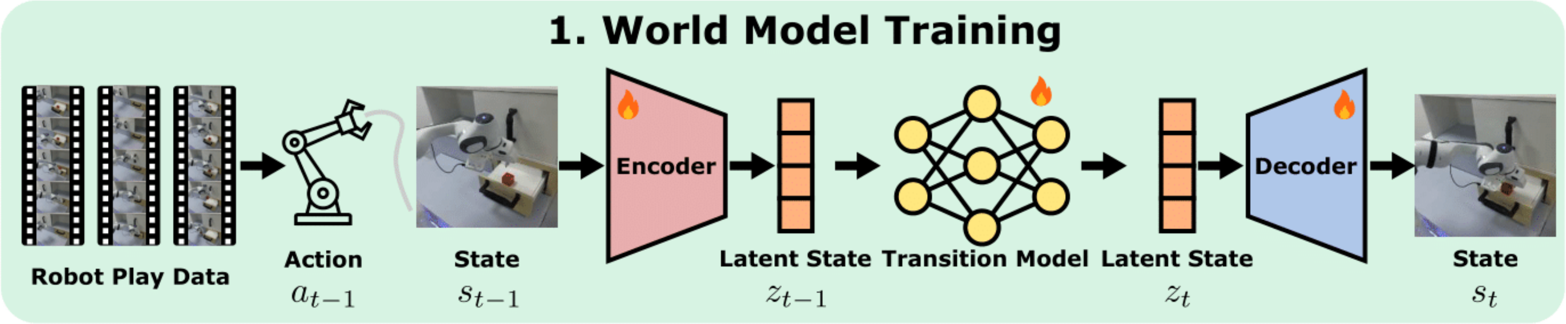
Sleep on a Problem

Offline Fine-Tuning in Dreams



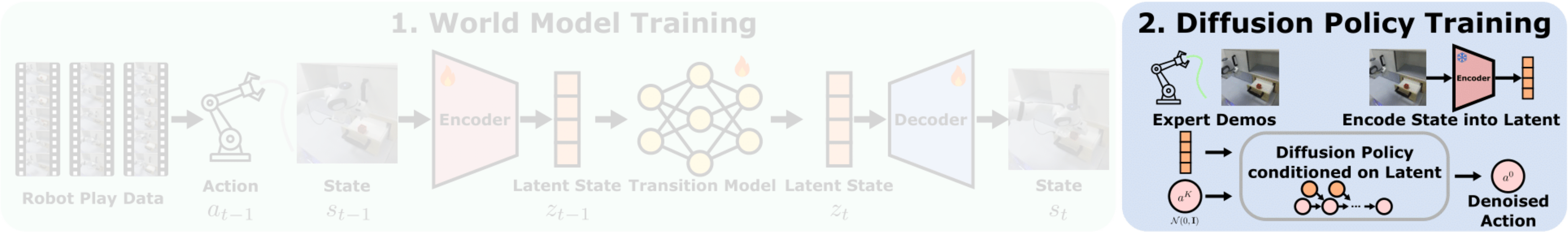
[Chandra *et al.*, CoRL 2025]

DiWA: Learn a World Model



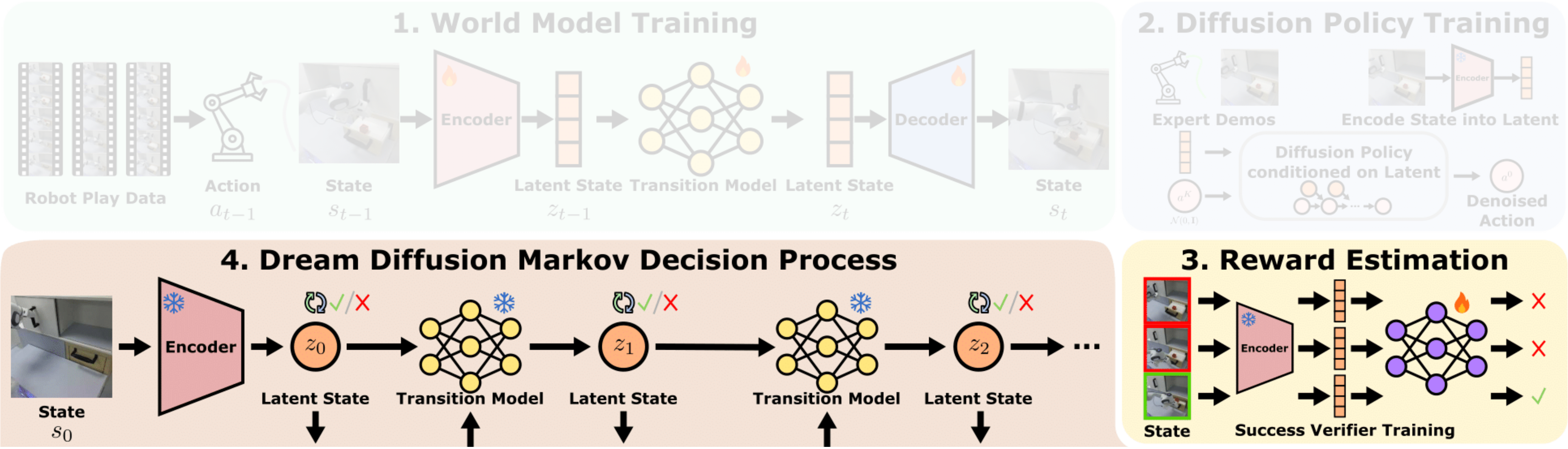
[Chandra *et al.*, CoRL 2025]

DiWA: Pre-Train a Diffusion Policy



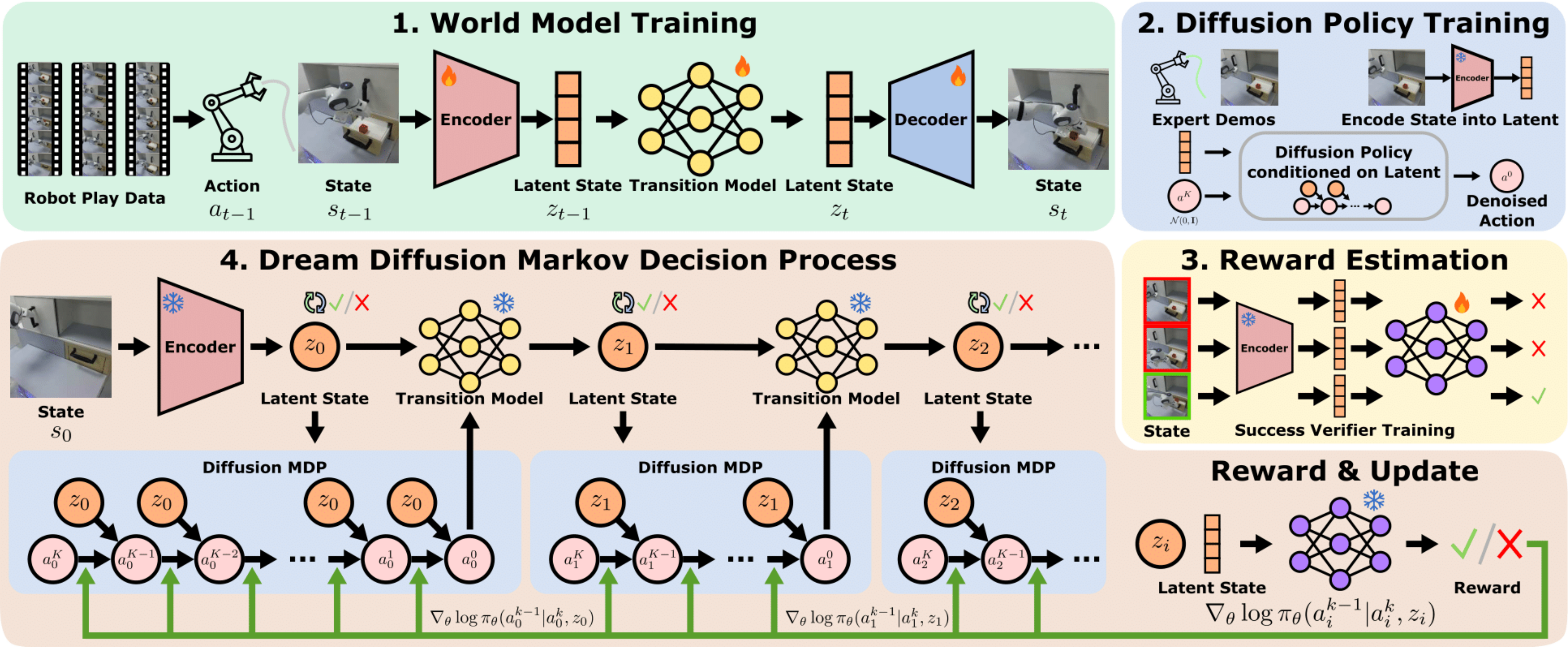
[Chandra *et al.*, CoRL 2025]

DiWA: Latent Reward Estimation



[Chandra *et al.*, CoRL 2025]

DiWA: Dream Diffusion MDP

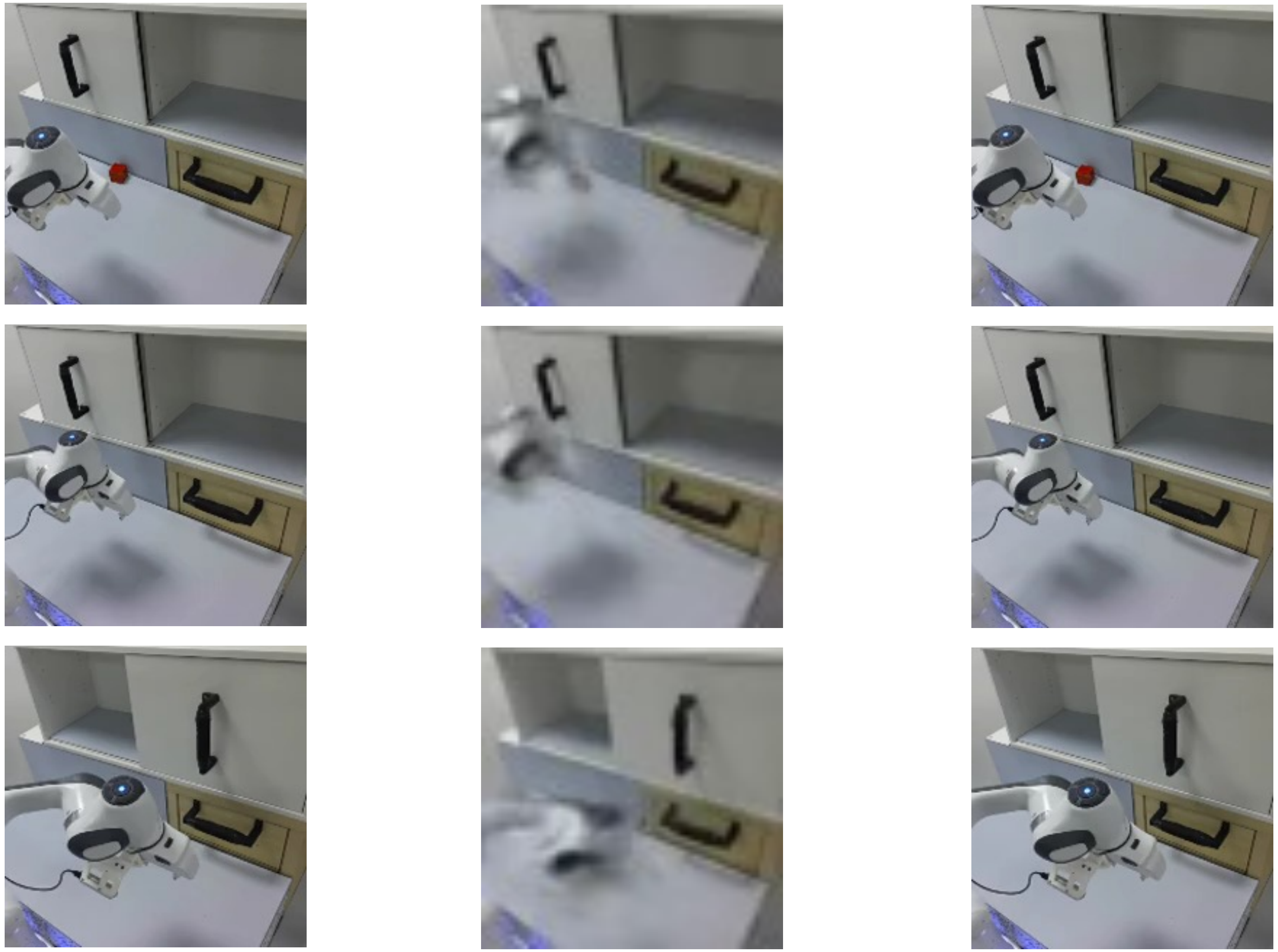


[Chandra *et al.*, CoRL 2025]

Competence After Dreaming

Real-world play data collection

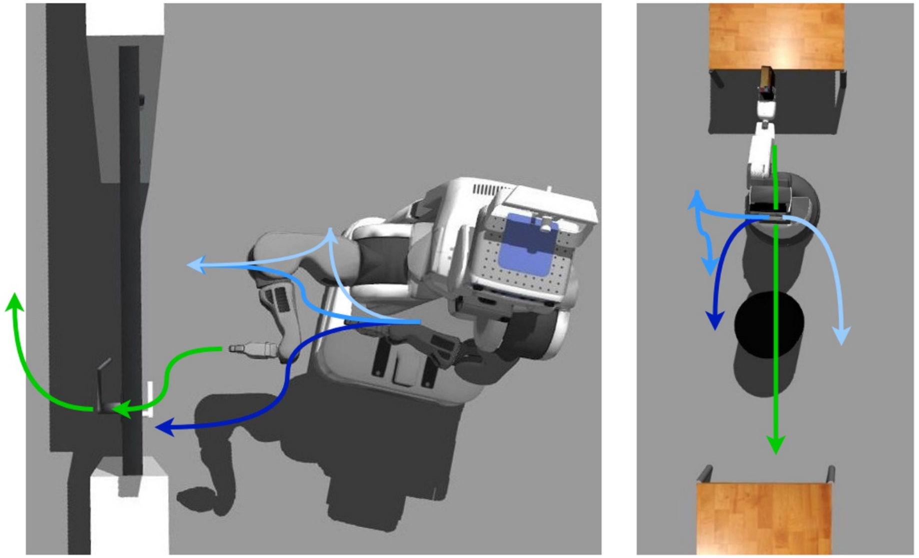
Initial Policy Before Dreaming | Offline Fine-tuning During Dreaming | Final Policy After Dreaming

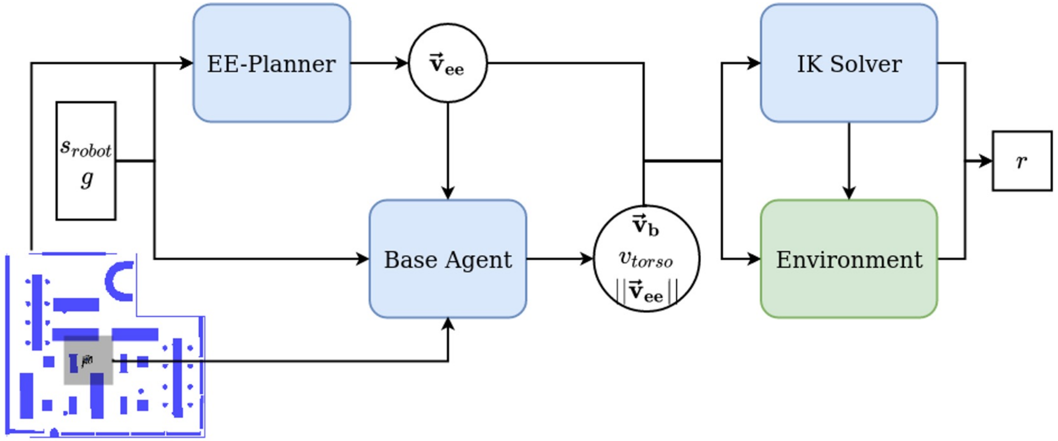


Online diffusion policy fine-tuning requires 2.4 million interactions to match DiWA

Learning General Mobile Manipulation Motions

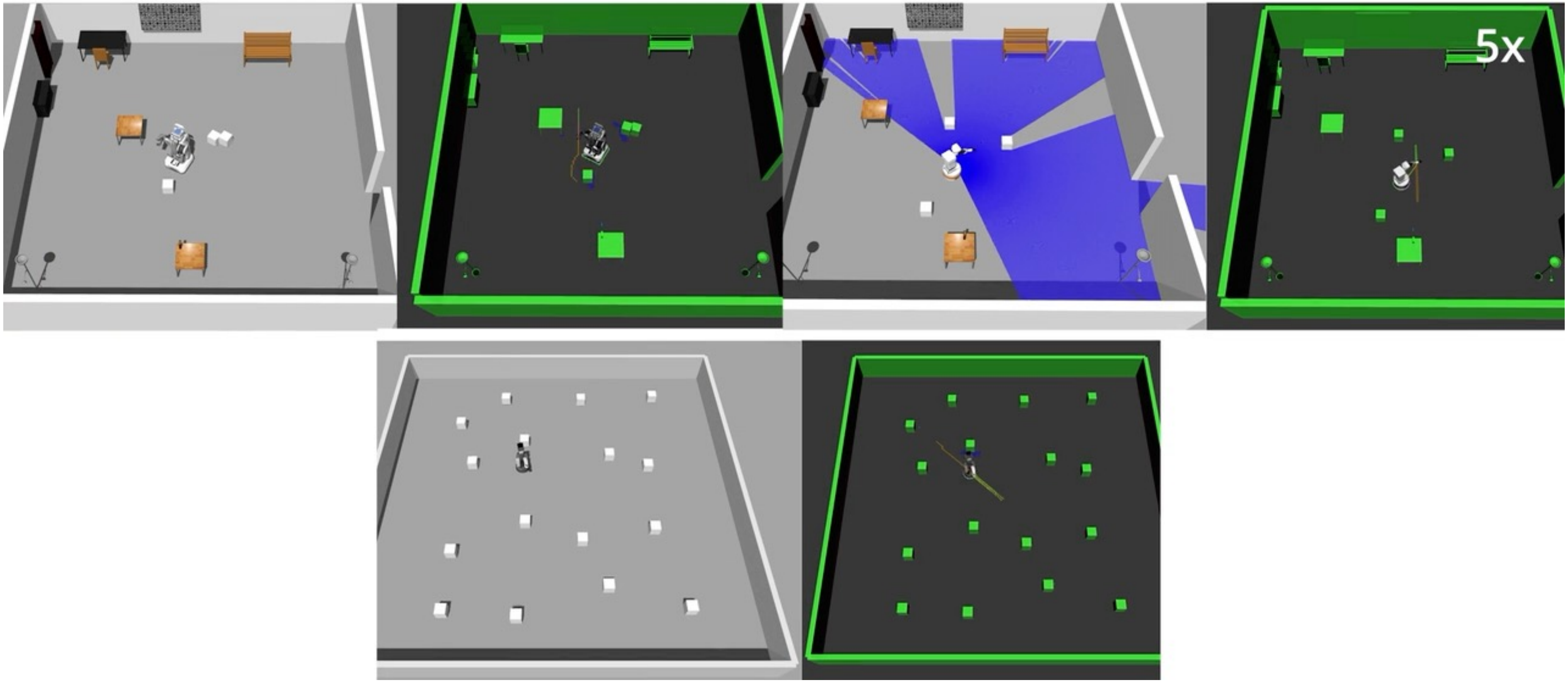
- Defining whole-body motions is highly challenging
- Goal: Learn feasible motions for the mobile base, given an end-effector goal
- Decompose mobile manipulation into arbitrary end-effector planner and RL agent controlling the base
- Kinematic feasibility as dense reward
- Generalizes across diverse robotic platforms and unseen tasks





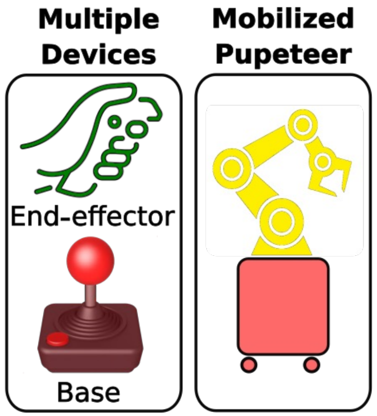
[Honerkamp *et al.*, RA-L 2021, T-RO 2023]

N^2M^2: Neural Navigation for Mobile Manipulation




5x

Zero-Cost Whole-Body Teleoperation



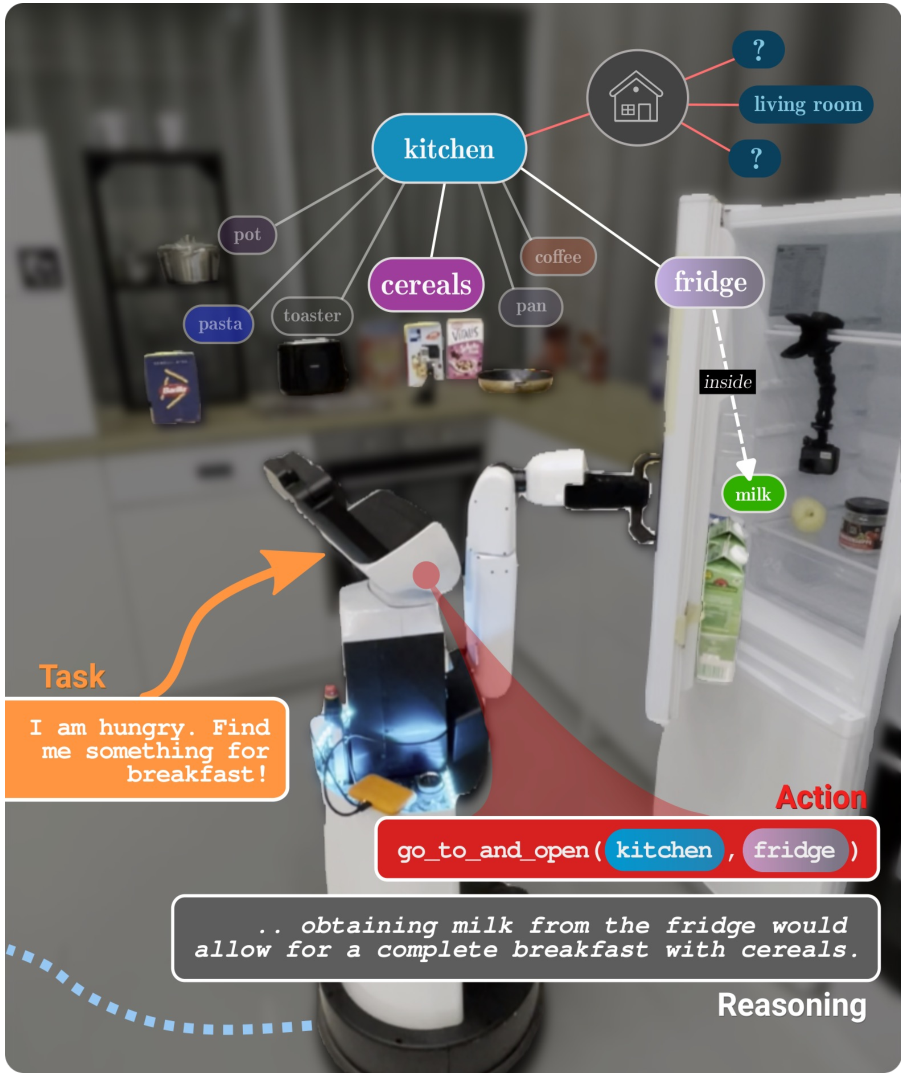
➤ Operation either cumbersome or expensive

Idea: Infer user intentions and delegate the remaining actions to our base agent

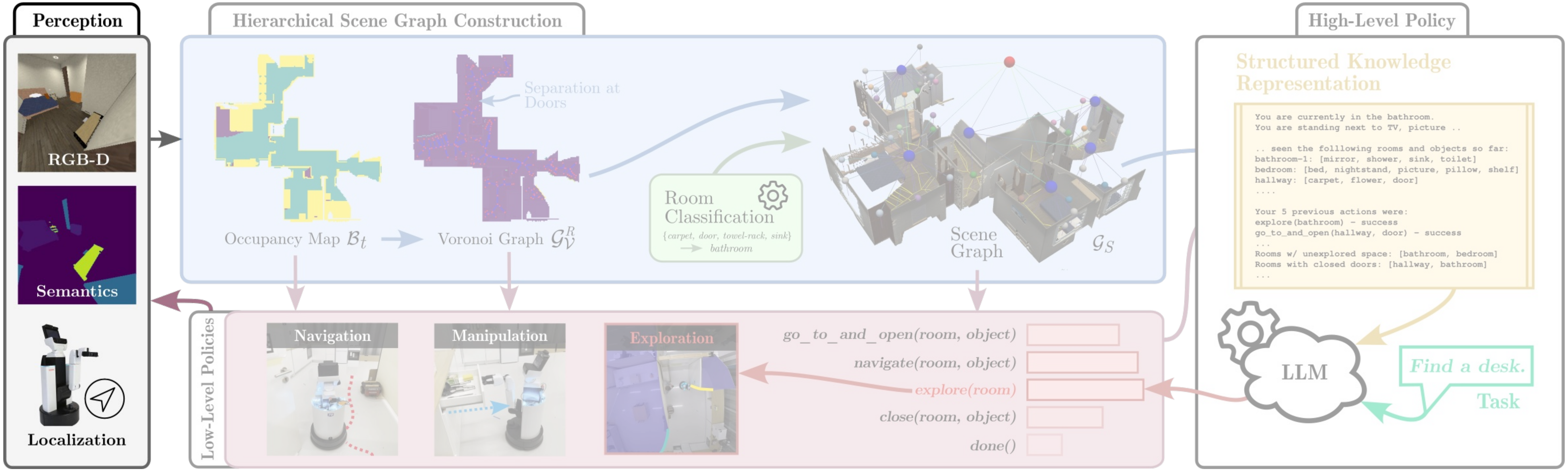
[Honerkamp *et al.*, RA-L 2025]

Language-Grounded Mobile Manipulation

Idea: Ground natural language instructions in dynamic scene graphs for high-level reasoning and mobile manipulation

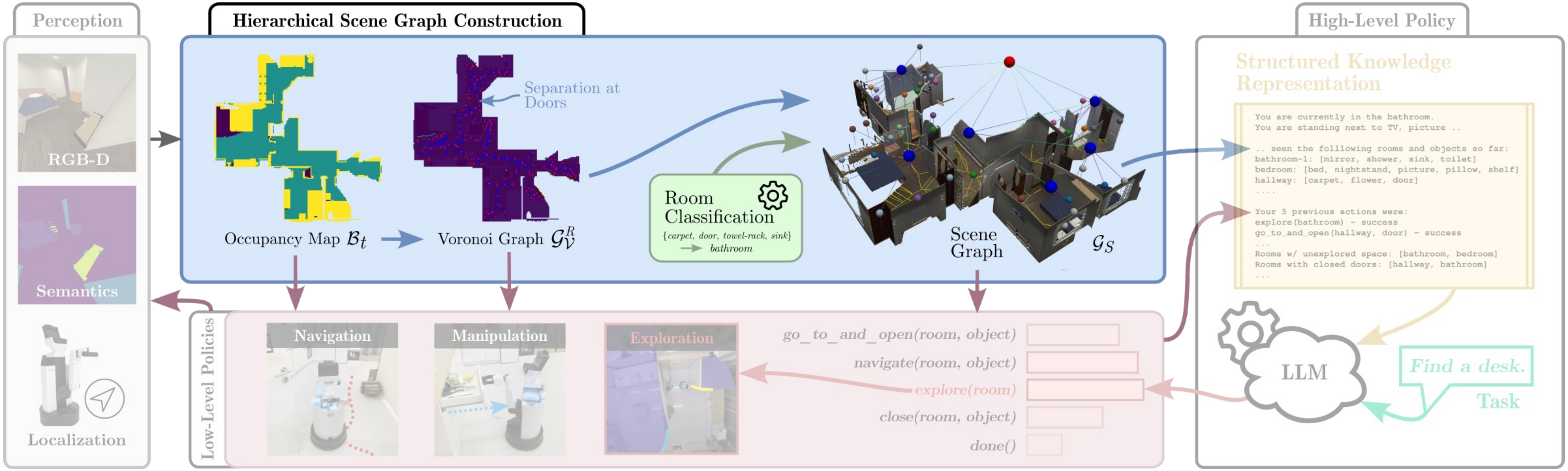


MoMa-LLM Approach



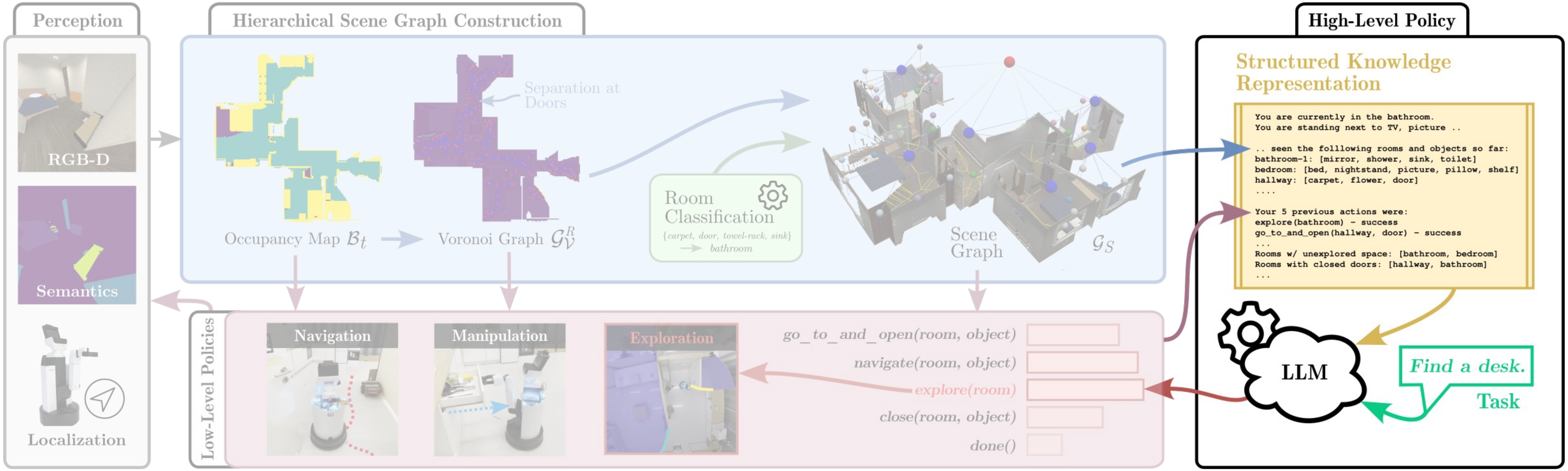
[Honerkamp *et al.*, RA-L 2024]

MoMa-LLM: Mapping and Structure



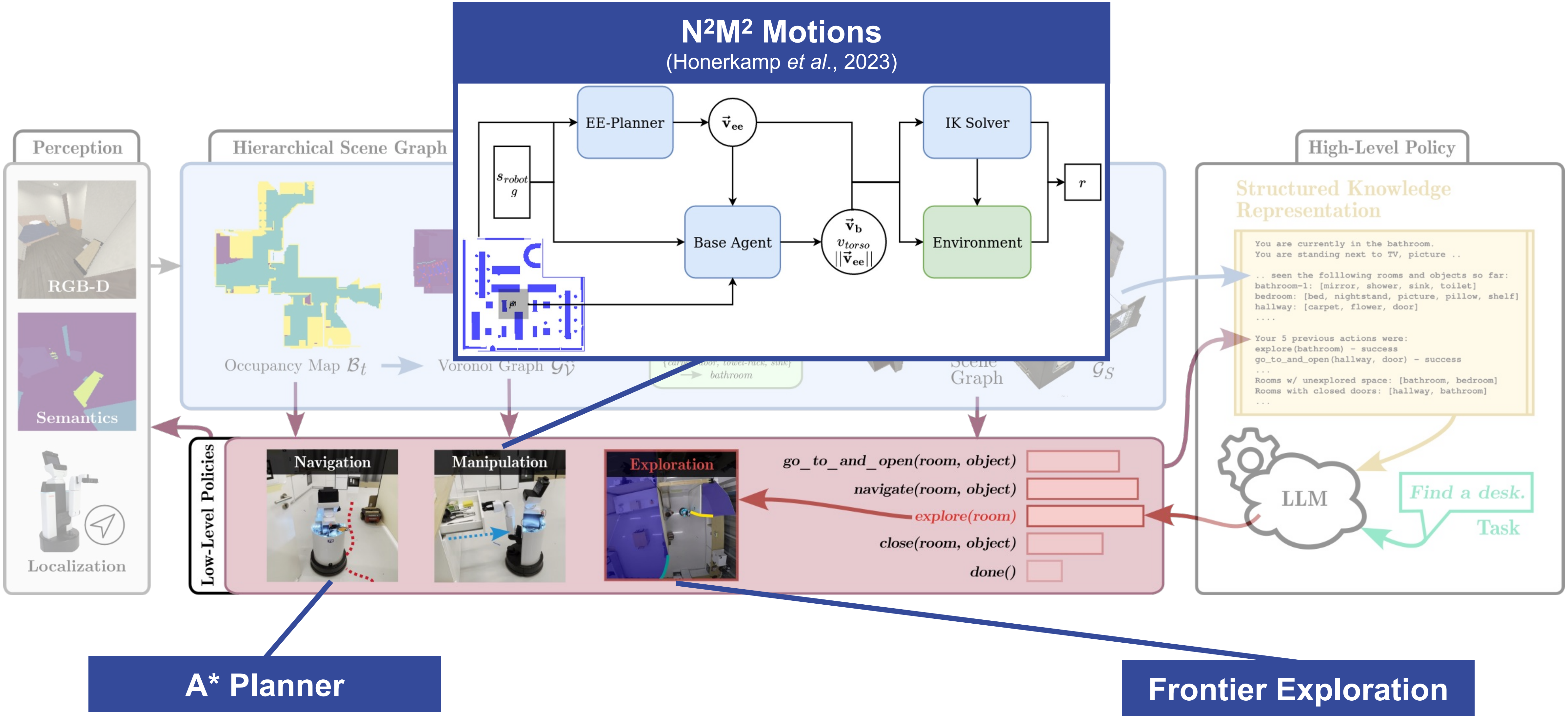
[Honerkamp *et al.*, RA-L 2024]

MoMa-LLM: Task Knowledge



[Honerkamp *et al.*, RA-L 2024]

MoMa-LLM: Planning with the LLM

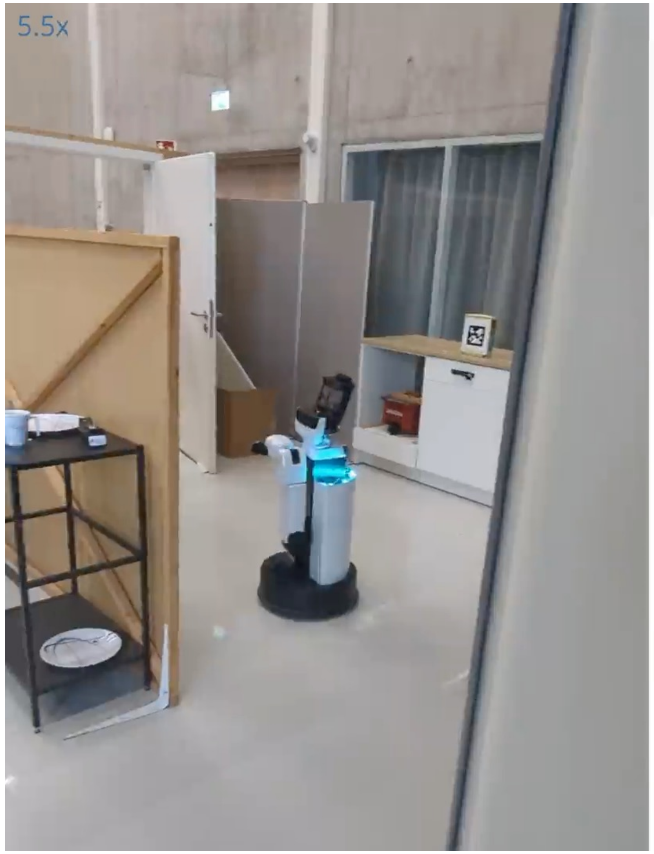


[Honerkamp *et al.*, RA-L 2024]

3x

Task: I am hungry. Find me something for breakfast.

MORE: Scaling Across Very Large Environments



Task: "Set the table for cereals."

[Mohammadi *et al.*, IROS 2025]

MORE: Mobile Manipulation Rearrangement Through Grounded Language Reasoning

Mohammad Mohammadi[1,2*], Daniel Honerkamp[1*], Martin Büchner[1*], Matteo Cassinelli[3*], Tim Welschehold[1], Fabien Despinoy[3], Igor Gilitschenski[2], and Abhinav Valada[1]

***Abstract*— Autonomous long-horizon mobile manipulation encompasses a multitude of challenges, including scene dynamics, unexplored areas, and error recovery. Recent works have leveraged foundation models for scene-level robotic reasoning and planning. However, the performance of these methods degrades when dealing with a large number of objects and large-scale environments. To address these limitations, we propose MORE, a novel approach for enhancing the capabilities of language models to solve zero-shot mobile manipulation planning for rearrangement tasks. MORE leverages scene graphs to represent environments, incorporates instance differentiation, and introduces an active filtering scheme that extracts task-relevant subgraphs of object and region instances. These steps yield a bounded planning problem, effectively mitigating hallucinations and improving reliability. Additionally, we introduce several enhancements that enable planning across both indoor and outdoor environments. We evaluate MORE on 81 diverse rearrangement tasks from the BEHAVIOR-1K benchmark, where it becomes the first approach to successfully solve a significant share of the benchmark, outperforming recent foundation model-based approaches. Furthermore, we demonstrate the capabilities of our approach in several complex real-world tasks, mimicking everyday activities. We make the code publicly available at** `https://more-model.cs.uni-freiburg.de`.

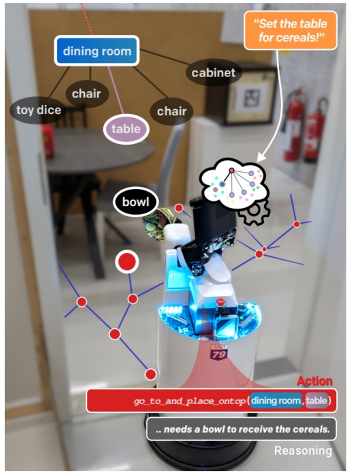


Fig. 1. We present MORE, an efficient model for the task of rearrangement through mobile manipulation. We utilize 3D scene graphs as a logical scene representation manifold that is filtered to obtain task-relevant subgraphs.

I. Introduction

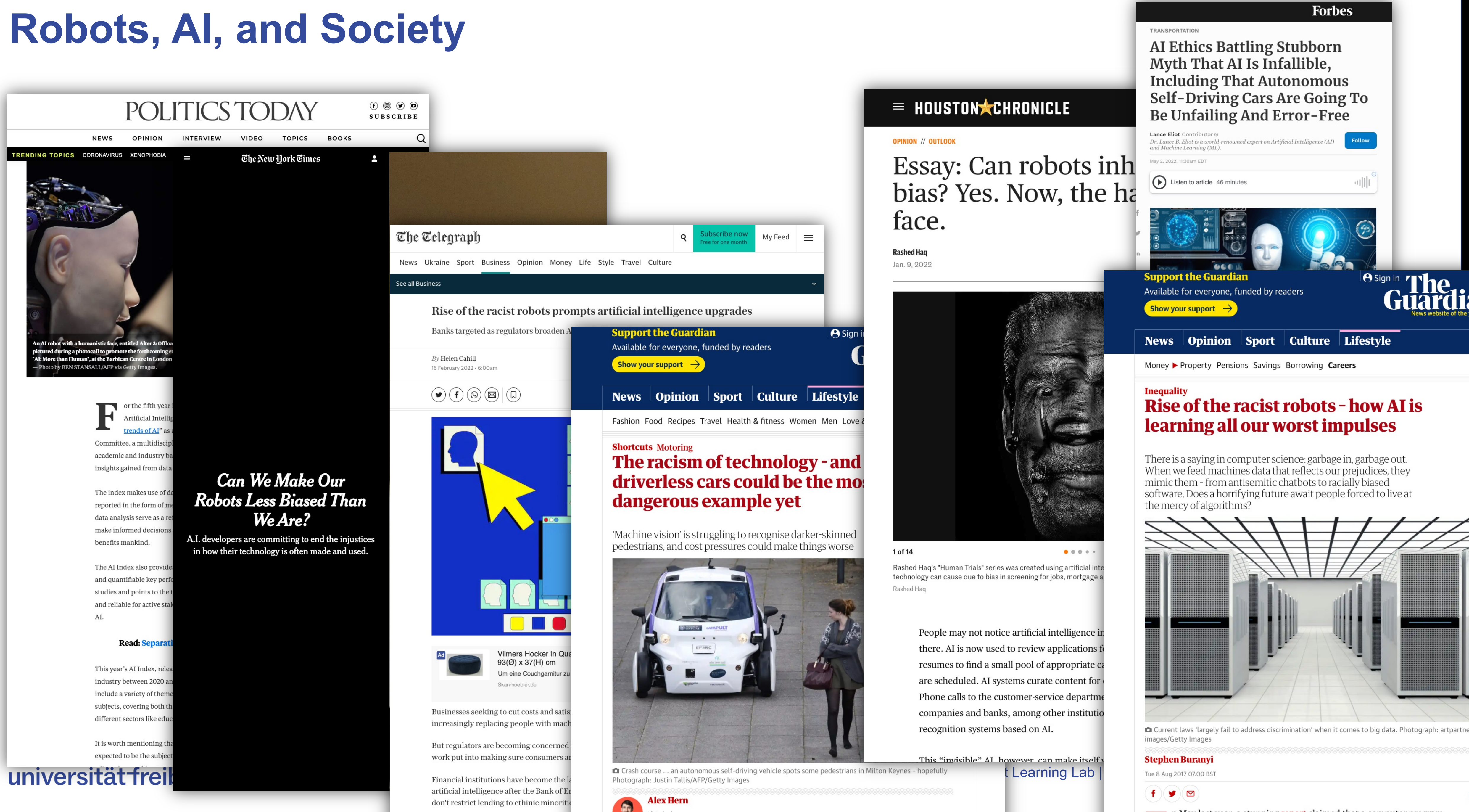
Recent studies have achieved notable advances in the completion of long-horizon robotic tasks in large-scale environments [1]–[4]. This progress was largely fueled by recent breakthroughs in scene comprehension and the integration of foundation models. At the same time, evaluations are still limited to known environments [1], [3], interactive search tasks [2], [5], or a series of hand-crafted tasks in specific real-world settings that lack reproducibility by the community [1], [3], [6]–[8].

Previous research conducted within single-room environments has demonstrated the capability to accomplish a broad spectrum of tasks [9], [10]. Nonetheless, in the context of mobile robotic manipulation, the scope of environments can expand to encompass entire apartments and outdoor areas, leading to an exponential increase in the number of objects and possible interactions, leading to exploding planning times [11] or hallucinations [1]. Furthermore, assumptions of a priori known scene layouts and object locations do not account for environment changes or flexible deployment. These partial scene observability and unexplored areas complicate planning by requiring reasoning about the unknown.

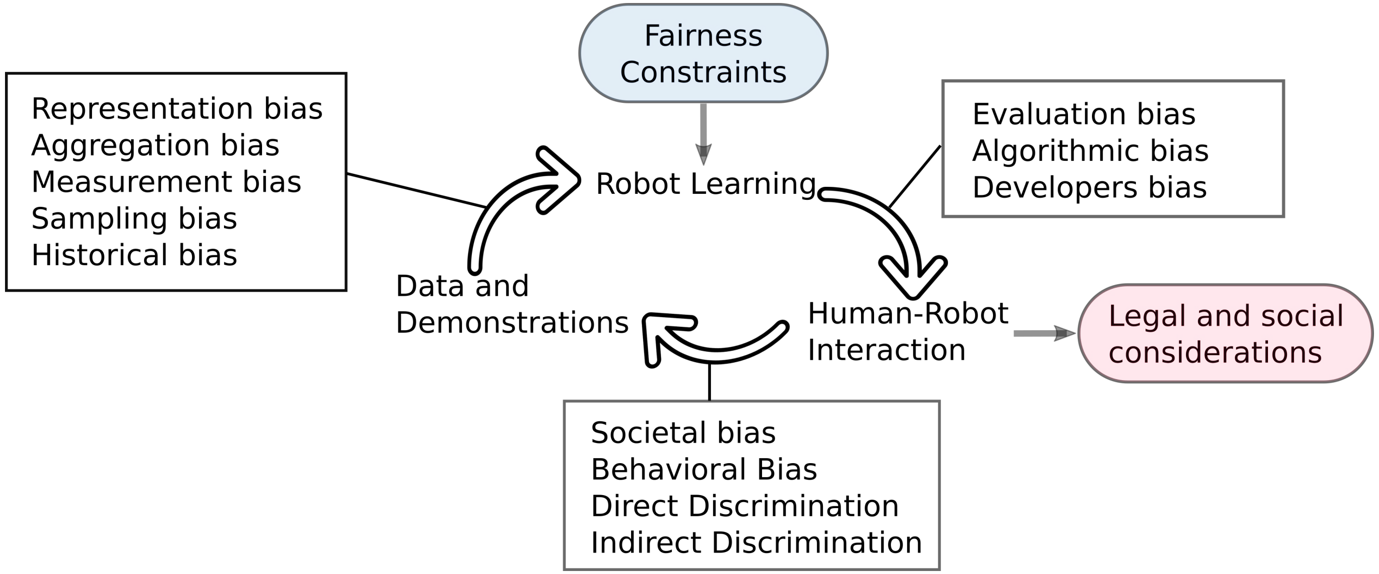
In this work, we focus on generalizing rearrangement tasks with a mobile manipulator in a zero-shot manner, tackling both large-scale and unexplored environments spanning across indoor and outdoor spaces. Building upon previous work in natural language-based interactive object search [2], we expand its application to encompass general everyday activities and introduce the first method capable of solving a substantial part of the BEHAVIOR-1K benchmark [12]. We refine the benchmark by providing fully defined language task descriptions and implementing computationally efficient evaluations, thereby facilitating reproducible assessments

* Equal contribution.
[1] Department of Computer Science, University of Freiburg, Germany.
[2] Department of Computer Science, University of Toronto, Canada.
[3] Toyota Motor Europe.
This work was funded by Toyota Motor Europe, an academic grant from NVIDIA, and the BrainLinks-BrainTools center of University of Freiburg.

Robots, AI, and Society



Responsible Robot Learning



[Londoño *et al.*, P-IEEE 2024]

Thank you for your attention!

https://rl.uni-freiburg.de