AAAI 2025 Tutorial T04
Time: 2025-02-25 8:30-12:30
Location: Room 118A

Foundation Models Meet Embodied Agents

Manling Li
Northwestern

Yunzhu Li
Columbia

Jiayuan Mao
MIT

Wenlong Huang
Stanford

AAAI 2025 Tutorial T04
Time: 2025-02-25 8:30-12:30
Location: Room 118A

Remaining Challenges

AAAI Tutorial: Foundation Models Meet Embodied Agents

Northwestern University

Stanford University

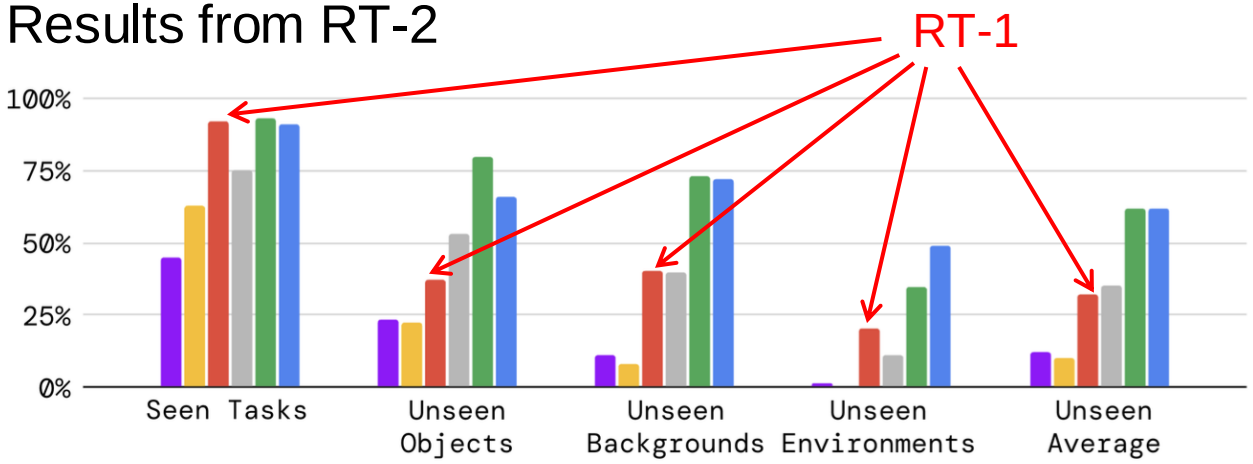
Evaluation of the Robotic Foundation Models

- Inconsistent results across papers

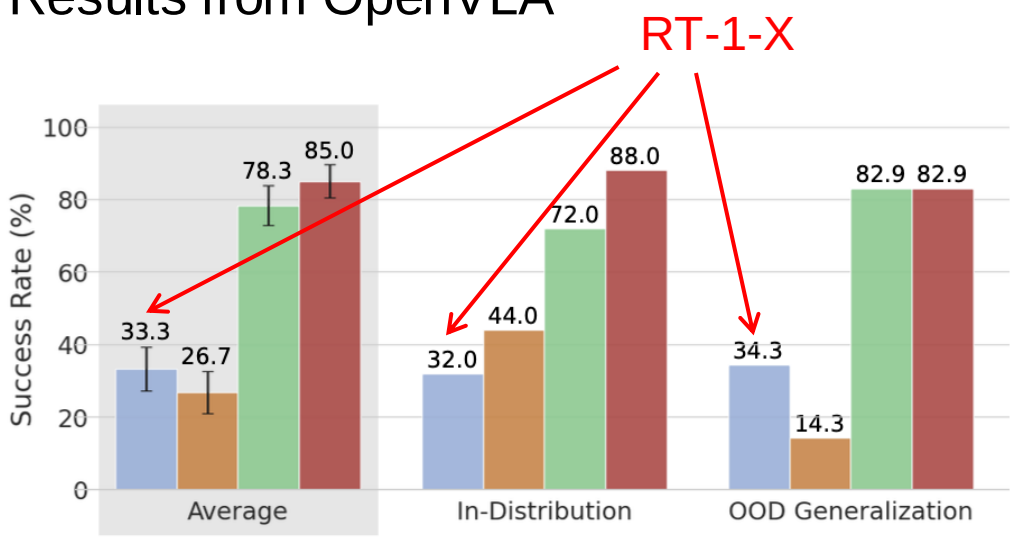
Results from RT-1

Model	Seen Tasks	Unseen Tasks	Distractors	Backgrounds
Gato (Reed et al., 2022)	65	52	43	35
BC-Z (Jang et al., 2021)	72	19	47	41
BC-Z XL	56	43	23	35
RT-1 (ours)	**97**	**76**	**83**	**59**

Results from RT-2



Results from OpenVLA



Evaluation of the Robotic Foundation Models

- Evaluation is primarily conducted in the real world
  - Real-world evaluation is costly and noisy
    - “We have large enough budget such that we can still make progress.”
  - Weak correlation between training loss and real-world success rate.
    - Training objectives vs task-specific metrics, training vs testing horizons

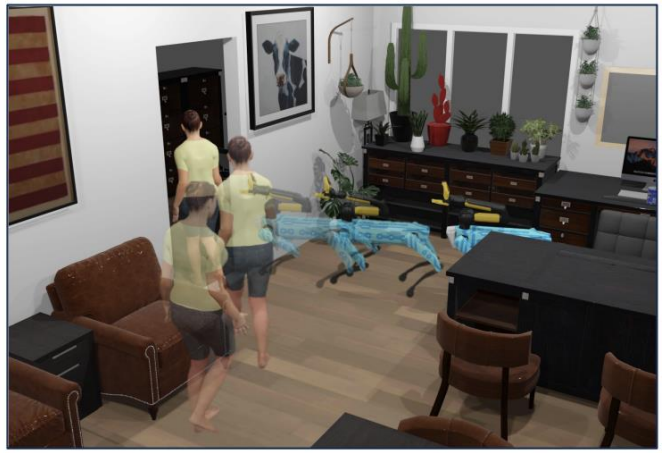
ALOHA 2

Evaluation of the Robotic Foundation Models

- What about evaluation in simulation?
  - Sim-to-real gap: rigid / deformable / cloth
  - Efficient asset generation
  - Digitalization of the real world
  - Procedural generation of realistic and diverse scenes
  - Correlation between sim and real

ImageNet in Embodied AI?

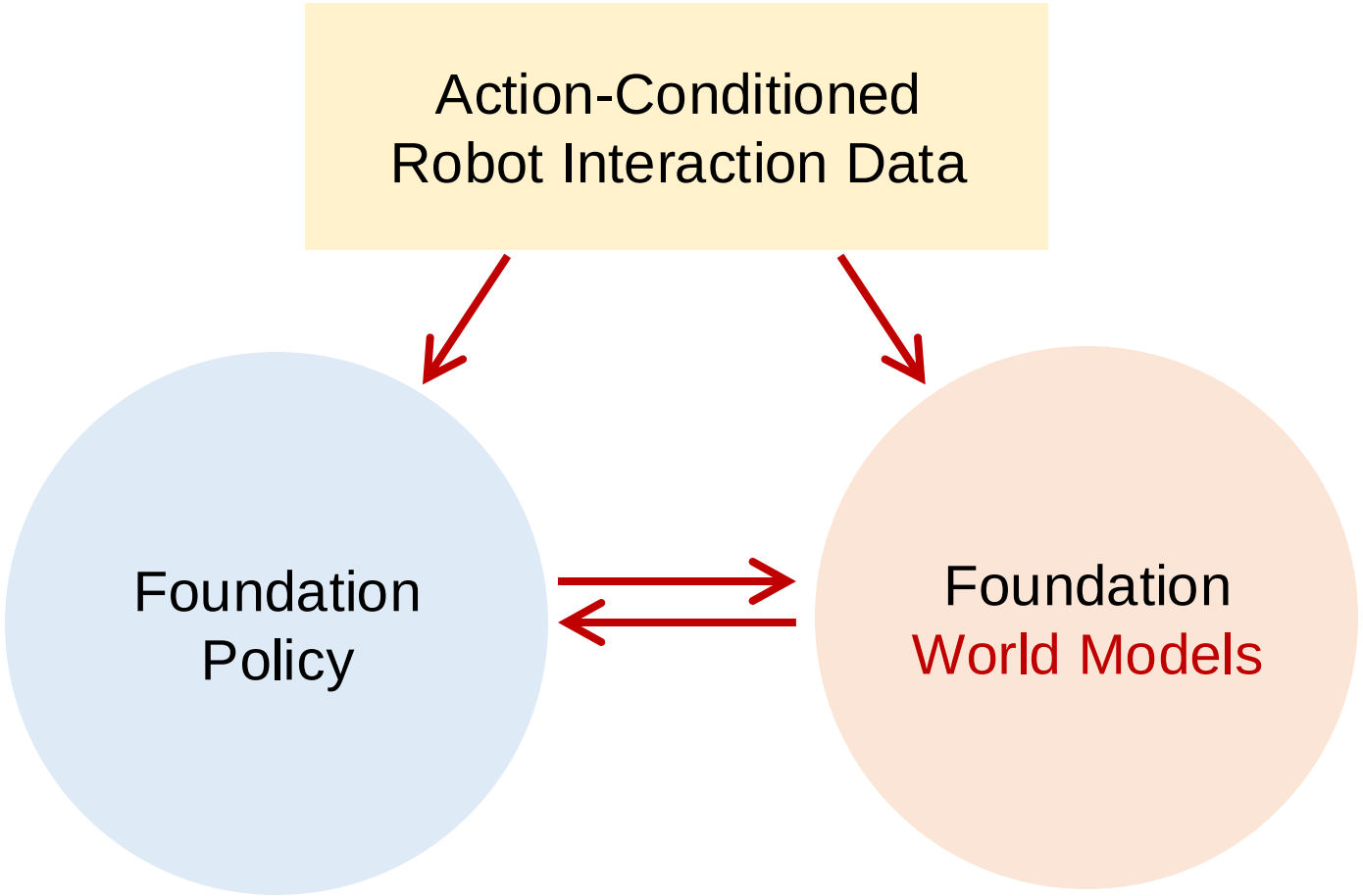
BEHAVIOR-1K



Habitat 3.0

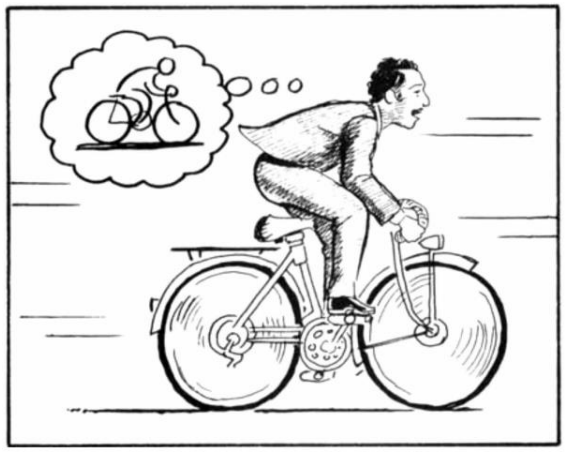
Foundation Policy → Foundation World Models

- My definition of world models: action-conditioned future prediction

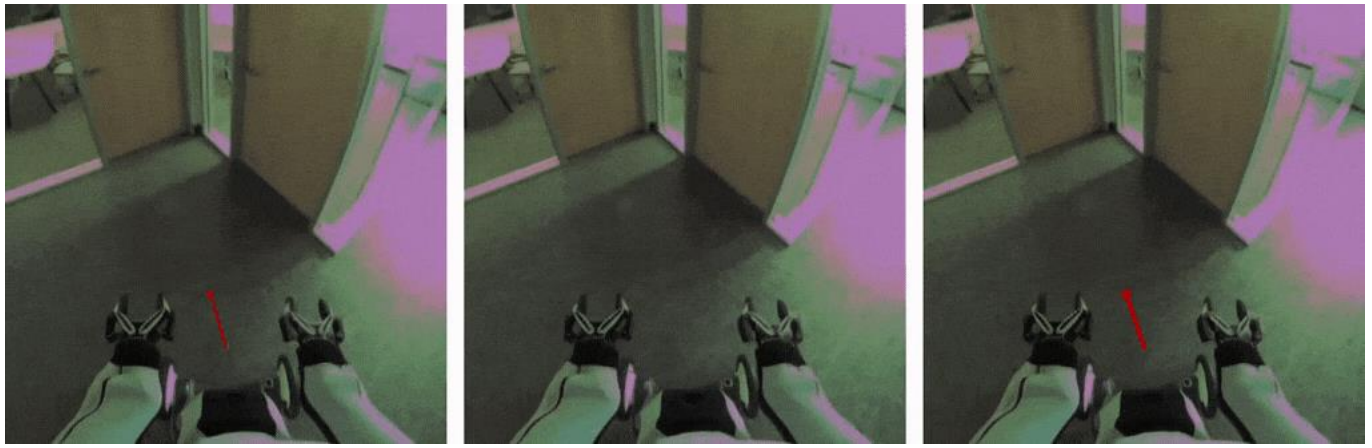


Foundation Policy → Foundation World Models

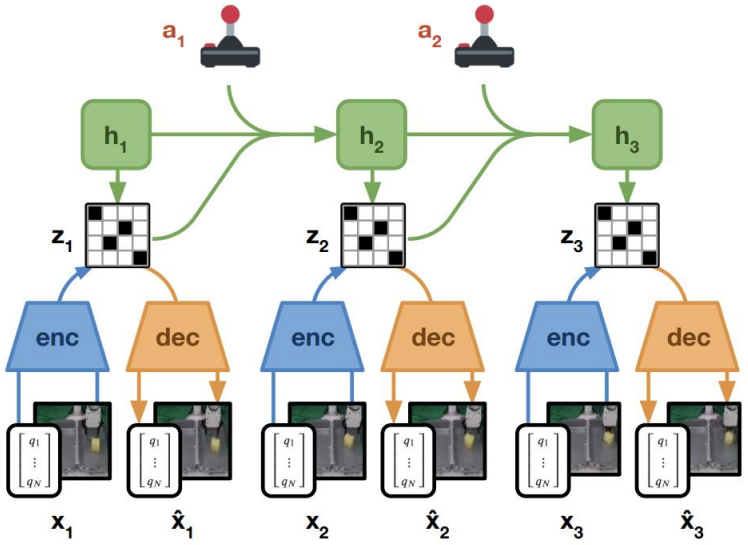
- My definition of world models: action-conditioned future prediction



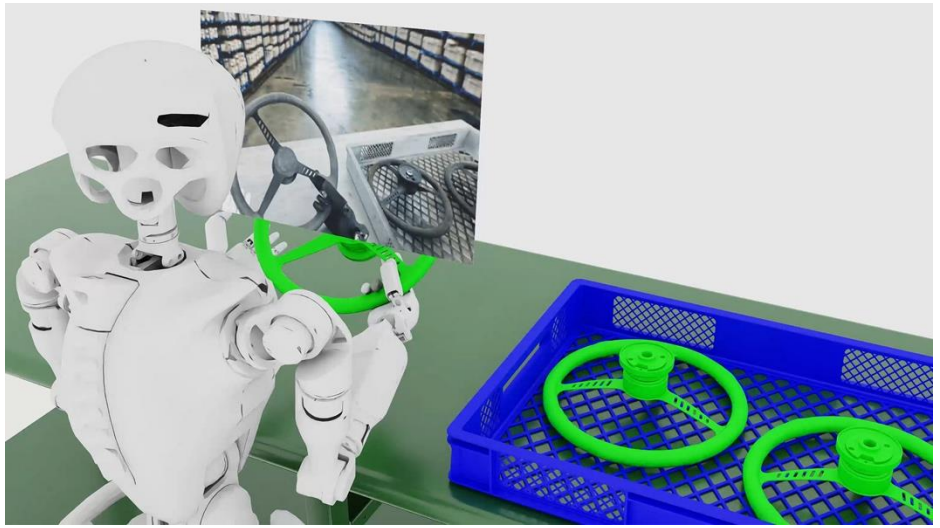
World Models



1X World Models



DayDreamer



Nvidia Cosmos - World Foundation Model

- 3D?
- Structural Prior?
- Learning + Physics?
- Corr. w/ Real World

Foundation Models for Embodied Agents

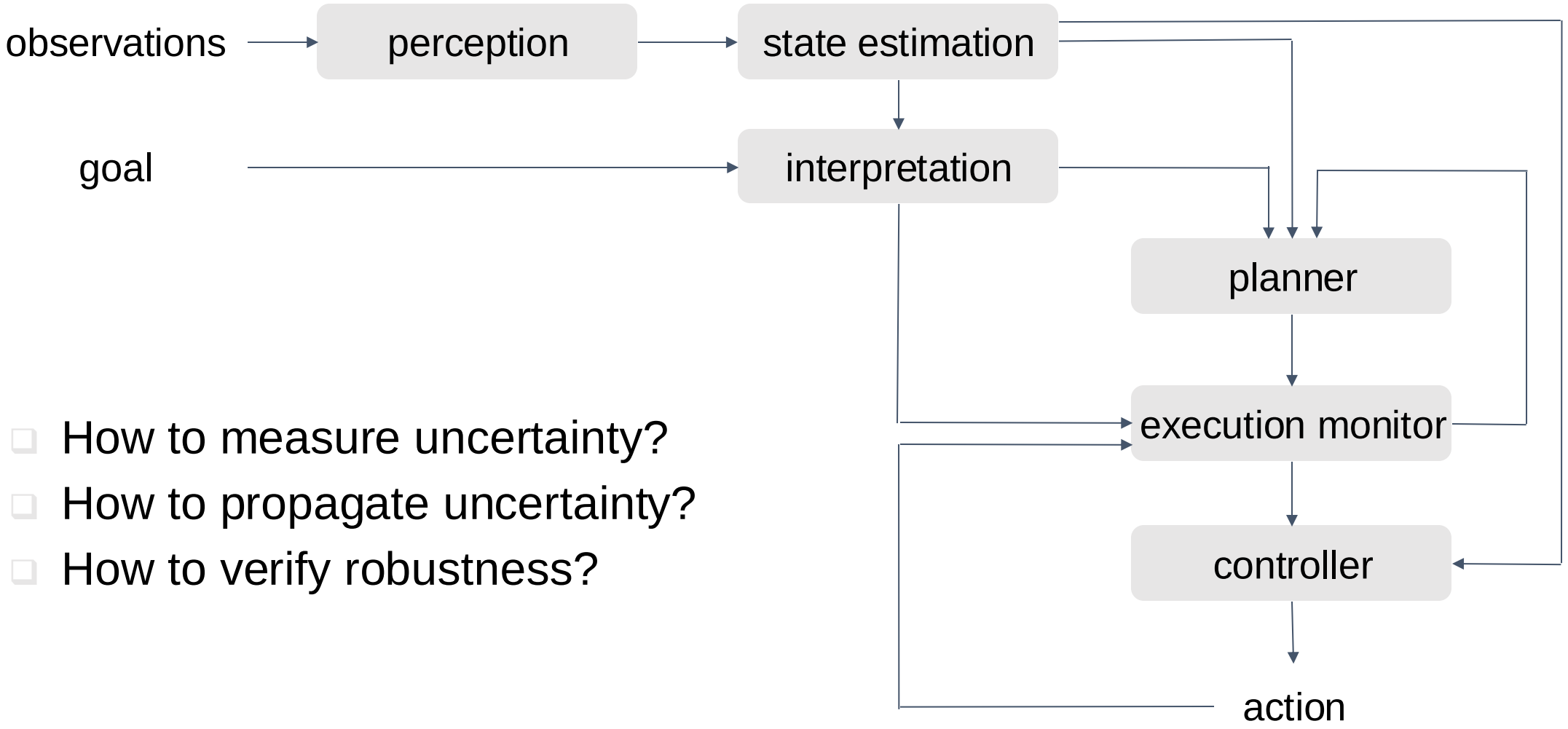
- Current foundation models are not tailored for embodied agents
  - LLM/VLM can fail in embodied-related tasks
  - Limited understanding of geometric / embodied / physical interactions
  - Reinforcement learning (RL) from human feedback → RL from Embodied Feedback

GPT

Segment Anything

DINOv2

Uncertainty Quantification & Propagation

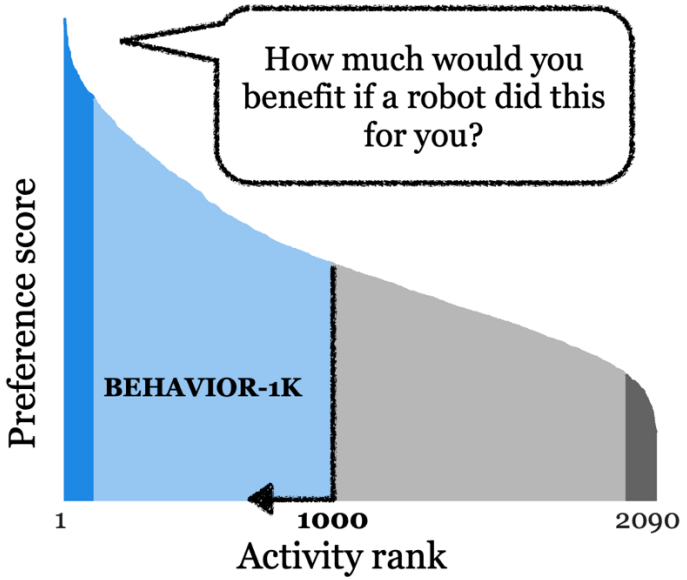


- How to measure uncertainty?
- How to propagate uncertainty?
- How to verify robustness?

Adaptation / Life-Long Learning

- Adapt to new scenarios
- Adapt to human preferences
- Self improve / life-long learning

Adapt to new scenarios

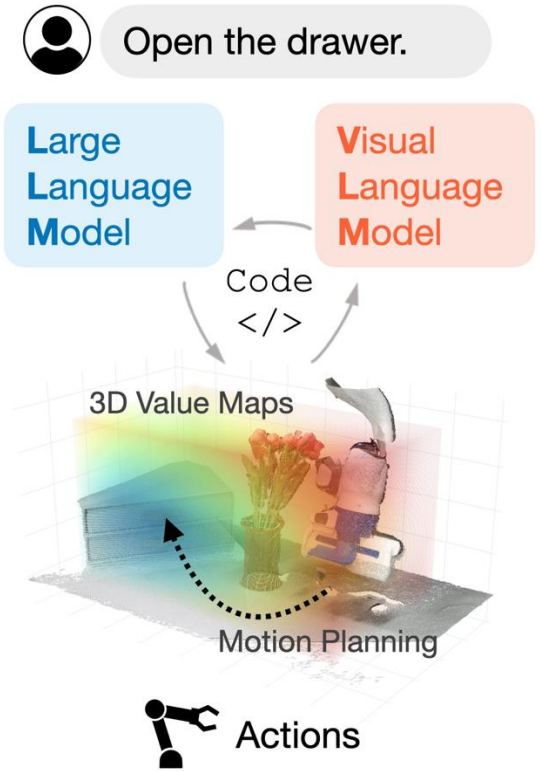


Adapt to human preferences

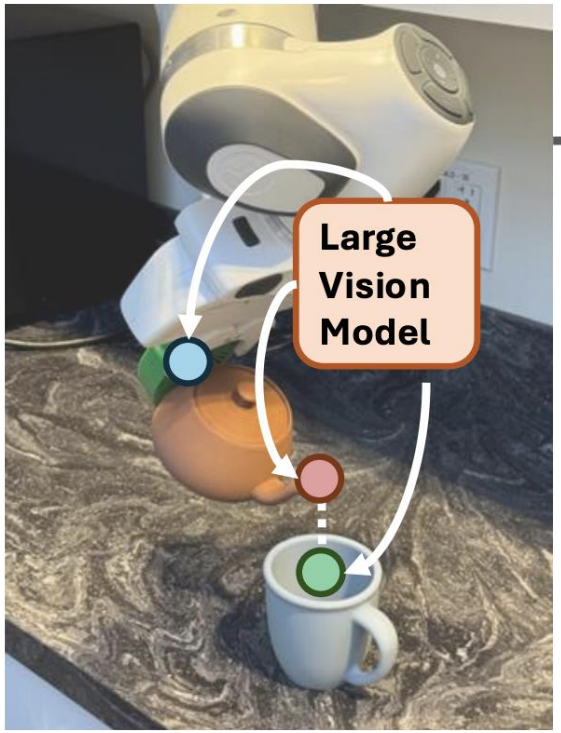
Improve through experience

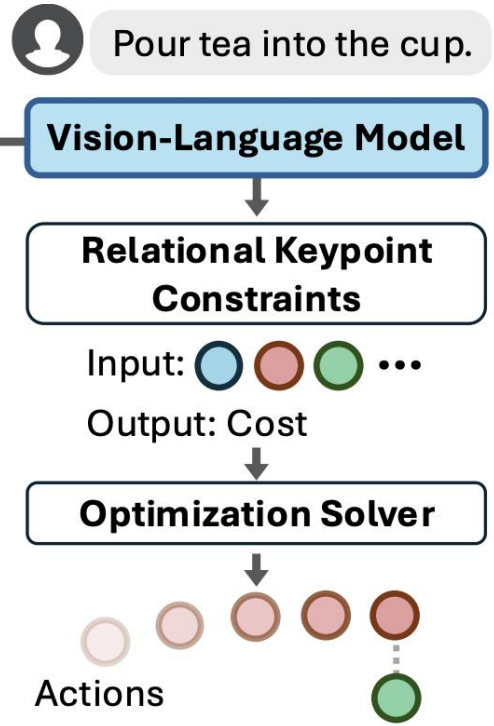
Practical Considerations of Foundation Models

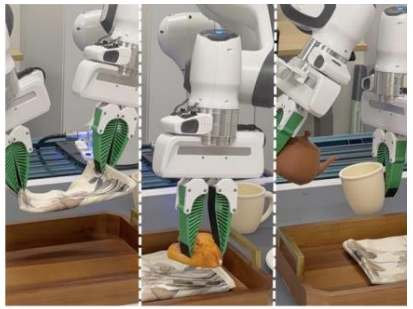
- Every robotics work is a system work
- System-level considerations:
  - delays / computing / modules talking to each other (high-level vs low-level)



VoxPoser







ReKep

Practical Considerations of Foundation Models

- Every robotics work is a system work
- System-level considerations: delays / computing / modules talking to each other

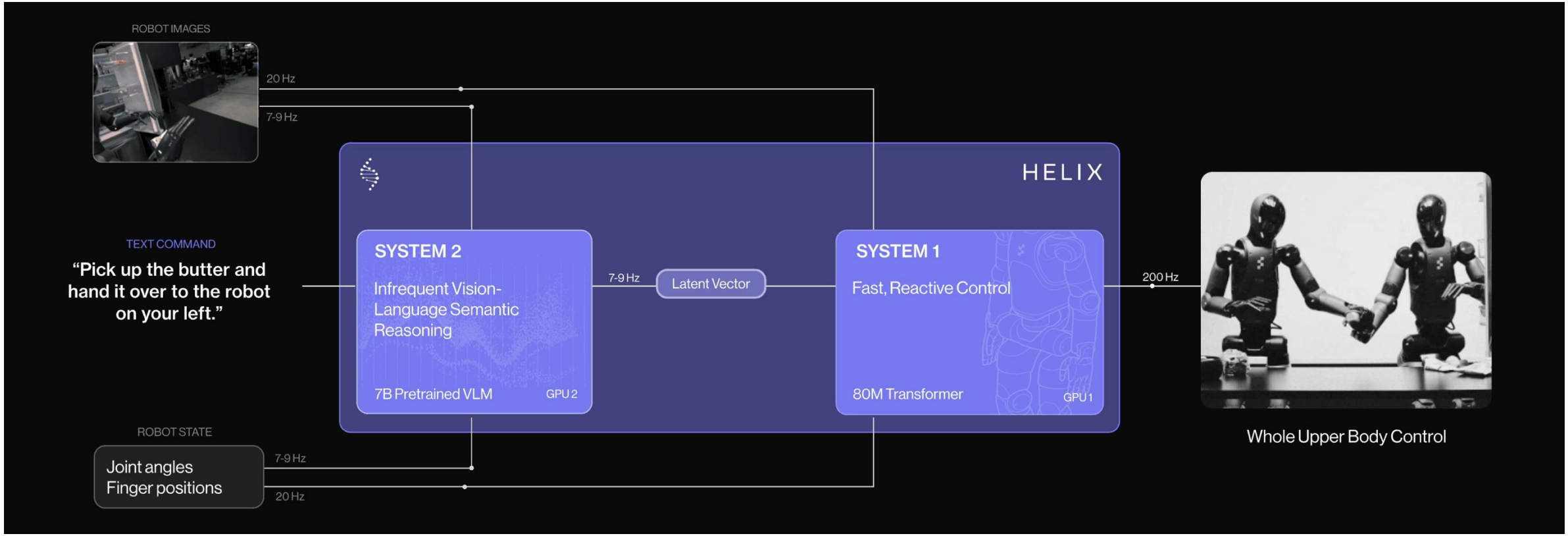
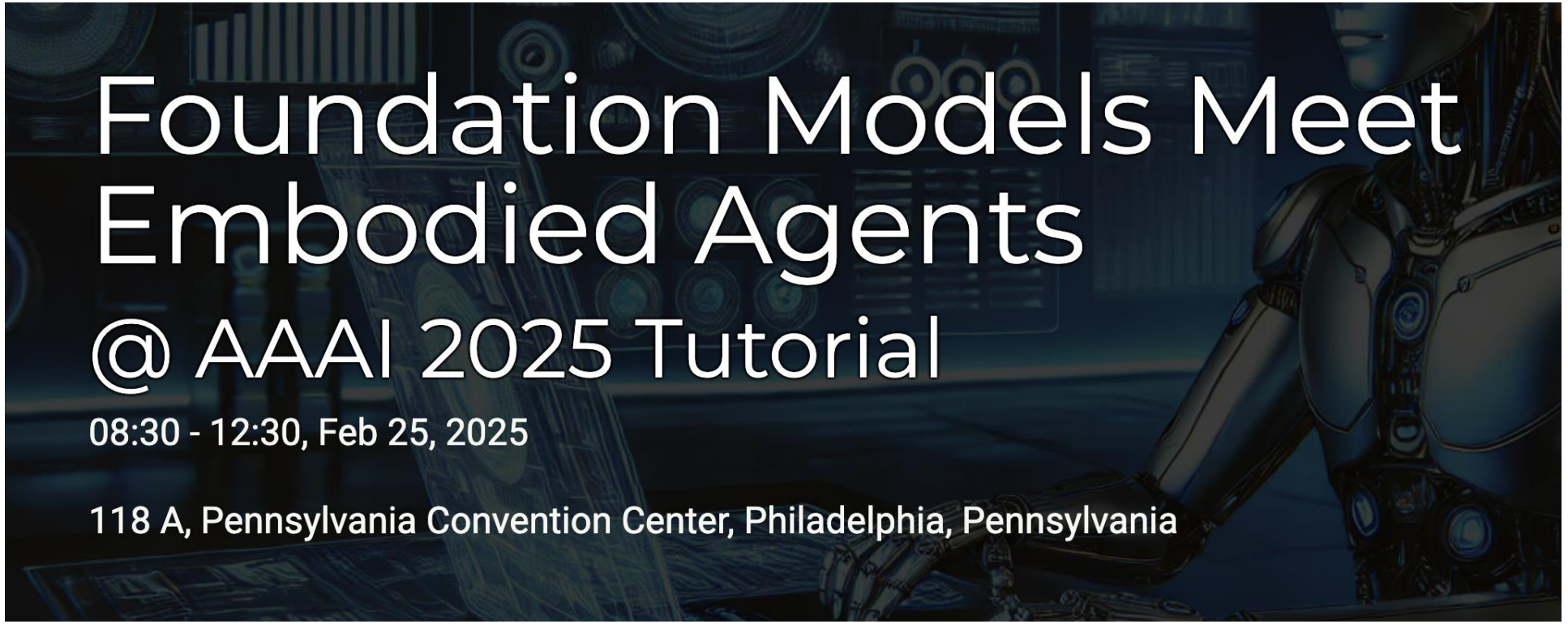


Figure AI: Helix

Virtual Agent & Physical Agent

- Bring together researchers working on both virtual & physical agents
- Approaching the problem from a more structured lens: MDP
- Key techniques, emerging opportunities, and notable challenges



AAAI 2025 Tutorial T04
Time: 2025-02-25 8:30-12:30
Location: Room 118A

Foundation Models Meet Embodied Agents

Manling Li
Northwestern

Yunzhu Li
Columbia

Jiayuan Mao
MIT

Wenlong Huang
Stanford