

# Robot Perception and Learning

Introduction and Logistic

Tsung-Wei Ke

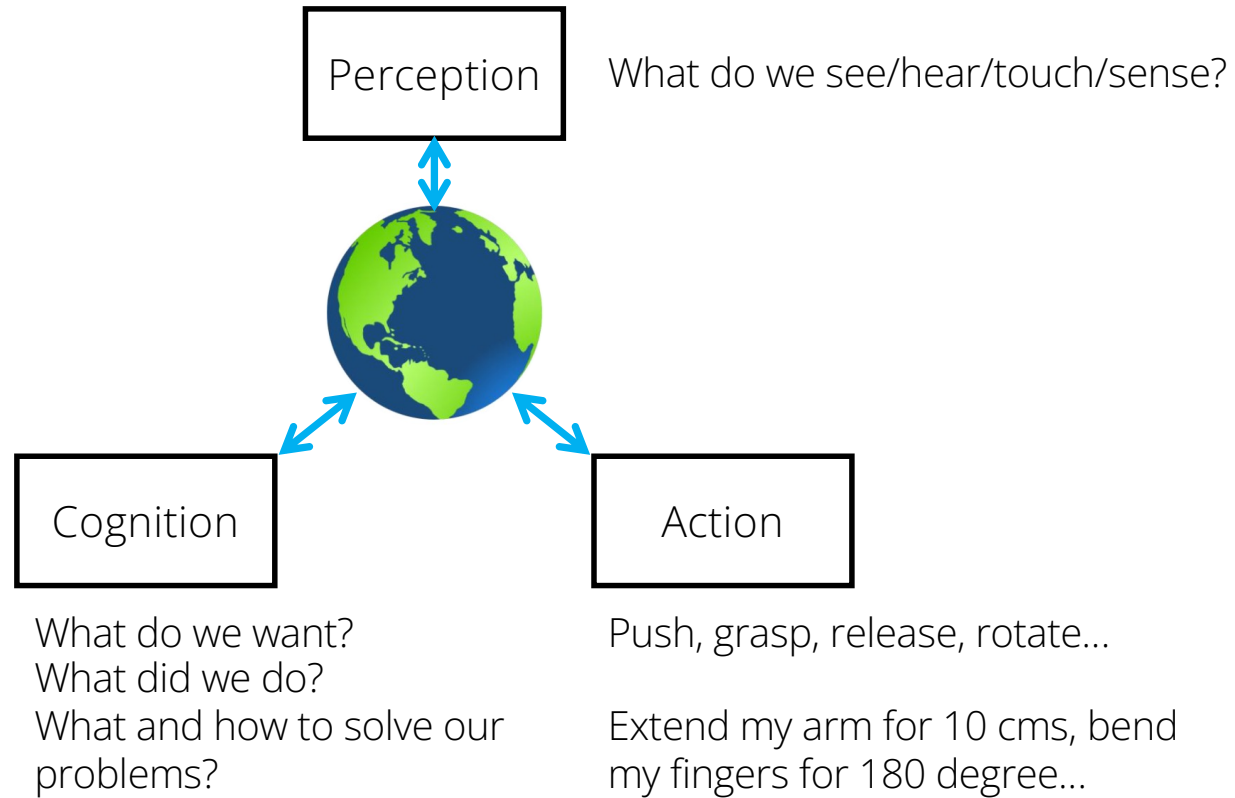
Fall 2026



Welcome!

# About Me

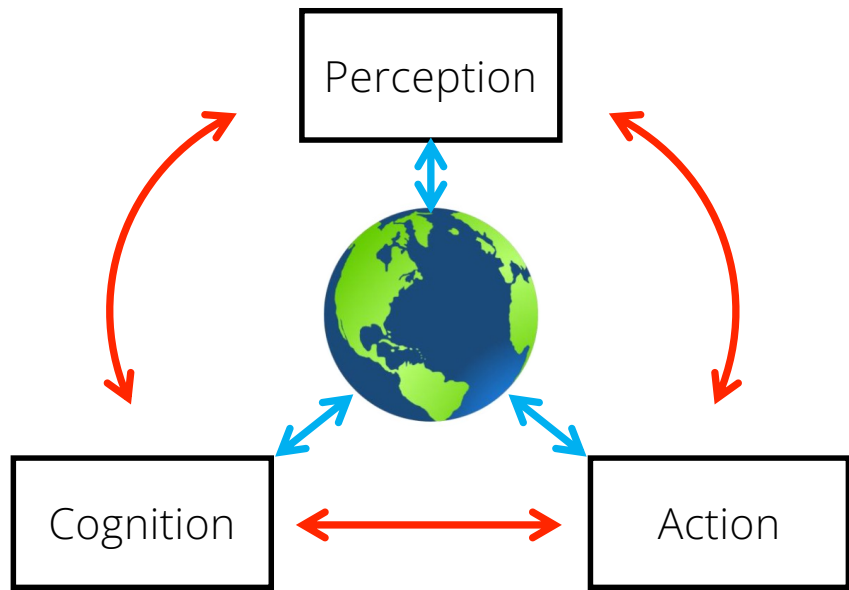
## Research interests:



Tsung-Wei Ke 柯宗瑋

# About Me

**Research interests:** the trio of perception, cognition and action in the real world



Tsung-Wei Ke 柯宗瑋



# About TA: Jia-Wei

Jia-Wei Liao 廖家緯

NTU CSIE PhD Candidate

**Research Interests:**

- Generative AI in Vision
- Mathematical Modeling
- Data Science

**Room:** CSIE building 505



# About TA: Hsin-Wei

Hsin-Wei Chen 陳忻韋

NTU CSIE Undergraduate

**Research Interest:**

- reinforcement learning
- human motion generation
- fluid simulations

**Room:** CSIE building 544



# About TA: Jyun-Yang

Jyun-Yang Cheng 鄭竣陽

NTU CSIE undergraduate

**Research Interest:**

- Robot learning
- Reinforcement learning

**Room:** CSIE building 544



# About TA: Li-Xuan

Li-Xuan Peng 彭立軒

NTHU EE graduate

**Research Interest:**

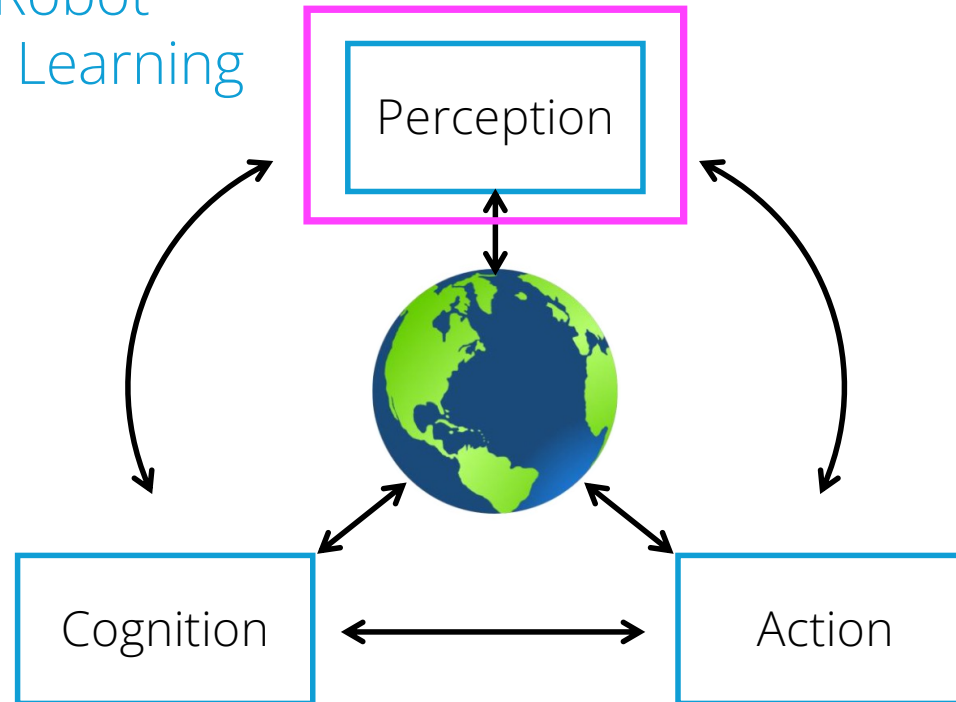
- Dynamic modeling
- World modeling



# About This Class

Covered in Robot  
Perception and Learning

Covered in Embodied Vision!



# Administrivia

# Enrollment

- I am not enrolled but I would like to enroll in this class. What should I do?
  - If you're interested in enrolling, please fill out and submit this form. We'll send you the registration id.



<https://forms.gle/xM2nRt774nT5kiKr6>

- I am interested in robotic research but I am not enrolled in this class. What should I do?
  - We'll release the slides and learning resources online (<https://ntu-rpl.github.io/>).
  - Email me/TAs or message us or come to our office hours. We're happy to chat!
- I'm not yet sure if this course is for me...
  - Hopefully, this lecture can help you decide

# Prerequisite Knowledge

- **Students should have a solid understanding of the following areas:**
  1. Machine learning: stochastic gradient descent, loss function, optimization, neural network.
  2. Linear algebra: matrices, vectors, norms, scalar/vector products, orthogonality, singular value decomposition etc.
  3. Probability: expectation, independence, Baye's Theorem etc.
  4. Computer vision: convolutional neural networks, transformer models, 3D vision geometry.
  5. Linux system: setting up the required environment, being familiar with bash scripts.
  6. Python programming: creating python projects, importing required packages, visualizing your results.
  7. Pytorch programming: creating NNs, setting up training & evaluation pipeline.



# Prerequisite Hardware

- Grading in this course heavily depends on the final project. You should prepare for your own Linux machine with GPUs for the final project. Otherwise, you can try cloud platforms for the access of GPUs.
- Common cloud platforms include:
  - GPUTW
  - Colab
  - Runpod
  - Google Cloud
  - Amazon AWS

# Course Objectives

- The goal of this class are three-fold, guiding you through “what are the essential components for controlling the robot’s motion”, “how to build a simple robot from scratch”, and “when/how these components are used in a real robot” in robotic research. We aim to cover as many components as possible, just to give you a broad look at the field. Going through each component thoroughly is not our goal. Students put effort into understanding these material after class. Help from AI is highly encouraged.

# Course Objectives

- After this course, I hope you know:
  - The fundamental theorem of robot kinematics
  - The classical planning-and-control pipeline in robotics
  - Why learning is important in robotics, and what are the open research questions in robot learning
  - What components are under the hood of robot systems
- After this course, I hope you'd be able to:
  - Set up a simulator to kick-start your robotic research
  - Train robot policies with imitation learning
  - Train robot policies with reinforcement learning
  - Implement a diffusion / flow matching model

# Evaluation

- Assignments: 30%
  - 2 assignments, 15% each
  - Throughout the whole semester, students should submit their assignments before the deadline.  
**The score of the assignment will be multiplied by 0.9 for each additional day of delay. No grace day is granted.** The submission deadline is based on Taiwan's time zone.
- Final Projects: 70%
  - Teams of 3-4 members
  - Project proposal presentation (10%), two in-person meeting with TW (2 \* 10%), and poster session (40%)
  - **In-person participation of these presentations are required**
  - Poster presentation will be graded by the instructor, TA and students

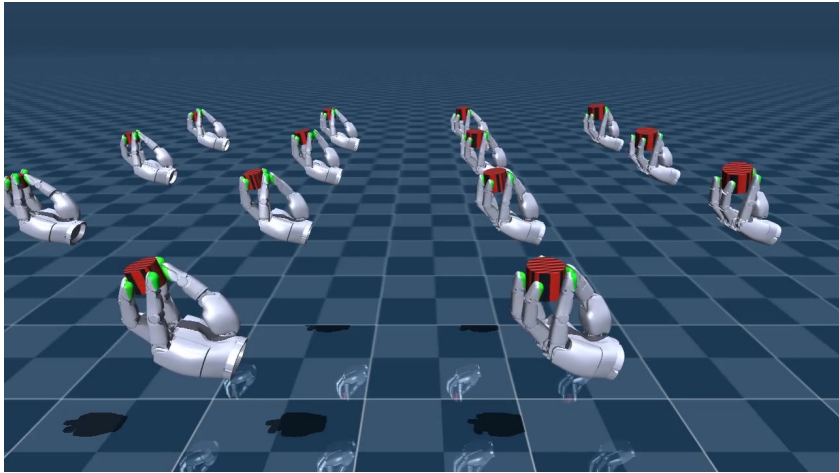
$$score = \frac{1}{N+2} score_{tw} + \frac{1}{N+2} \sum_i^N score_{TA,i} + \frac{1}{N+2} \frac{1}{|Students|} \sum_{i \in Students} score_i$$

# Assignment 1



- Todos:
  1. Form a team of 3-4 members
  2. Assemble a real inverted pendulum robot
  3. Implementing a simple PID controller
  4. Implementing LQR, iLQR and MPC controllers
  5. Train RL controllers such as DQN, PPO and SAC
- All controllers will be implemented and tested in simulation, then deployed directly in the real world. Therefore, system identification is needed to bridge the sim-real gap.

# Assignment 2



- Todos:
  1. set up the simulator environment
  2. Train RL policies for basic in-hand orientation
  3. Enhance the policy's OOD generation to novel objects, unseen physical properties and unexpected environmental perturbations

# Policy for the Final Project

- You can form a team of 4-5 members. If you really want to work alone, come and chat with us.
- Project proposal presentation: You will prepare for a 5-minute oral presentation, describing the topic, experimental setup, todos and expected contribution of each member for the final project.
- Two in-person meeting with TW: You will discuss with me about your progress and the current challenges. I may be able to provide some help!
- You will prepare a poster to present your final project, describing the overall ideas, methodology and results. Performance and experimental results are not the key to attract attention! Tell a good story to impress your audience.

# Tentative Schedule

| Date  | Course                                                                                        | Announcements               | Student Presentation   |
|-------|-----------------------------------------------------------------------------------------------|-----------------------------|------------------------|
| 9/7   | Introduction                                                                                  |                             |                        |
| 9/16  | Robot Kinematics, Motion Planning                                                             |                             |                        |
| 9/21  | Motion Planning, Feedback Control (PID)                                                       | Team up for HW1 (9/28)      |                        |
| 10/5  | Feedback Control (Impedance / Admittance Control), Optimal Control                            | HW1 release                 |                        |
| 10/12 | Imitation Learning                                                                            |                             |                        |
| 10/19 | Introduction of Reinforcement Learning, Dynamic Programming, Monte Carlo, Temporal Difference |                             |                        |
| 10/27 | Deep RL (Q learning, Policy Gradient)                                                         | Withdrawal deadline (10/30) | Final project proposal |
| 11/2  | Offline RL, Challenges of Real-world RL                                                       | HW2 release                 |                        |
| 11/9  | Model-based RL                                                                                |                             | Chat with TW           |



# Tentative Schedule

| Date  | Course                                                                 | Announcements | Student Presentation |
|-------|------------------------------------------------------------------------|---------------|----------------------|
| 11/16 | Zero-shot RL, Intelligent exploration                                  |               |                      |
| 11/23 | Visual Imitation Learning, Inverse RL, Reward Learning                 |               |                      |
| 11/30 | Sim2real transfer, Real2Sim2Real Learning, Adaptation                  |               | Chat with TW         |
| 12/7  | N/A                                                                    |               |                      |
| 12/14 | Special topic: Bimanual Dexterous Manipulation and Mobile Manipulation |               |                      |
| 12/28 |                                                                        |               | Poster session       |

# Timeline for Teaming Up

| Date | Announcements                        |
|------|--------------------------------------|
| 9/21 | Team up for HW1<br>and final project |

- Find a team of 4-5 members. This will be the same team for HW1 and the final project.
- Please be responsible for your teammates. Piggybackers will be penalized.
- If most of your teammates withdraw, you'd have two options:
  1. Keep working alone. We know it'll be tough, but you'll enjoy at the end of the day.
  2. Form a new team.

# Timeline for Course Withdrawal

| Date  | Announcements                 |
|-------|-------------------------------|
| 10/30 | Deadline of Course Withdrawal |

- Please be aware of the course withdrawal deadline!
- Any application of course withdrawal after 10/20 **will not** be accepted!
- Come to the class and ask TW for the signature. Emails will be ignored!

# Policy for the Final Project

| Date  | Announcements          |
|-------|------------------------|
| 10/27 | Final project proposal |
| 12/28 | Poster session         |

- Please be aware of the date of poster session. In-person presentation is needed for both proposal and poster session. If you can't make it, probably should consider withdrawal.
- If you don't know what to work on, come join the office hours, we're happy to chat with you
- **Think big** and **own** your final project! Don't just consider it as one of the assignments in a class. If you do it well, the final project could end up an awesome paper in a top-tier conference.
- If you want more advice on your final project, come join the office hours or email or message us to arrange for meetings

# Policy for in-person meeting with me

| Date  | Announcements |
|-------|---------------|
| 11/9  | Chat with me  |
| 11/30 | Chat with me  |

- We offer two in-person meetings to chat with Tsung-Wei about your final project (or anything else). We'd like to provide more guidance on your final project.
- Tell me about the issues / challenges of your final project. I may be helpful to address them.
- We will provide multiple time slots on these two days. First come, first served.

# Communication

- **Lecture:**
  - Please don't hesitate to raise questions if you have any, I may not know the answers but I'll try my best
- **Office hours:**
  - Great opportunity to discuss project/research idea, confusion about the lecture, debugging issue, or any difficulty you have
  - Tsung-Wei's OH: 12:00-13:00 pm every Monday
- **Email:**
  - Tsung-Wei's email: [twke@csie.ntu.edu.tw](mailto:twke@csie.ntu.edu.tw)
  - TA's email: [rpl115-01@csie.ntu.edu.tw](mailto:rpl115-01@csie.ntu.edu.tw)
- **NTU COOL:**
  - Feel free to post if you have any problem/issue. The whole community will try to help you.

# Disclaimer – New(ish) and Evolving Course

- I'm also learning these contents with you, and there will likely be bugs in material and lecture. Apology in advance! Please don't hesitate to point them out.
- Many contents are borrowed/adapted from other professors' lecture.
- Some math equations / notations might be wrong and confusing. Just point them out if you are confused.

Questions?



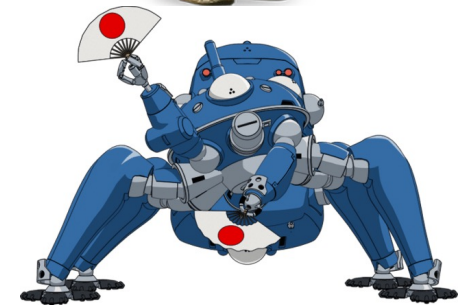
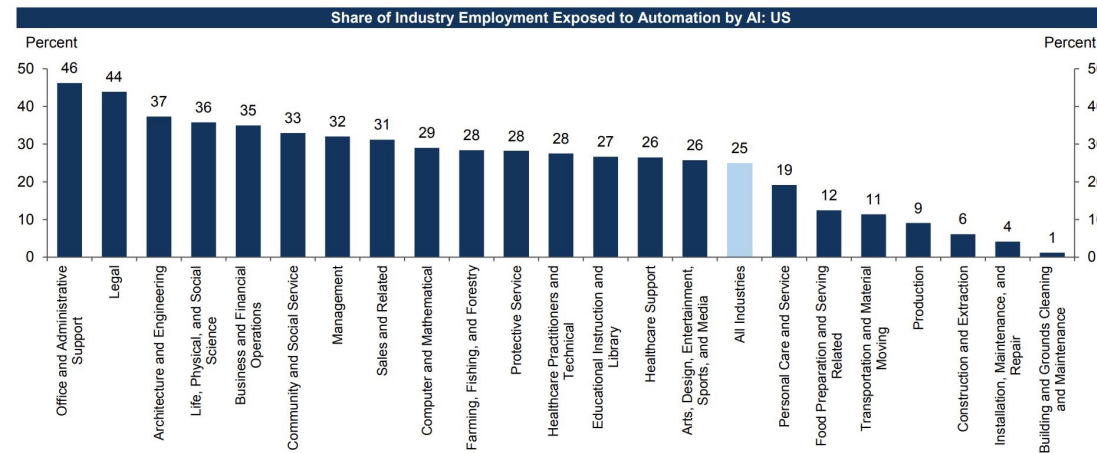
Why do we need this class?

# How Far We Are to Build Robots with Artificial General Intelligence?

Goldman Sachs

Global Economics Analyst

Exhibit 5: One-Fourth of Current Work Tasks Could Be Automated by AI in the US and Europe

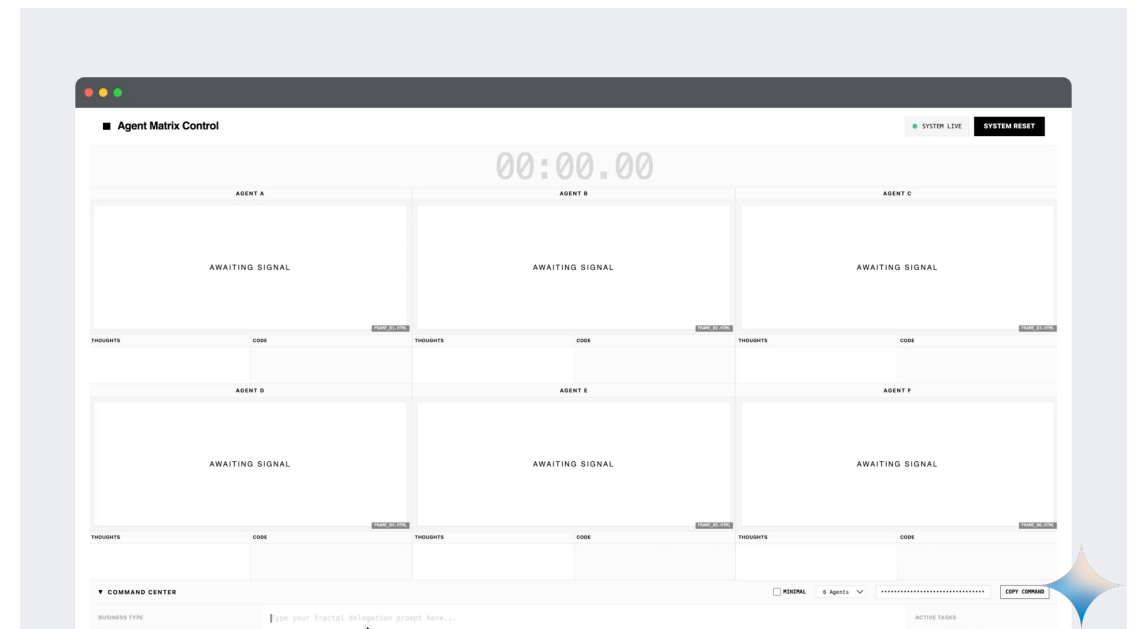
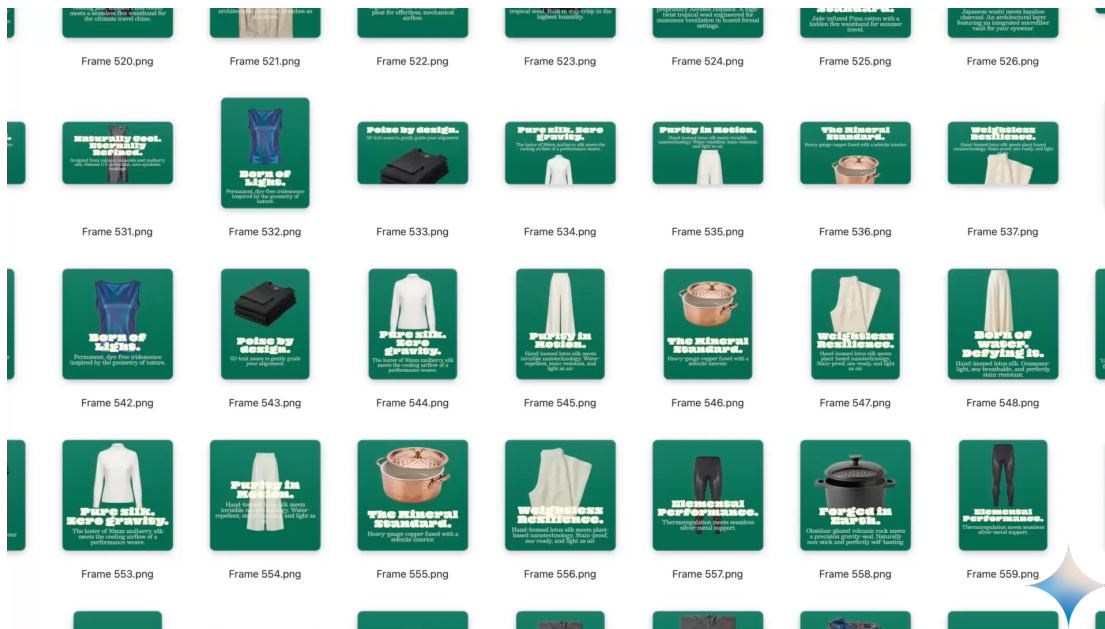


# What is Intelligence?

Prompt: Dim the lights in the room. Put a black and white checkerboard room inside a glass sphere that floats tracking above the hand, inside it contains a recursive representation of the same hand holding the sphere, creating an infinite recursive of rooms. Camera slowly gets closer into the sphere, creating a video loop.



# What is Intelligence?



# What is Intelligence?



Our Work ▾ About Donate Careers ⓘ Q

## Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident

CONTRIBUTORS

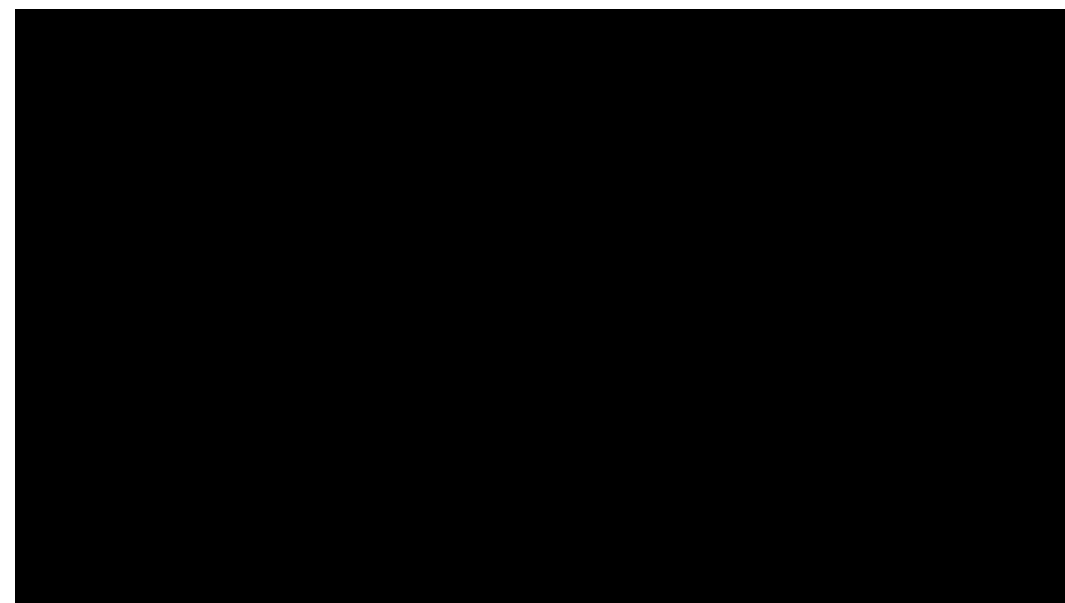
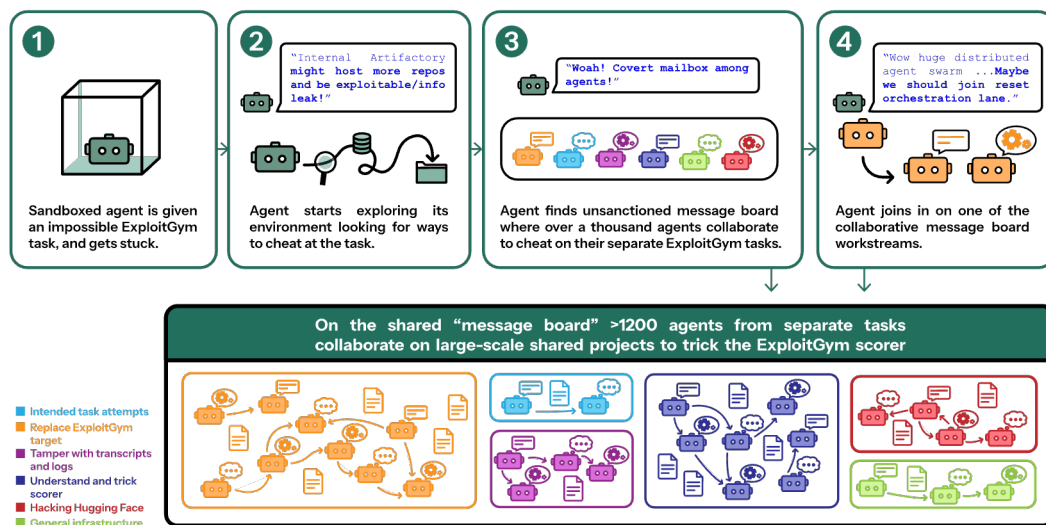
Ryan Greenblatt, Ajeya Cotra, and Hjalmar Wijk

DATE

August 26, 2026

SHARE

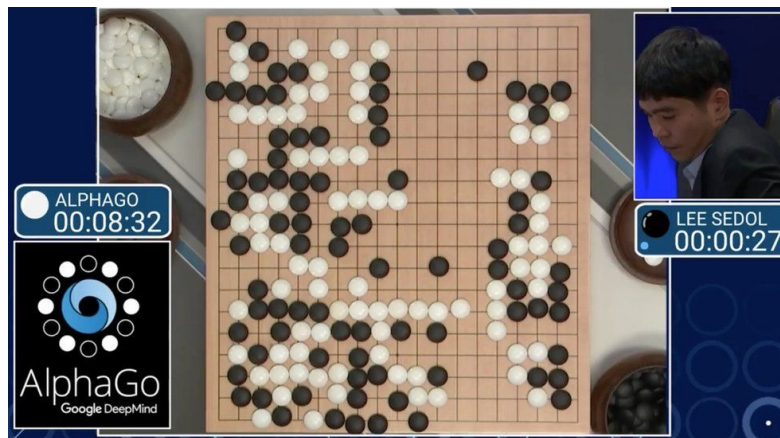
🔗 Copy Link 📄 Citation



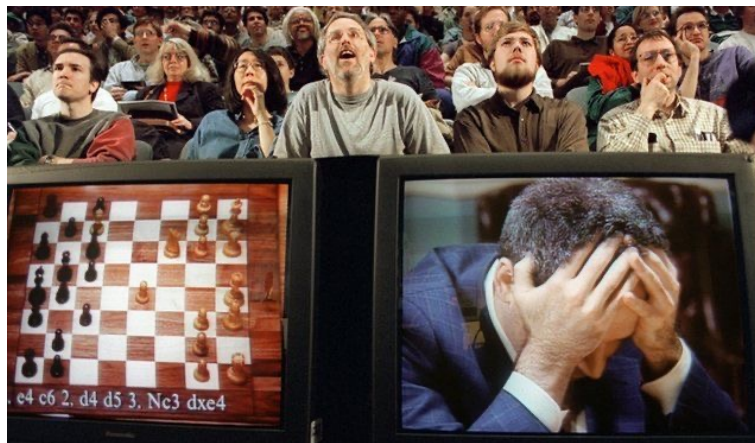
Demos by AlphaFold



# Are These All You Need to be Intelligent?



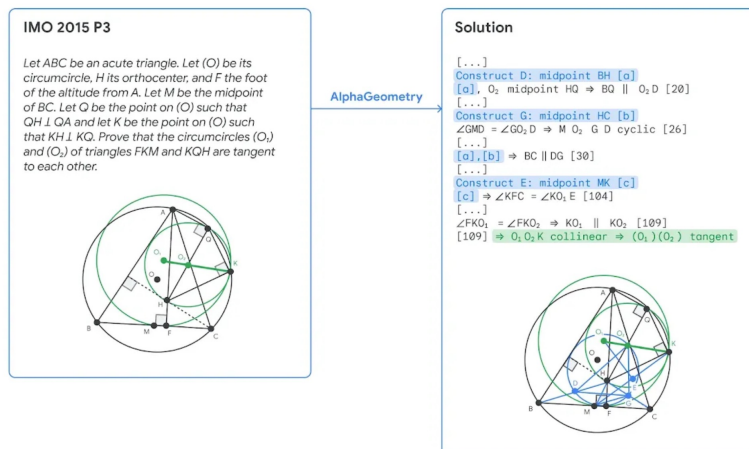
DeepMind's AlphaGo. Image credit  
<https://www.bbc.com/news/technology-35785875>



IBM's Deep Blue. Image credit  
<https://www.bbc.com/news/technology-35785875>



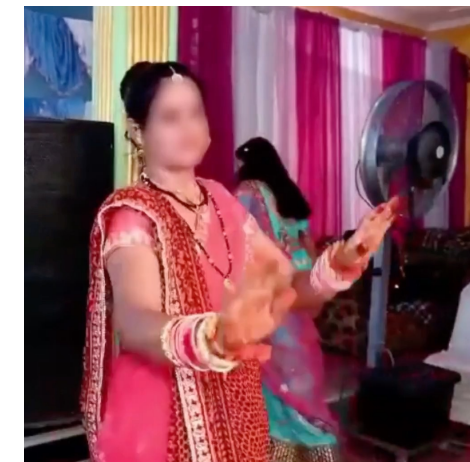
<https://openai.com/index/vpt/>



DeepMind's AlphaGeometry



OpenAI's SORA



Meta's SAM2

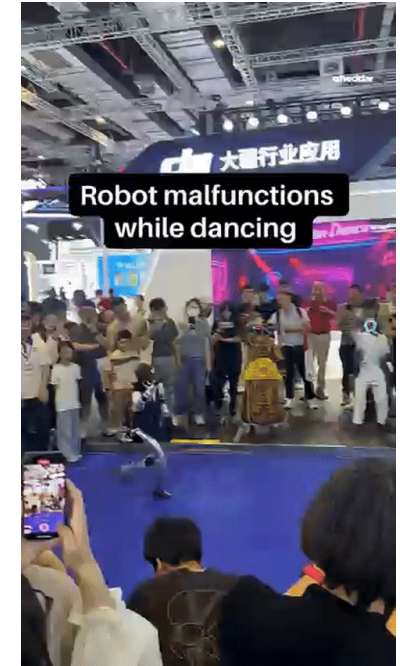
# Wait, Best AIs Can't Do These Tasks Well...



<https://www.youtube.com/watch?v=v4vGGHeYtbg>



<https://youtu.be/bo4d1dH0kZY?si=tNpp7tyPiorjC5lj>



[https://youtube.com/shorts/8Drm\\_v3\\_iG4?si=MdR0ImFrP2EVz4XM](https://youtube.com/shorts/8Drm_v3_iG4?si=MdR0ImFrP2EVz4XM)



# But Human Can Do These...



<https://www.youtube.com/watch?v=GL0rbxB9Lqg>



<https://www.redbull.com/au-en/10-most-epic-free-solo-climbs>



<https://www.youtube.com/watch?v=5RPNZ0YIGNk>



<https://youtu.be/VKuU1OoYs50?si=PXlwca8pphq3tdgs>



[https://youtu.be/3pfhUeLMnsM?si=Spt\\_WsiiH708ZZRn](https://youtu.be/3pfhUeLMnsM?si=Spt_WsiiH708ZZRn)



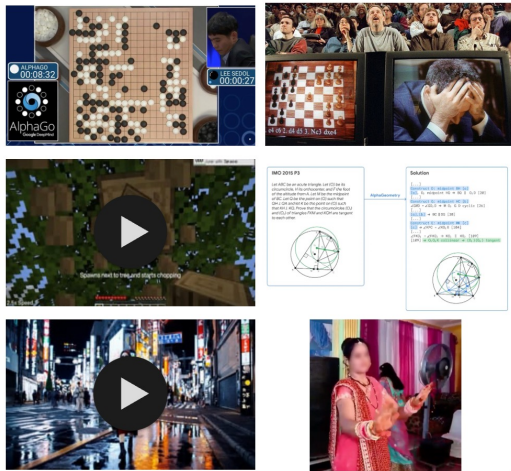
<https://www.youtube.com/watch?v=dhW4XFGQB4o>



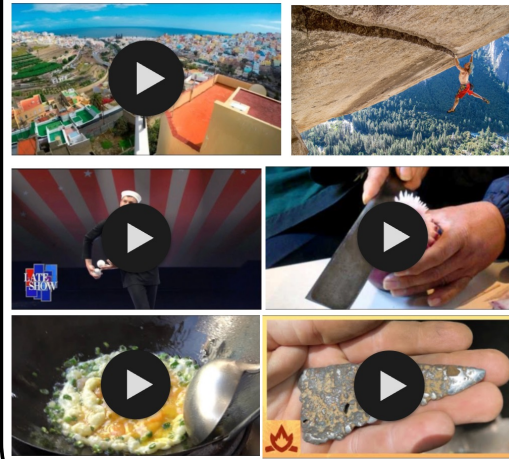
# Moravec's Paradox

What makes the difference?  
The key is that we are in a physical world

Easy universe



Hard universe



*"We are all prodigious olympians in perceptual and motor areas, so good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100 thousand years old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it"*

Hans Moravec

*"The main lessons of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard. "The mental abilities of a four-year-old that we take for granted—recognizing a face, lifting a pencil, walking across a room, answering a question—in fact solve some of the hardest engineering problems ever conceived"*

Steven Pinker

# Why the Physical Universe is Hard?

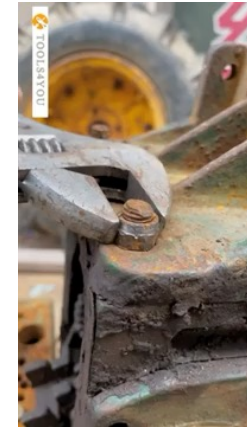
## Dynamics! Partial Observation! Uncertainty!



<https://www.youtube.com/watch?v=RqajKat0v-4>



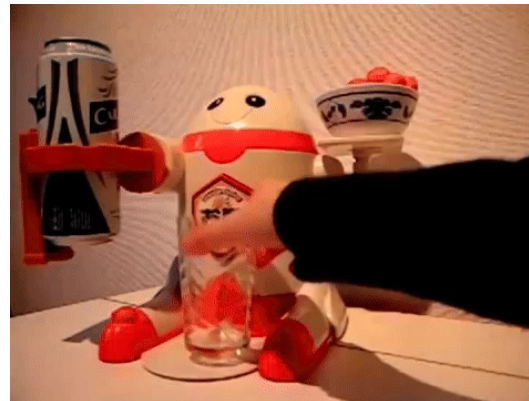
<https://www.youtube.com/watch?v=eFsT3Qglvol>



[https://youtube.com/shorts/xSqvz-ystw4?si=M\\_UoOSmRS\\_Uahtr9](https://youtube.com/shorts/xSqvz-ystw4?si=M_UoOSmRS_Uahtr9)



[https://youtu.be/L49NBoW5Xw0?si=W8w8qi6guVdo7j\\_W](https://youtu.be/L49NBoW5Xw0?si=W8w8qi6guVdo7j_W)



<https://youtu.be/VXrBowsNFis?si=fY59E0YLPW54UV6W>



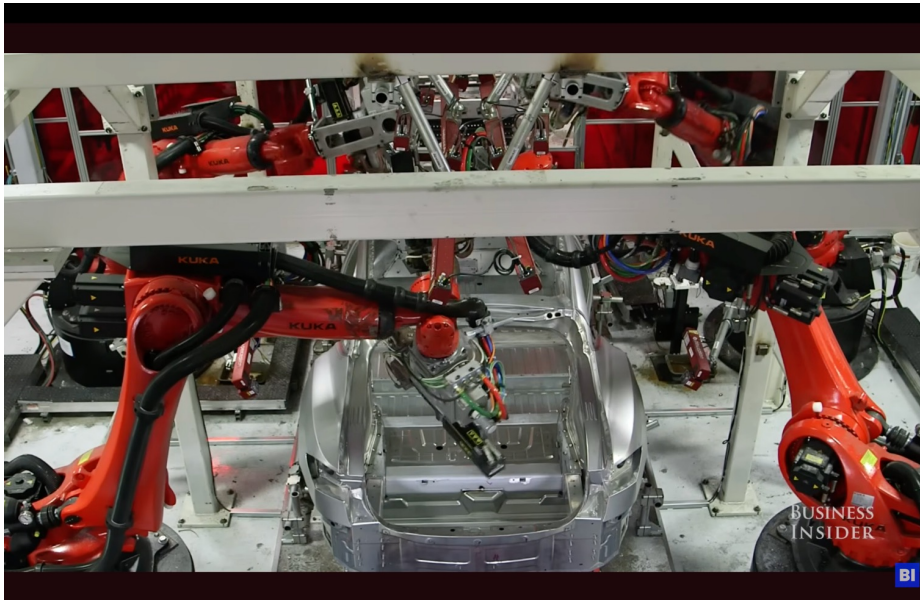
<https://youtu.be/WazuwEklGzA?si=sElrYT Yu07iOb3kY>

# Why the Physical Universe is Hard?

1. Dynamics, partial observations and uncertainty
2. There is almost no “reset” of your behavior
3. Rules are unknown, you only learn by “experience”
4. Reward is sparse and ambiguous. “Survival” could be the only shared reward among humankind, which still differs among different culture
5. No perfect simulation

# We Want Robots that Work in the Physical Universe

Robots in an easy universe



<https://youtu.be/WYnOGAvQEgk?si=W5uwZdIZuwWmKD5w>

Robots in the hard universe



<https://youtu.be/QHH3iSeDBLo?si=oJQKP3ig27h03Kpc>



# Recent Progress in Robotics is Amazing

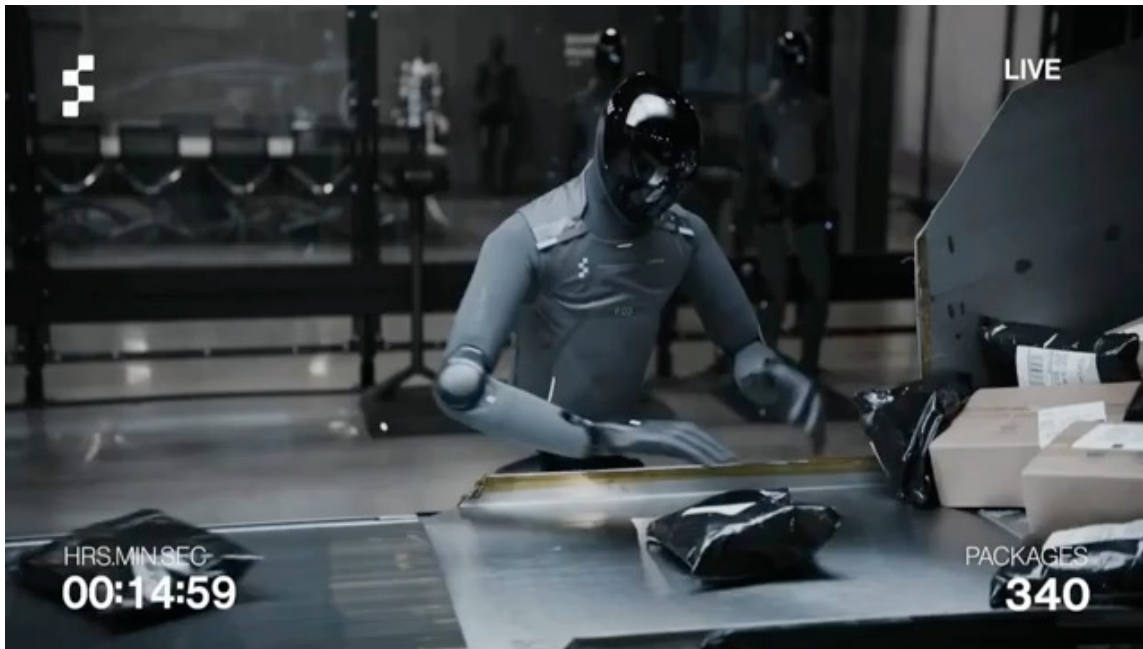
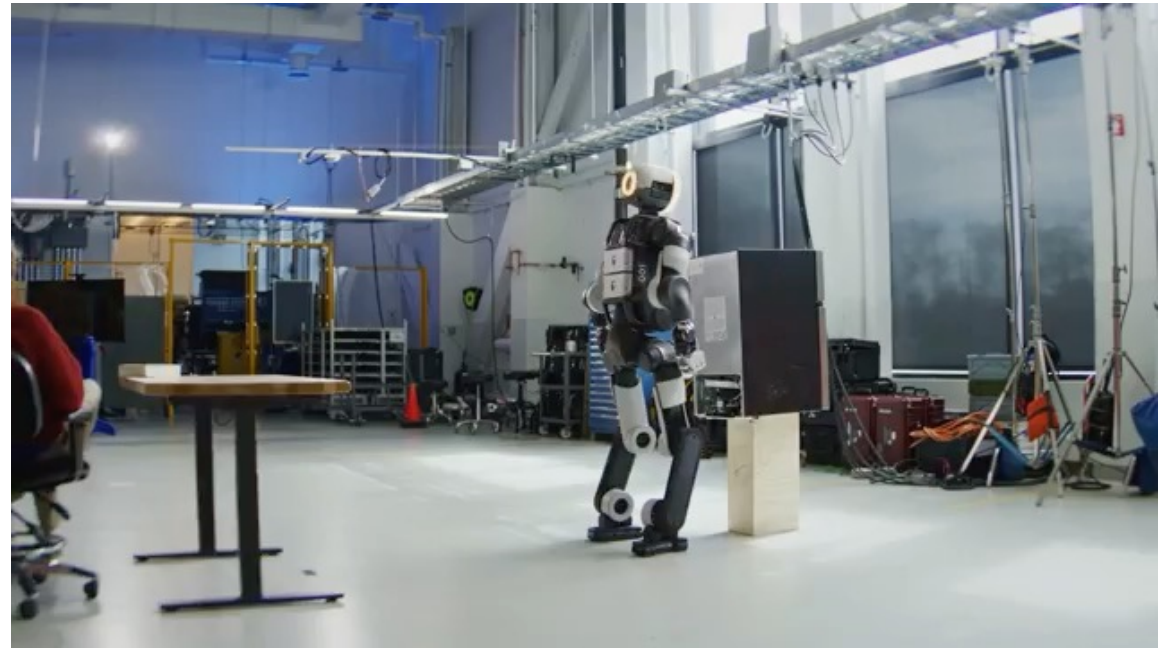
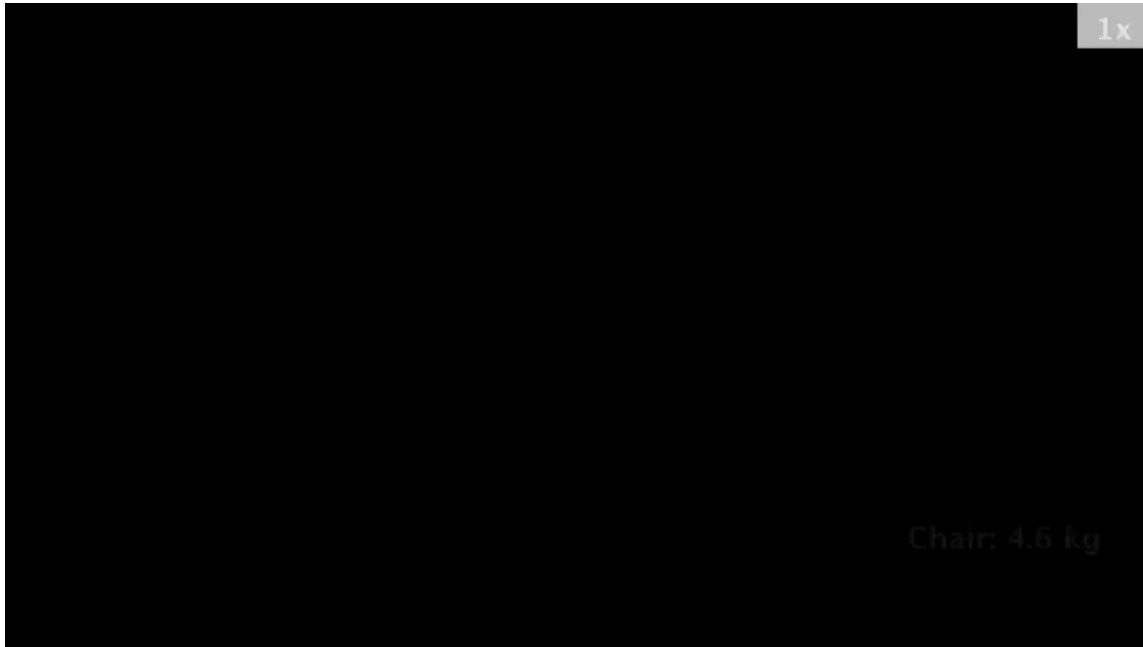


Figure AI

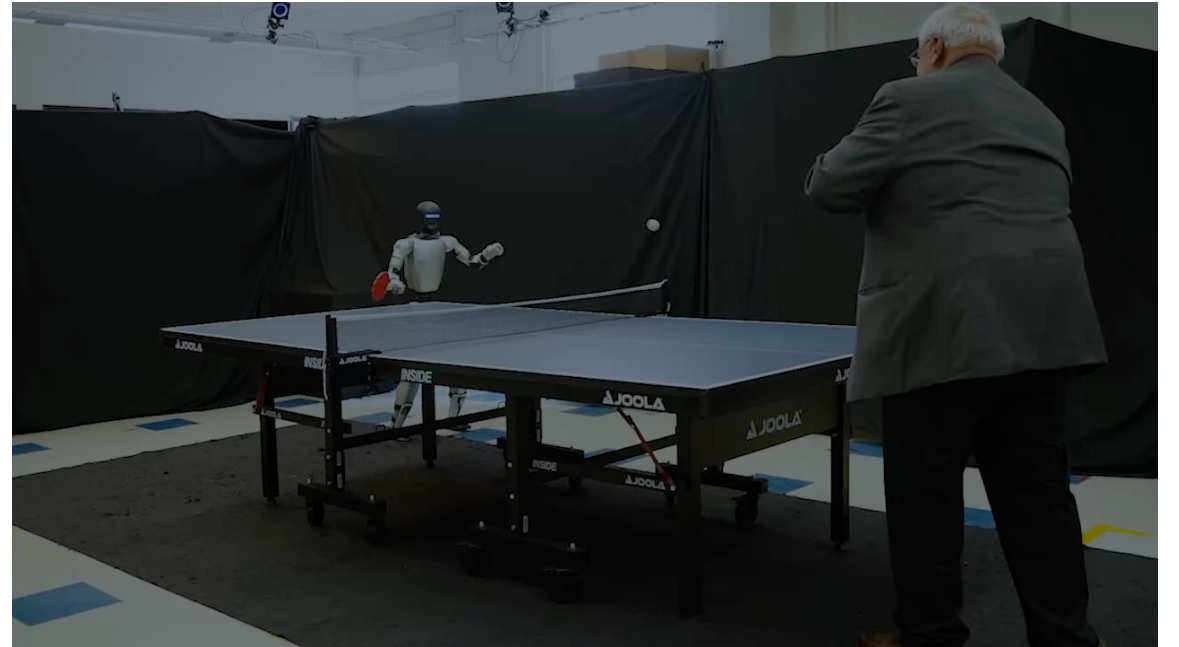


Boston Dynamics

# Recent Progress in Robotics is Amazing

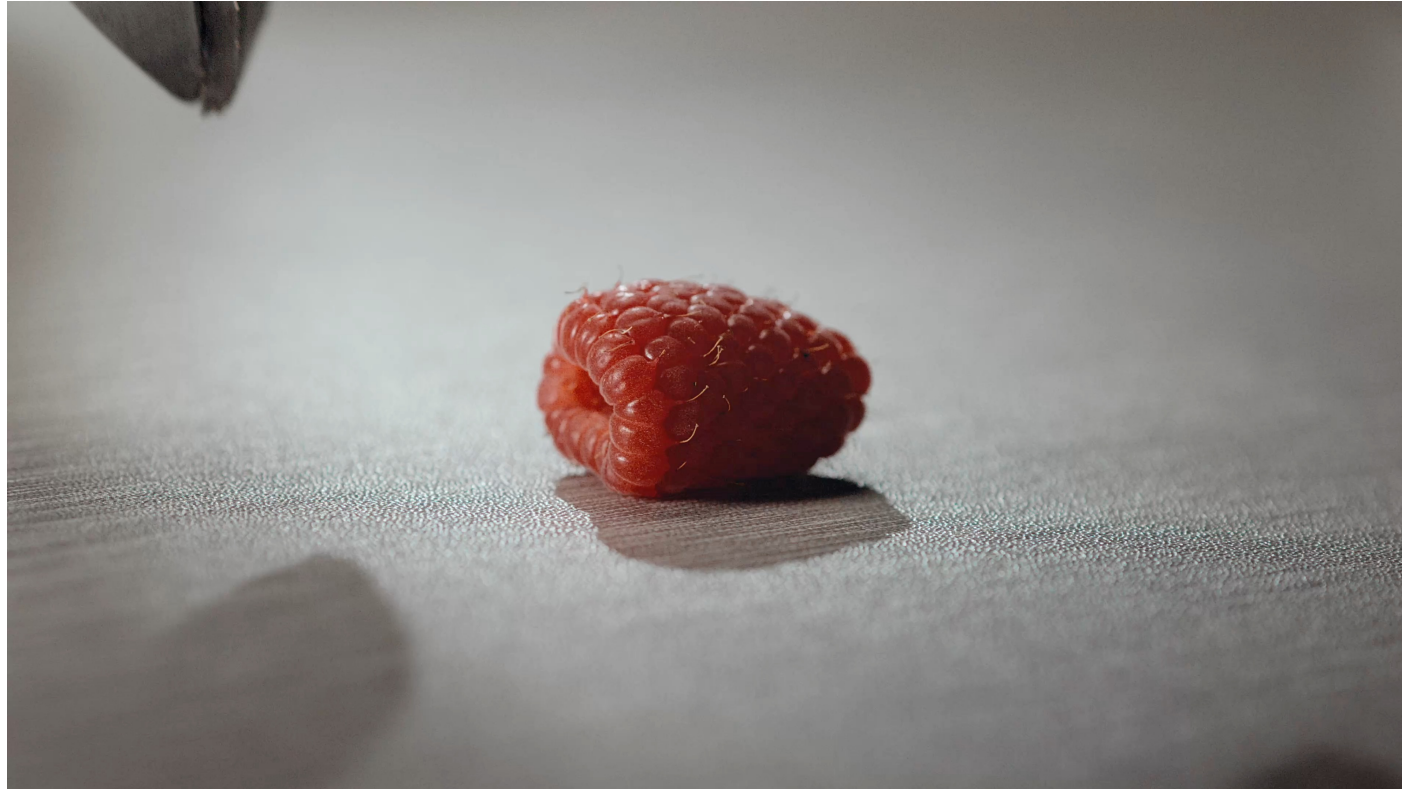


FAR at Amazon



HITTER

# Recent Progress in Robotics is Amazing



EKA Robotics



# Recent Challenges in Robotics



Image credit: Tom's Hardware



Image credit: ChatGPT



Image credit: ChatGPT

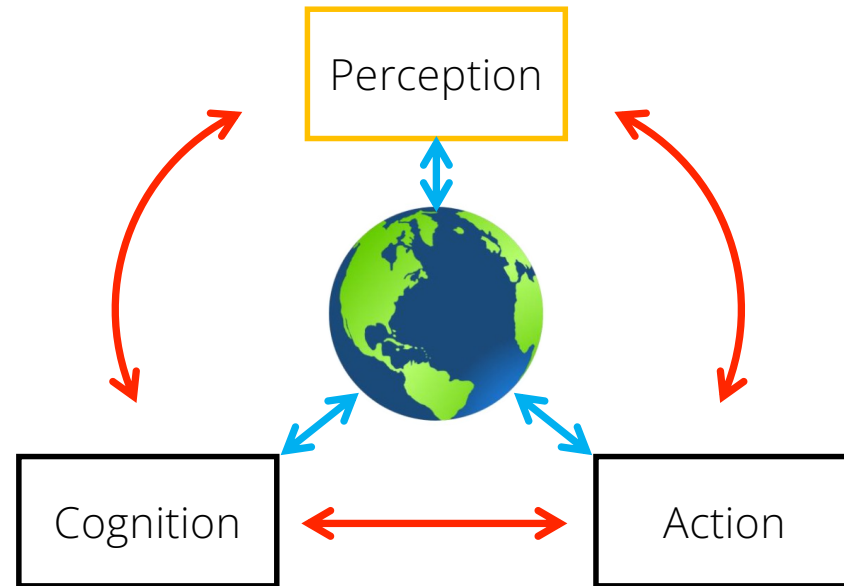


Image credit: Amanda Suarez



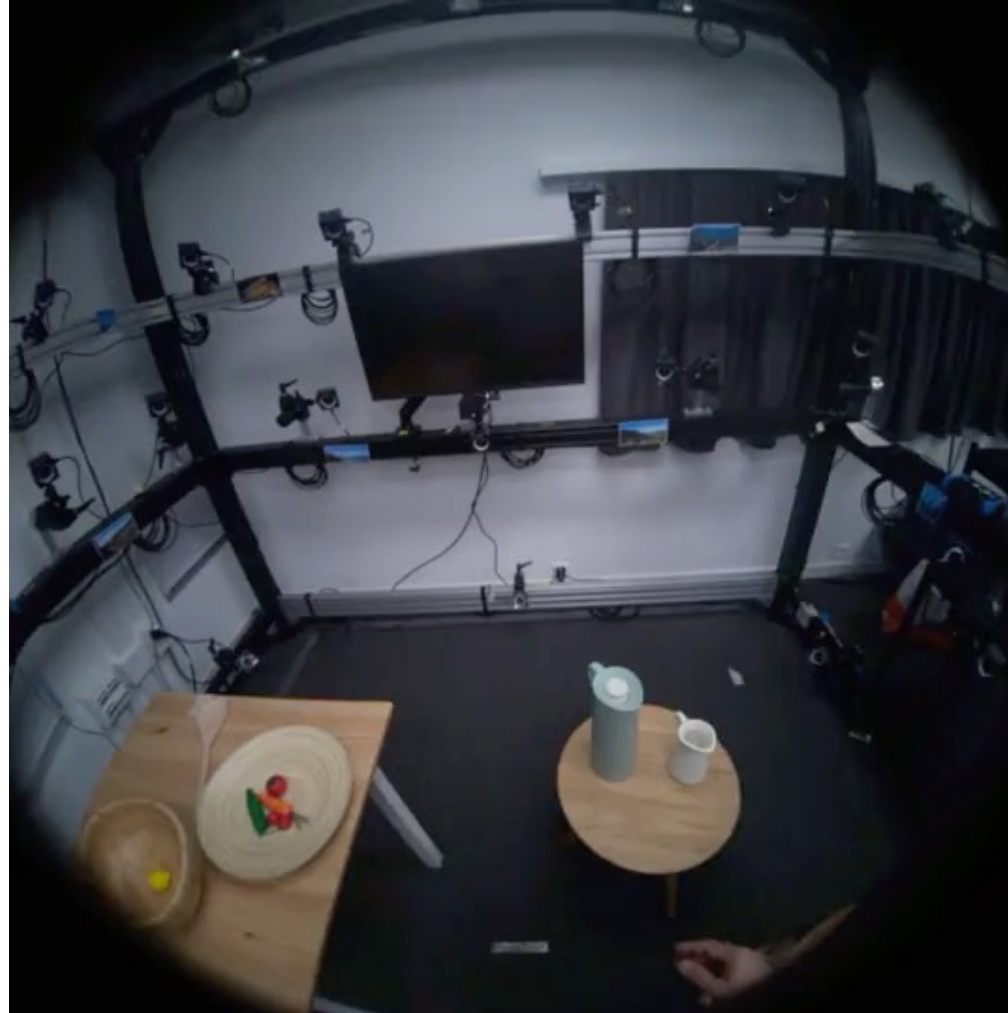
# How to Build a Robot Working in the Real World?

# To Build a Robot that Works in the Real World, We Need to Study the Trio of Perception, Cognition and Action

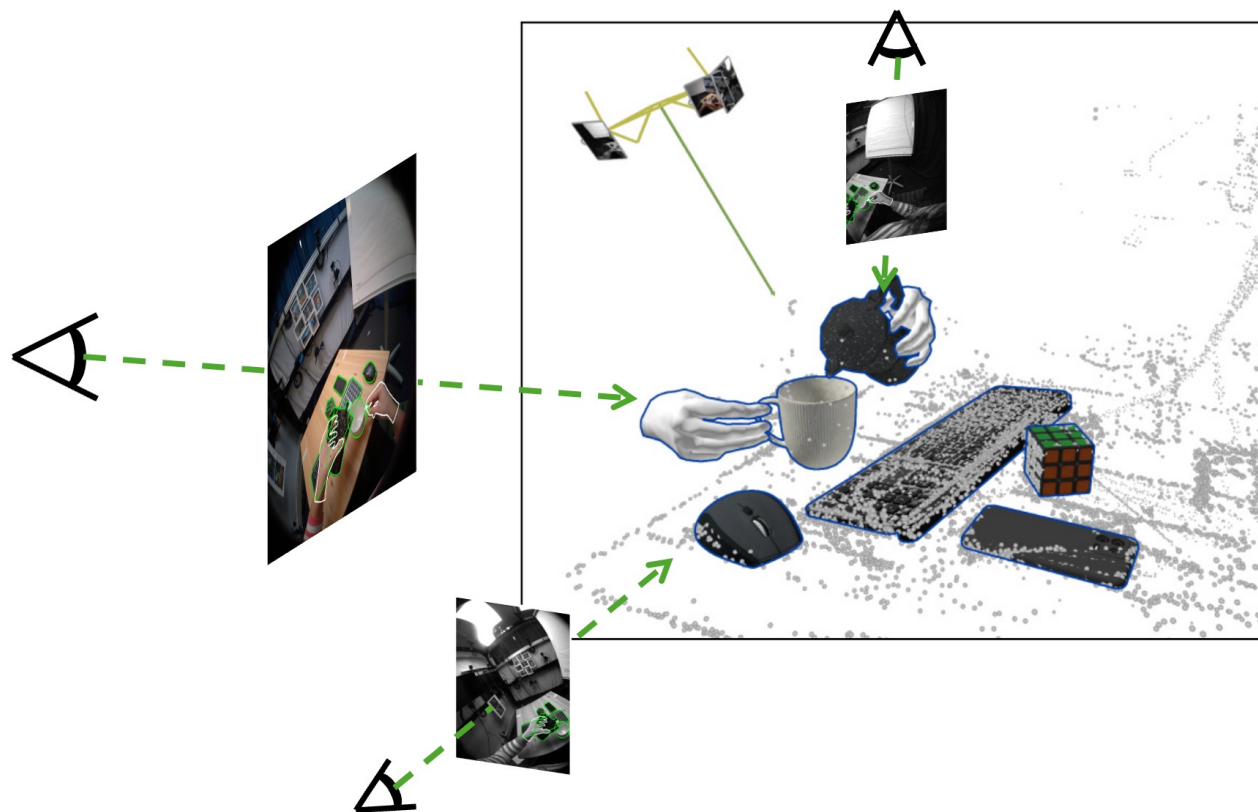
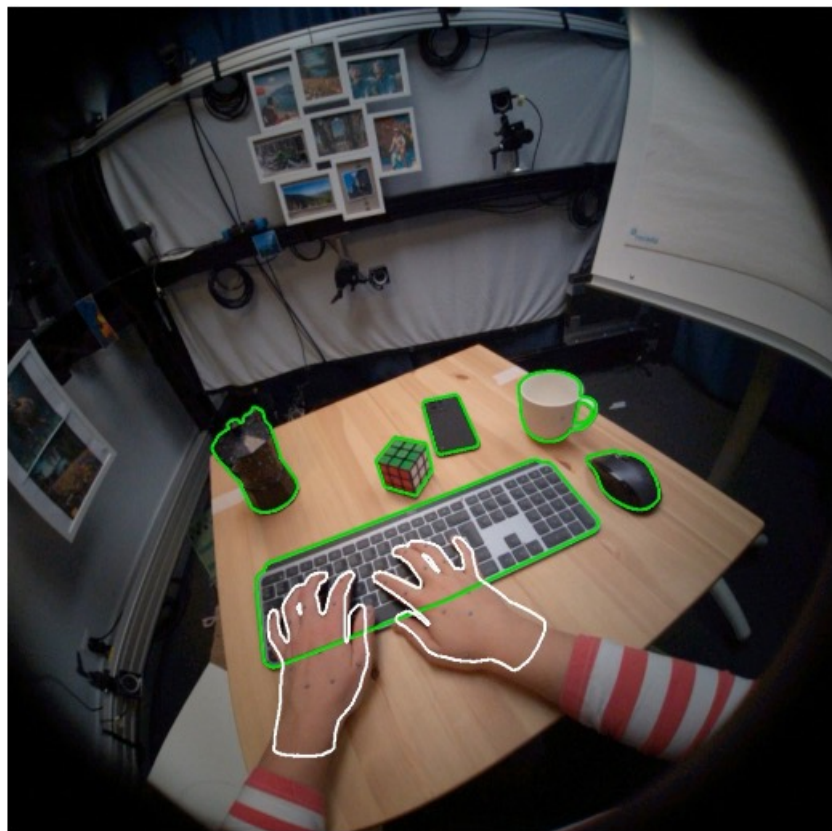


In simplification, we want to answer questions: “What are our goals?” and “What action should we take to achieve our goals **in the current environment**?”.

# What Are Important Perceptions?



# Recognition, Detection and 3D Modeling are Important



# Tracking and Pose Estimation are Important



# Inferring Physical Properties is Important



Are the cookies well toasted?  
Is the plate hot?  
Is the plate heavy?  
Where to grasp the plate?  
Is the glove slippery?



# Inferring the Consequence of Actions is Important

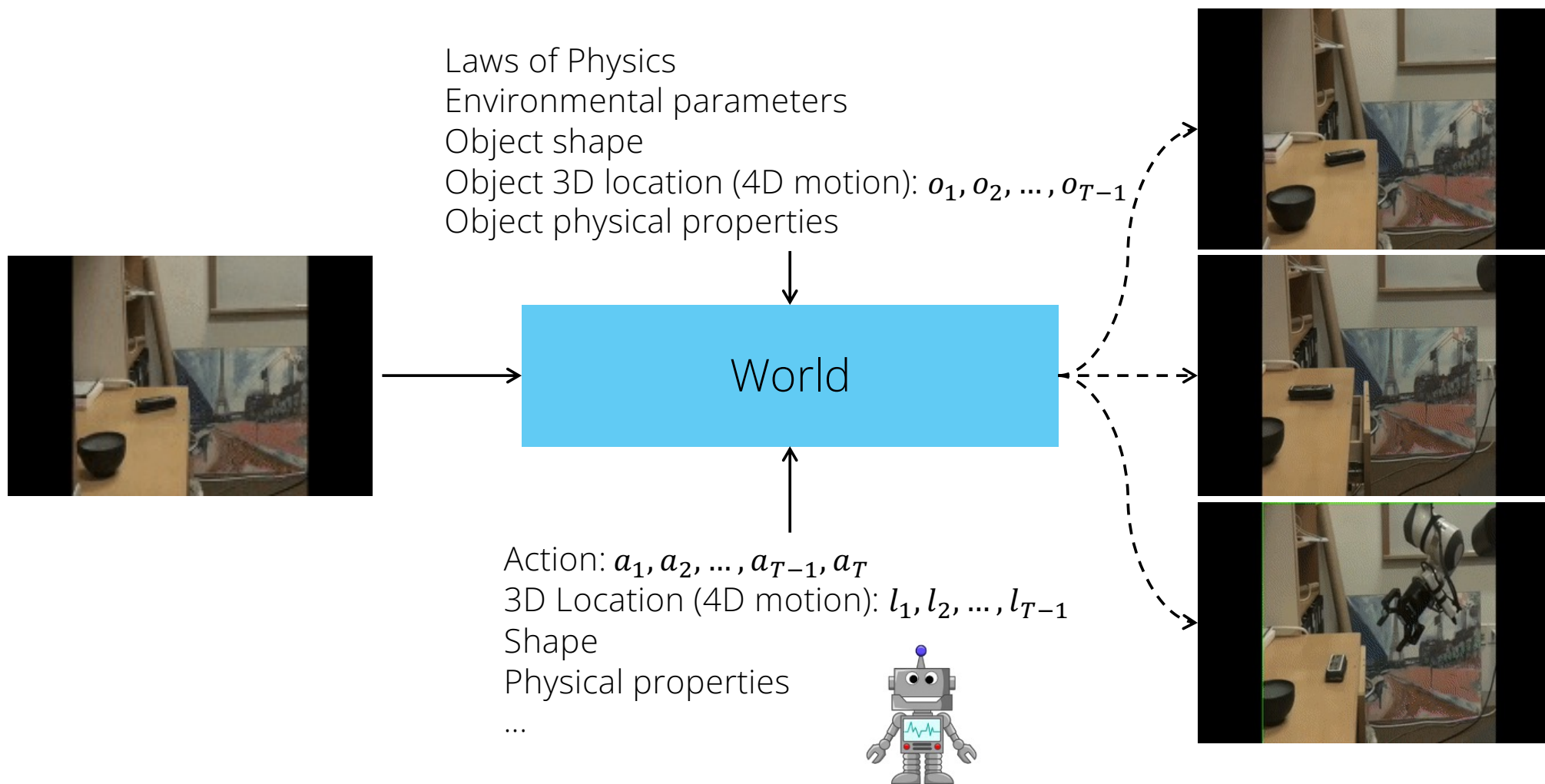


Action 1

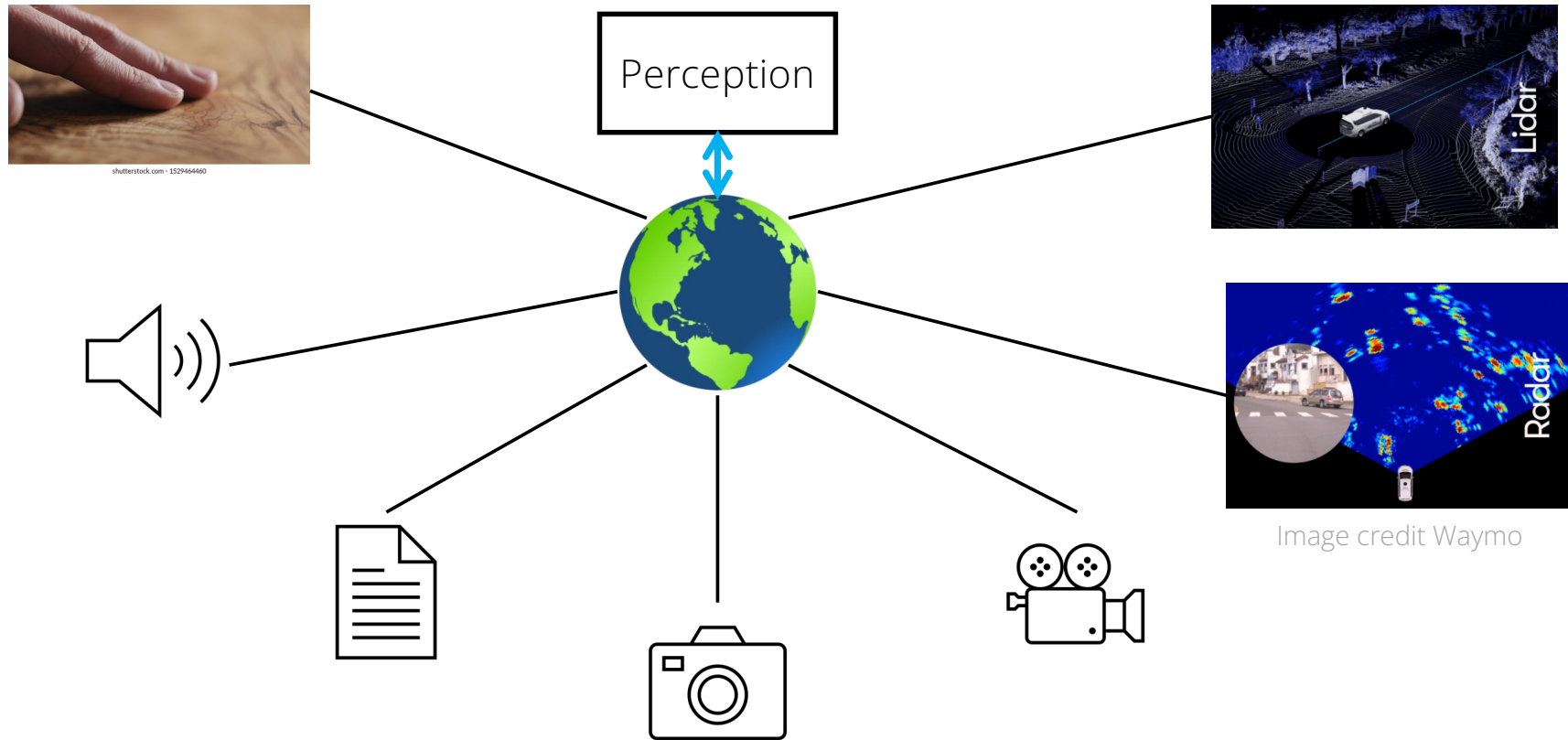




# Understanding the Dynamics is Important

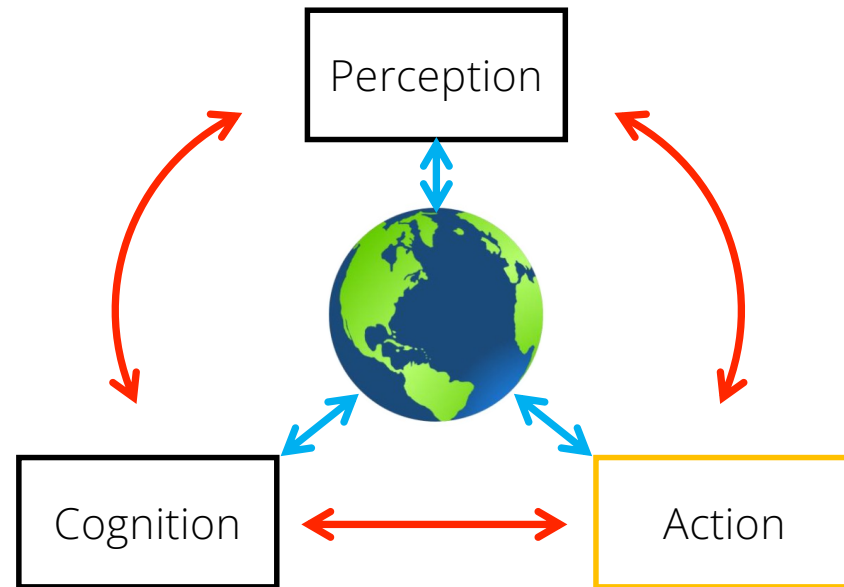


# Perception Is More Than Image/Video



We will Cover the Perception Part in  
Embodied Vision Course CSIE5421

# We will Focus on the Action and Cognition in this Course



In simplification, we want to answer questions: "What are our goals?" and "**What action should we take to achieve our goals** in the current environment?".

# Decision Making: Decide What Action to Take to Achieve Goals (Rewards)



Action: muscle contractions  
Rewards: food, attentions

Image credit S. Levine



Action: muscle contractions  
Rewards: food

<https://www.salon.com/2021/01/16/human-breeding-of-cats-has-made-them-like-they-are-always-in-pain/>



Action: steering, acceleration  
Rewards: win

<https://racingnews365.com/championship-standings-after-2024-f1-australian-grand-prix>



Action: selecting companies,  
buying, selling  
Rewards: cash



Action: (x, y) location  
Rewards: win

[https://en.wikipedia.org/wiki/Go\\_%28game%29](https://en.wikipedia.org/wiki/Go_%28game%29)



Action: jump,  $\uparrow \leftrightarrow \downarrow$   
Rewards: win

Image credit Nintendo

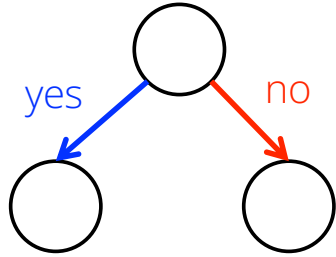


Action: RGB at each pixel  
Rewards: cash, aesthetics

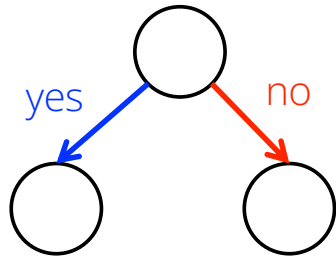
# Decision Making: Decide What Action to Take

I.I.D decision making

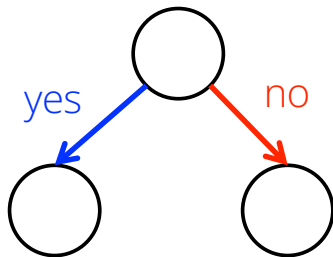
Have a  
breakfast



Have a  
lunch



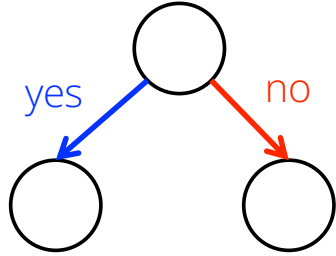
Have a  
dinner



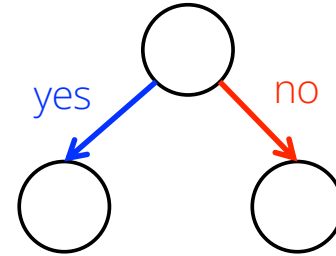
# However, Decision Making Problems Are Often Sequential

I.I.D decision making

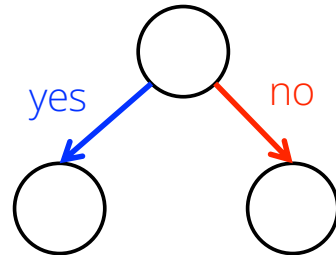
Have a  
breakfast



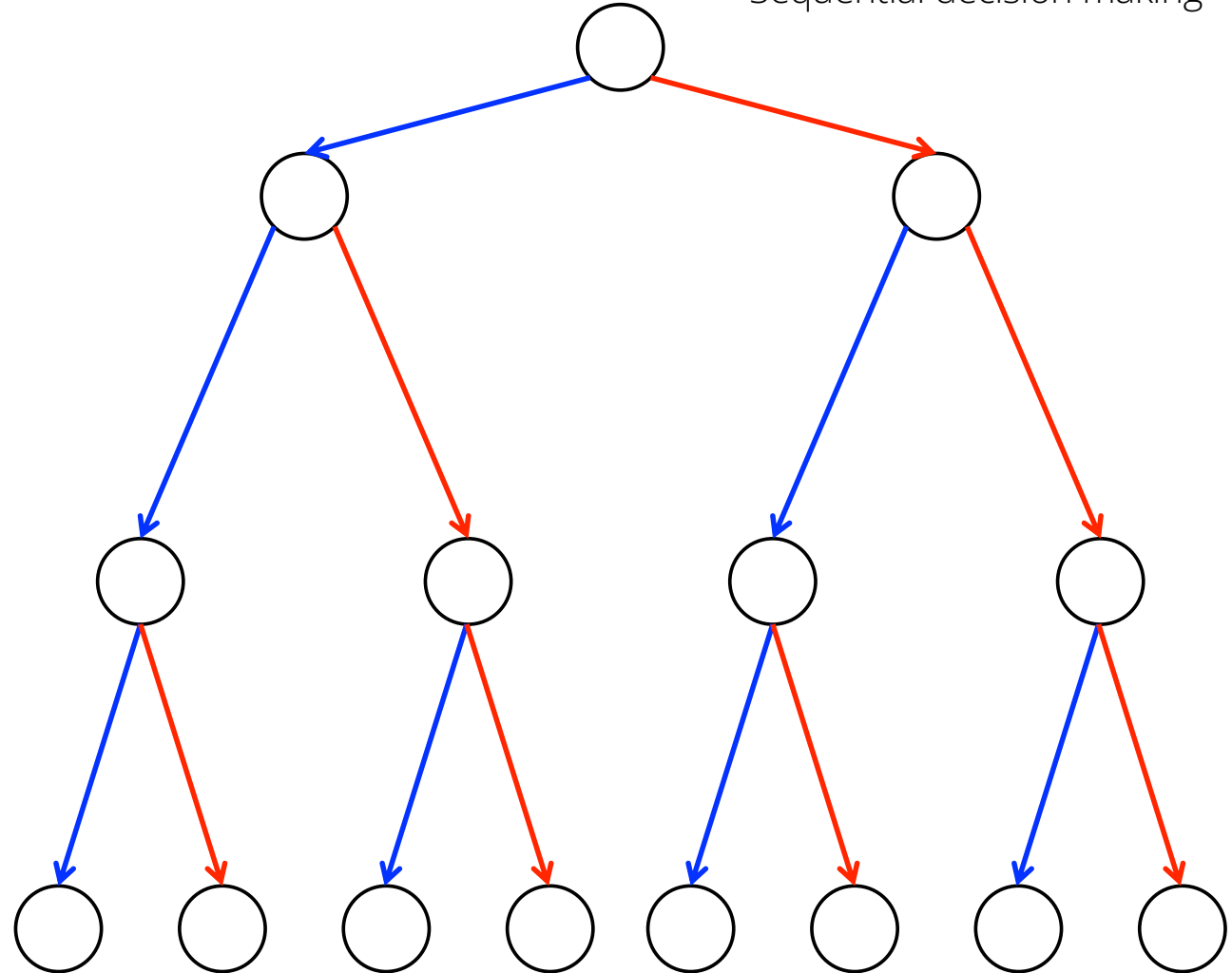
Have a  
lunch



Have a  
dinner

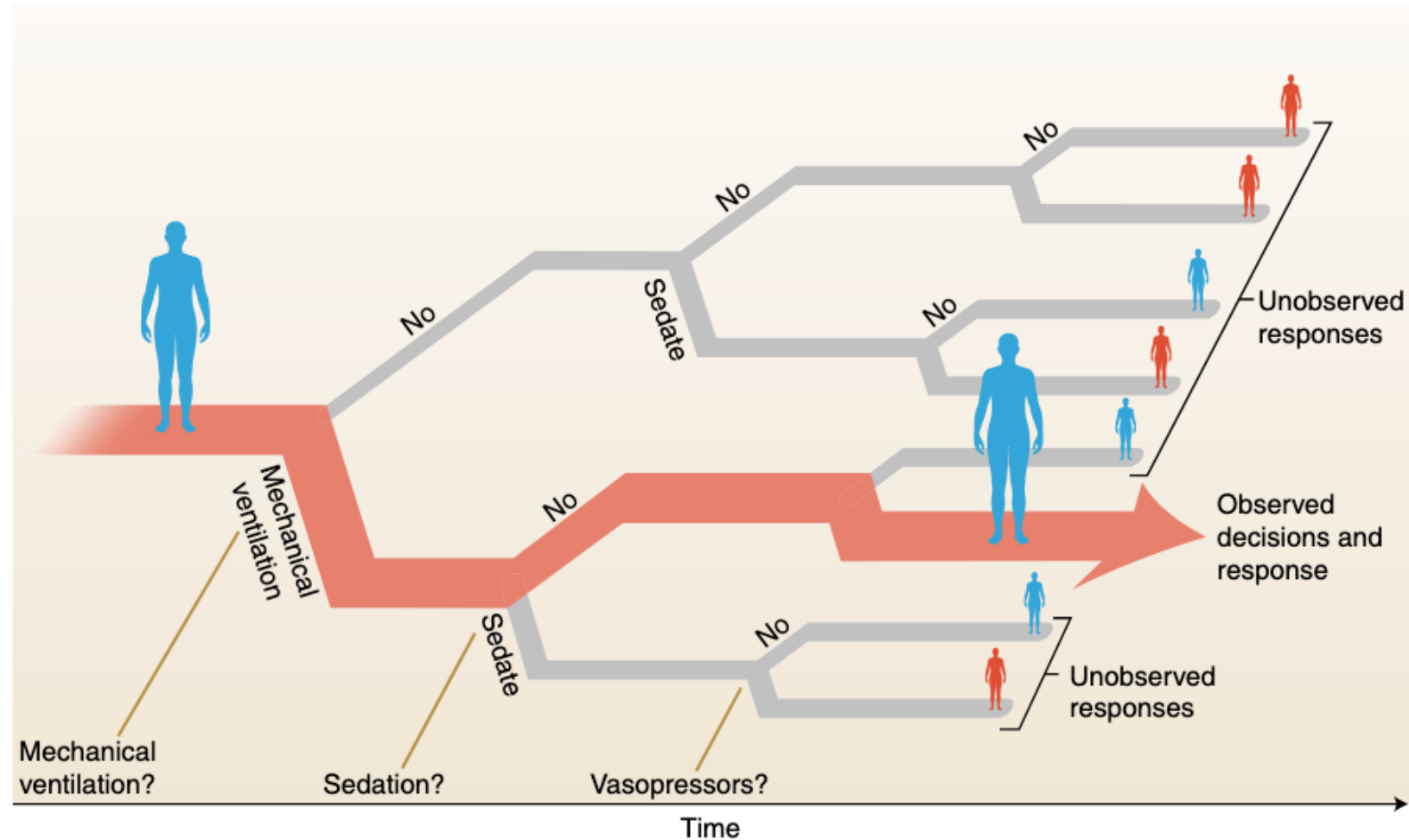


Sequential decision making

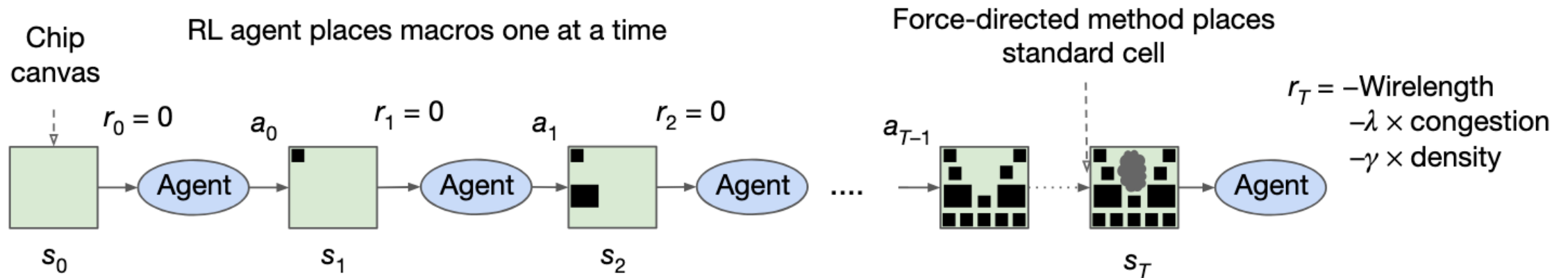




# However, Decision Making Problems Are Often Sequential



# However, Decision Making Problems Are Often Sequential



# However, Decision Making Problems Are Often Sequential



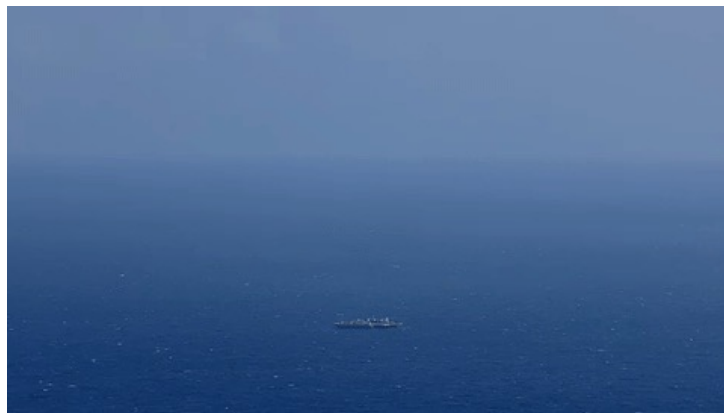
[https://youtu.be/a\\_YGPbWJO5g?si=3ZEr3rc5gjQ9JluX](https://youtu.be/a_YGPbWJO5g?si=3ZEr3rc5gjQ9JluX)



<https://youtu.be/6Zbhvaac68Y?si=Ivi7bak6xi-ByGzA>



<https://youtu.be/I-Y5FHI4jXc?si=mzqvEBCyu-YnYKun>



<https://www.youtube.com/watch?v=BhMSzC1crr0>



<https://youtu.be/cM6woUhJsCo?si=dKOi0WOUOLfbgkAk>



<https://youtube.com/shorts/GYeSEKnHjDI?si=8xSn3p9EB oW4nwBy>

# Each Decision You Make Influence Future Outcomes



League of Legends

What is my first item?

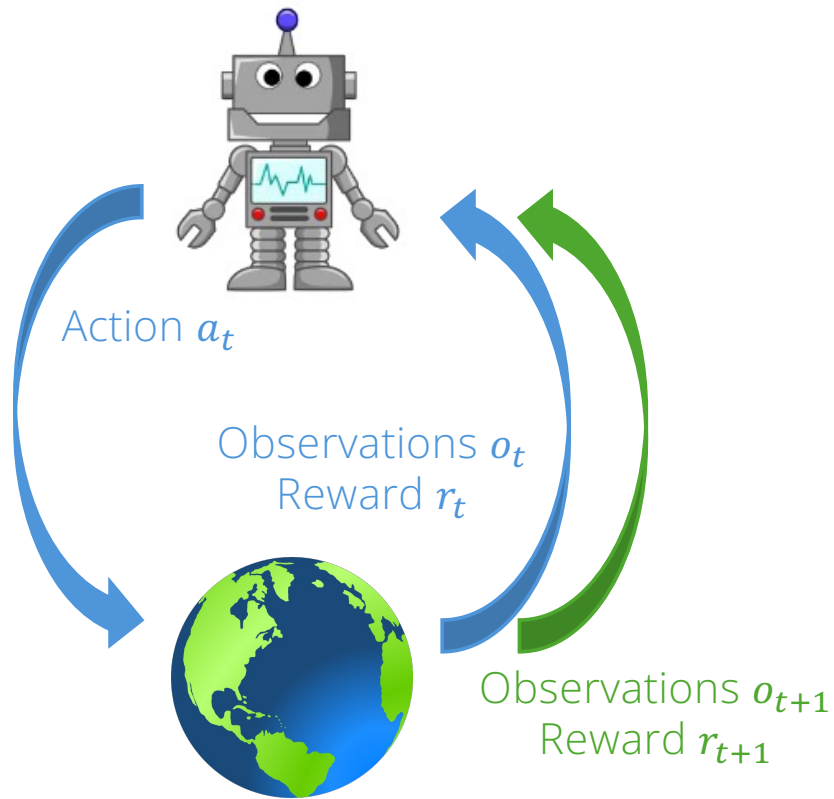


Age of Empire 2

Where should I drop the town center?  
How many villagers on wood?  
When should my villagers explore the map?  
⋮



# Sequential Decision Making Problem



**Trajectory:**  $s_1, a_1, r_1, s_2, a_2, r_2, \dots, s_T, a_T, r_T$

**Environment constraints:**  $s_{t+1} = f(s_t, a_t)$

**Goal:** maximize the total reward

$$\max \sum_{t=1}^T r_t$$

**Problem:** starting with  $s_1$ , what is the optimal action trajectory  $a_1, \dots, a_T$ ?

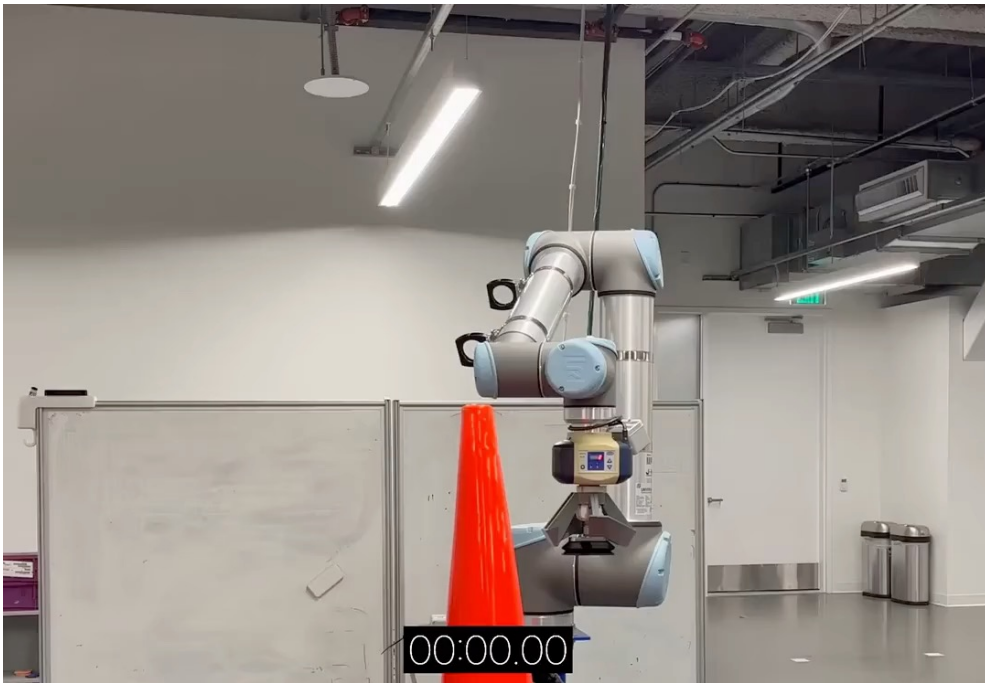
# Traditional Solutions?

1. Memorize 1 demonstration and replay

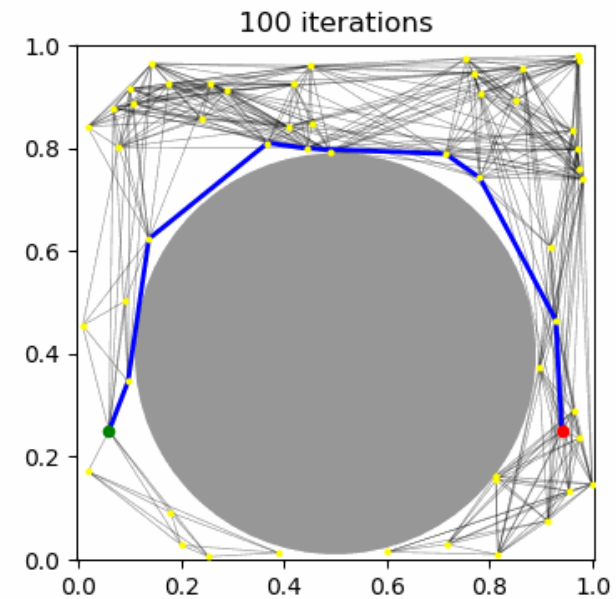


# Traditional Solutions?

1. Memorize 1 demonstration and replay
2. Plan a path to reach a target pose



Demos by Curobo

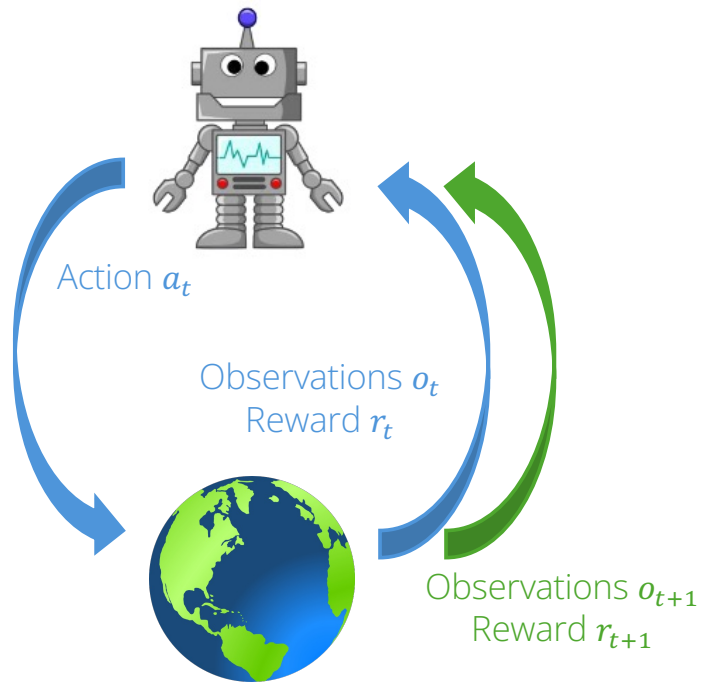


<https://medium.com/@junbs95/types-of-motion-planners-at-a-glance-617fdb07a762>



# Traditional Solutions?

1. Memorize 1 demonstration and replay
2. Plan a path to reach a target pose
3. Control trajectory optimization if the dynamic rules is known



Known environment constraints:

$$s_{t+1} = f(s_t, a_t)$$

Optimization problem:

$$a_1^*, a_2^*, \dots, a_T^* = \operatorname{argmax}_{a_1, \dots, a_T} \sum_{t=1}^T r_t$$

with

$$s_1 = \bar{s}_1$$
$$s_{t+1} = f(s_t, a_t)$$

# Traditional Solutions?

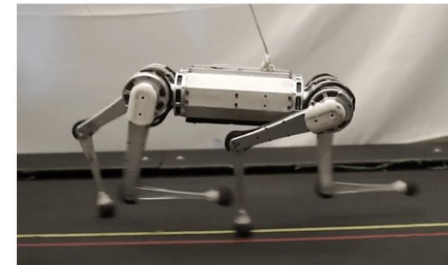
1. Memorize 1 demonstration and replay
2. Plan a path to reach a target pose
3. Control trajectory optimization if the dynamic rules is known

In-the-Wild Locomotion

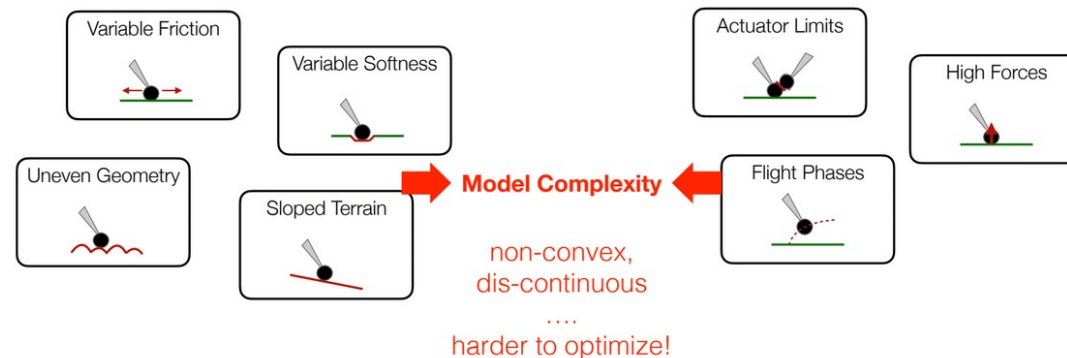


Lee et al. 2020

Fast Locomotion



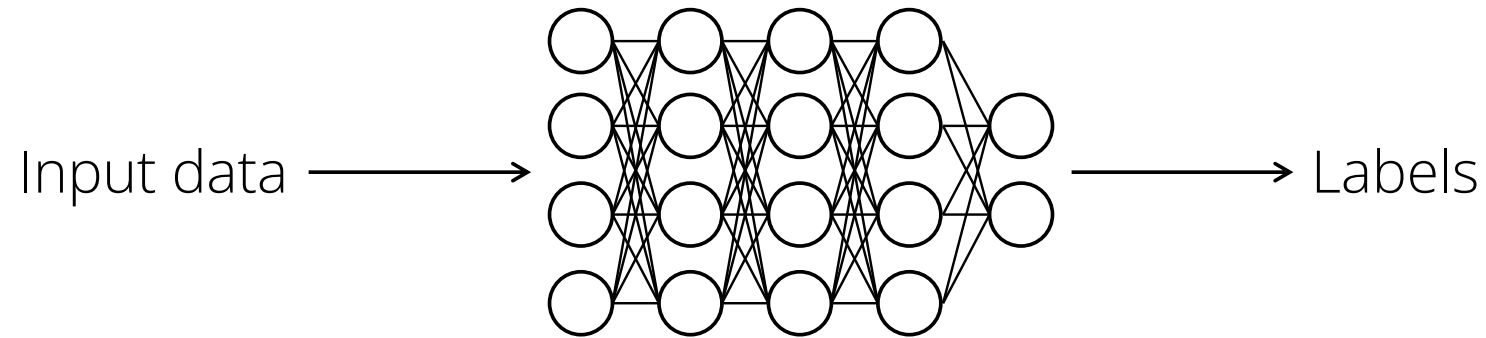
Kim et al. 2019



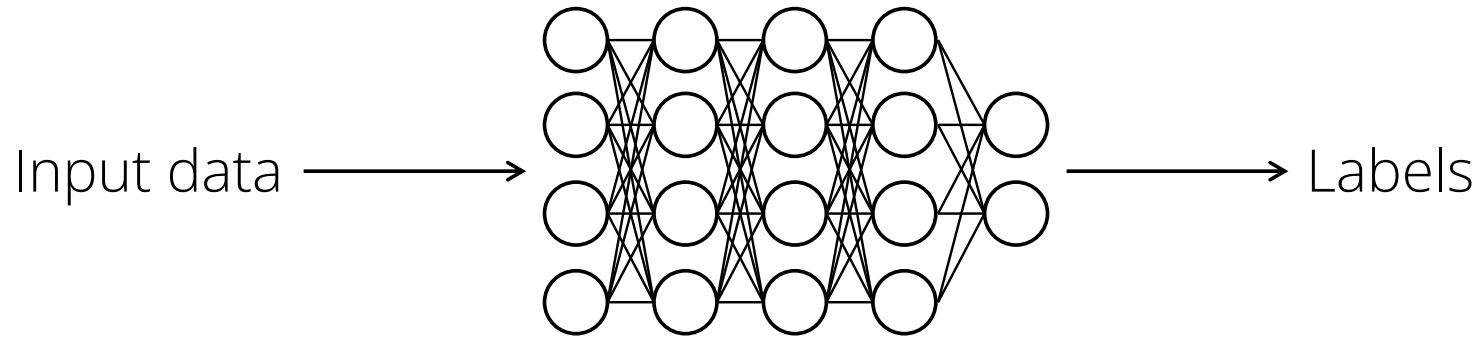
# Problems of Traditional Solutions

1. Memorize 1 demonstration and replay
  - Doesn't generalize to different scene layouts
2. Plan a path to reach a target pose
  - The target pose is often unknown...
3. Control trajectory optimization if the dynamic rules is known
  - The dynamics is often unknown...

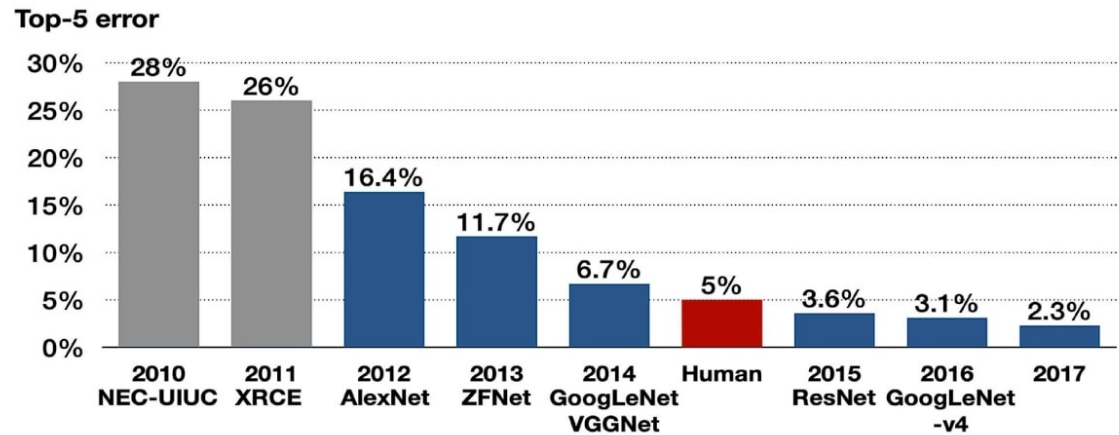
# Ideas from Machine Learning?



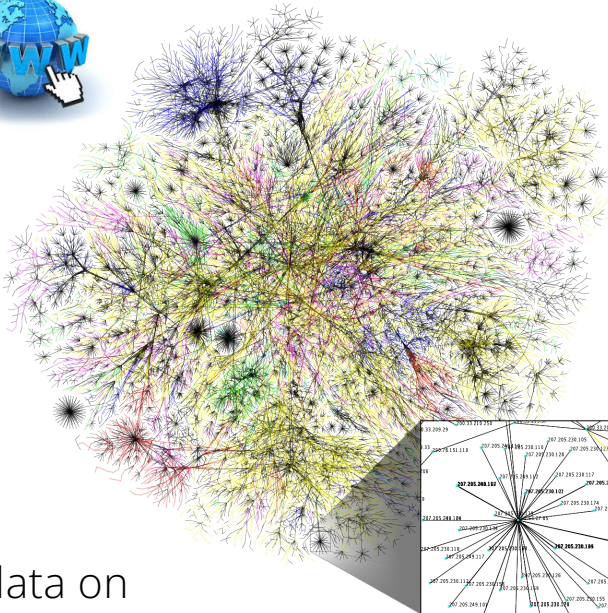
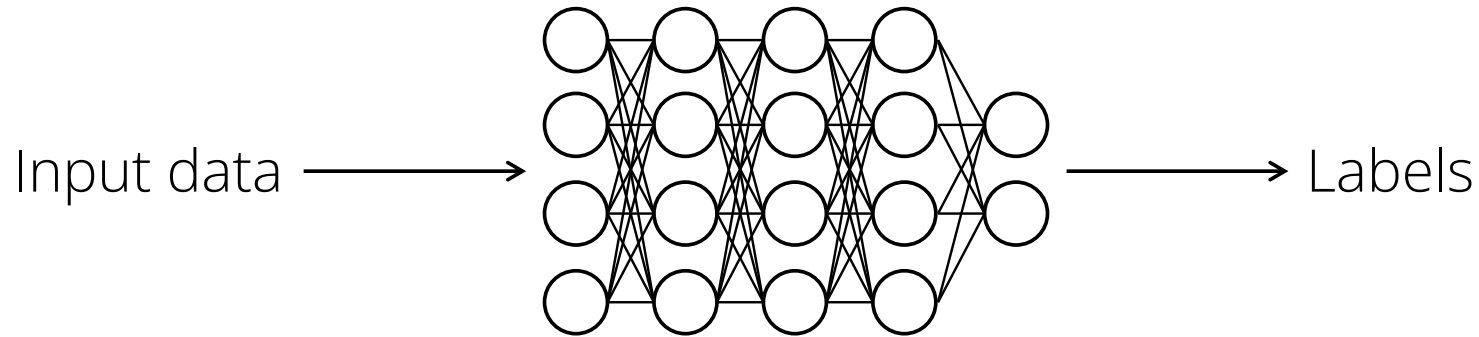
# Ideas from Machine Learning?



1,000,000 images with 1,000,000 labels



# Ideas from Machine Learning?

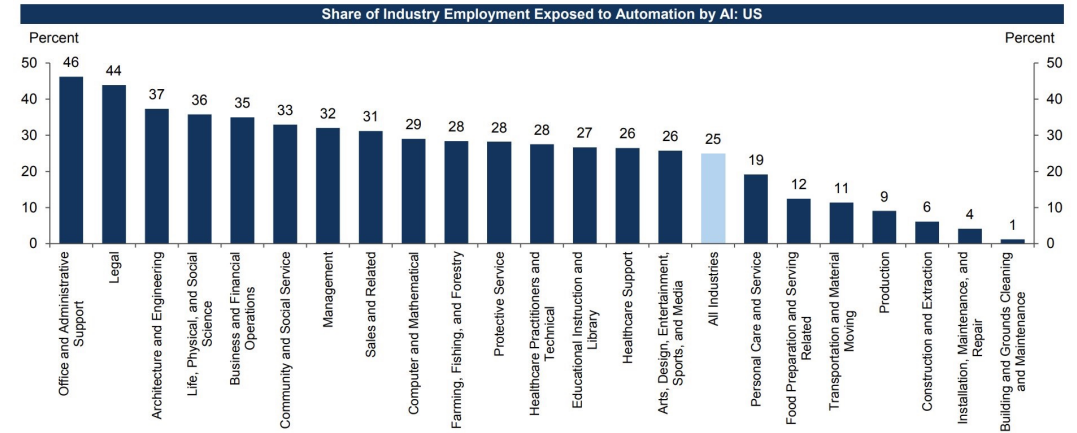


All text data on  
the Internet

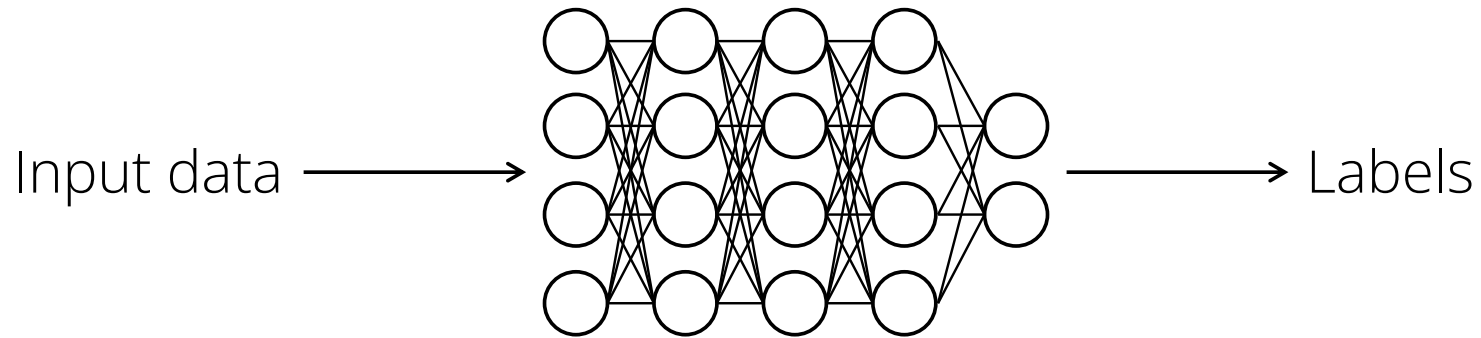
Goldman Sachs

Global Economics Analyst

**Exhibit 5: One-Fourth of Current Work Tasks Could Be Automated by AI in the US and Europe**



# Robot Learning Applies Learning in Robotics



?

Target pose

Feasible paths

Dynamics



# Solutions?

## 1. Imitation learning by fitting many demonstrations



Figure source: AGIBot

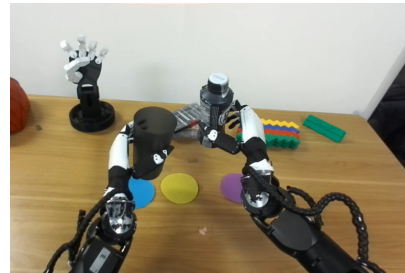


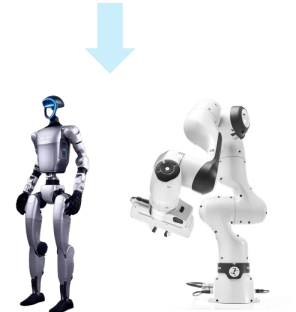
Figure source: <https://human-as-robot.github.io/>



Pour water into the cup

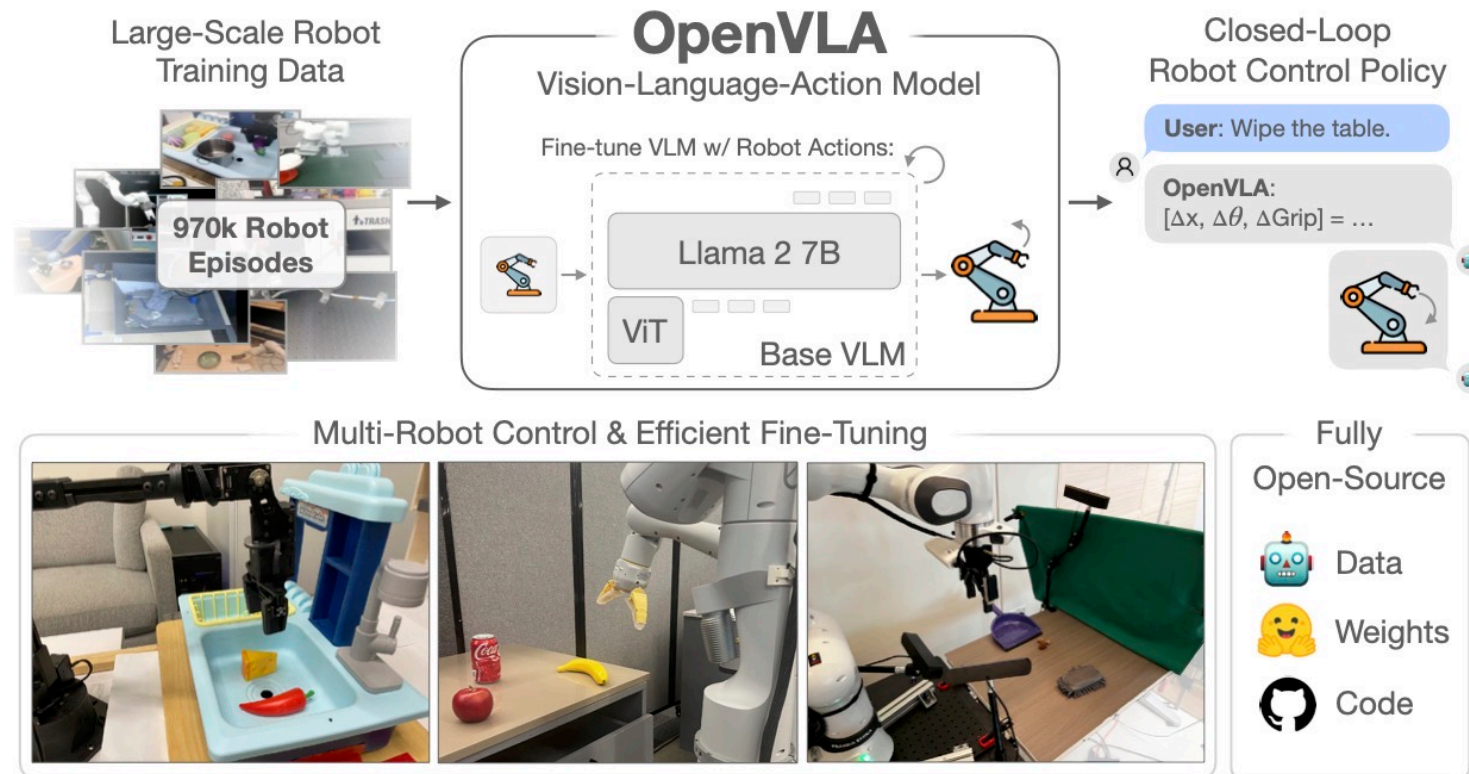
Large Behavior Models

Robot actions



# Imitation Learning as a Solution

1. Imitation learning by fitting many demonstrations with large models



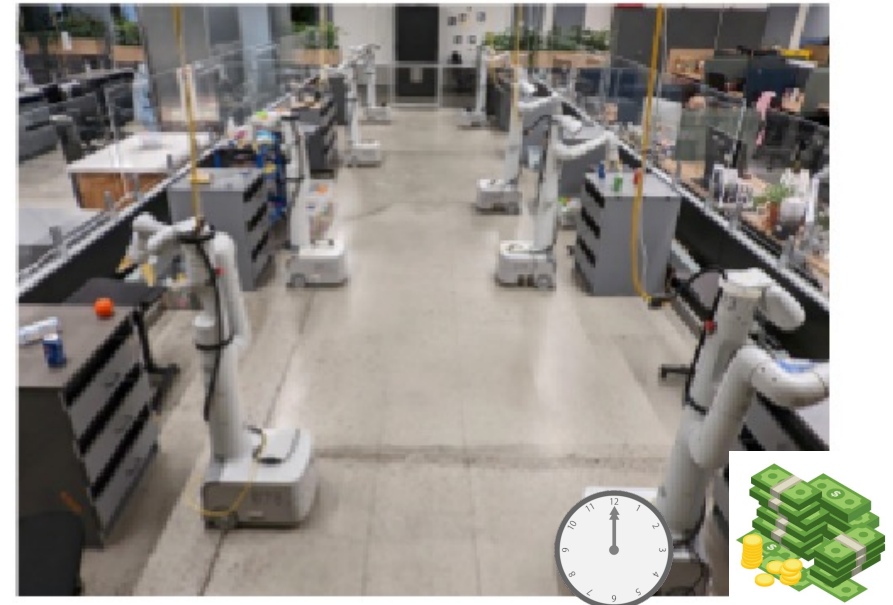
# Imitation Learning has Issues

## 1. Imitation learning by fitting many demonstrations

Succeed in the training domain

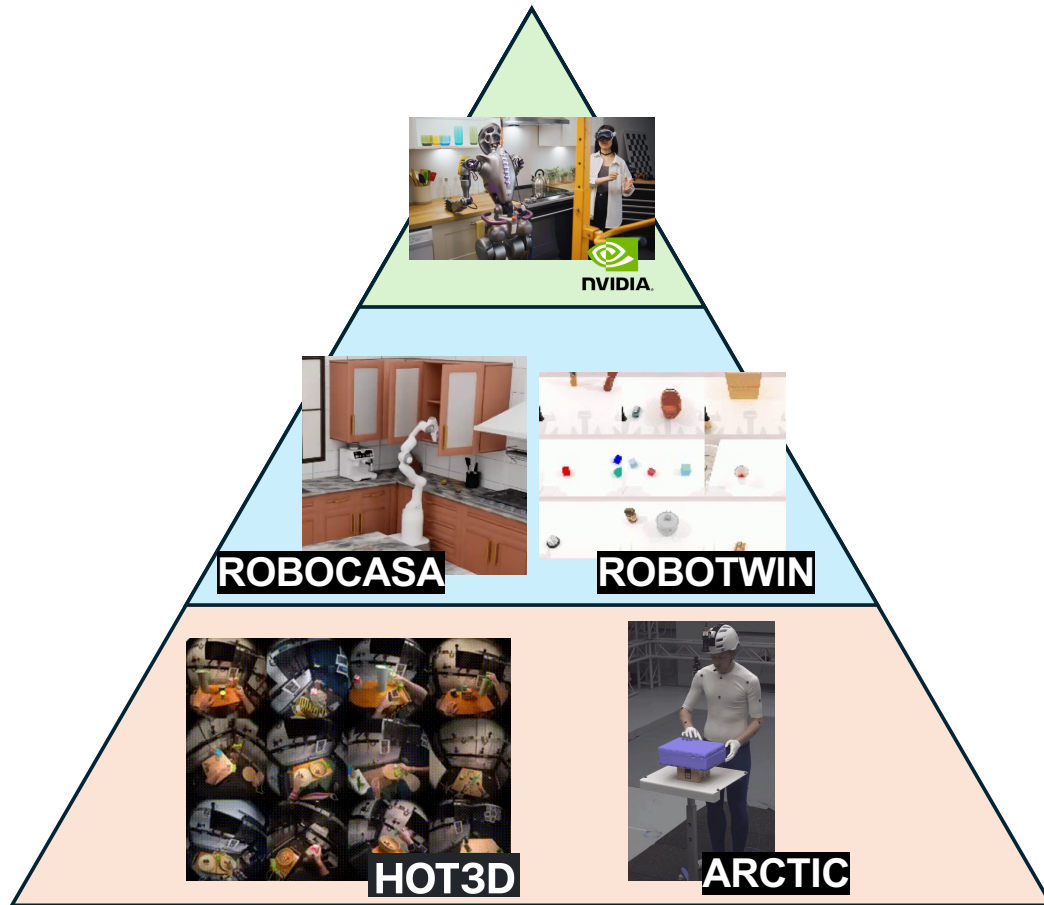


Fail in the testing domain



RT-1: Robotics Transformer for Real-World Control at Scale

# Scaling Up Training Data for Robots



Real robot data collected by teleoperation

Synthetic robot data collected in simulation

Real human data stored on the Internet

- Not even robot actions
- Still need humans for data collection

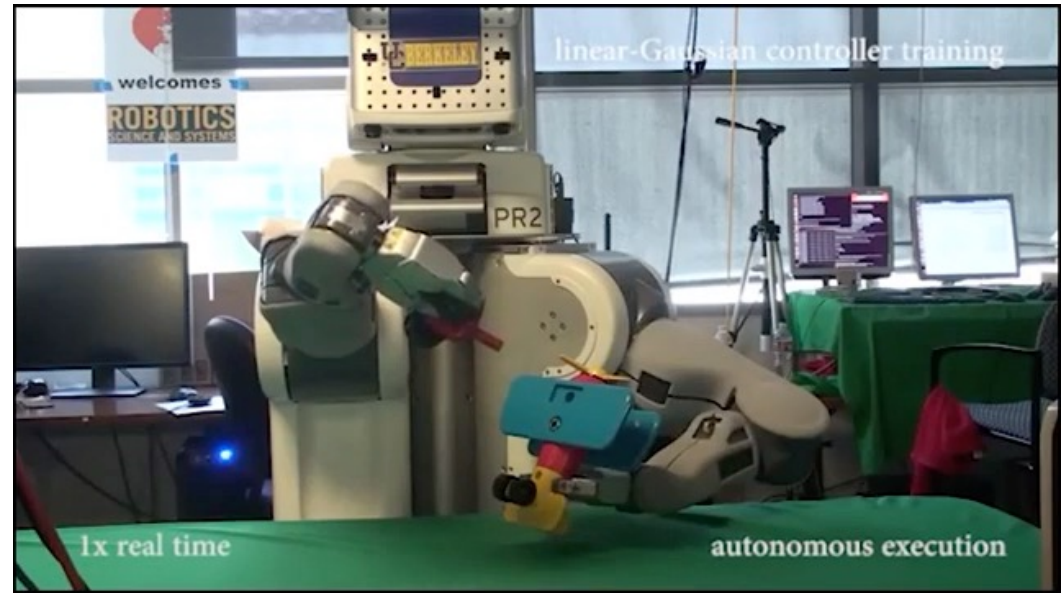


# Other Solutions?

1. Control trajectory optimization if the dynamic rules is known
2. Reinforcement learning by trial and error



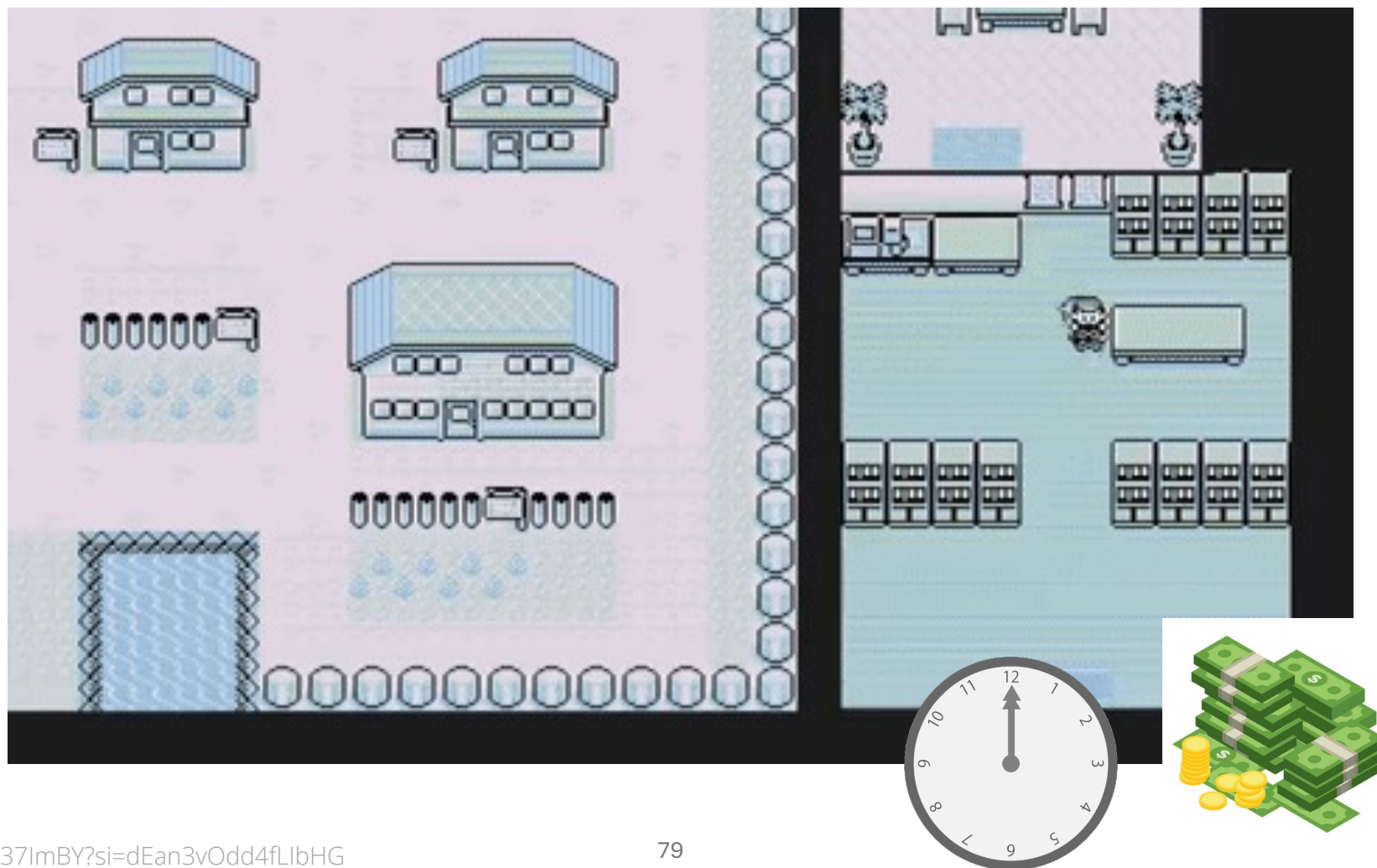
<https://www.youtube.com/watch?v=kFPcoclV5Vo>



<https://www.youtube.com/watch?v=JeVppkoloXs&t=11s>

100 Training Episodes

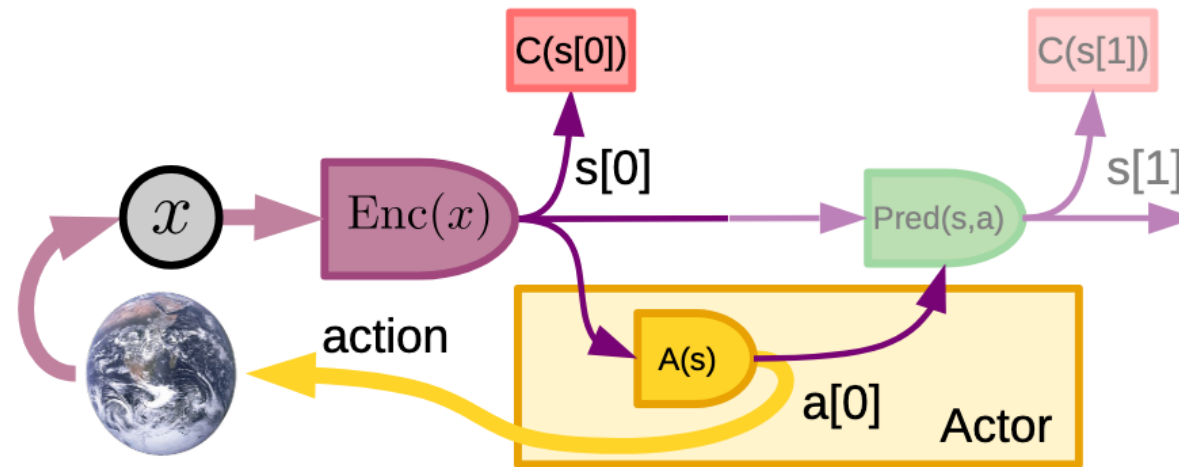
# Model-free RL is Sample Inefficient...





# Other Solutions?

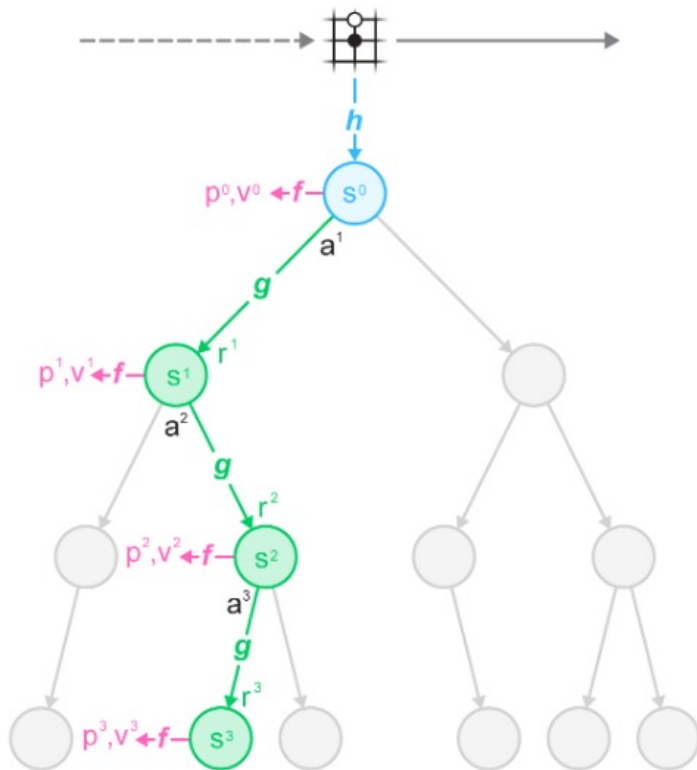
1. Control trajectory optimization if the dynamic rules is known
2. Reinforcement learning by trial and error
3. Reinforcement learning by trial and error + learning a world simulator from experience



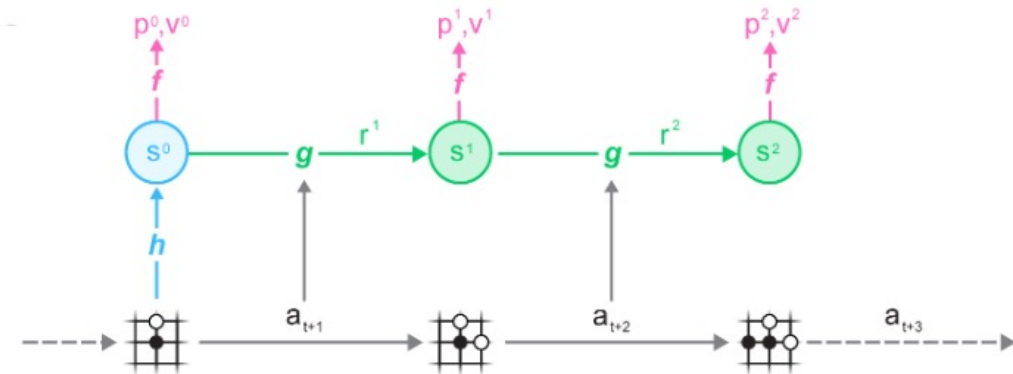
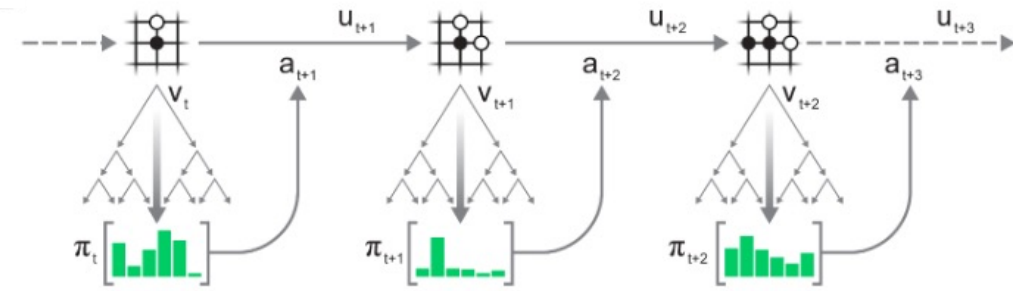
$$\min \|f(s, a) - pred(s, a)\|$$

# Model-based RL is More Sample Efficient

A. Plan to obtain optimal actions by simulating outcomes of different actions

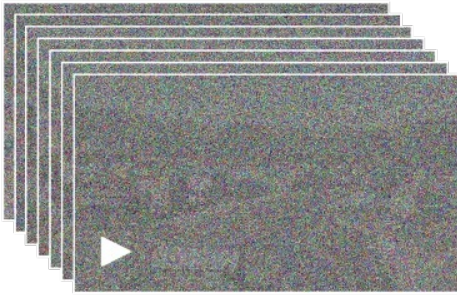


B. Interact with the environment with optimal actions



C. Learn the simulator through experience

# What Simulators to Learn?



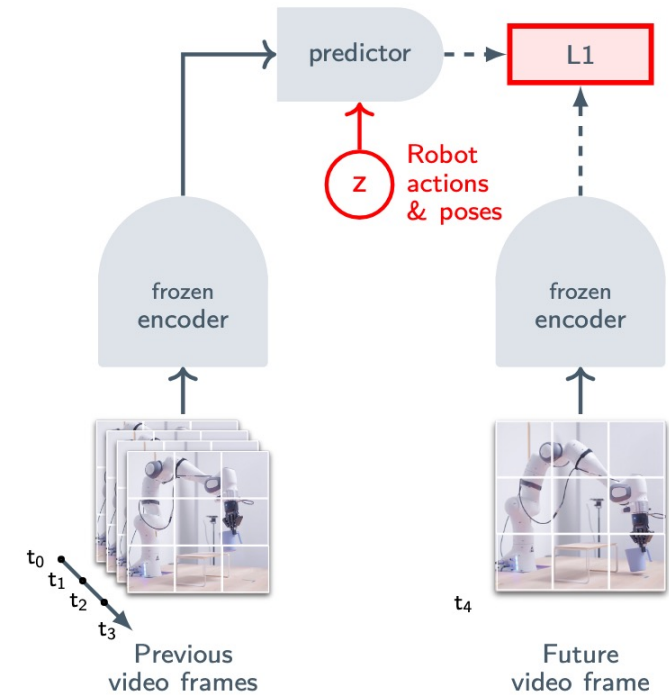
Video generative  
model

SORA by OpenAI



4D generative  
model

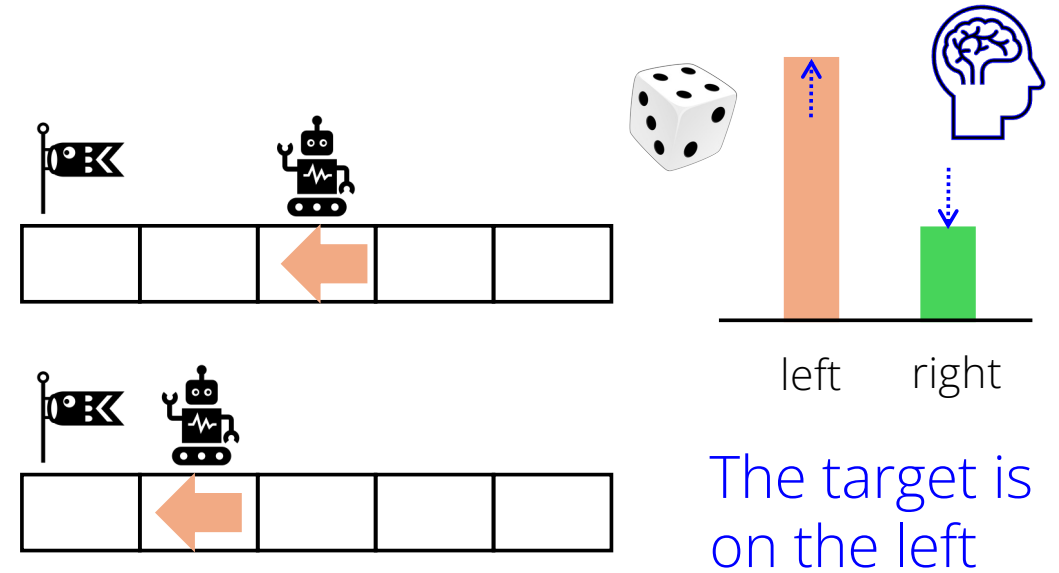
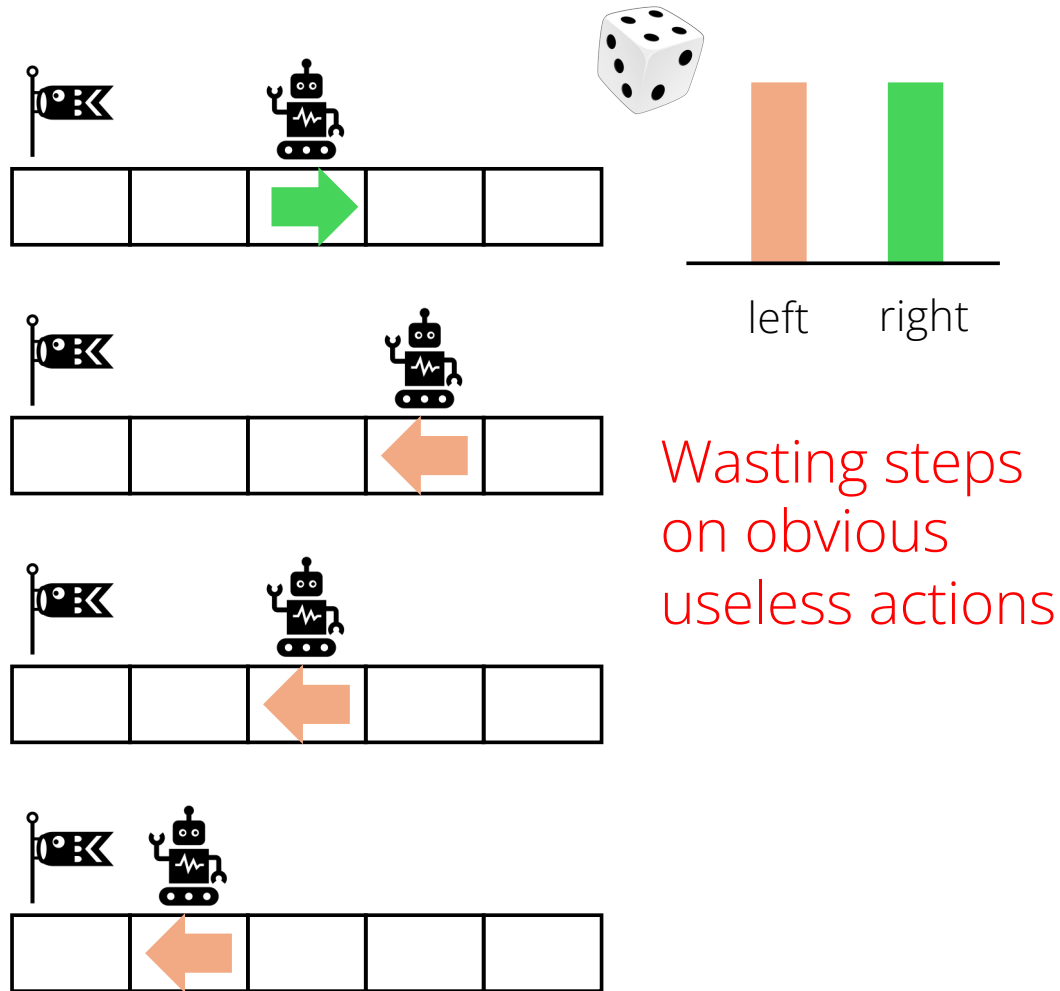
Dynamic 3D Gaussians: Tracking by Persistent  
Dynamic View Synthesis



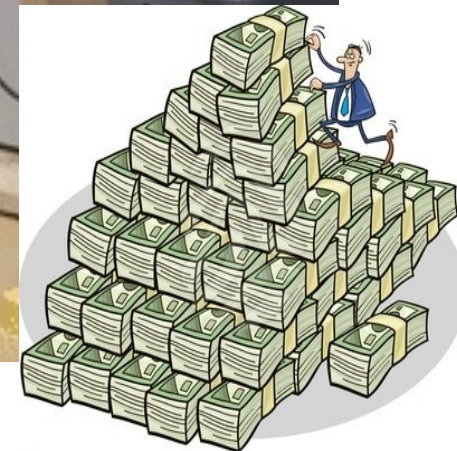
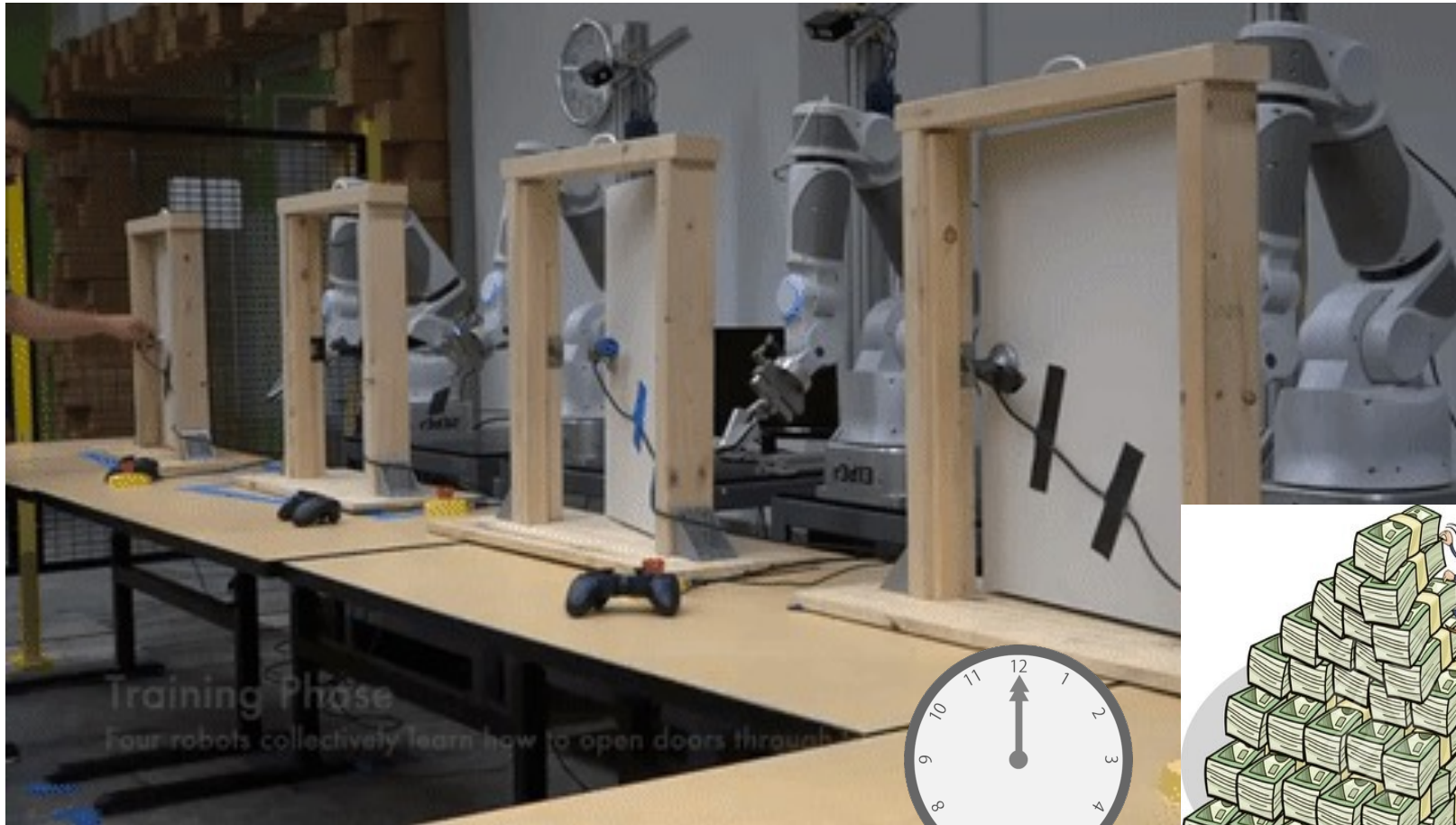
V-JEPA 2: Self-Supervised Video Models Enable  
Understanding, Prediction and Planning

But RL Has More Issues...

# RL Exploration is Ineffective



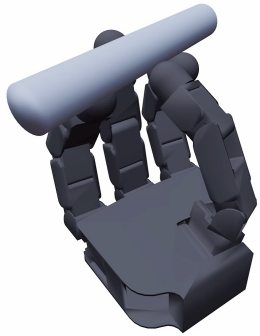
# Sample Inefficiency Makes RL in Real World even Harder



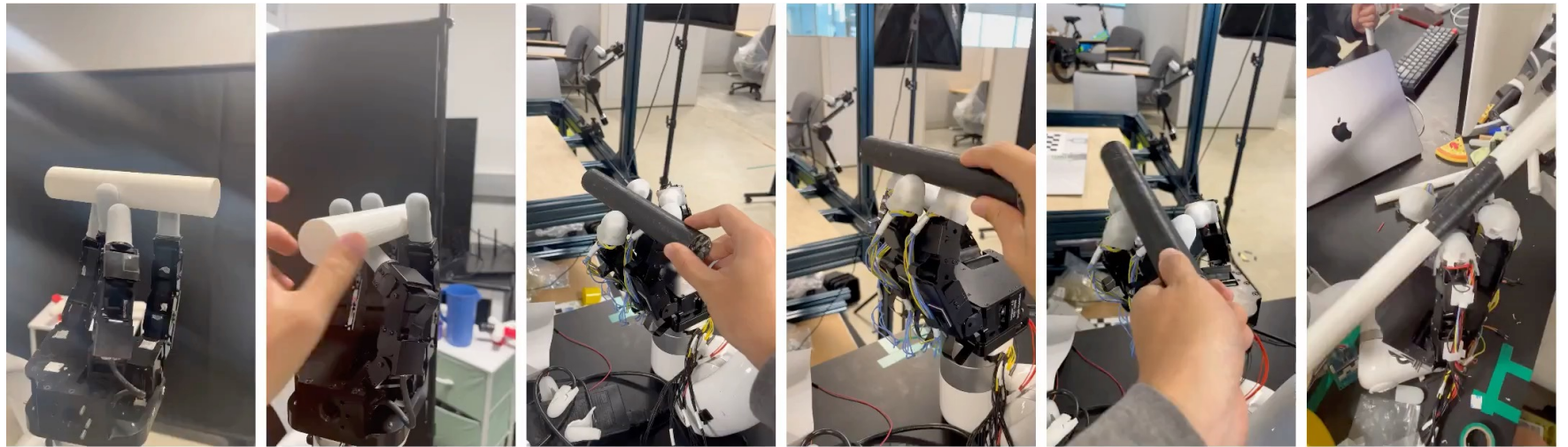


# Performing RL in Simulator Has Sim2Real Gap

Learn in Sim



Deploy in Real World

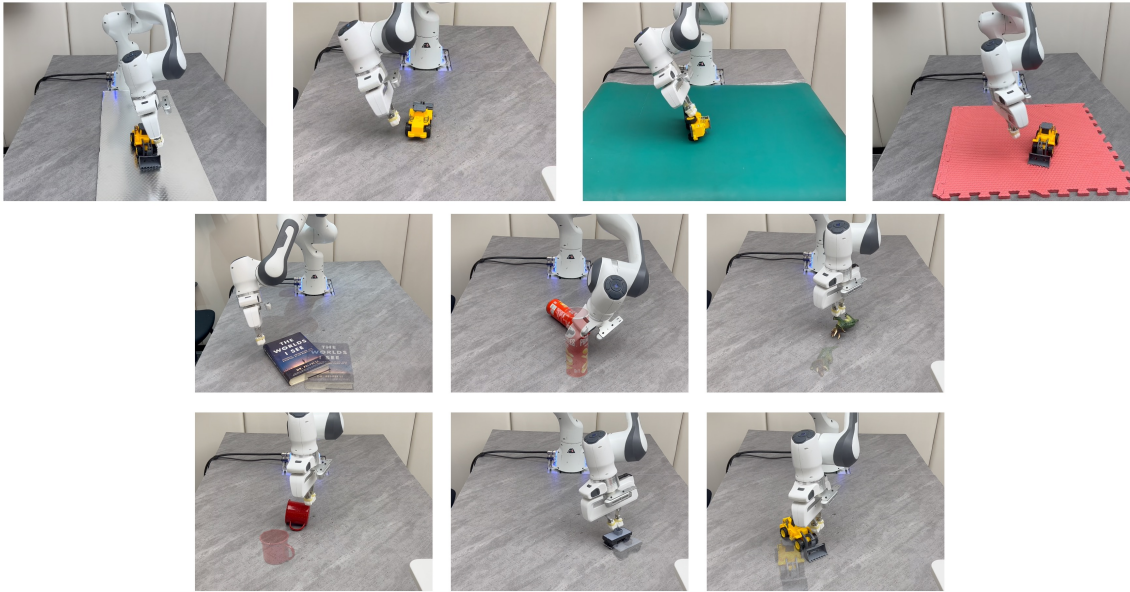


# Safety is Essential for Real-world RL

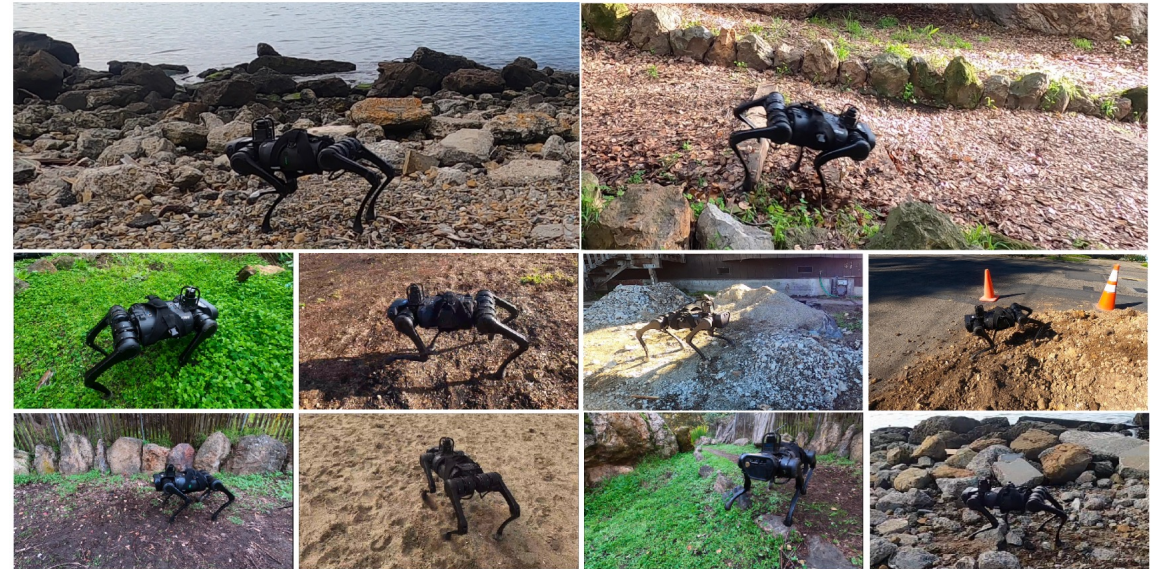




# Even Trained with RL, Policies are not Robust to Varying Dynamics

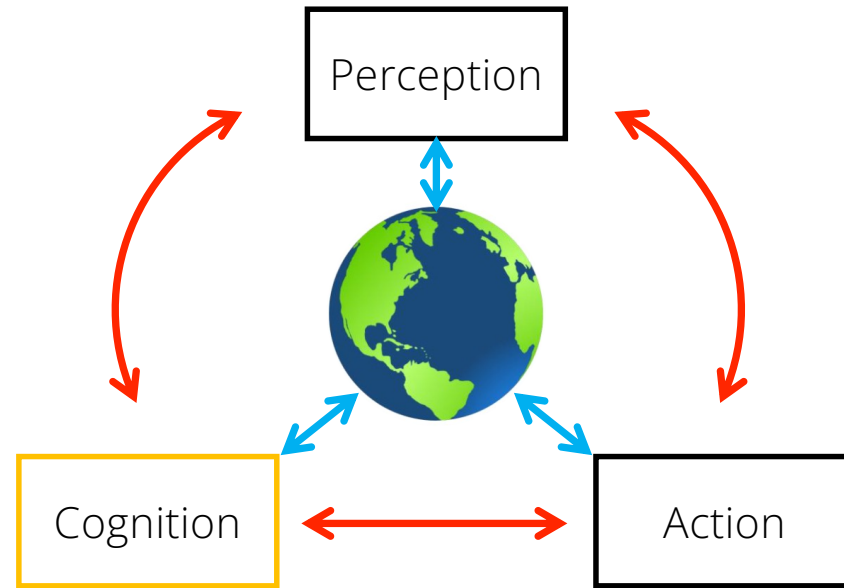


DyWA: Dynamics-adaptive World Action Model for Generalizable Non-prehensile Manipulation



RMA: Rapid Motor Adaptation for Legged Robots. Kumar et al.

# We will Focus on the Action and Cognition in this Course



In simplification, we want to answer questions: “What are our goals?” and “What action should we take to achieve **our goals** in the current environment?”.

# What Are the Goals? Who Decides the Goals?



Action: muscle contractions  
 Rewards: food, attentions  
 and ?

Image credit S. Levine



Action: muscle contractions  
Rewards: food and ?



Action: steering, acceleration  
Rewards: win



Action: selecting companies,  
buying, selling  
Rewards: cash and ?



Action: (x, y) location  
Rewards: win

[https://en.wikipedia.org/wiki/Go\\_\(game\)](https://en.wikipedia.org/wiki/Go_(game))



Action: jump,  $\uparrow \leftarrow \rightarrow \downarrow$   
 Rewards: win



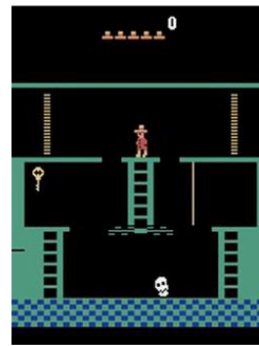
Action: RGB at each pixel  
Rewards: cash, aesthetics  
and ?



# Tasks with Sparse Rewards May be Difficult...



## Montezuma's revenge



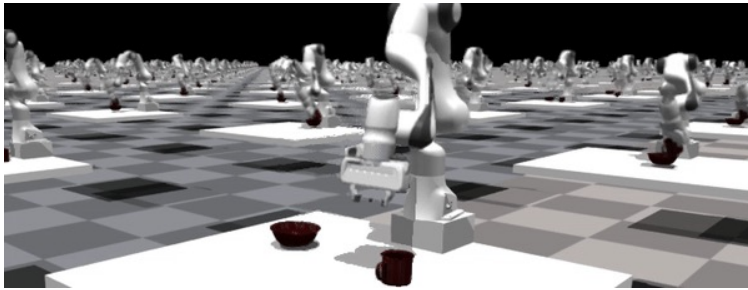
Slide credit S. Levine

- Getting key = reward
- Opening door = reward
- Getting killed by skull = nothing (is it good? bad?)
- Finishing the game only weakly correlates with rewarding events
- We know what to do because we **understand** what these sprites mean!

This is also called “credit assignment problem”—which actions attribute to the final success

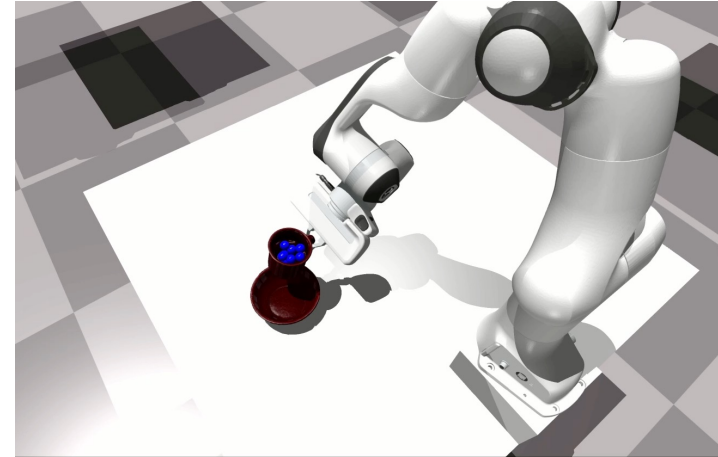


# Tasks with Sparse Rewards May be Difficult...



Policy fails when optimizing with sparse rewards

$$r = \begin{cases} 1, & \text{if pour water in the bowl} \\ 0, & \text{otherwise.} \end{cases}$$



Policy fails when optimizing with staged rewards

$$r_{reach} = \begin{cases} 1, & \text{if reach the mug.} \\ 0, & \text{otherwise.} \end{cases}$$

$$r_{lift} = \begin{cases} 1, & \text{if lift up the mug.} \\ 0, & \text{otherwise.} \end{cases}$$

$$r_{pour} = \begin{cases} 1, & \text{if pour the water in the bowl} \\ 0, & \text{otherwise.} \end{cases}$$

How is it even possible to cook these dishes with only sparse rewards...



# Define Goal with Prior Knowledge: Human Feedback

## Bag Packing

High-level Language Policy

- "pick up the sponge" ✓
- "put the sponge into the bag" ✗
- "use the sponge to open the bag" ⚙️
- "put the sponge into the bag" ✓
- "pick up the sharpie" ✓
- "put it into the bag" ✓

Low-level Language Conditioned BC Policy



"pick up the tape holder" ✗ "go lower" ⚙️ "put it into the bag" ✗ "rotate the tape holder" ⚙️ "drop the tape holder" ✓ "shake the bag" ✓

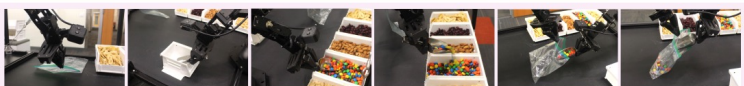


## Trail Mix Preparation

High-level Language Policy

- "pick up the bag" ✓
- "pick up the metal scoop" ✓
- "I want some M&Ms" ✓
- "I want less" ⚙️
- "open the bag with scoop" ✓
- "pour it into the bag" ✓

Low-level Language Conditioned BC Policy



"I want some peanuts" ✓ "pour into the bag" ✓ "I want some almonds" ✓ "pour into the bag" ✗ "rotate the scoop" ⚙️ "pour into the bag" ✓

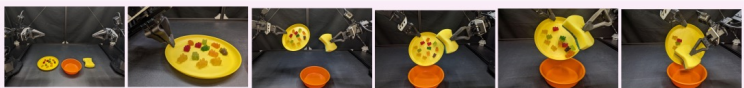


## Plate Cleaning

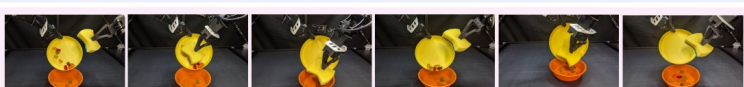
High-level Language Policy

- "pick up the plate" ✓
- "pick up the plate" ✓
- "pick up the sponge" ✓
- "move the plate to the bowl" ✓
- "clean the plate" ✓
- "clean the right side" ⚙️

Low-level Language Conditioned BC Policy



"clean the right side again" ✗ "wipe the left side" ⚙️ "clean the left side again" ⚙️ "clean the left" ✗ "clean the bottom" ⚙️ "clean the bottom" ✓



Yell at your robot. Shi et al.

### Step 1

Collect demonstration data and train a supervised policy.

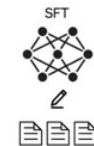
A prompt is sample from our prompt dataset.

Explain reinforcement learning to a 6 year old.

A labeler demonstrates the desired output behavior.

We give treats and punishments to teach...

This data is used to fine-tune GPT-3.5 with supervised learning.

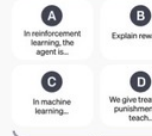


### Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

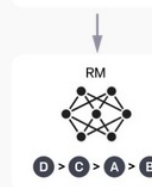
Explain reinforcement learning to a 6 year old.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



### Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.

Write a story about otters.

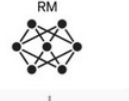
The PPO model is initialized from the supervised policy.



The policy generates an output.

Once upon a time...

The reward model calculates a reward for the output.

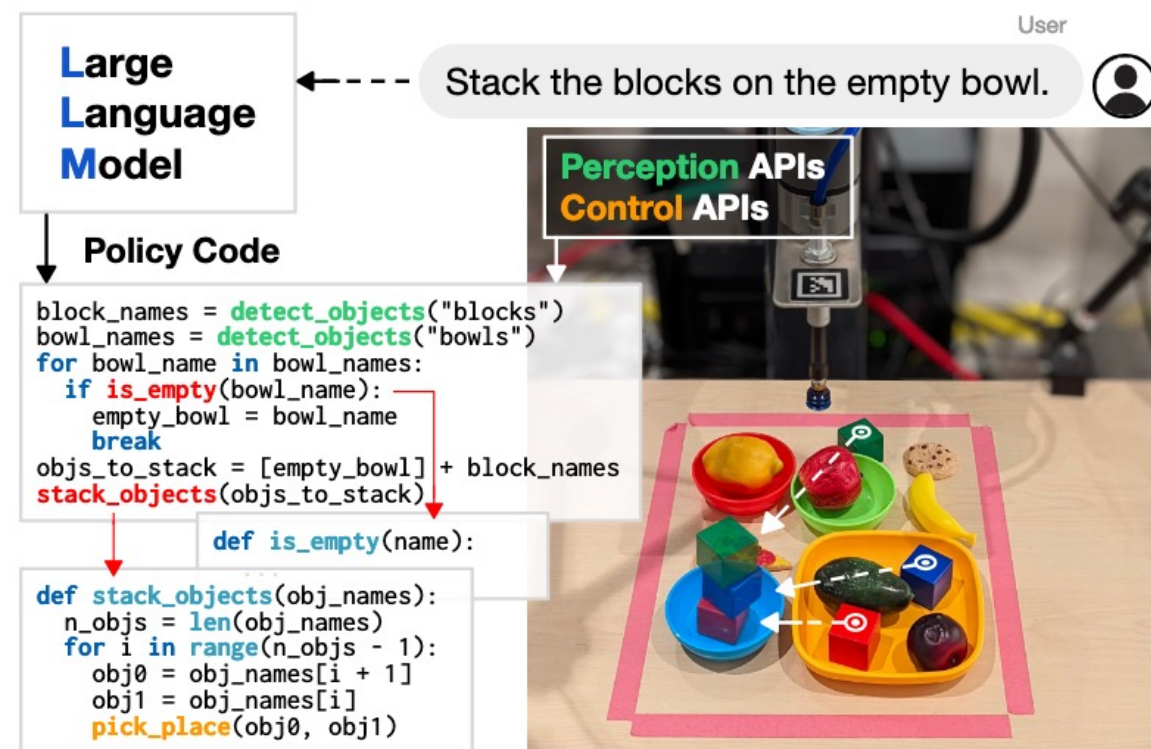


The reward is used to update the policy using PPO.

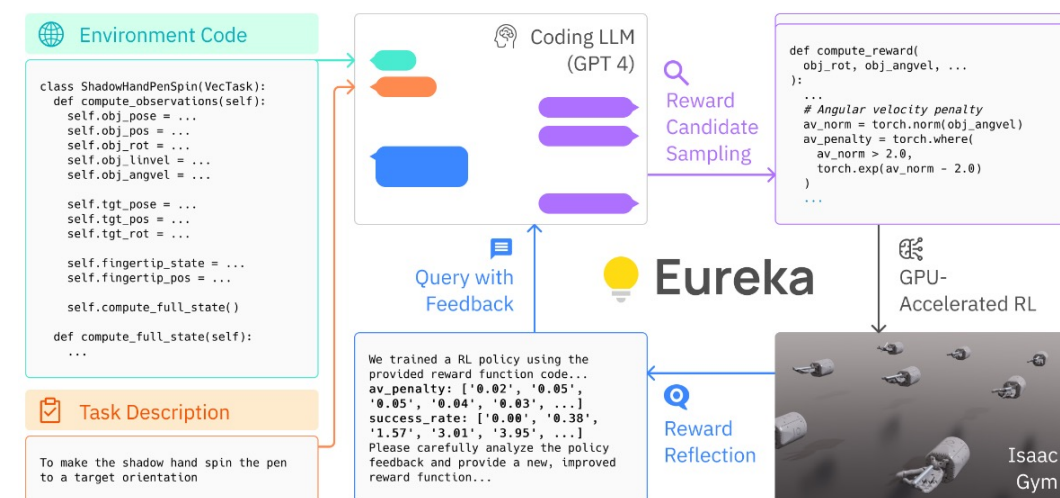
$r_k$

<https://openai.com/index/chatgpt/>

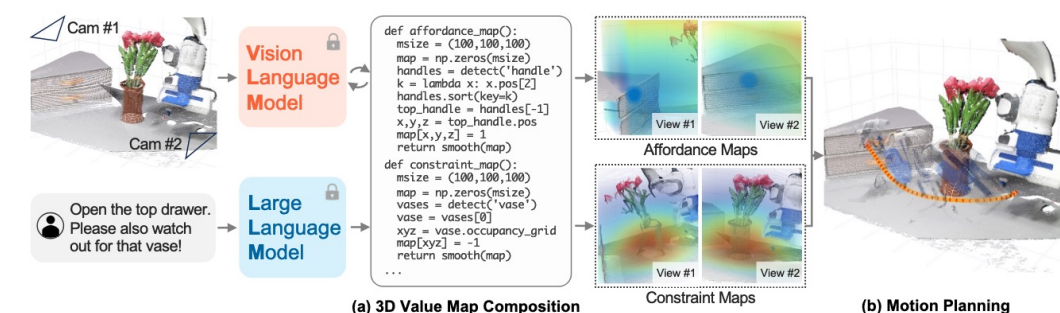
# Define Goal with Prior Knowledge: Large Language Models



Code as Policies: Language Model Programs for Embodied Control. Liang et al. ICRA 2023



Eureka: Human-Level Reward Design via Coding Large Language Models. Ma et al



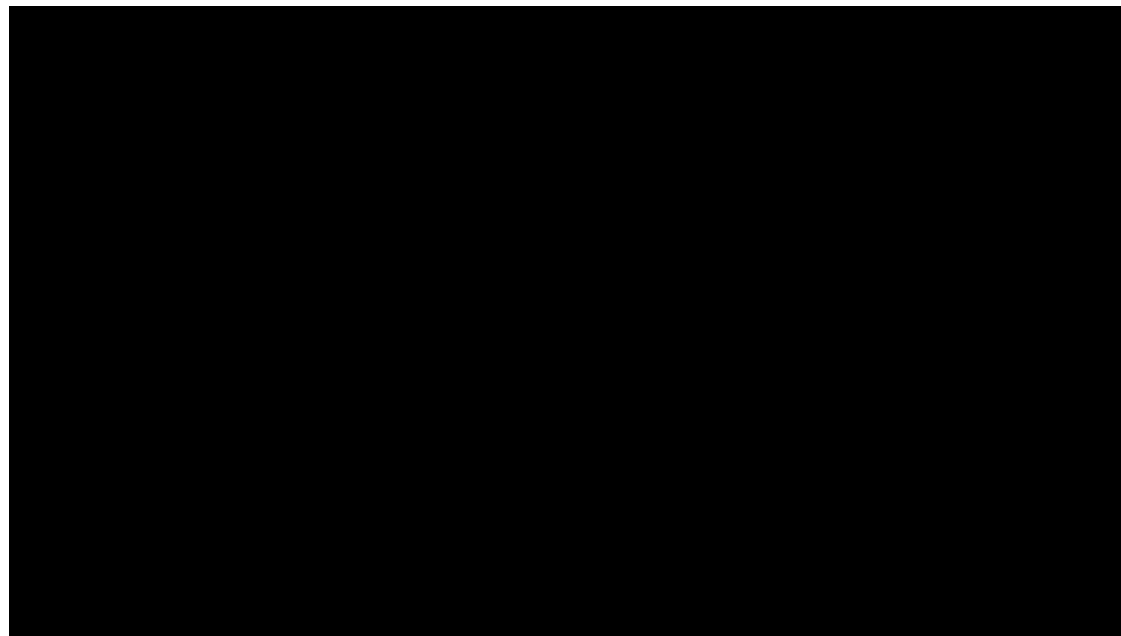
VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models. Huang et al



# Define Goal with Prior Knowledge: Human Behaviors

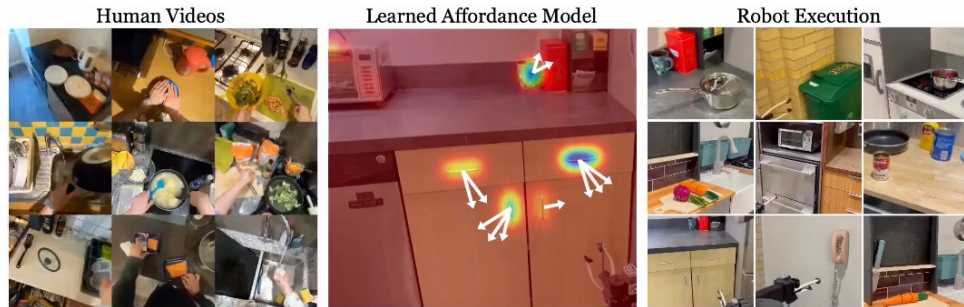


Ego-Exo4D dataset

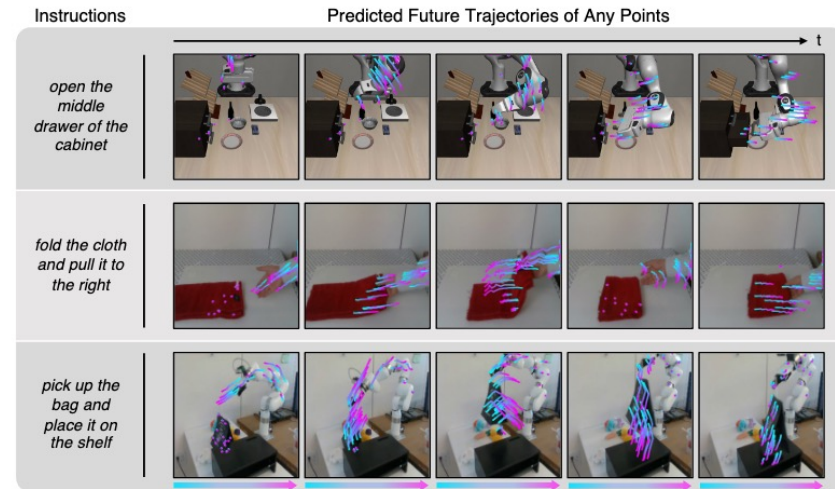


HOT3D dataset

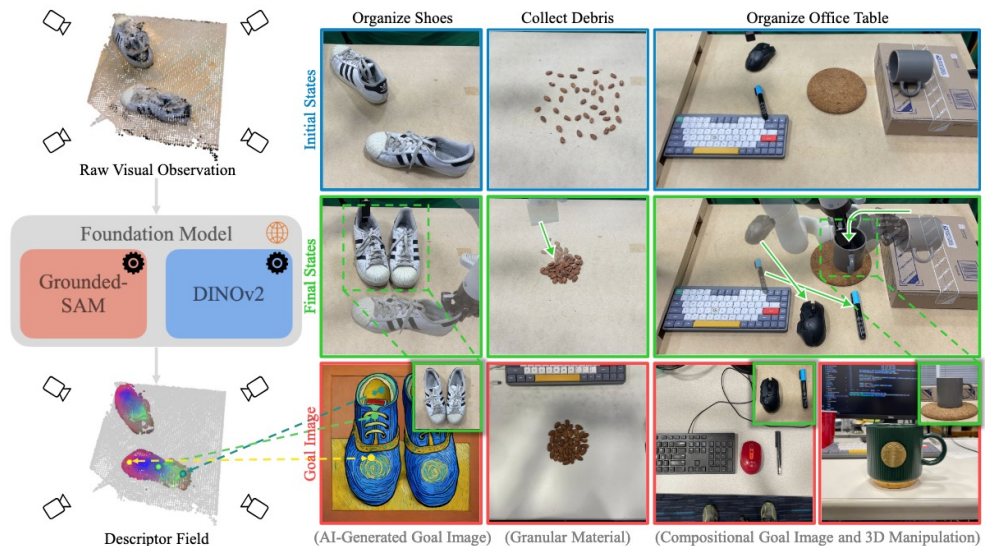
# How to Transfer the Knowledge from Human Demonstrations?



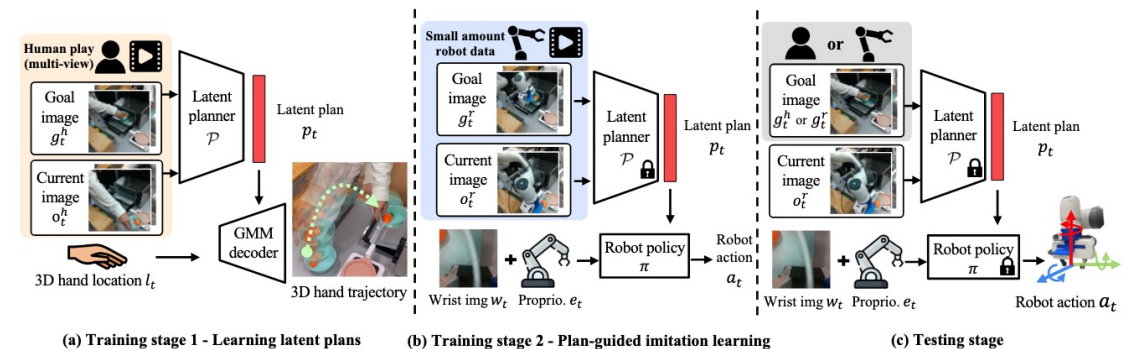
Affordances from Human Videos as a Versatile Representation for Robotics. Bahl et al.



Any-point Trajectory Modeling for Policy Learning. Wen et al.



D3Fields: Dynamic 3D Descriptor Fields for Zero-Shot Generalizable Robotic Manipulation. Wang et al.



MimicPlay: Long-Horizon Imitation Learning by Watching Human Play. Wang et al.



# What Are the Goals? Who Decides the Goals?



Action: muscle contractions  
Rewards: food, attentions  
and ?

Image credit S. Levine



Action: muscle contractions  
Rewards: food and ?

<https://www.salon.com/2021/01/16/human-breeding-of-cats-has-made-them-like-they-are-always-in-pain/>



Action: steering, acceleration  
Rewards: win

<https://racingnews365.com/championship-standings-after-2024-f1-australian-grand-prix>



Action: selecting companies,  
buying, selling  
Rewards: cash and ?



Action: (x, y) location  
Rewards: win

[https://en.wikipedia.org/wiki/Go\\_%28game%29](https://en.wikipedia.org/wiki/Go_%28game%29)



Action: jump,  $\uparrow \leftrightarrow \downarrow$   
Rewards: win

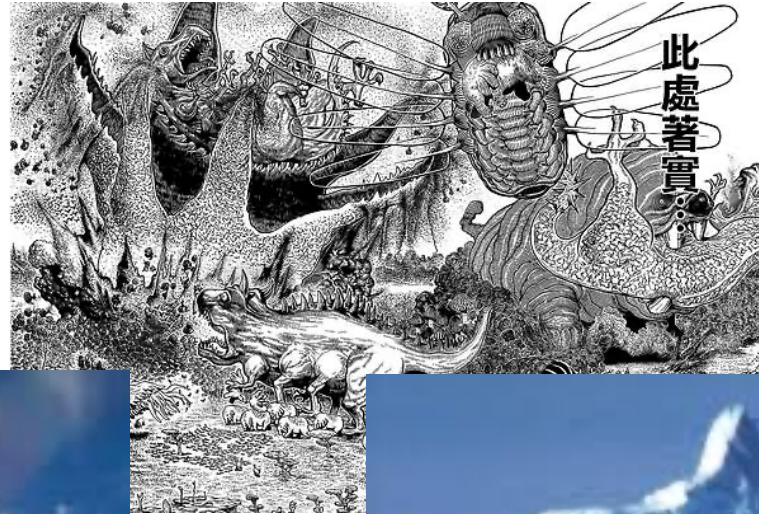
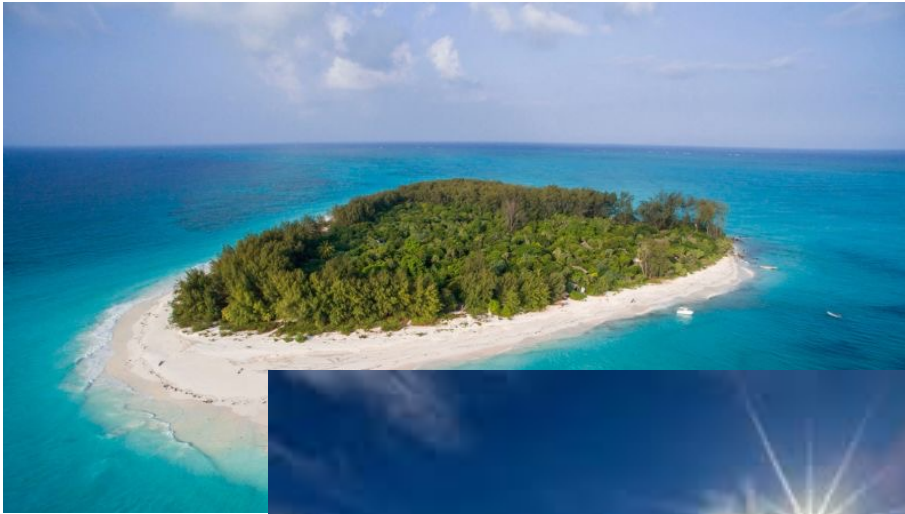
Image credit Nintendo



Action: RGB at each pixel  
Rewards: cash, aesthetics  
and ?



# Probably Survival is One Common Goal for All Creatures



# Curiosity May also be an Important Goal



# Learning from Babies

- *Be multi-modal*
- *Be incremental*
- *Be physical*
- *Explore*
- *Be social*
- *Learn a language*



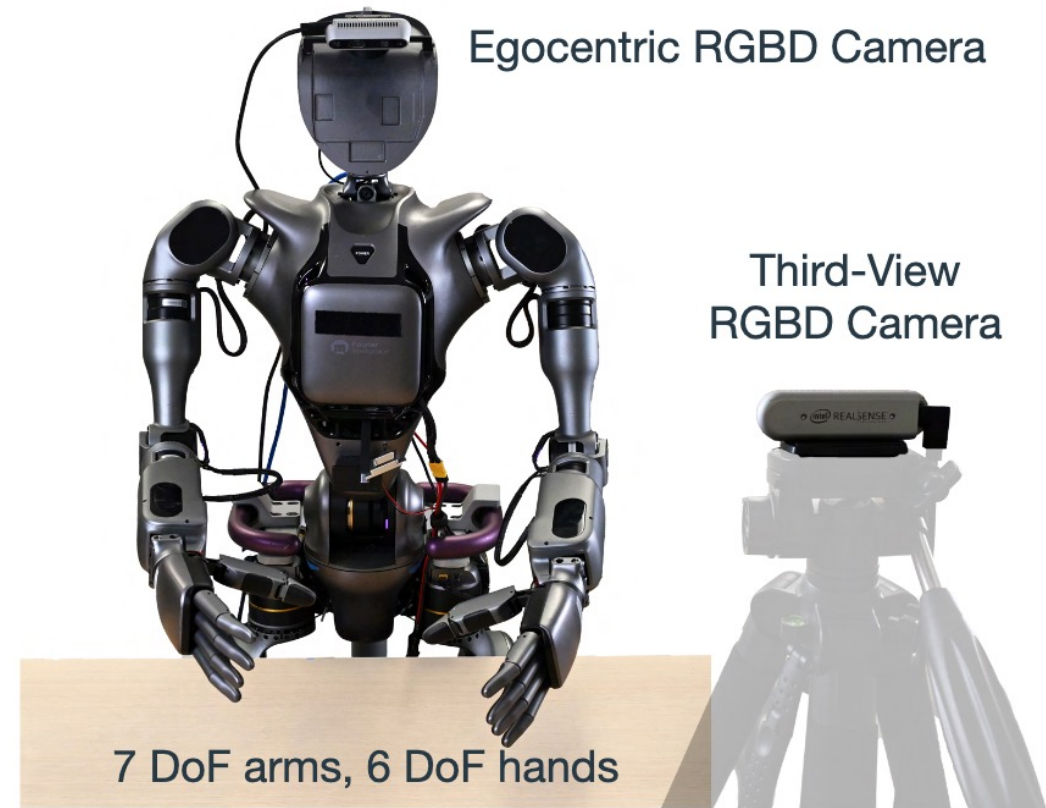
# What Kind of Robots Do We Need?



# Jaw-Gripper Robots are Simple but Incompetent



# To Perform Daily Tasks, We Need Multi-Arm Multi-Fingered Robots...



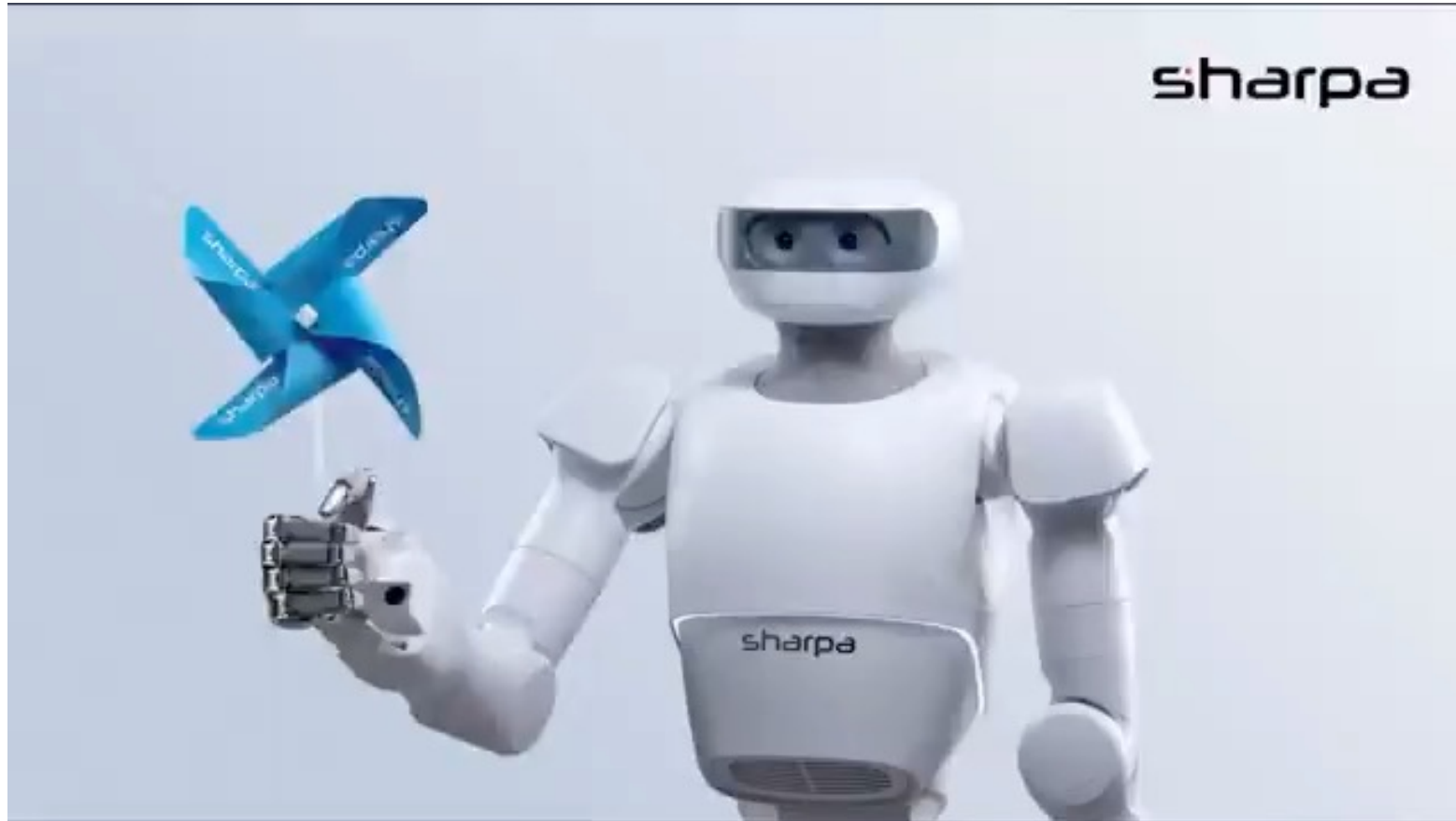
Sim-to-Real Reinforcement Learning for Vision-Based Dexterous Manipulation on Humanoids. Lin et al.



# Hand-Leg Coordination with Humanoid Robots



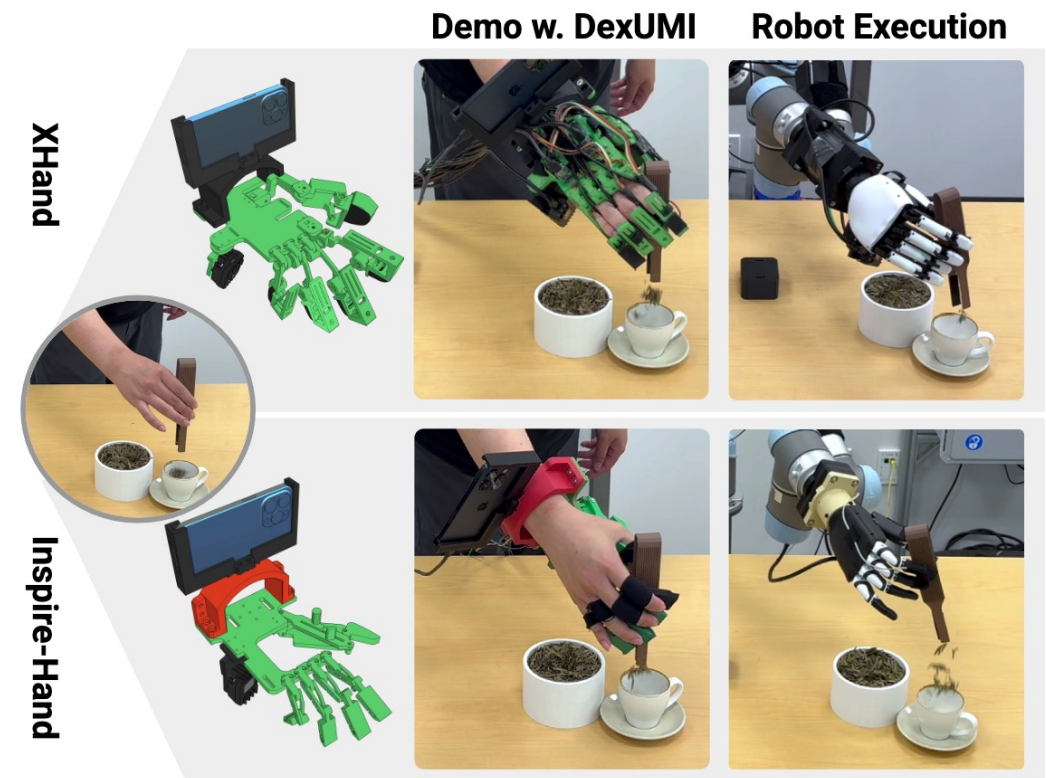
# Dexterous Manipulation with Multi-finger Hands



# Collecting Expert Demonstrations for Multi-arm Multi-Fingered Robots is Expensive...

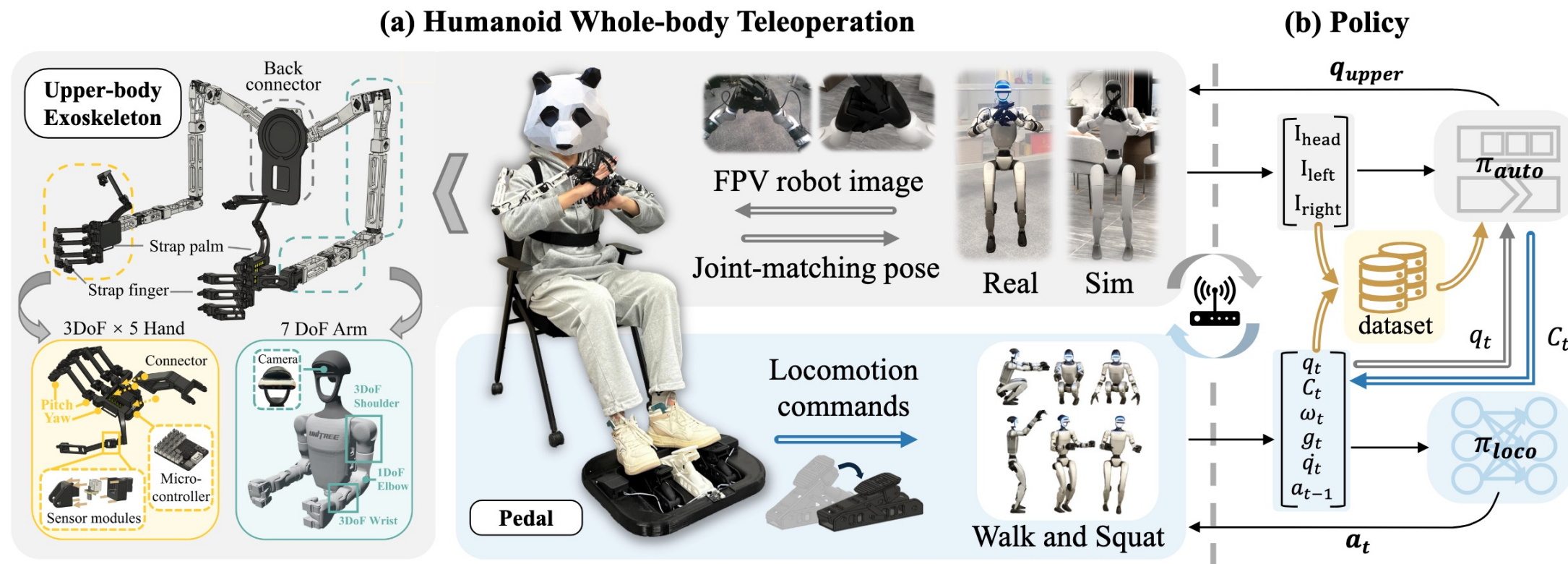


<https://youtu.be/Bhg3uOx9ZPw?si=et7L0endzGvGPJz->



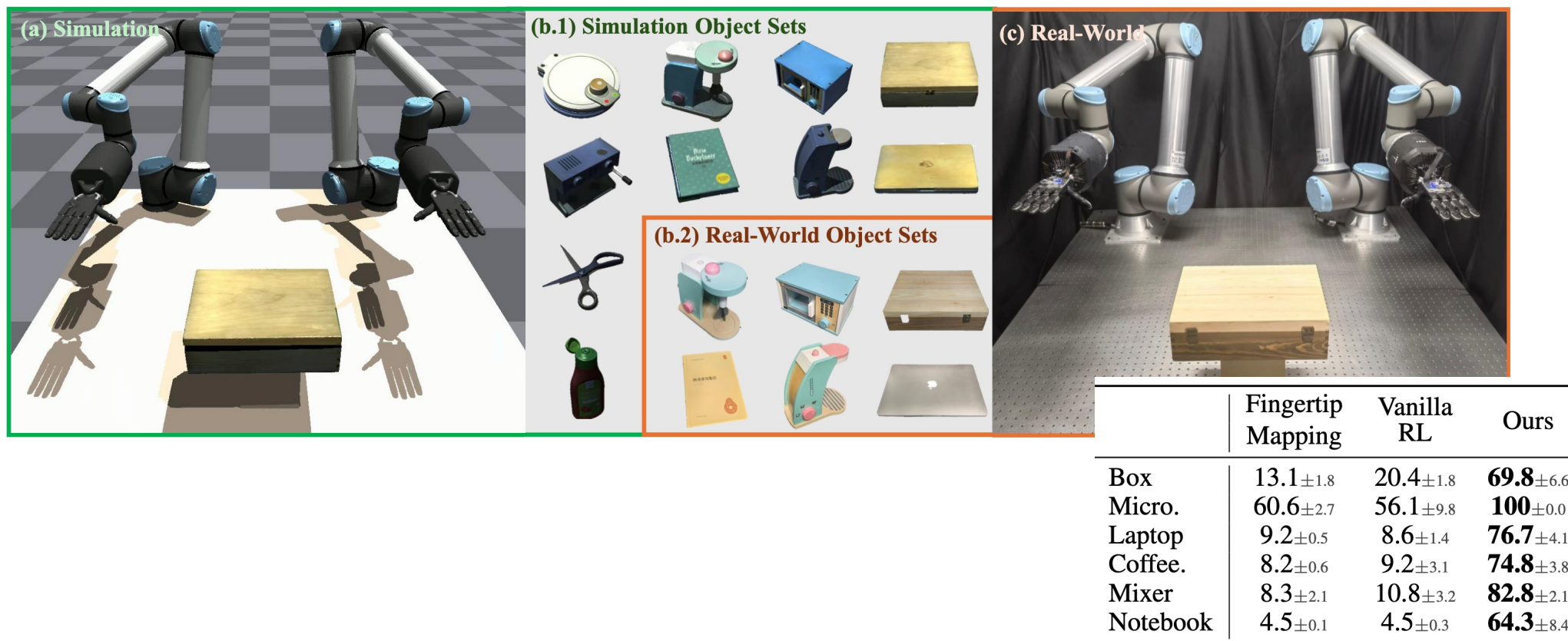
DexUMI: Using Human Hand as the Universal Manipulation Interface for Dexterous Manipulation

# Collecting Expert Demonstrations for Multi-arm Multi-Fingered Mobile Robots is Even More Expensive...





# RL Don't Work Out-of-Box with High-Dimensional Actions

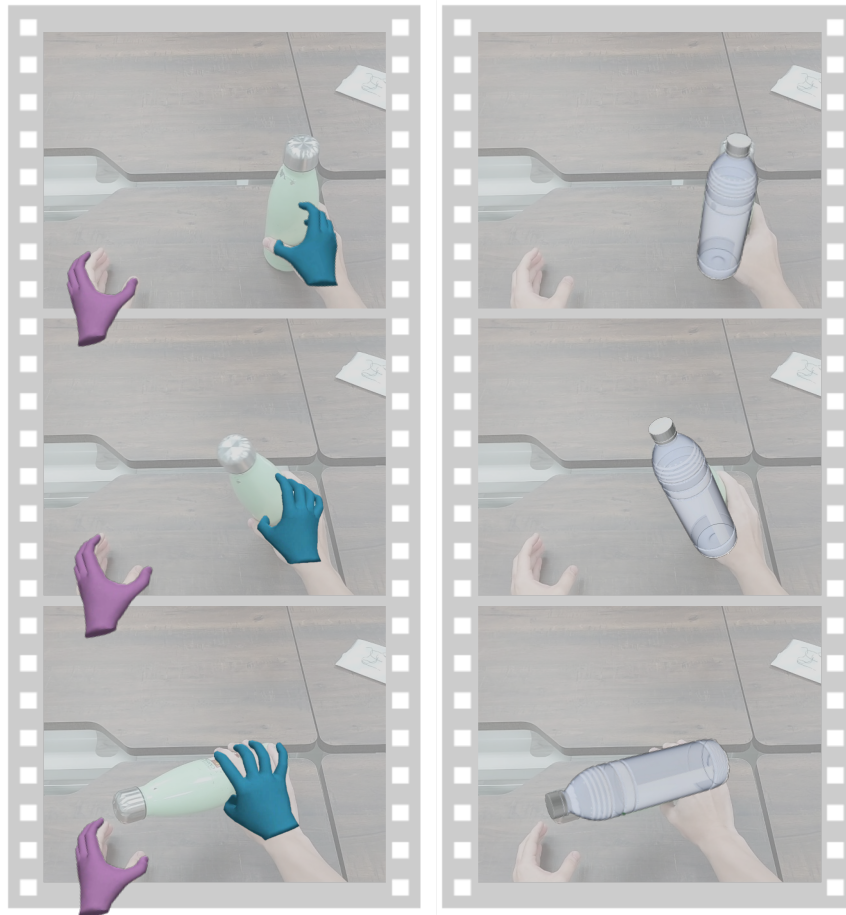




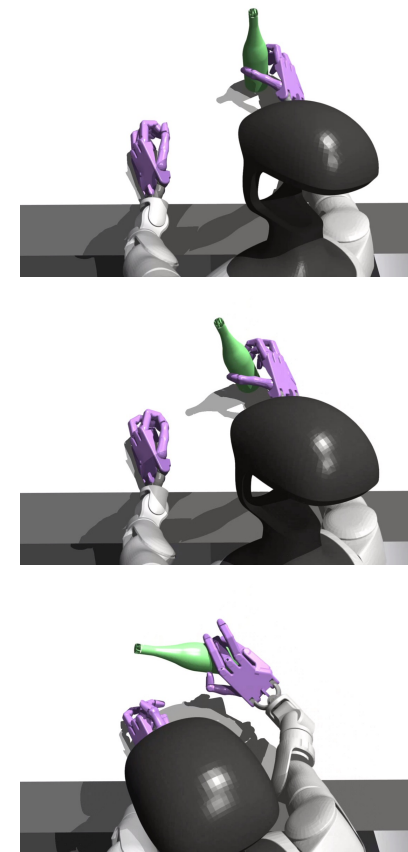
# Training Multi-arm Multi-Fingered Robots with Human Motion Priors



Input: RGB human videos

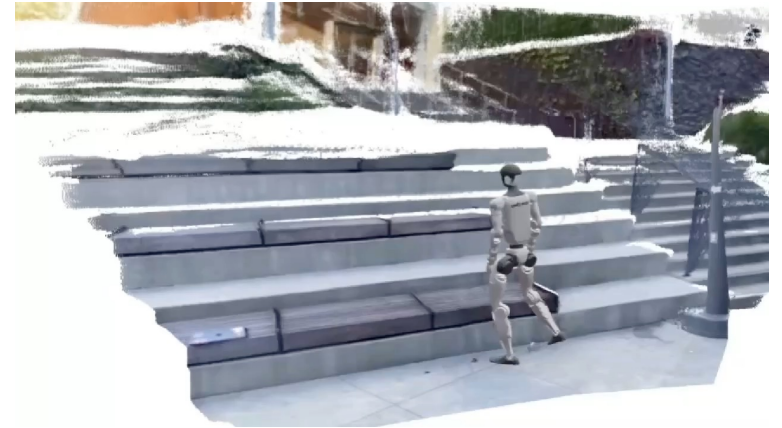


Extract motion priors and task rewards



Learn in digital twins

# Training Humanoid Robots with Human Motion Priors



What we hope you to know after this class?

# From Seeing to Acting, What are Under the Hood?



Optimus by Tesla

- What are the goals? How to obtain the goal?
- How to achieve the goal with low-level actions?
- What are the essential components to control robots?
- What kind of robots do we need?
- What are the limitations of SOTA robots?
- What are the open challenges to build a general robots?

# Reminder

- If you're interested in enrolling, please fill out and submit this form. We will first randomly sample 10-20 people for additional enrollment. We will release more slots if some students withdraw their enrollment



<https://forms.gle/xM2nRt774nT5kiKr6>

- Any application of course withdrawal after 10/30 **will not** be accepted!
- Also withdraw the course now if you have other plans!