2nd **Class** / Jan 13 (Mon)

Modern Robot Learning:
Hands-on Tutorial

Haoshu Fang, Younghyo Park, Jagdeep Bhatia, Lars Ankile, Pulkit Agrawal

Last Week...

Course Overview

Hands-on Tutorial

Robot Data Collection

- What is **robot data**?
- What/how do we collect?
- How do we use it?

Policy Training

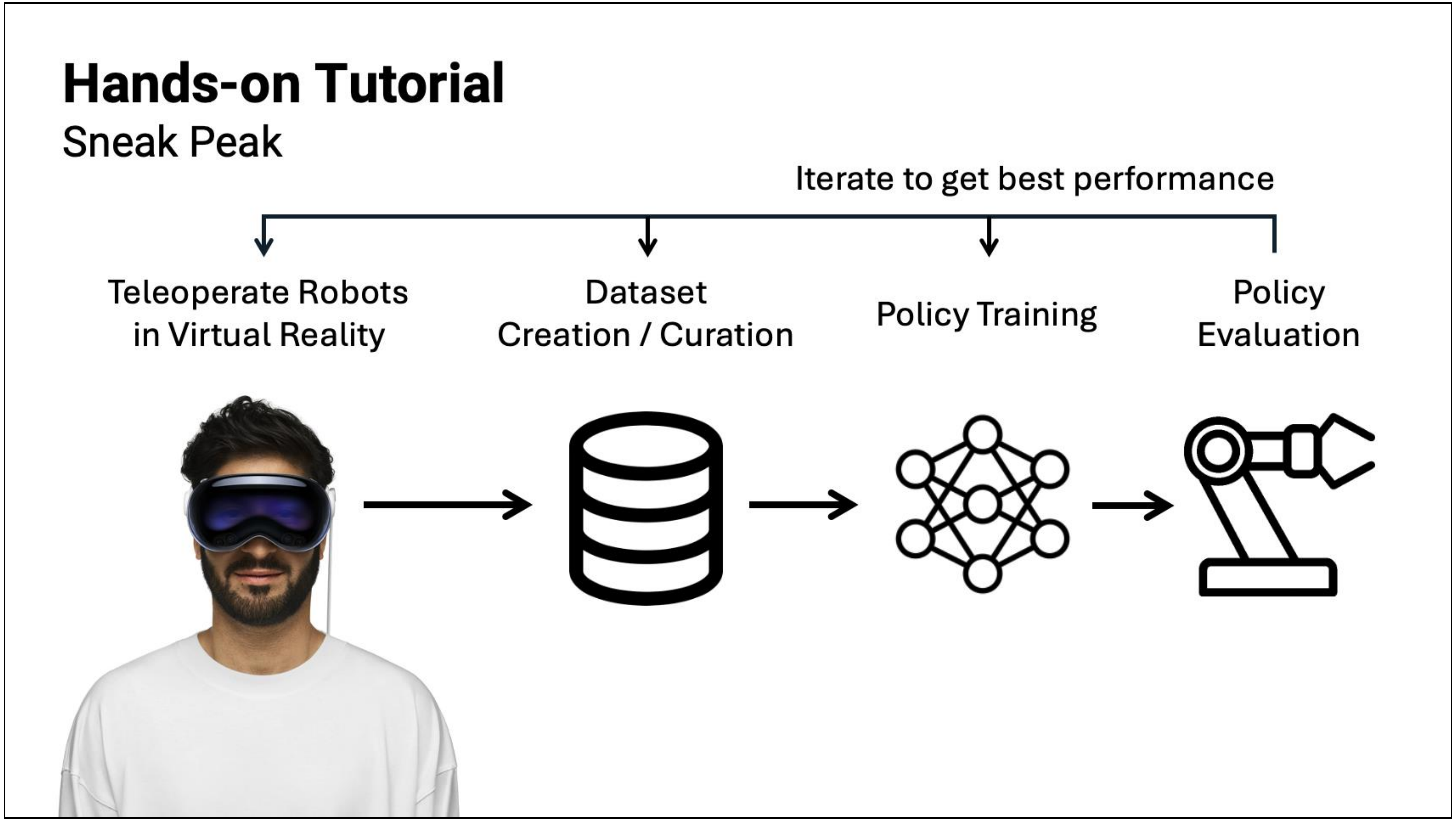
- Training Methods
- Policy Architectures
- Policy Evaluation

Simulation for Robotics

- Role of simulation
- Designing environments in simulation world
- Transfer to real-world

Course Overview

Last Week…



Hands-on Tutorial Sneak Peak

Last Week...

Stark contrast

artificial intelligence of various sorts
will become an accepted
part of daily life by the year 2020

Stanford Law School

GPT-4 Passes the Bar Exam: What That Means for Artificial Intelligence Tools in the Legal Profession | Stanford Law ...

CodeX–The Stanford Center for Legal Informatics and the legal technology company Casetext recently announced what they called "a watershed...

Apr 19, 2023

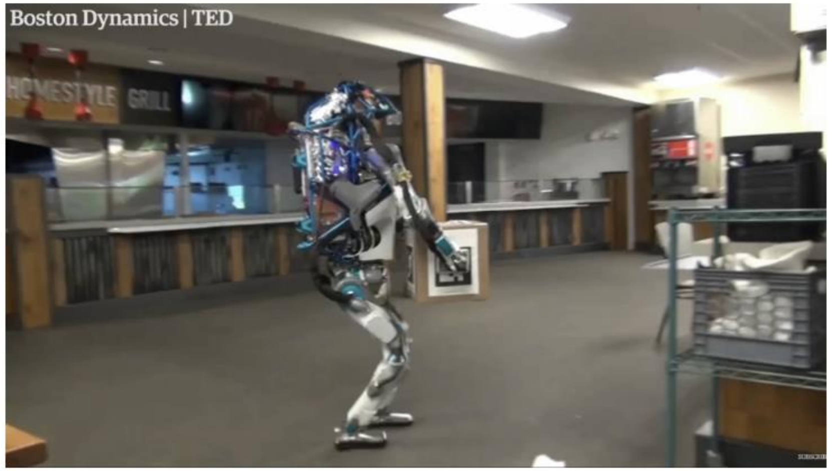
PCMag

ChatGPT Passes Google Coding Interview for Level 3 Engineer With $183K Salary

Google fed coding interview questions to ChatGPT and, based off the AI's answers, determined it would be hired for a level three engineering...

Feb 1, 2023

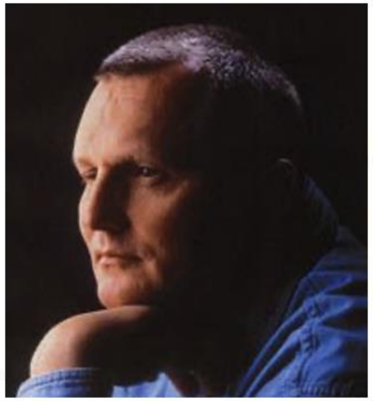
robots
will almost completely take over physical work,



So ··· We can make machines **pass bar exams**, but cannot make it **move boxes**?

Non-Physical vs *Physical* *Intelligence*

Last Week...



Hans Moravec

"<u>reasoning</u> requires very little computation, but <u>sensorimotor and perception</u> skills require enormous computation resources" (1980)

Moravec's Paradox

Steven Pinker

"… the main lesson of 35 years of research is that the <u>hard problems are easy</u> and the <u>easy problems are hard</u> … " (1994)

Slide from Pulkit Agrawal

Moravec's Paradox

Last Week...

strategy: **massive dataset** **with the** **right training method**

	massive dataset	right training method
Vision/Language Models	Scraped datasets from the web	Next token (word) Prediction
Robot Models	**?**	Action Prediction

Why data matters for generalist robot intelligence models

Last Week...

Road to <u>Large-Scale Robot Dataset</u>

What's a **Robot Dataset**?

- Data recorded by **robot embodiments** solving diverse tasks in real-world.
- Any data from **any embodiments** (including humans) that contains useful knowledge about manipulation strategies.

Two types of robot datasets

Last Week...

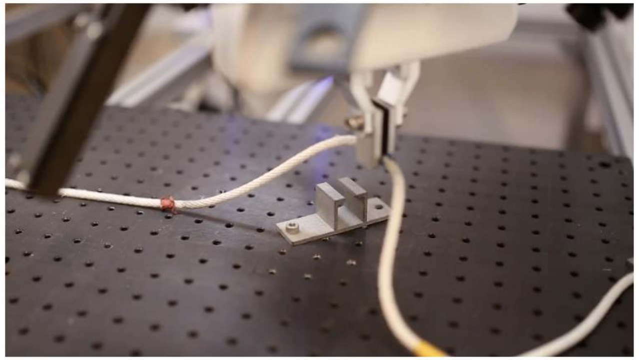
Two types of robot datasets

Last Week...

Road to Large-Scale Robot Dataset

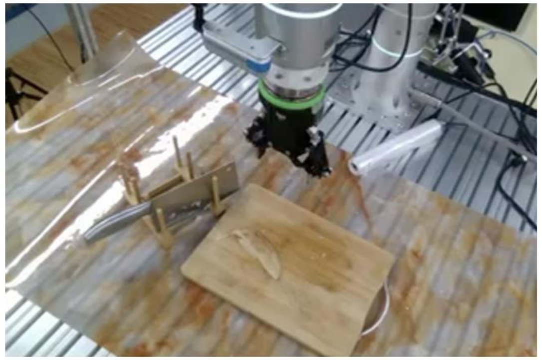
What's a **Robot Dataset**?

- Data recorded by **robot embodiments** solving diverse tasks in real-world.



O'Neill, Abby, et al. "**Open x-embodiment**: Robotic learning datasets and rt-x models." *arXiv:2310.08864* (2023).

Khazatsky, Alexander, et al. "**DROID**: A large-scale in-the-wild robot manipulation dataset." *arXiv preprint arXiv:2403.12945* (2024).



Fang, Hao-Shu, et al. "**RH20t**: A robotic dataset for learning diverse skills in one-shot." *RSS 2023 Workshop on Learning for Task and Motion Planning*. 2023.

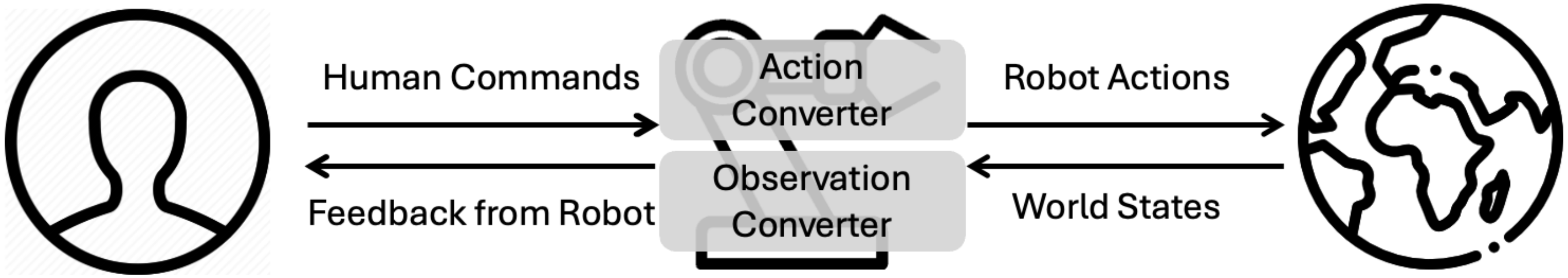
Two types of robot datasets

Last Week...

Robot Teleoperation

4 Key Elements of Teleoperation System

1. **Designing** command space for humans
2. **Converting** commands to robot actions
3. **Designing** feedback space for humans
4. **Converting** robot perceptions to human feedback

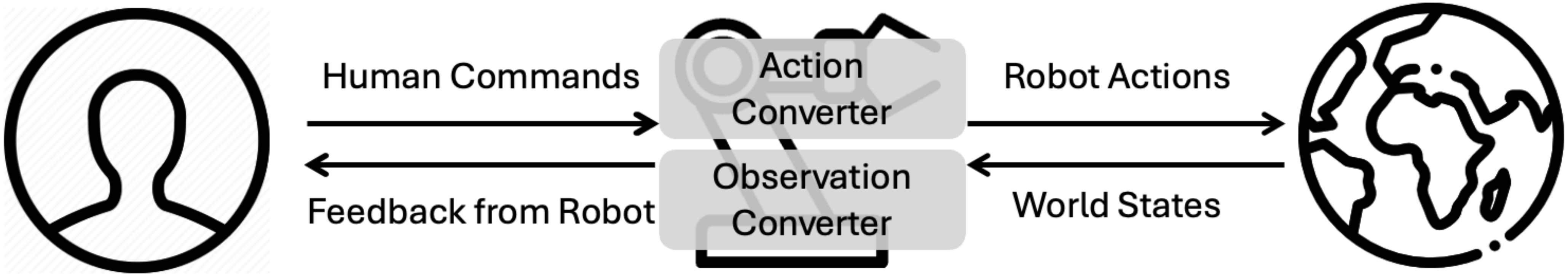


*Most of the robot datasets are created by **"teleoperation"***

Today...

4 Key Elements of Teleoperation System

1. **Designing** command space for humans
2. **Converting** commands to robot actions
3. **Designing** feedback space for humans
4. **Converting** robot perceptions to human feedback



Teleoperation System Case Studies: In-Depth Analysis

Today…

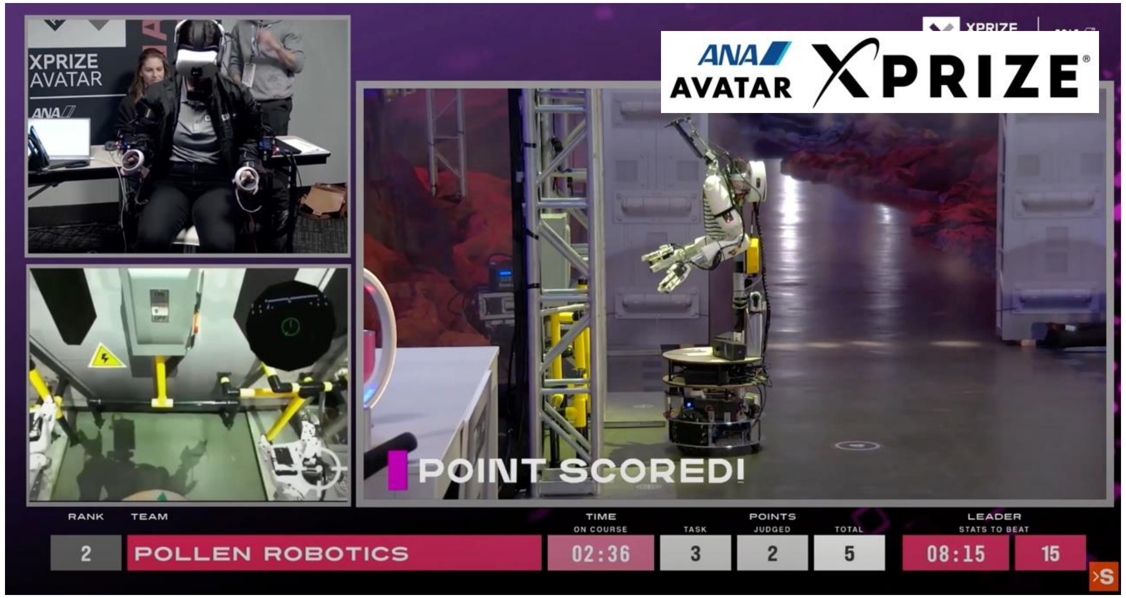
- Teleoperation System Case Studies: In-Depth Analysis
- Policy Training with Teleoperated Datasets
  - Policy Architectures
  - Policy Training Methods

Today…

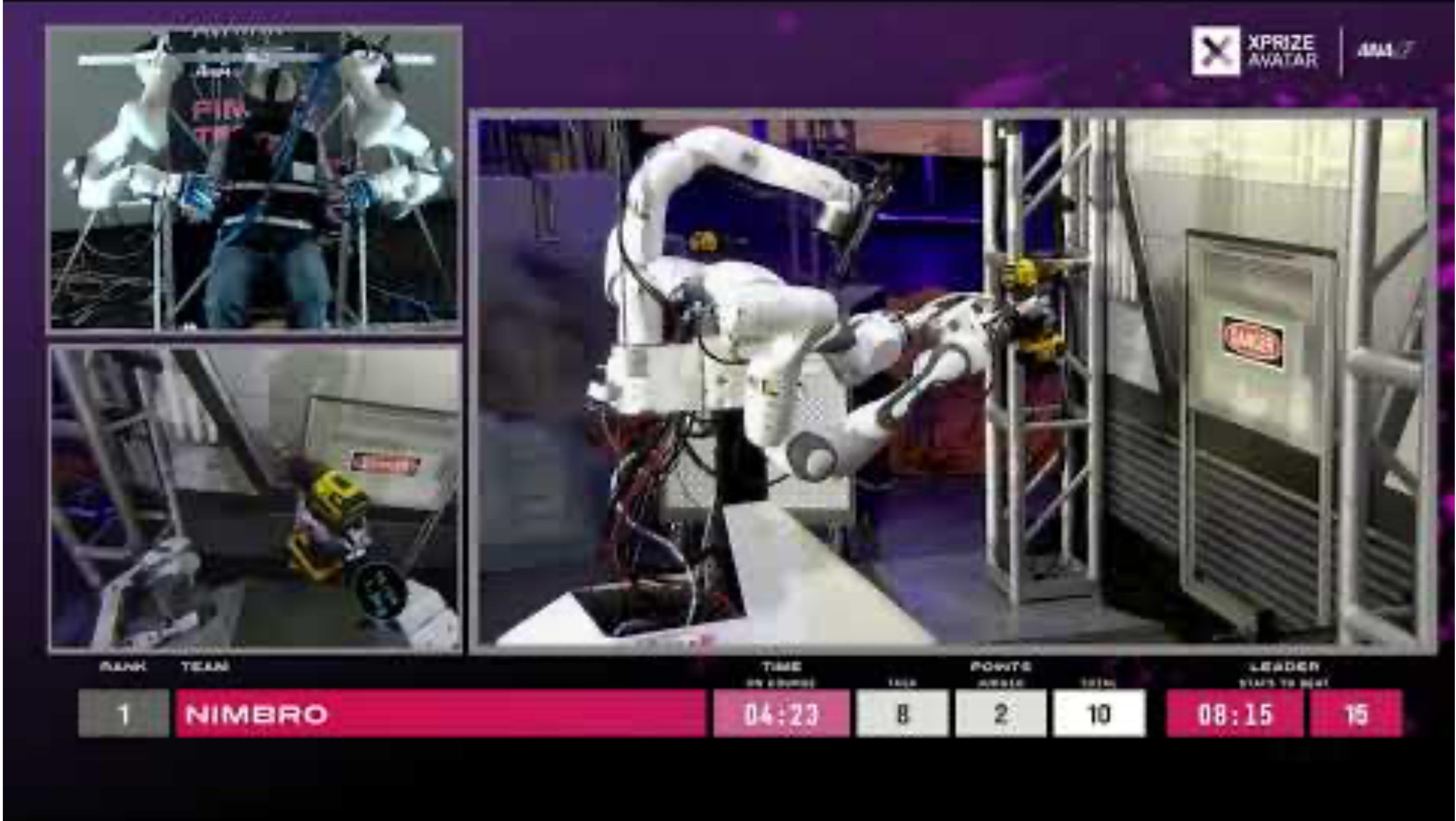
- Teleoperation System Case Studies: In-Depth Analysis
- Policy Training with Teleoperated Datasets
  - Policy Architectures
  - Policy Training Methods
- Role of Simulation
  - Real2Sim: Simulation Environment Design
  - Sim2Real

Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

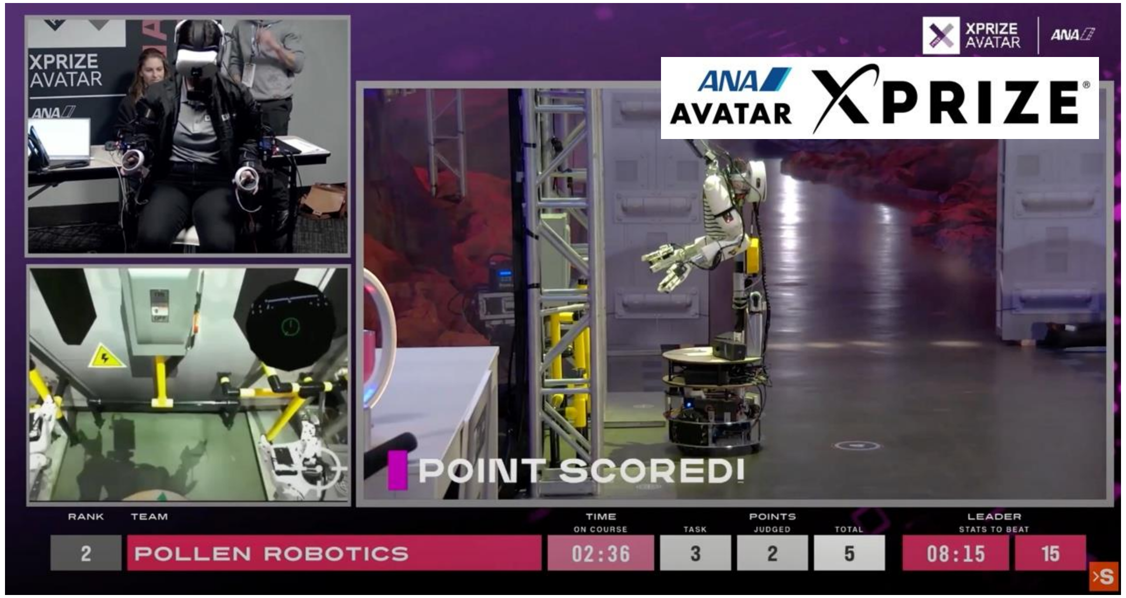


[B] ALOHA

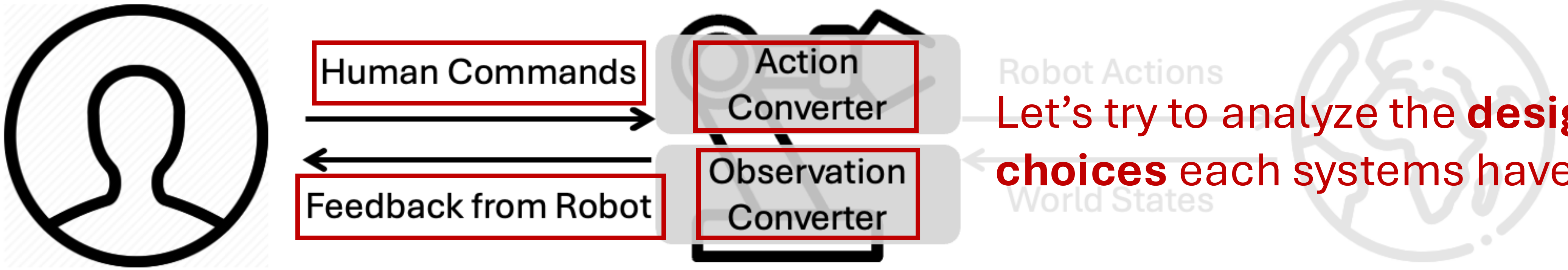

XPRIZE
AVATAR
RANK
TEAM
TIME
POINTS
LEADER
1
NIMBRO
04:23
8
2
10
08:15
15

Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize



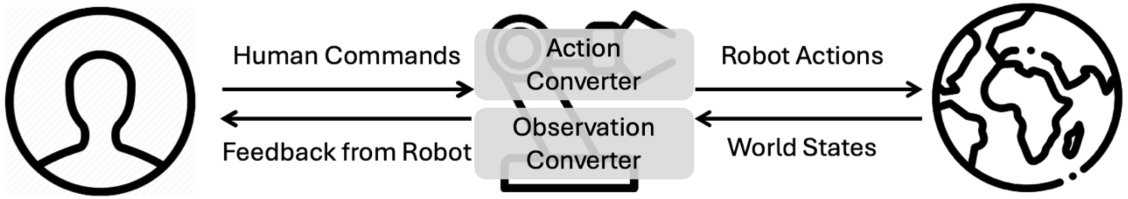
[B] ALOHA



Let's try to analyze the **design choices** each systems have made!

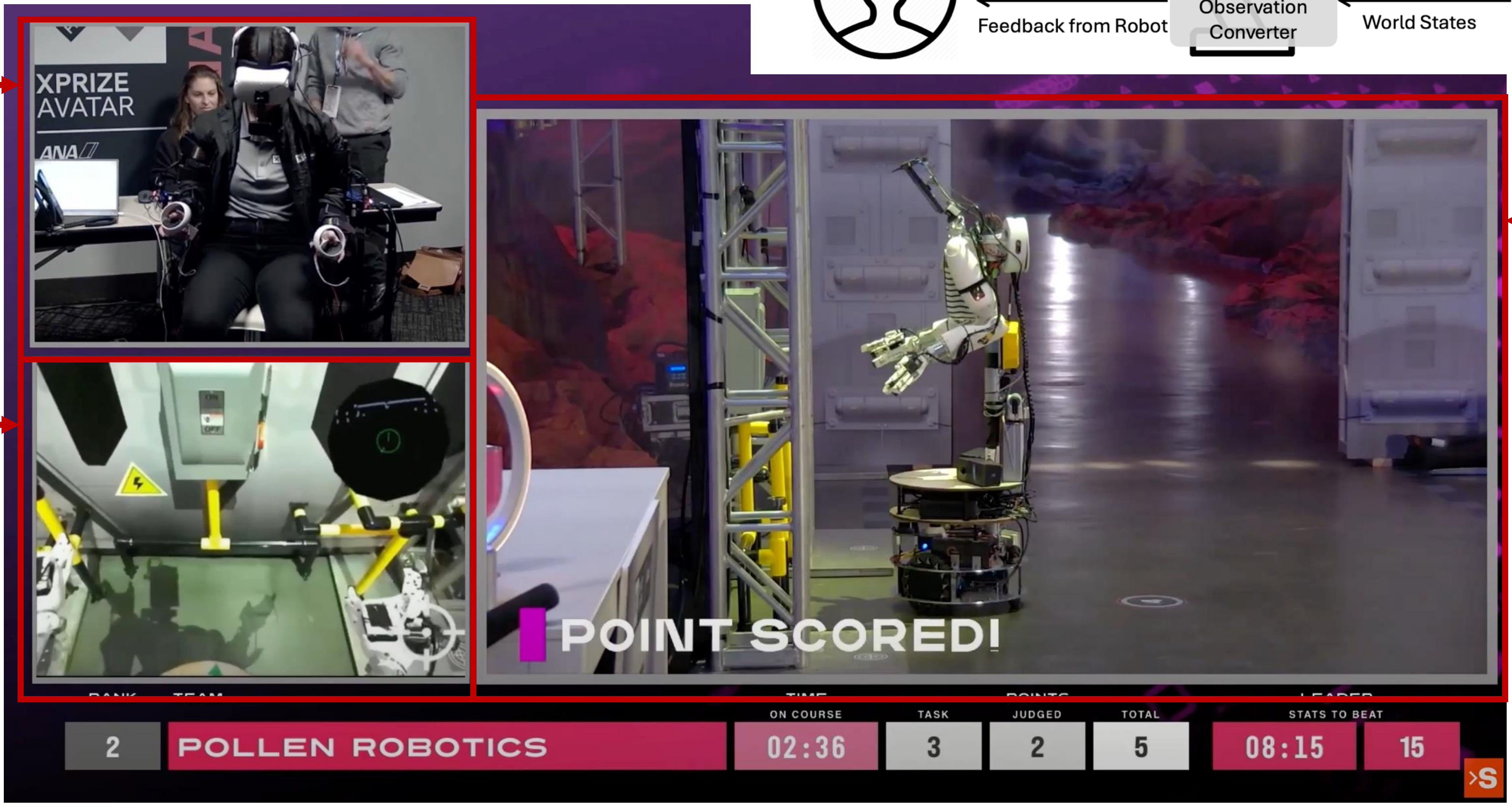
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize



What Human Does

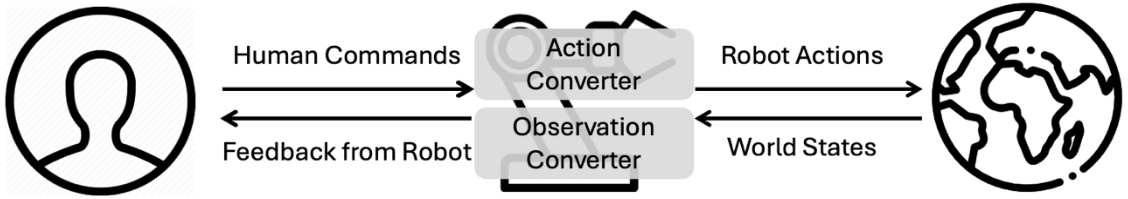
What Human Sees



What Robot Does

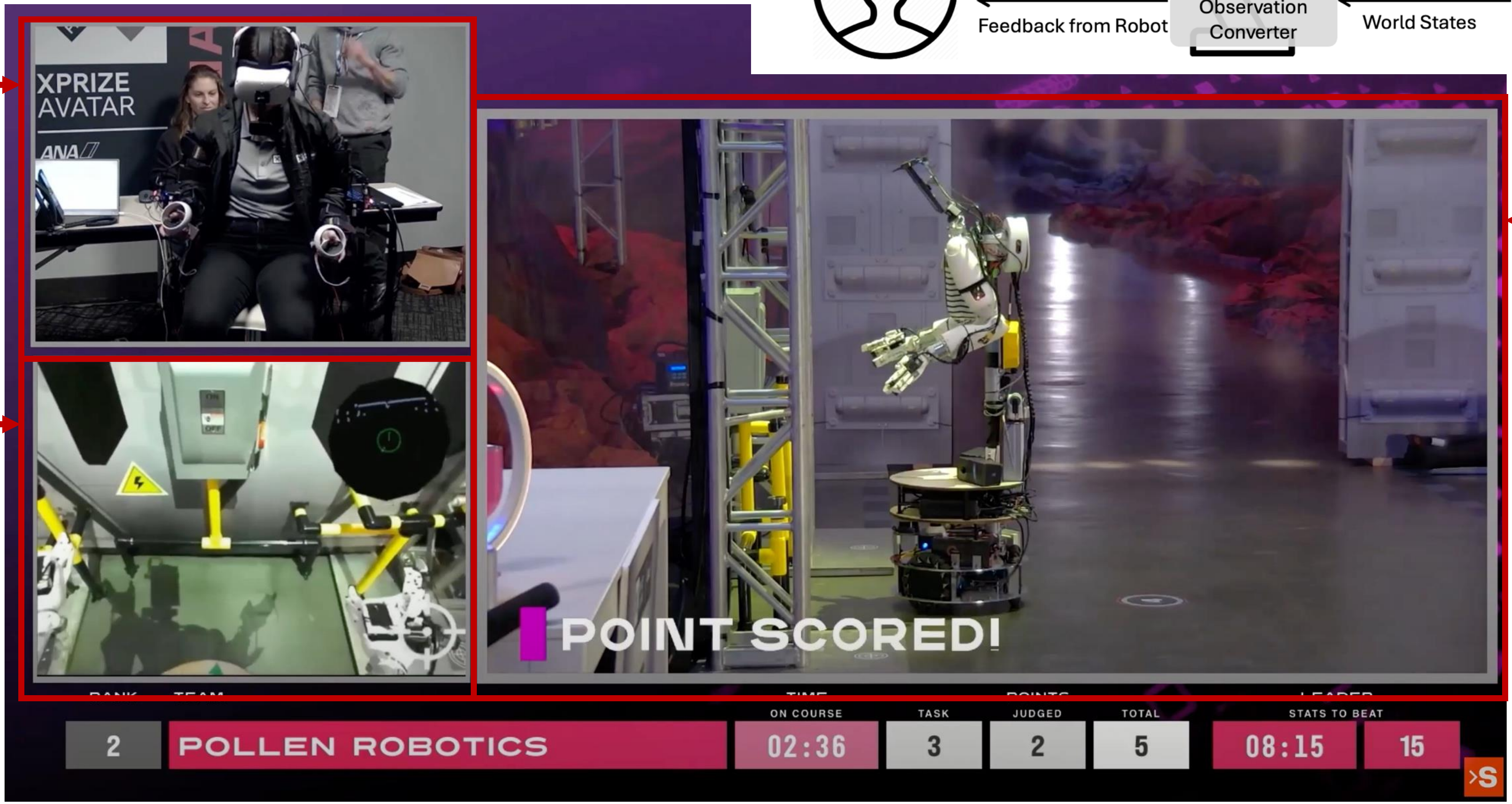
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize



What Human Does

What Human Sees

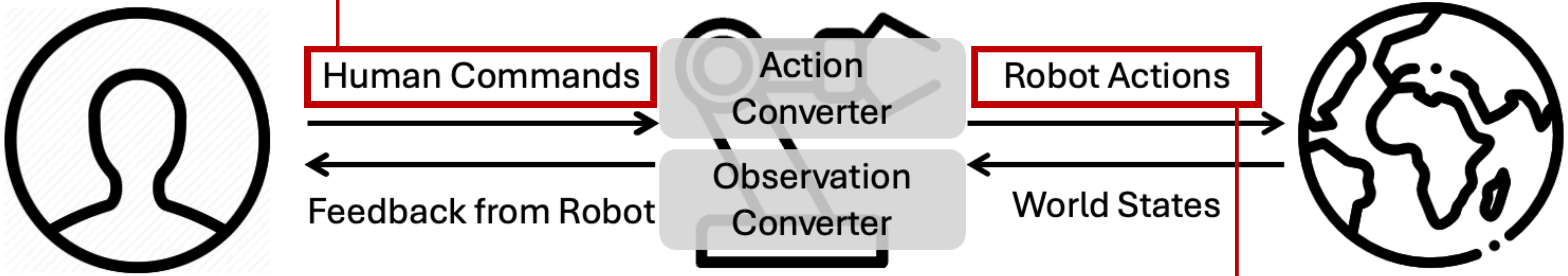


What Robot Does

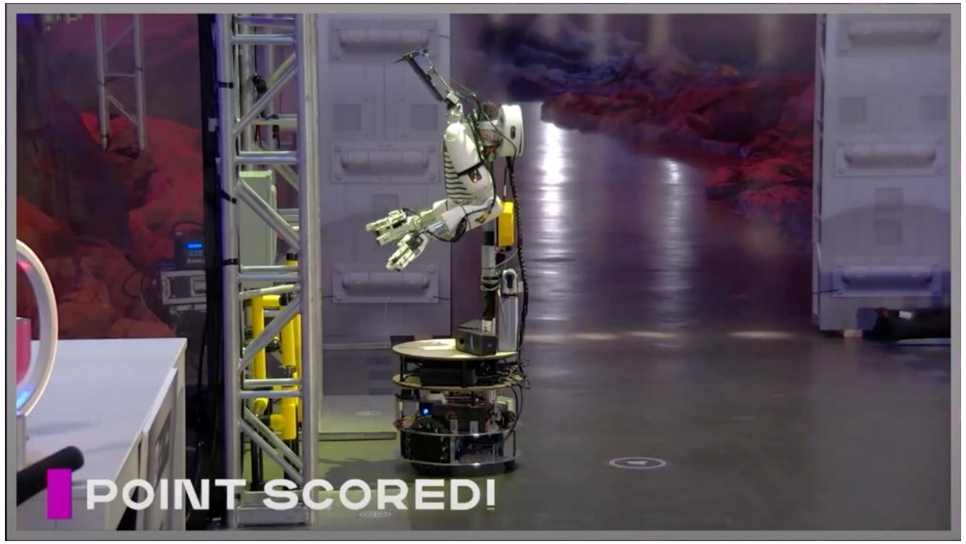
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

What's the main **command interface** of human controlling the robot?



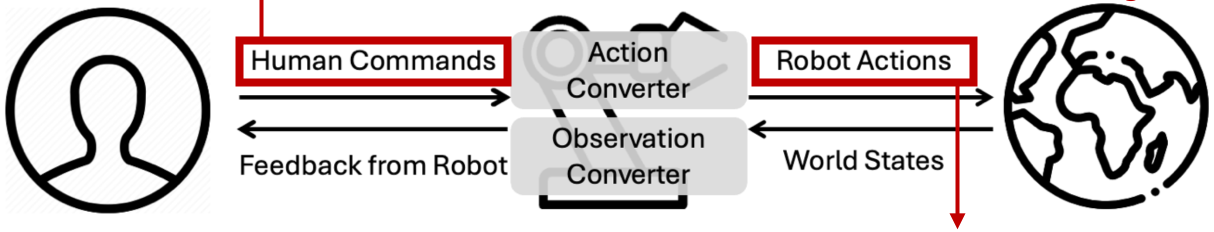
What **exact part of the robot** is the human controlling with each interface?



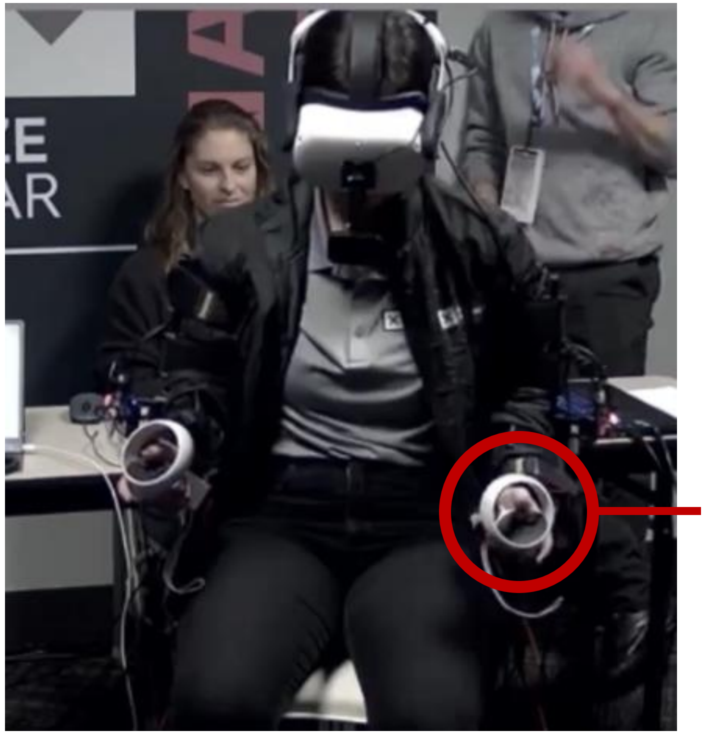
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

What's the main **command interface** of human controlling the robot?



What **exact part of the robot** is the human controlling with each interface?



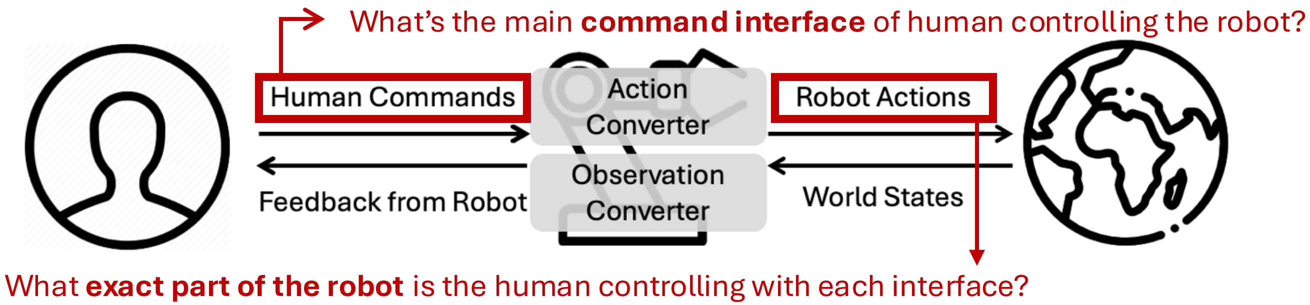
Spatial SE(3) Pose of Human Hand

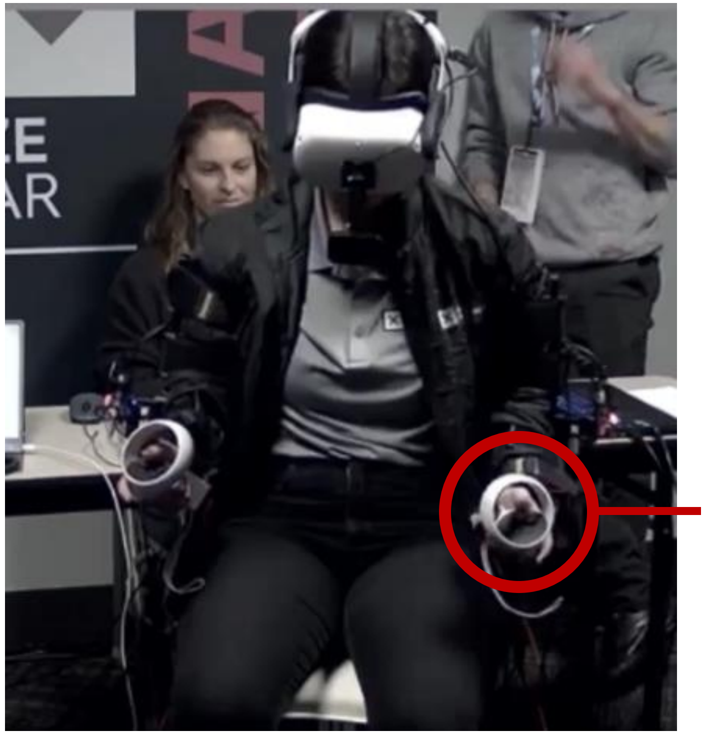
Inverse Kinematics

Joint Targets for Robot Arm

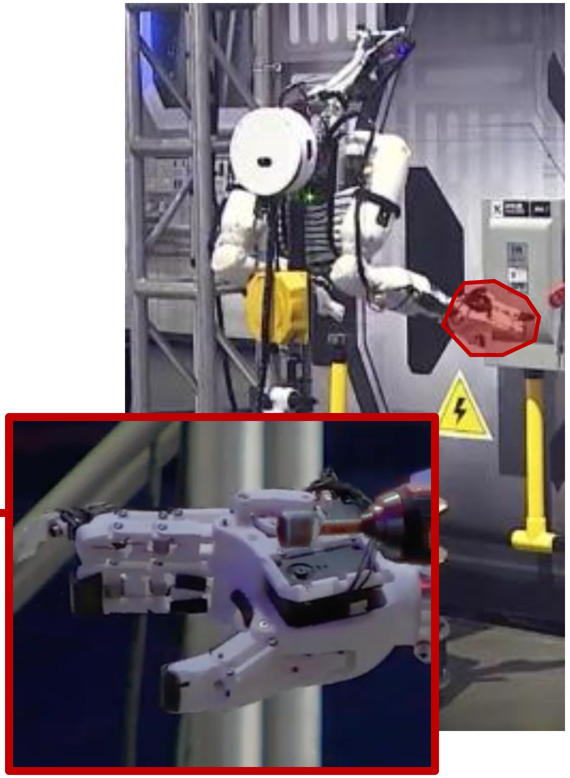
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize





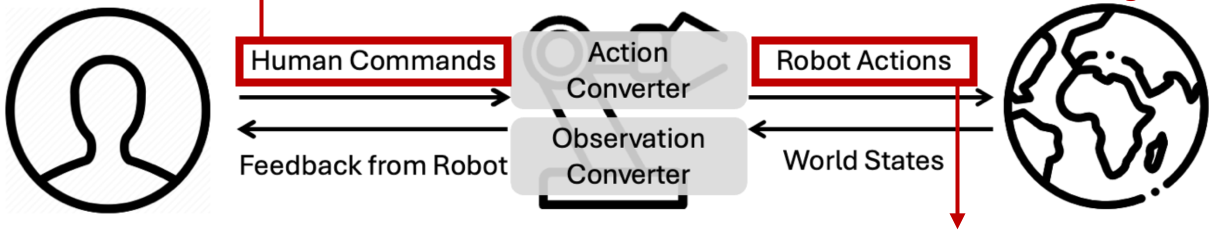
Button Press — Manual Mapping — Joint Targets for all Fingers



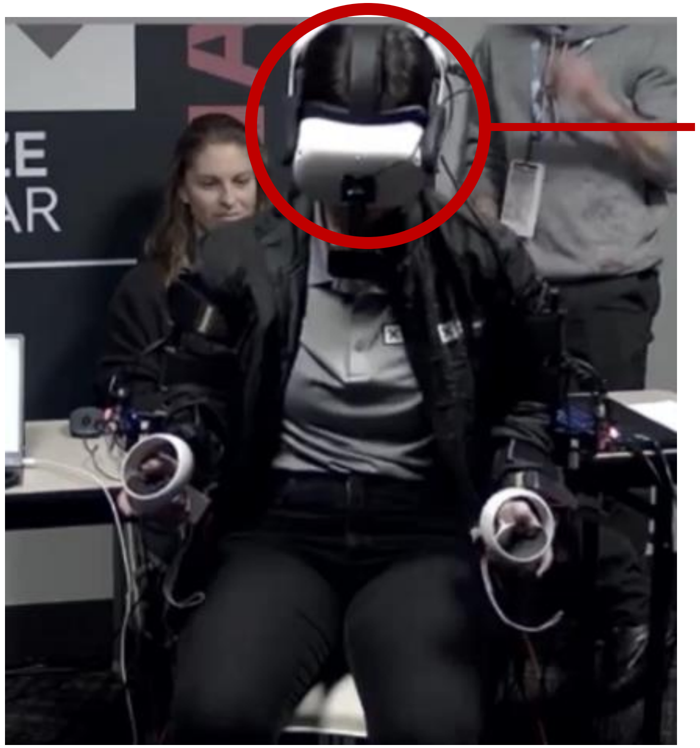
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

What's the main **command interface** of human controlling the robot?



What **exact part of the robot** is the human controlling with each interface?



SE(3) Pose of Human Head

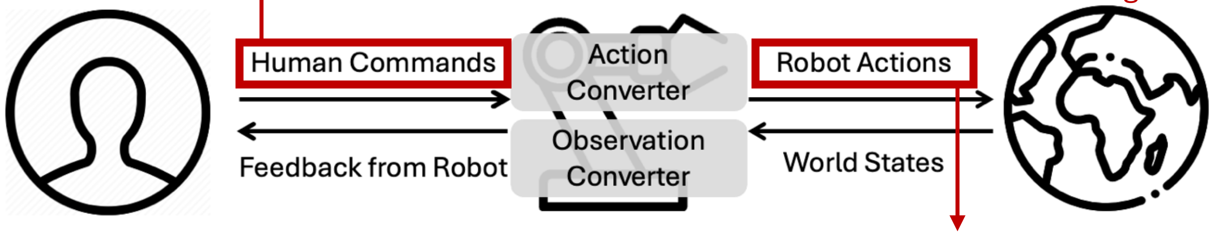
Inverse Kinematics

Joint Targets for Robot's Neck

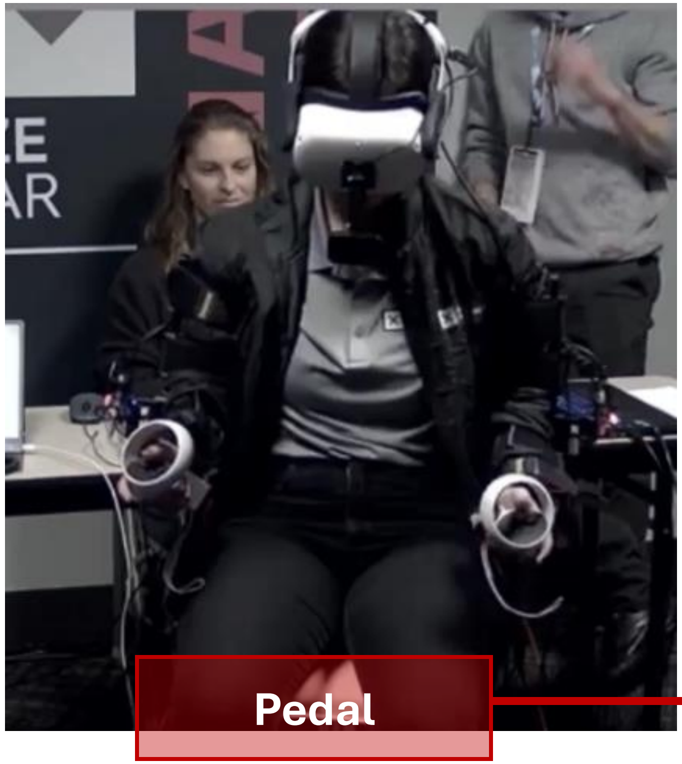
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

What's the main **command interface** of human controlling the robot?



What **exact part of the robot** is the human controlling with each interface?



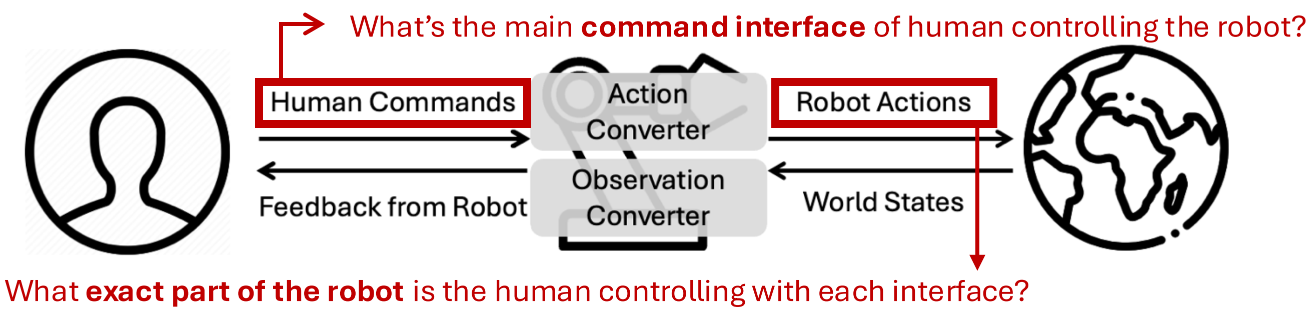
Translational Speed

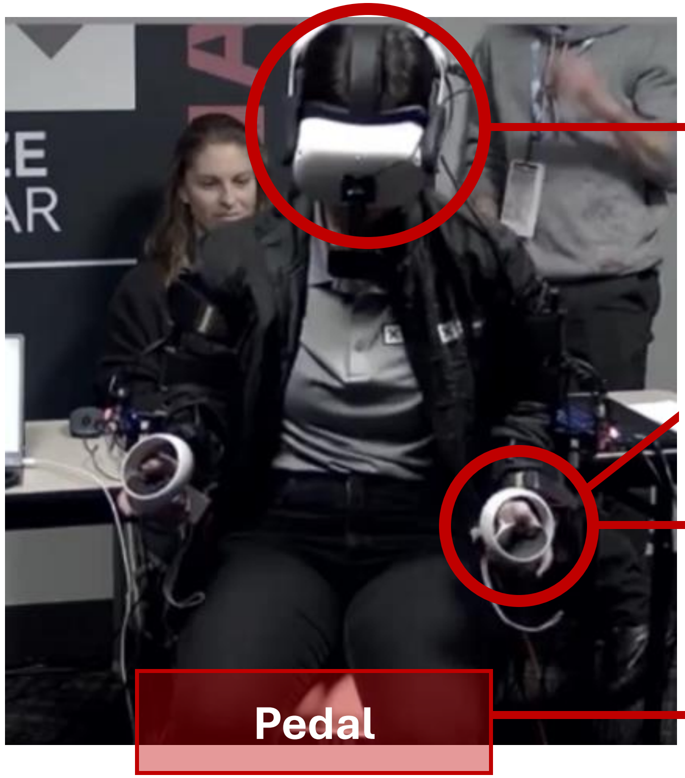
Differential Drive Controller

Mobile Base Wheel Control

Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize





SE(3) Pose of Human Head	Inverse Kinematics	Joint Targets for Robot's Neck
SE(3) Pose of Human Hand	Inverse Kinematics	Joint Targets for Robot Arm
Button Press	Manual Mapping	Joint Targets for all Fingers
Translational Speed	Differential Drive Controller	Mobile Base Wheel Control

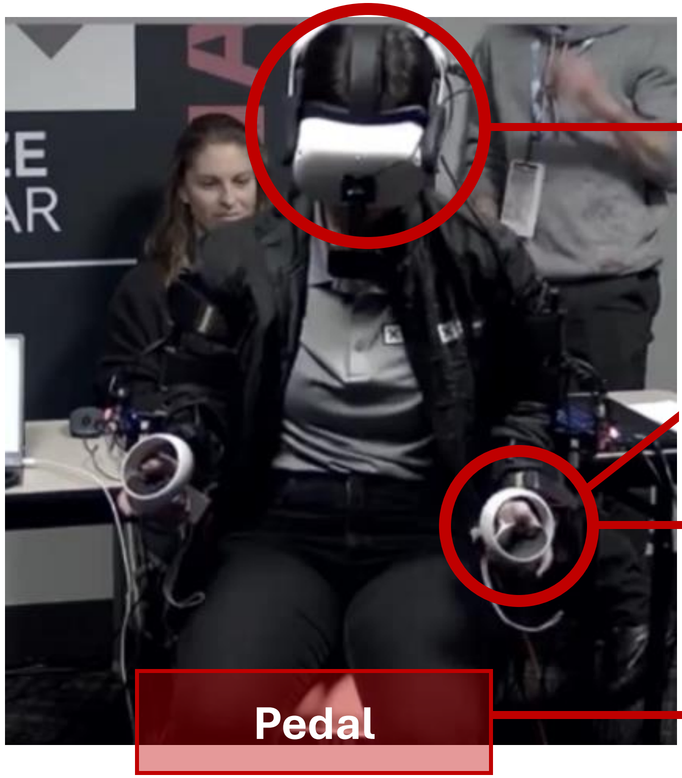
Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

What's the main **command interface** of human controlling the robot?

Any expected issues?

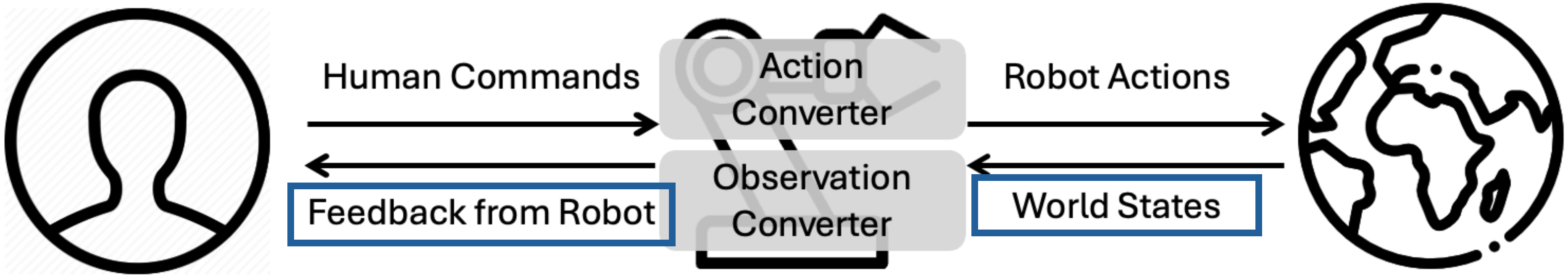
What **exact part of the robot** is the human controlling with each interface?



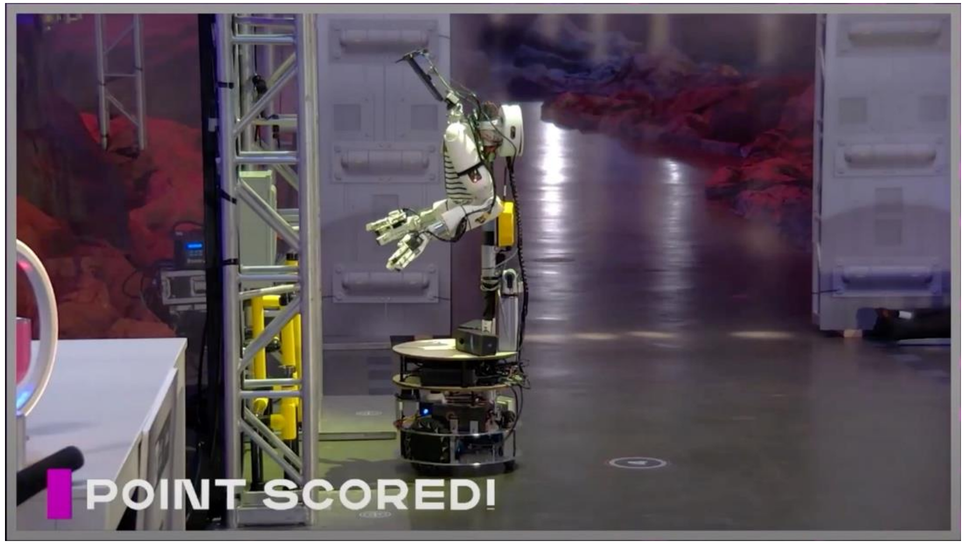
SE(3) Pose of Human Head	Inverse Kinematics	Joint Targets for Robot's Neck
SE(3) Pose of Human Hand	Inverse Kinematics	Joint Targets for Robot Arm
Button Press	Manual Mapping	Joint Targets for all Fingers
Translational Speed	Differential Drive Controller	Mobile Base Wheel Control

Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

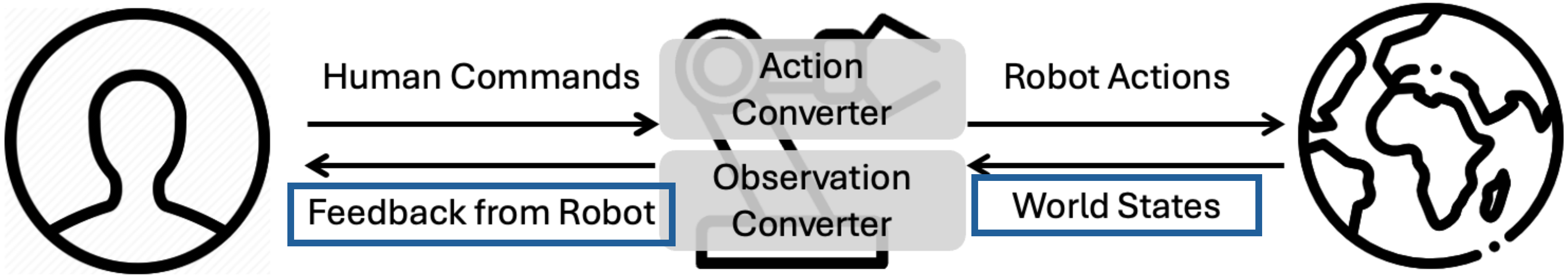


How does the state around the robot + state of the robot itself presented to the operator?

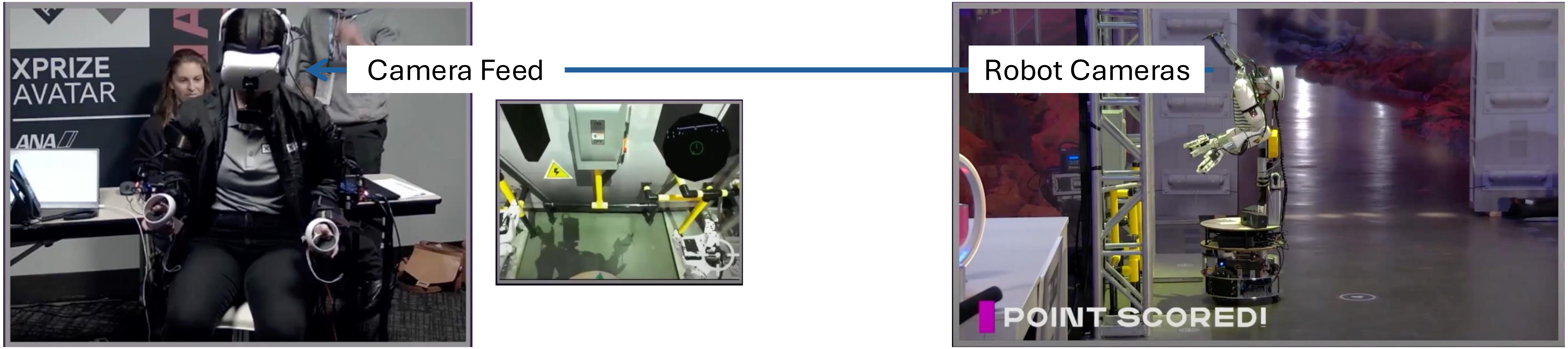


Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize



How does the state around the robot + state of the robot itself presented to the operator?

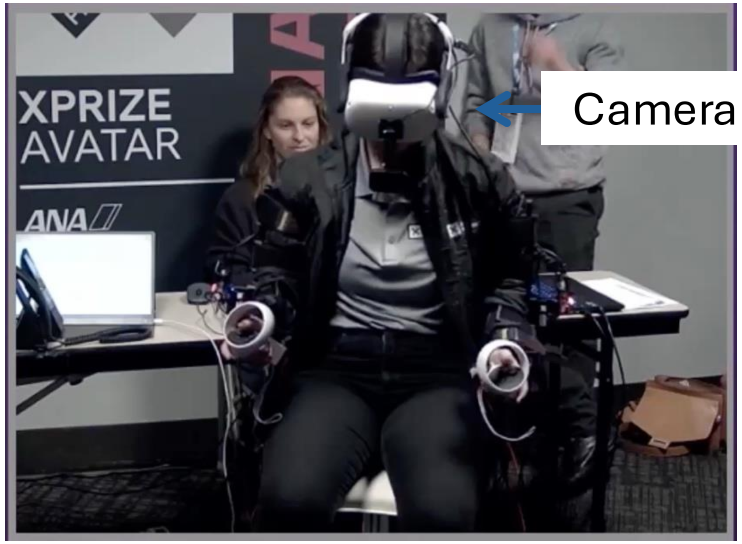


Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

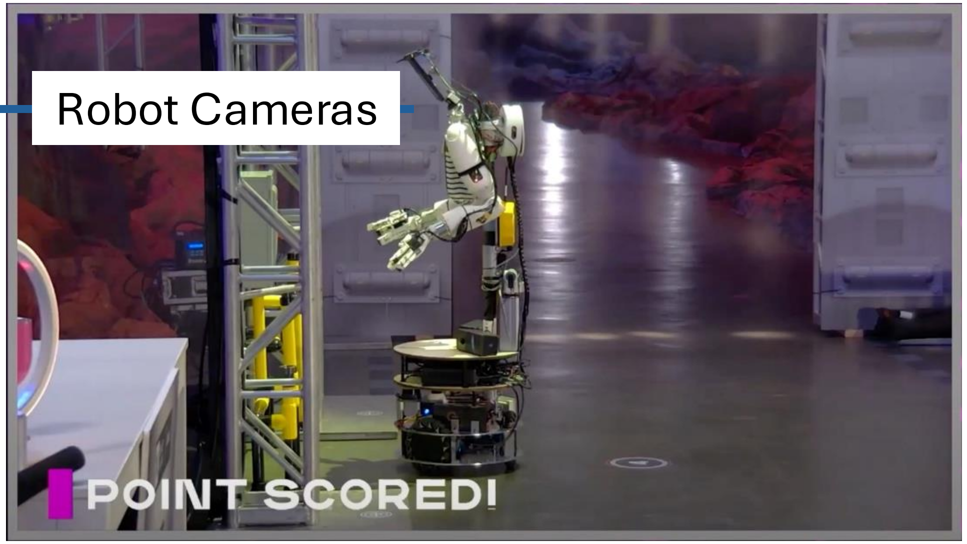
Expected Issues?

How does the state around the robot + state of the robot itself presented to the operator?



Camera Feed

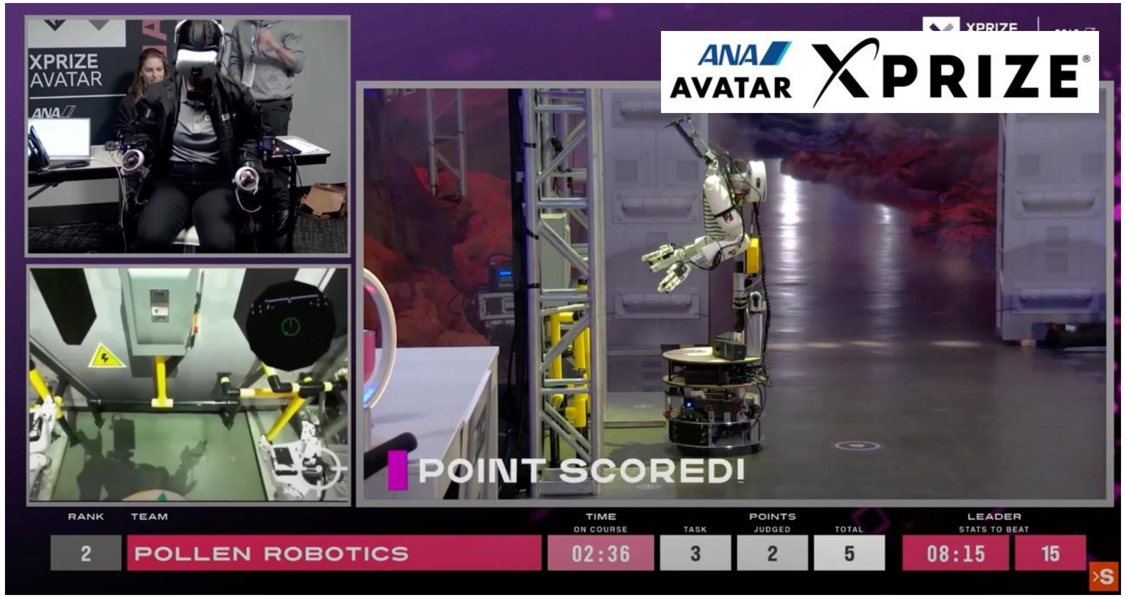
Robot Cameras



XPRIZE AVATAR
DANGER
RANK
TEAM
TIME
POINTS
LEADER
2
POLLEN ROBOTICS
08:42
8
2
10
08:15
15

Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize



Things got complicated as it involved an assumption of operator being **"remotely located"**

[B] ALOHA

Teleoperation System **Case Studies**: In-Depth Analysis

[A] Pollen Robotics @AVATAR XPrize

[B] ALOHA

Things can be quite simpler
if we remove the “remote” assumption

Teleoperation System **Case Studies**: In-Depth Analysis

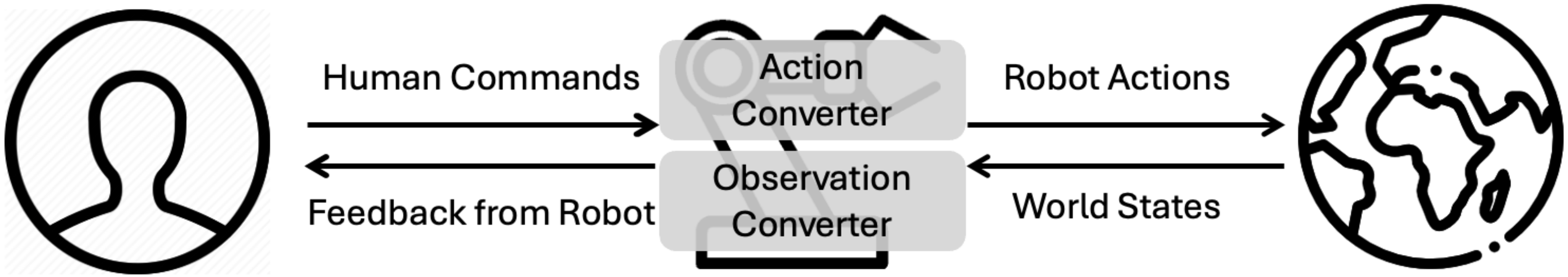
[B] ALOHA

Commander
Robot

Follower
Robot

Teleoperation System **Case Studies**: In-Depth Analysis

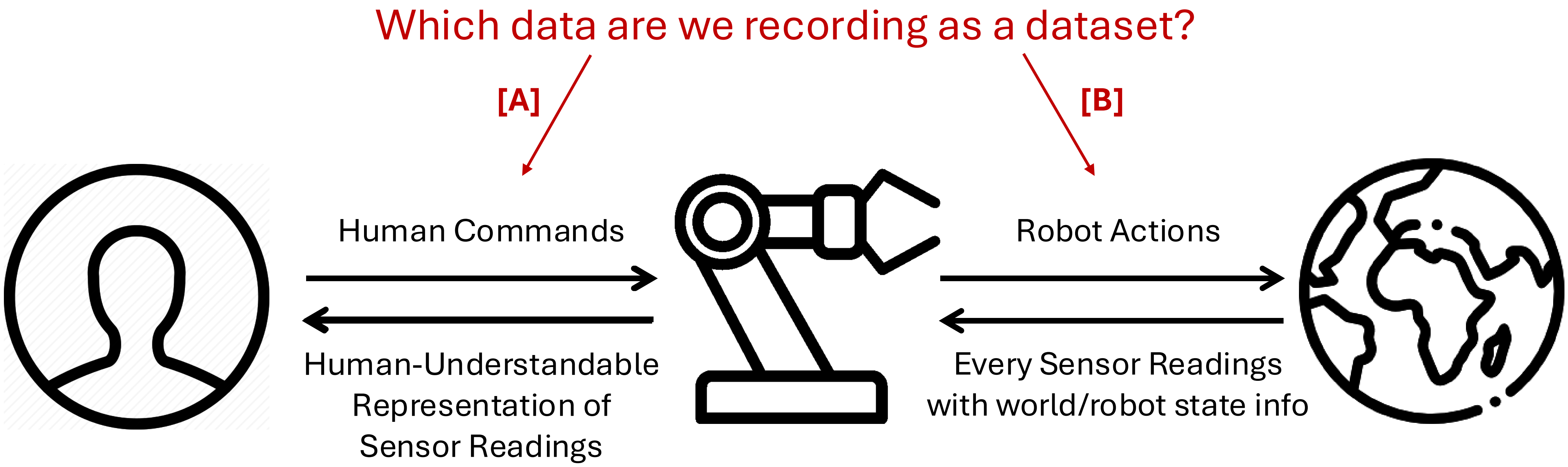
[B] ALOHA



Today...

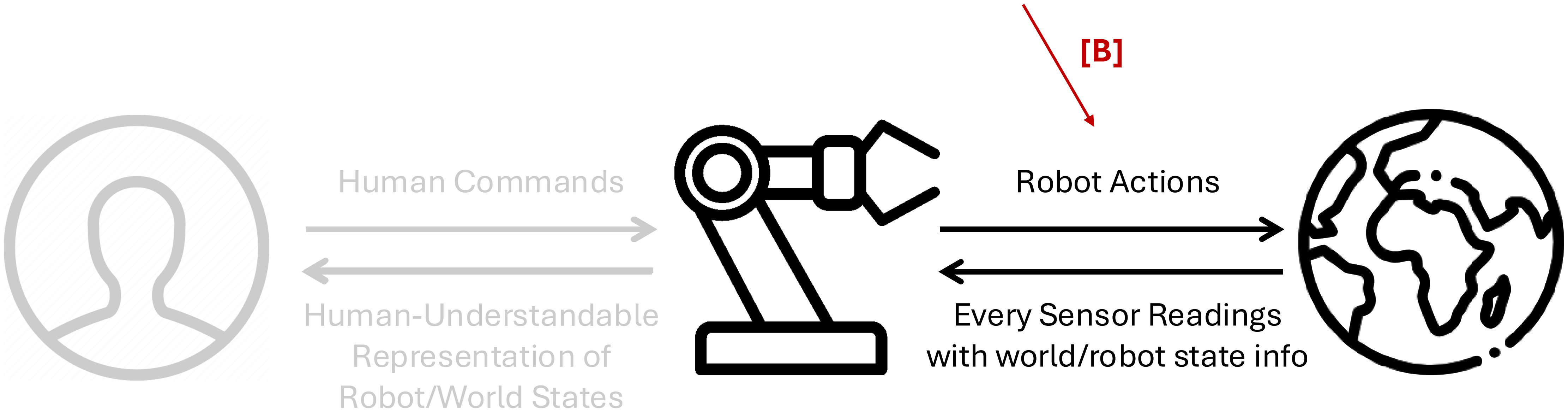
- ~~Teleoperation System Case Studies: In-Depth Analysis~~
- Policy Training with Teleoperated Datasets
  - Policy Architectures
  - Policy Training Methods
- Role of Simulation
  - Real2Sim: Simulation Environment Design
  - Sim2Real

Policy Training with Teleoperated Datasets



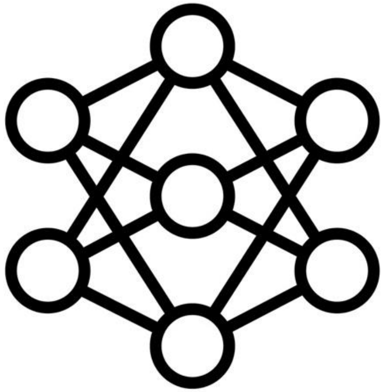
Policy Training with Teleoperated Datasets

Which data are we recording as a dataset?



$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\}$$

Policy Training with Teleoperated Datasets

Every Sensor Readings $\boldsymbol{s}$ →  → Robot Actions $\boldsymbol{a}$

$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\}$$

Robot control becomes a **supervised learning** problem.

Policy Training with Teleoperated Datasets

Imitation Learning in general ...

[1] Behavior Cloning

= directly learning the mapping of the paired state/actions from teleoperated datasets

$$D = \{(s_1, a_1, s_2, a_2, \ldots, s_n, a_n)\}$$

[2] Inverse Optimal Control (Inverse RL)

= learning the rewards from the dataset, then run RL

Robot control becomes a **supervised learning** problem.

Policy Training with Teleoperated Datasets

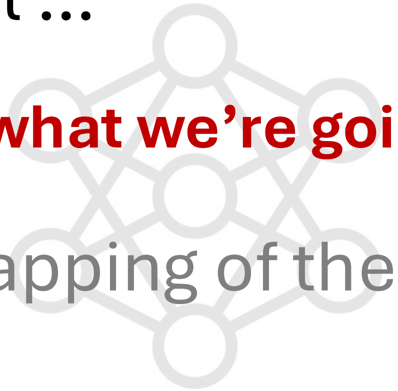
Imitation Learning in general …

[1] **Behavior Cloning** ← **what we're going to be focusing today**

= directly learning the mapping of the paired state/actions from teleoperated datasets

[2] Inverse Optimal Control (Inverse RL)

= learning the rewards from the dataset, then run RL

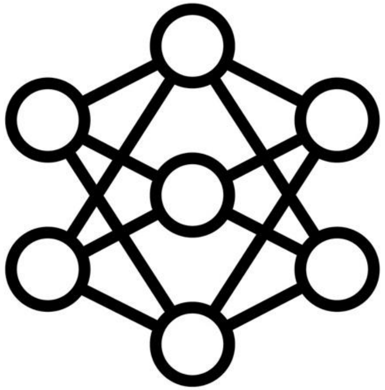
World / Robot States from Sensors →  → Robot Actions

$$D = \{(s_1, a_1, s_2, a_2, \ldots, s_n, a_n)\}$$

Robot control becomes a **supervised learning** problem.

Policy Training with Teleoperated Datasets

World / Robot States
from Sensors



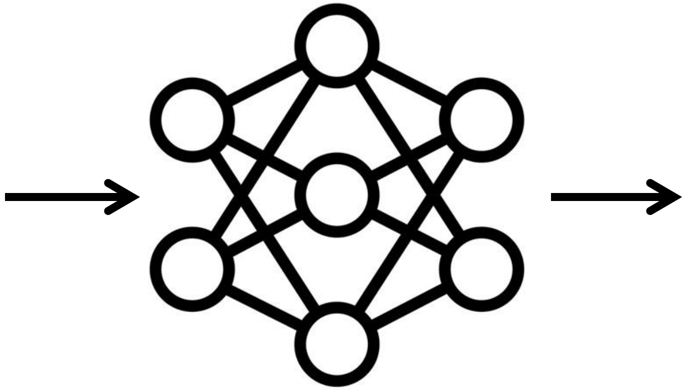
Robot Actions

$$D = \{(s_0, a_0, s_1, a_1, \ldots, s_n)\}$$

$$\max_{\theta} \mathbb{E}_{(s_t, a_t) \sim D}[\log \pi_\theta(\boldsymbol{a}|\boldsymbol{s})]$$

Policy Training with Teleoperated Datasets

World / Robot States
from Sensors



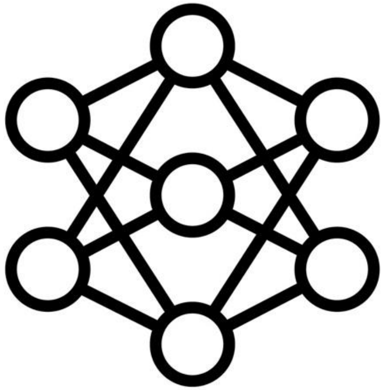
Robot Actions

$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\} \qquad \max_{\theta} \mathbb{E}_{(s_t, a_t)\sim D}[\log \pi_\theta(\boldsymbol{a}|\boldsymbol{s})]$$

Two main **<u>design decisions</u>** for policy training

Policy Training with Teleoperated Datasets

World / Robot States
from Sensors →



→ Robot Actions

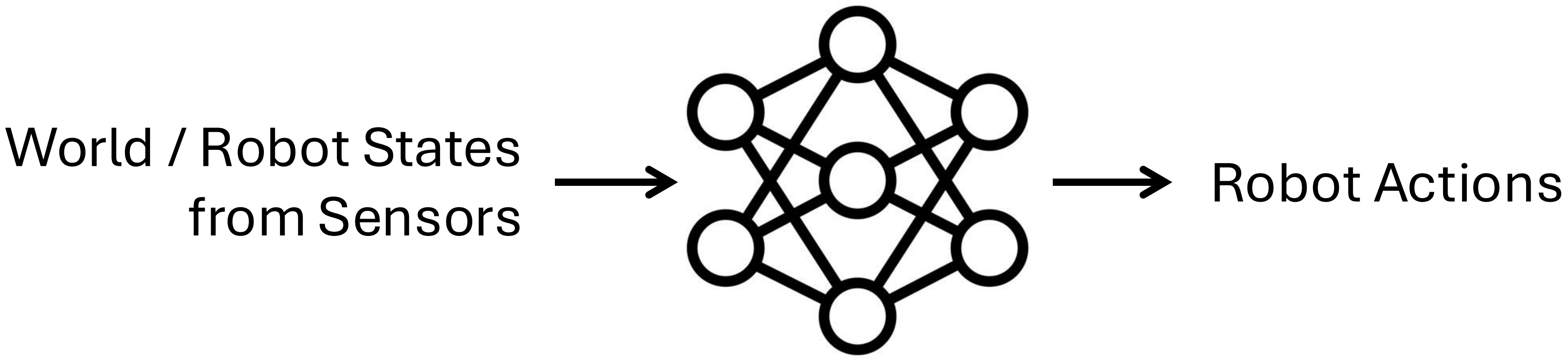
$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\}$$

$$\max_{\theta} \mathbb{E}_{(s_t, a_t)\sim D}[\log \pi_\theta(\boldsymbol{a}|\boldsymbol{s})]$$

Two main **design decisions** for policy training

1. Engineering input / output space of neural network policy

Policy Training with Teleoperated Datasets



$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\}$$

$$\max_{\theta} \mathbb{E}_{(s_t, a_t)\sim D}[\log \pi_{\theta}(\boldsymbol{a}|\boldsymbol{s})]$$

Two main **design decisions** for policy training

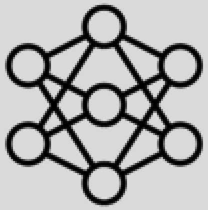
1. Engineering input / output space of neural network policy
2. Engineering policy architectures

Policy Training with Teleoperated Datasets

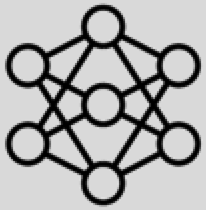
1. Engineering input / output space of neural network policy

$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\}$$

Type A $(m = 1 \text{ and } n = 1)$

$\hat{s}_t \longrightarrow$  $\longrightarrow a_t$

Type B $(m > 1 \text{ or } n > 1)$

$\underbrace{\hat{s}_t, \hat{s}_{t+1}, \cdots, \hat{s}_{t+m-1}}_{m \text{ timesteps}} \longrightarrow$  $\longrightarrow \underbrace{a_t, \cdots, a_{t+n-1}}_{n \text{ timesteps}}$

Notice any weird hats?

Policy Training with Teleoperated Datasets

1. Engineering input / output space of neural network policy

$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\}$$

$$\hat{s}_t = \boldsymbol{g}(s_t)$$

$\boldsymbol{g}$: a function that **massages the state** (combination of various sensor readings) for better learnability

- Exclusion of certain sensor readings (i.e., joint pose vs end-effector pose)
- Transformation of certain data types (i.e., SE(3))
- Dropping out certain modalities to prevent over-attention

Type A

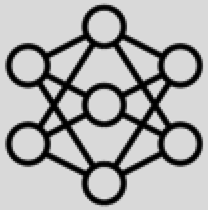
$\hat{s}_t$

Policy Training with Teleoperated Datasets

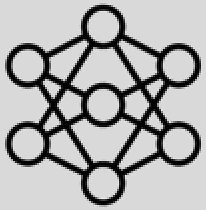
1. Engineering input / output space of neural network policy

$$D = \{(s_0, a_0, s_1, a_1, \dots, s_n)\}$$

Type A $(m = 1 \text{ and } n = 1)$

$\hat{s}_t \longrightarrow$  $\longrightarrow a_t$

Type B $(m > 1 \text{ or } n > 1)$

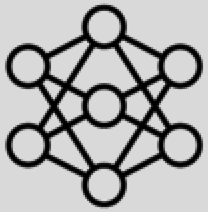
$\underbrace{\hat{s}_t, \hat{s}_{t+1}, \cdots, \hat{s}_{t+m-1}}_{m \text{ timesteps}} \longrightarrow$  $\longrightarrow \underbrace{a_t, \cdots, a_{t+n-1}}_{n \text{ timesteps}}$

Any guesses for an ideal combination of m and n?

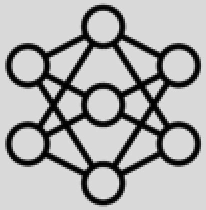
Policy Training with Teleoperated Datasets

1. Engineering input / output space of neural network policy

Type A $(m = 1 \text{ and } n = 1)$

$\hat{s}_t \longrightarrow$  $\longrightarrow a_t$

Type B $(m > 1 \text{ or } n > 1)$

$\overbrace{\hat{s}_t, \hat{s}_{t+1}, \cdots, \hat{s}_{t+m-1}}^{m \text{ timesteps}} \longrightarrow$  $\longrightarrow \overbrace{a_t, \cdots, a_{t+n-1}}^{n \text{ timesteps}}$

Any guesses for an ideal combination of m and n?

Pros and Cons of $\underline{m = 1}$

- 👍 Policy faces less out-of-distribution inputs
- 👍 Enables Reactive / Failure recovery behaviors
- 👎 Multimodal output distributions

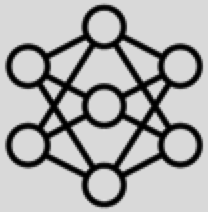
Pros and Cons of $\underline{n > 1}$

- 👍 Less prone to compounding errors*
- 👍 Less vulnerable to disturbances
- 👎 Less reactive / jerky behaviors

Policy Training with Teleoperated I

1. Engineering input / output space of

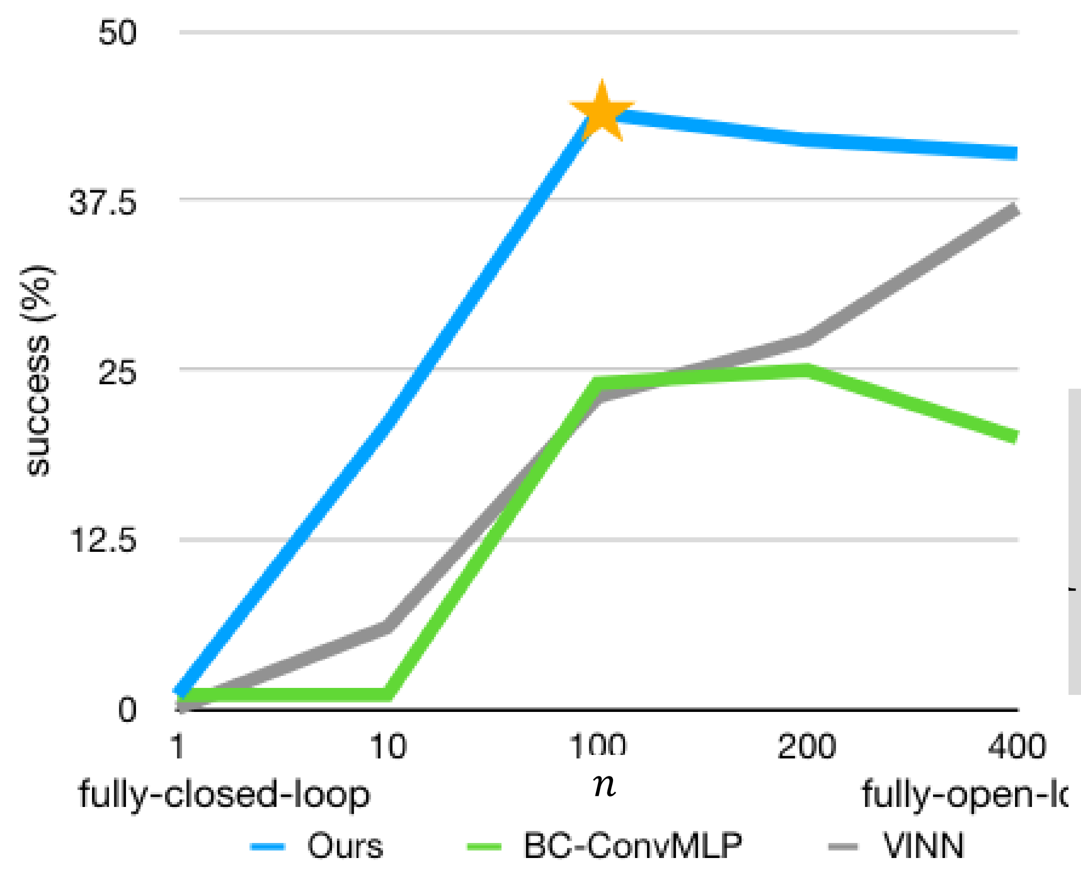
Type A $(m = 1 \text{ and } n = 1)$

$\hat{s}_t \longrightarrow$  $\longrightarrow a_t$

Type B $(m > 1$ or

m timeste

$\hat{s}_t, \hat{s}_{t+1}, \cdots,$



Any guesses for an ideal cc

Pros and Cons of $\underline{m = 1}$

- Policy faces less out-of-distribution inputs
- Enables Reactive / Failure recovery behaviors
- Multimodal output distributions

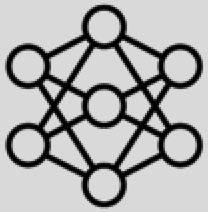
Pros and Cons of $\underline{n > 1}$

- Less prone to compounding errors*
- Less vulnerable to disturbances
- Less reactive / jerky behaviors

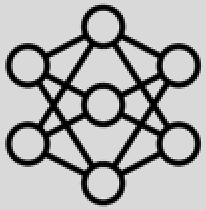
Policy Training with Teleoperated Datasets

1. Engineering input / output space of neural network policy

Type A $(m = 1 \text{ and } n = 1)$

$\hat{s}_t \longrightarrow$  $\longrightarrow a_t$

Type B $(m > 1 \text{ or } n > 1)$

$\overbrace{\hat{s}_t, \hat{s}_{t+1}, \cdots, \hat{s}_{t+m-1}}^{m \text{ timesteps}} \longrightarrow$  $\longrightarrow \overbrace{a_t, \cdots, a_{t+n-1}}^{n \text{ timesteps}}$

Any guesses for an ideal combination of m and n?

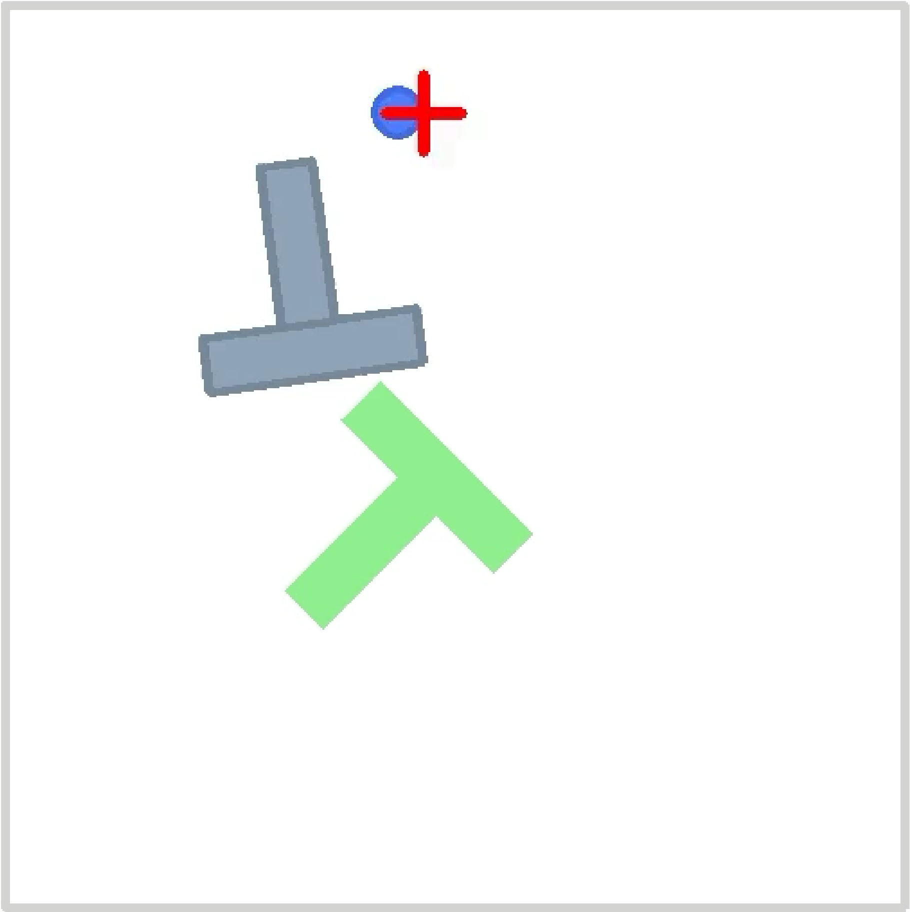
Pros and Cons of $\underline{m = 1}$

- 👍 Policy faces less out-of-distribution inputs
- 👍 Enables Reactive / Failure recovery behaviors
- 👎 **Multimodal output distributions**

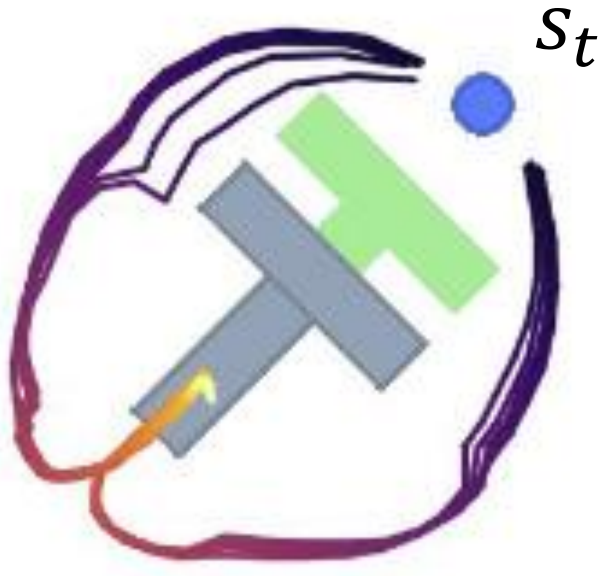
Pros and Cons of $\underline{n > 1}$

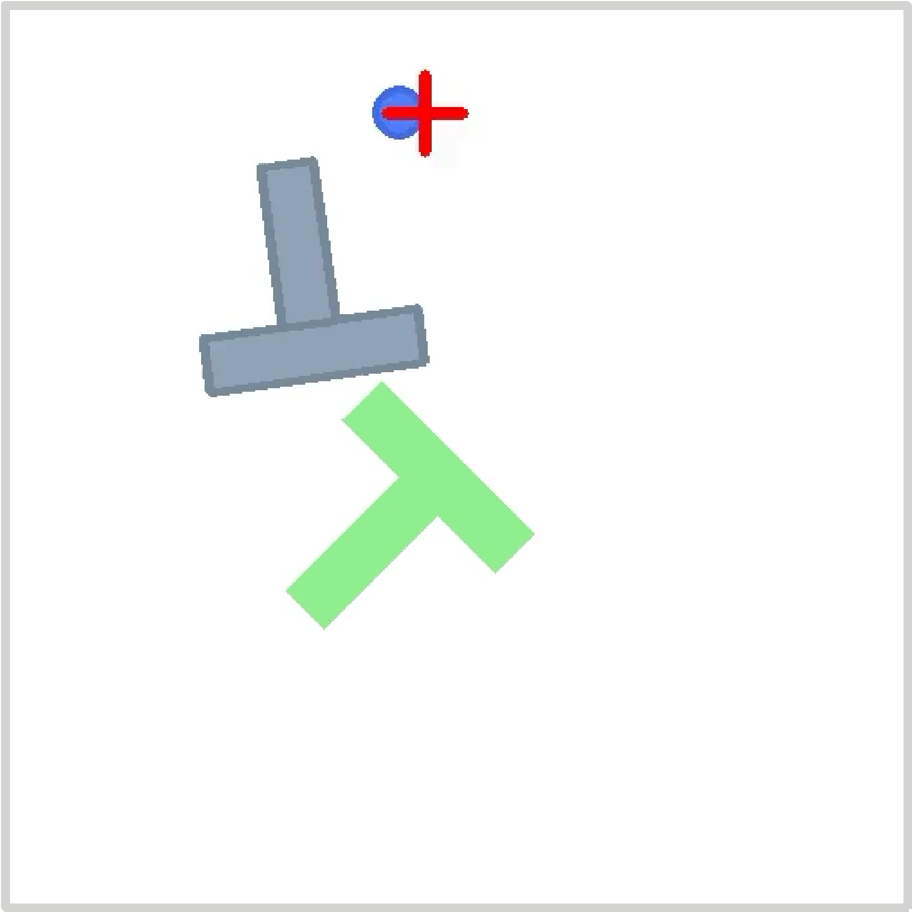
- 👍 Less prone to compounding errors*
- 👍 Less vulnerable to disturbances
- 👎 Less reactive / jerky behaviors

Policy Training with Teleoperated Datasets



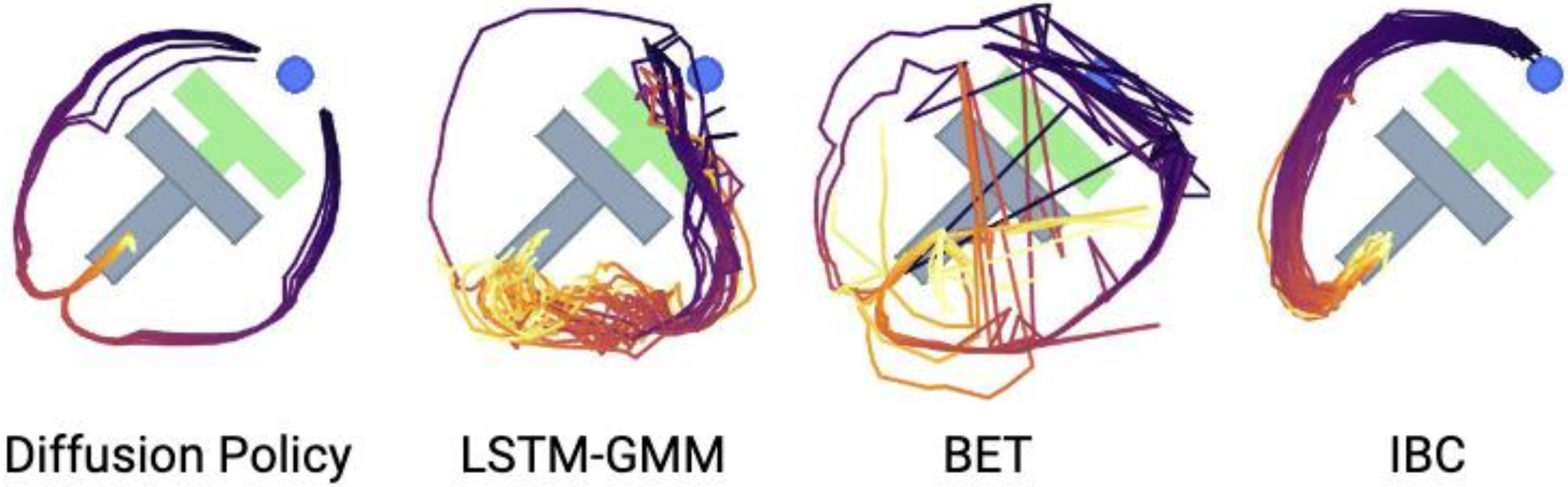
Policy Training with Teleoperated Datasets





Policy Training with Teleoperated Datasets

2. Engineering policy architectures



Policy architectures does matter

Policy Training with Teleoperated Datasets

2. Engineering policy architectures

Policy architectures vs **Neural Network** architectures

Difference in **mathematical formulations** used to generate action predictions

(i.e., diffusion vs VAE vs GAN)

Difference in **how the input data is encoded and decoded** to generate predictions

(i.e., MLP vs Transformers)

Policy Training with Teleoperated Datasets

2. Engineering policy architectures

Variational Autoencoders (VAEs)

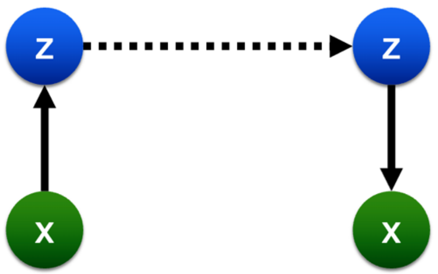
- We introduce an **inference model** q(z|x)

$$q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_\phi(\mathbf{x}), \boldsymbol{\Sigma}_\phi(\mathbf{x}))$$

- This allows us to efficiently optimize the log-likelihood, through the **evidence lower bound** (ELBO).

$$\log p_{\theta,\phi}(\mathbf{x}) \geq \mathrm{ELBO}(\mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right]$$

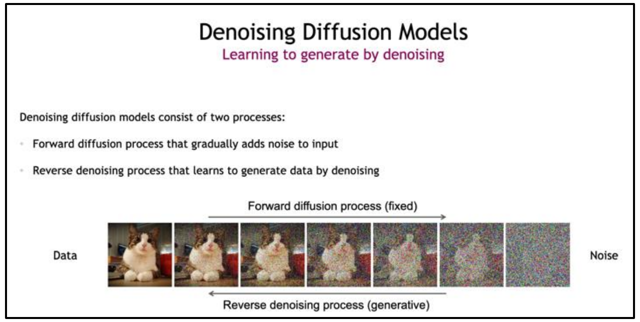
- We optimize q(z|x) and p(x,z) jointly w.r.t. ELBO
- Bound is tight with the right q(z|x)



Inference model **q(z|x)**

Generative model **p(x,z)**

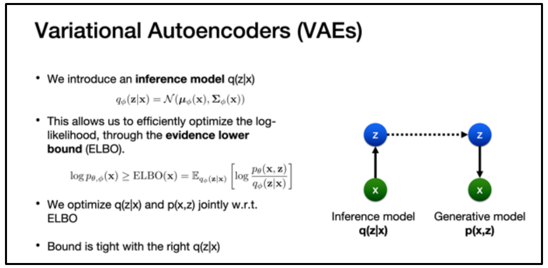
- Pros: cheap compute cost; one-step prediction
- Cons: cannot model extreme multimodality

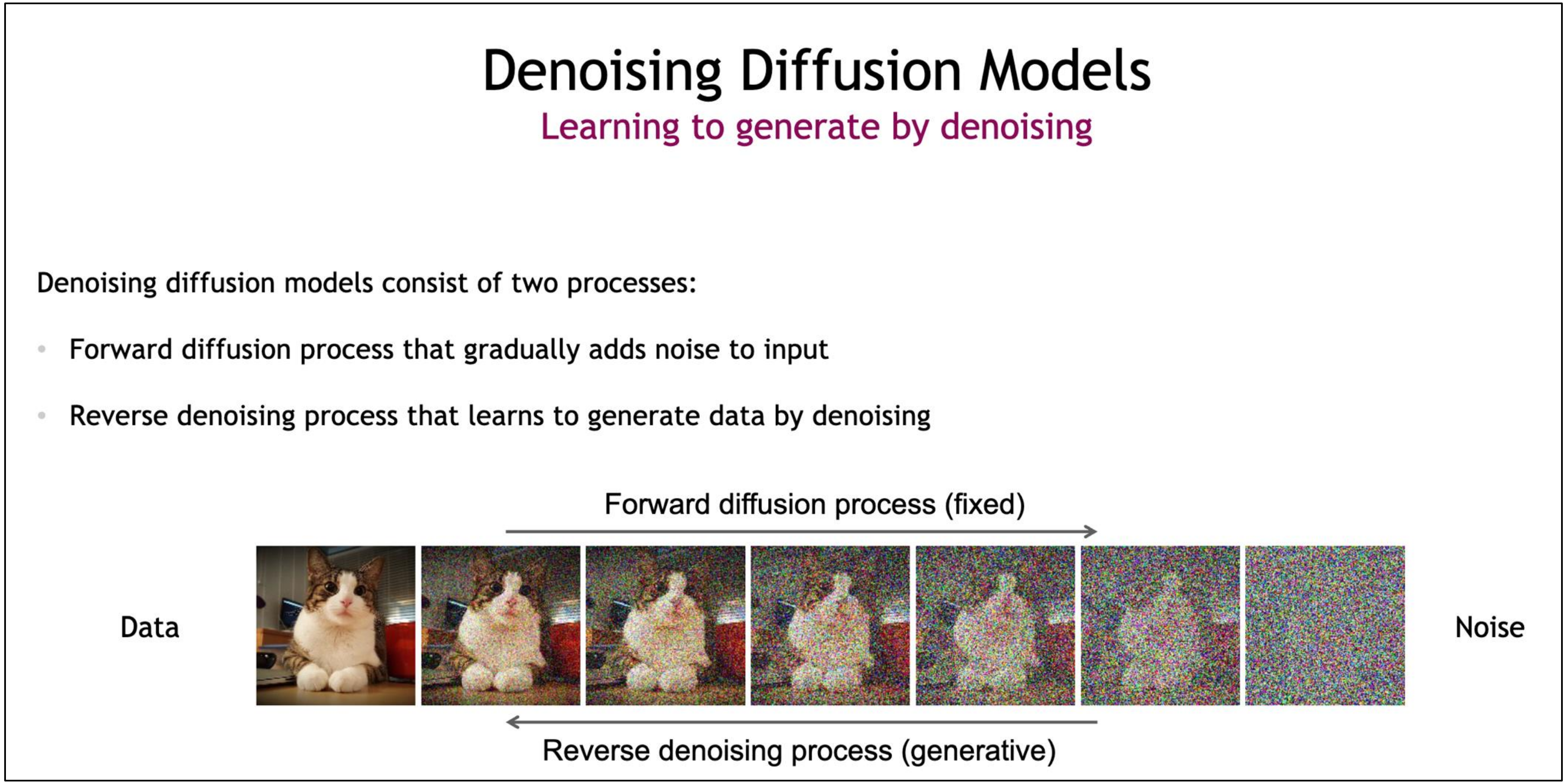


Policy Training with Teleoperated Datasets

2. Engineering policy architectures

- Pros: powerful expressivity
- Cons: expensive compute; multi-step inference required





Today...

- ~~Teleoperation System Case Studies: In-Depth Analysis~~
- ~~Policy Training with Teleoperated Datasets~~
  - ~~Policy Architectures~~
  - ~~Policy Training Methods~~

Andy Zeng's MIT CSL Seminar, April 4, 2022

"Dirty Laundry"

Symptoms of a larger problem

The not-so-secret recipe to making a rockstar behavior cloning demo on **real** robots

Step 1. collect your own "expert" data and don't trust anyone else to make it perfect

Step 2. avoid "no action" data so your policy doesn't just sit there

Step 3. It's not working? Collect more data until "extrapolation" becomes "interpolation"

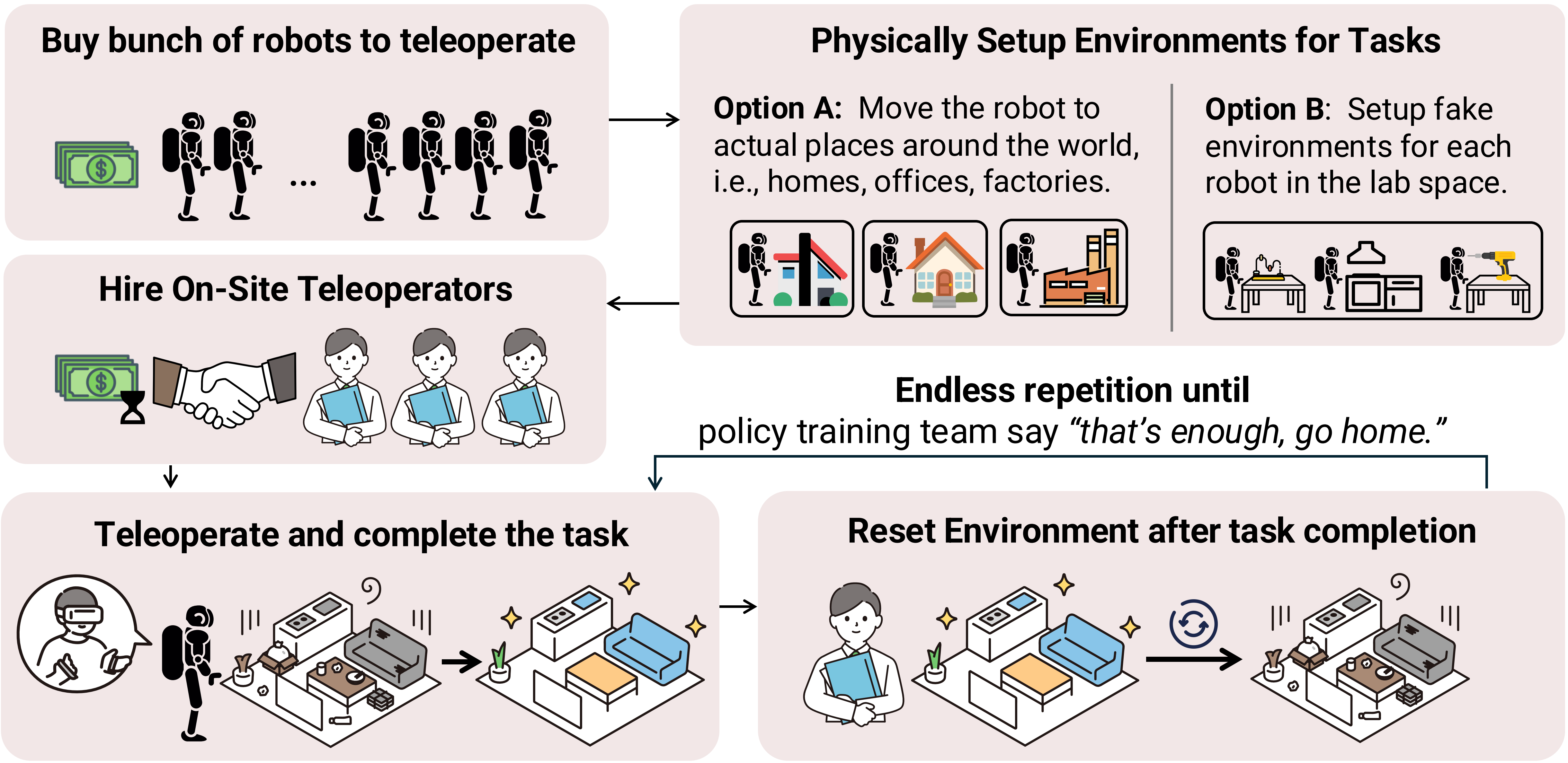
Step 4. Train and test on the same day because your setup might change tomorrow

Mostly because we don't have a lot of data

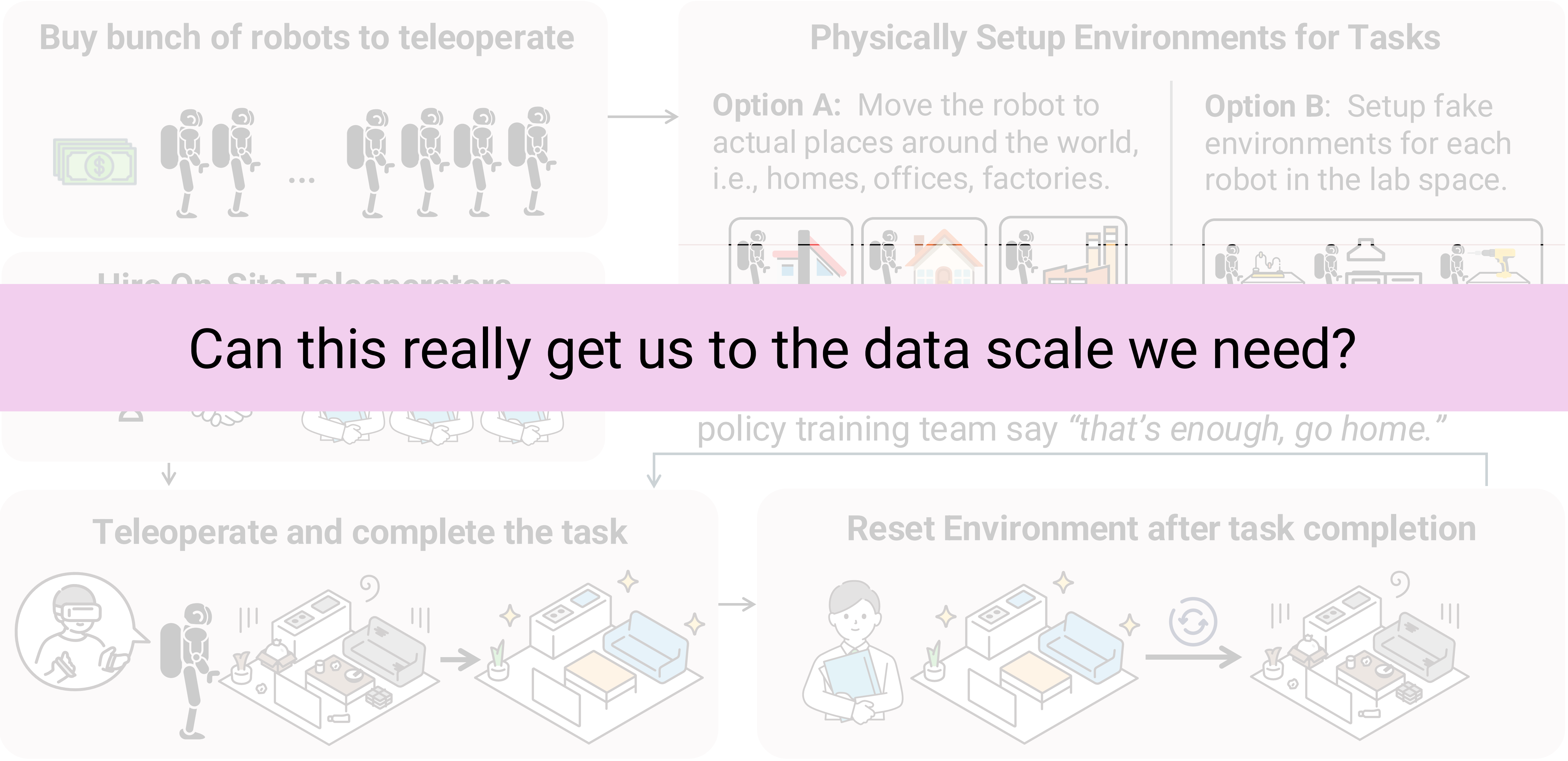
Today...

- ~~Teleoperation System Case Studies: In-Depth Analysis~~
- ~~Policy Training with Teleoperated Datasets~~
    - ~~Policy Architectures~~
    - ~~Policy Training Methods~~
- Role of Simulation
    - Real2Sim: Simulation Environment Design
    - Sim2Real

Role of Simulation: Cost of Real-world Teleop Data Collection



Role of Simulation: Cost of Real-world Teleop Data Collection



Can this really get us to the data scale we need?

Role of Simulation

Learning from Human Videos

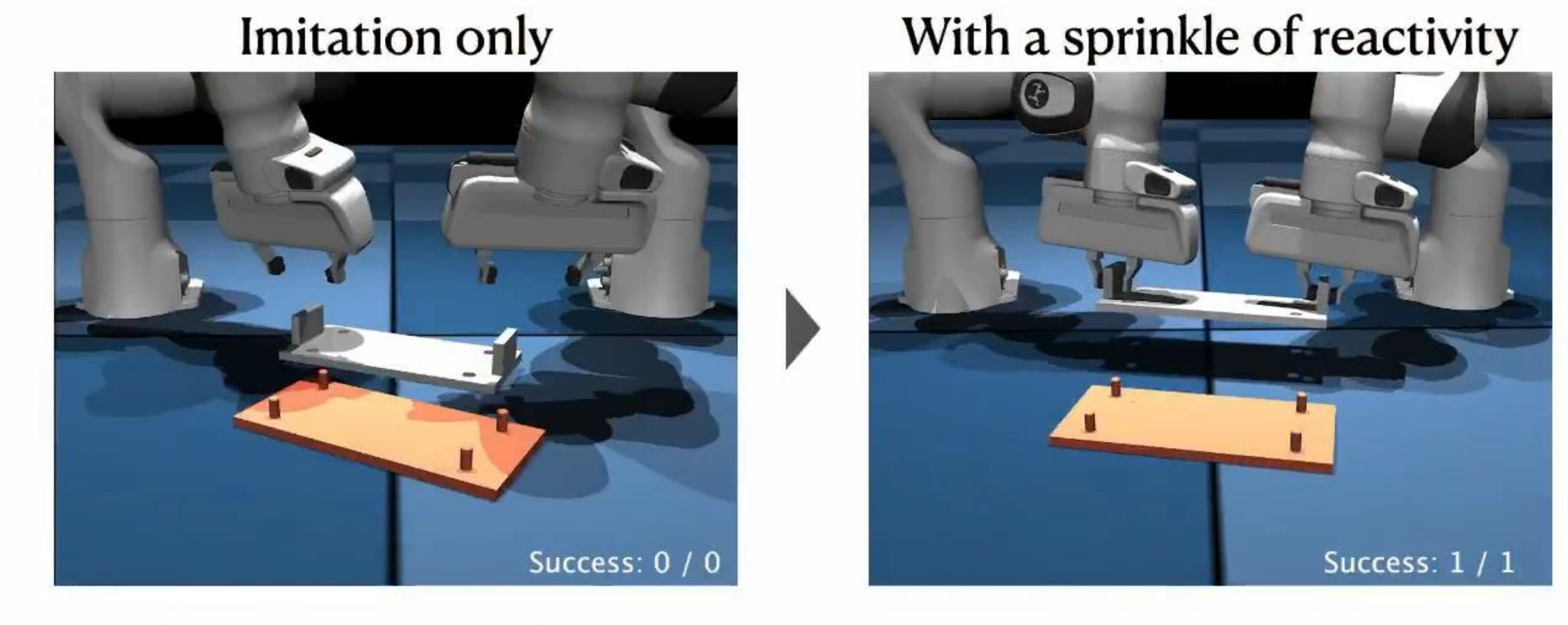
Passive Data with wearables

Collecting Robot Data in Virtual World


ACHVSETTS INSTITVTE OF TECHNOLOGY

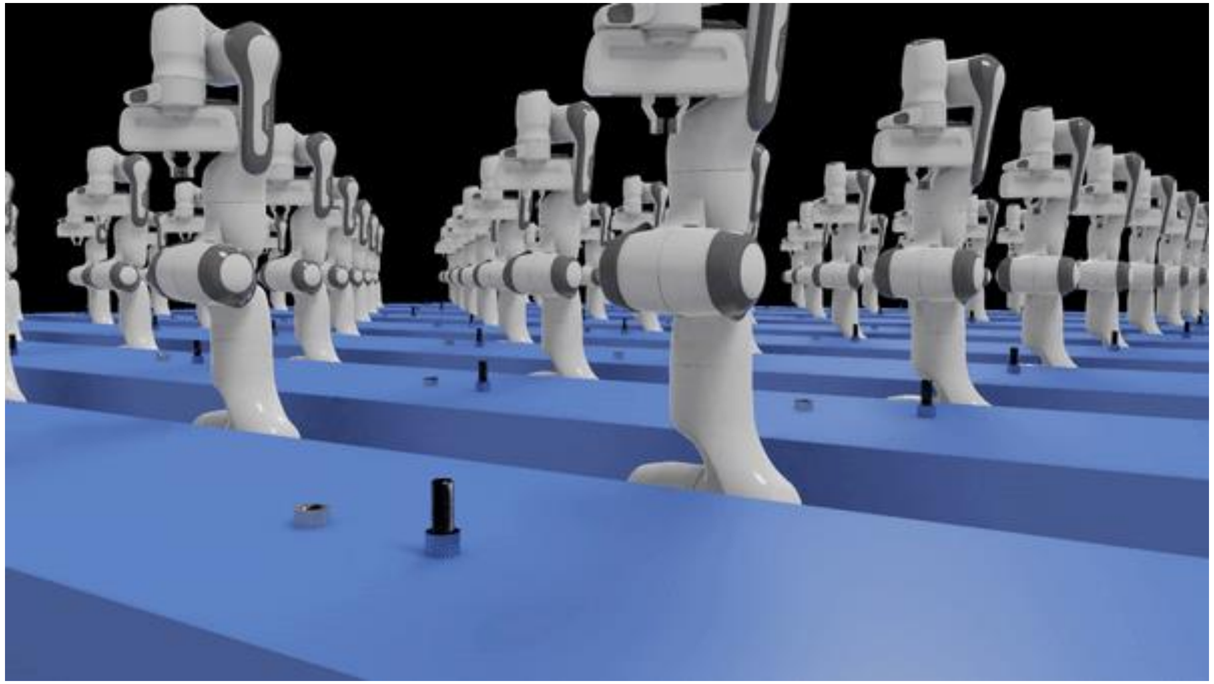
Role of Simulation

Demos collected in simulation supports
last-mile performance improvement through **RL finetuning.**



Role of Simulation

Massively parallelizable simulation
with randomizable parameters



Access to
Oracle (Privileged) States*

States that are hard to retrieve from real-world sensors, for instance:

- Object Poses / Velocities
- Contact Force / Pairs
- etc...

Role of Simulation with a small cost

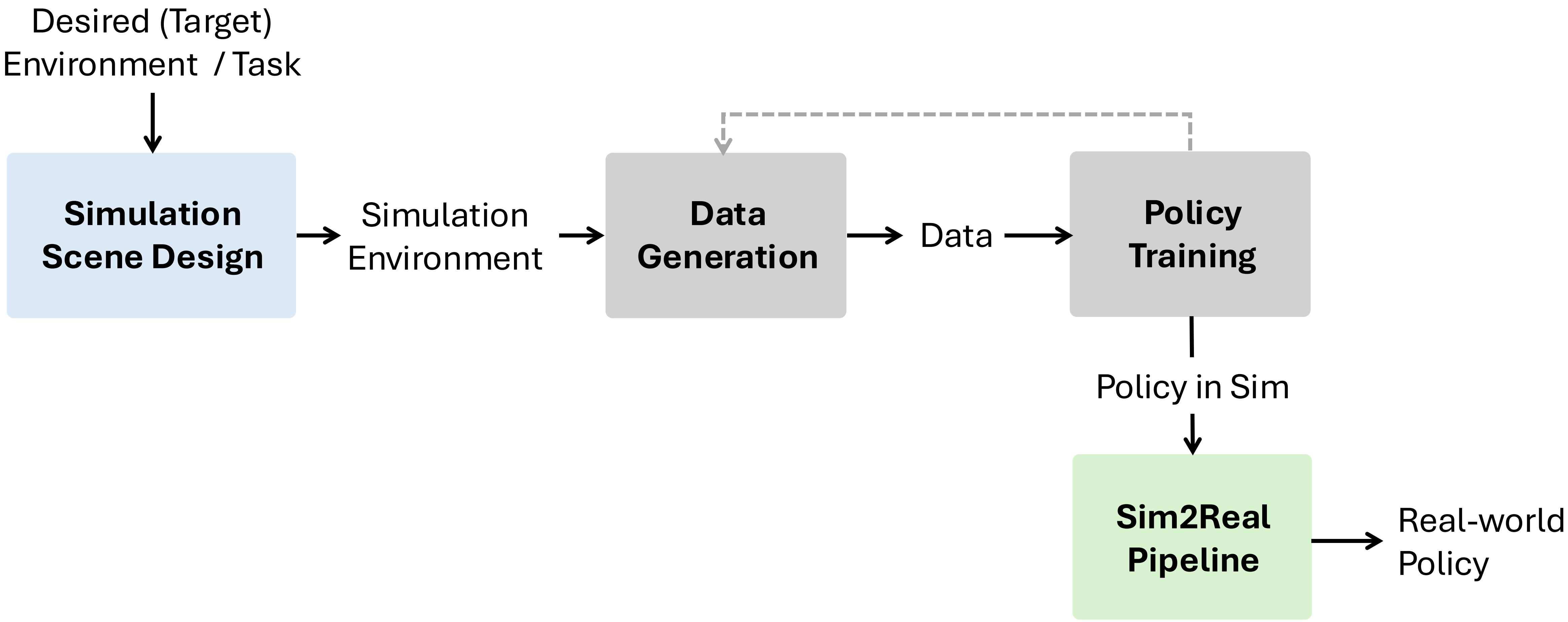
Simulation Scene Design

Generating realistic enough simulation scenes that captures the essence of real-world environments

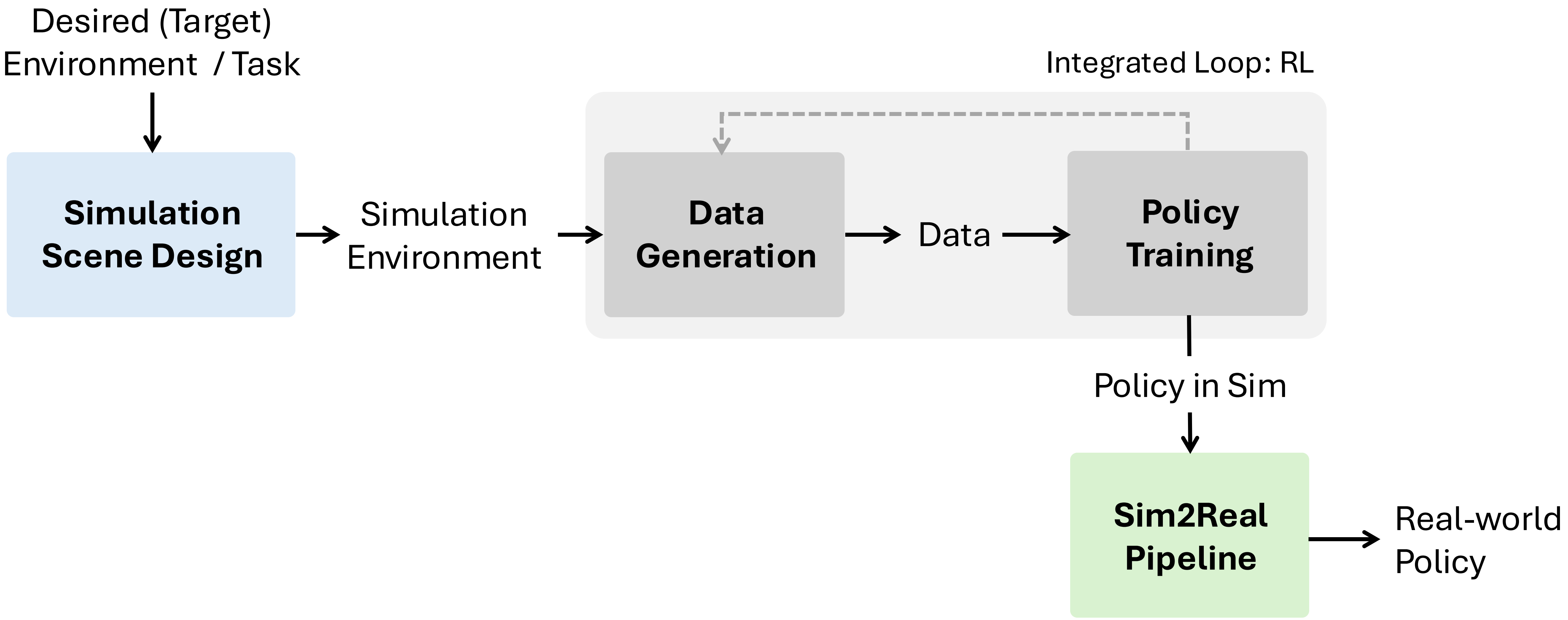
Sim2Real Pipeline

Transferring policies trained with simulated experiences back to real-world evaluation environment.

Role of Simulation with a small cost



Role of Simulation with a small cost



Role of Simulation with a small cost

Simulation Scene Design : Real2Sim

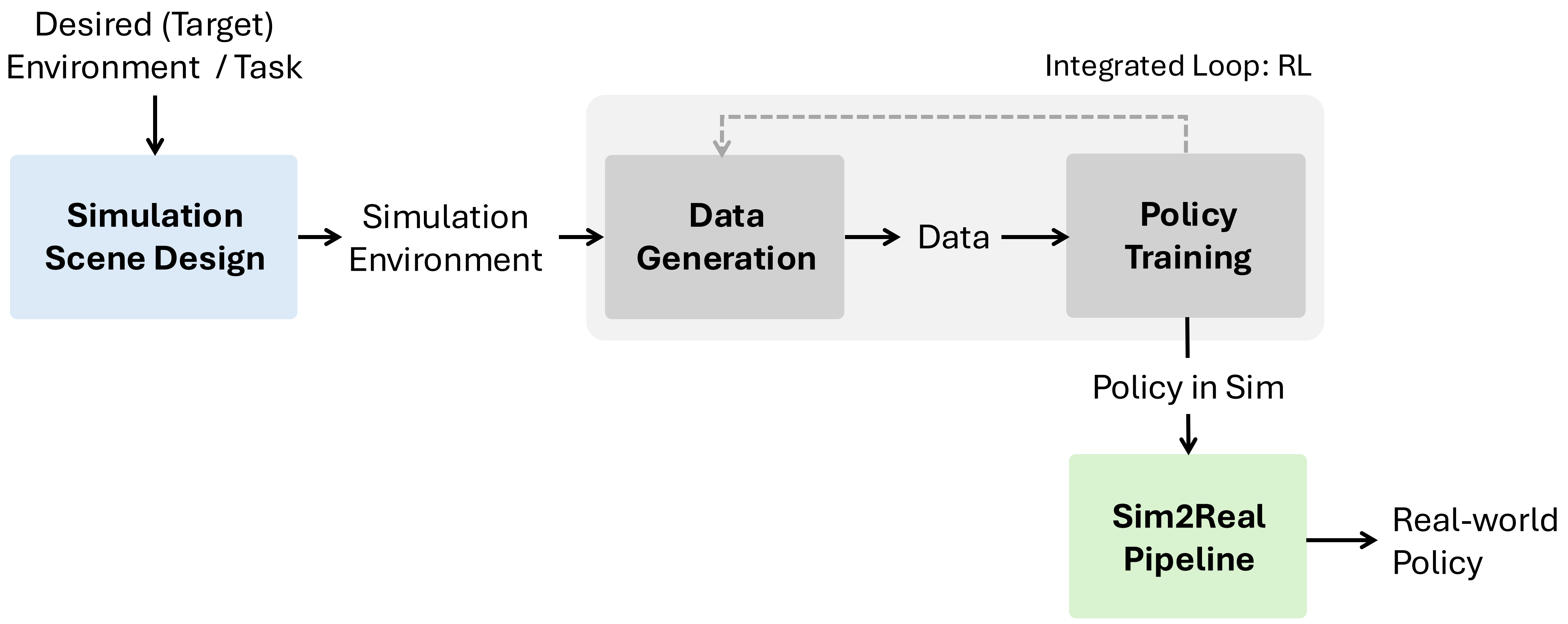
Environments you will experience during tutorial session

Role of Simulation with a small cost

Simulation Scene Design : Generative Simulation

https://gen2sim.github.io/

Role of Simulation with a small cost

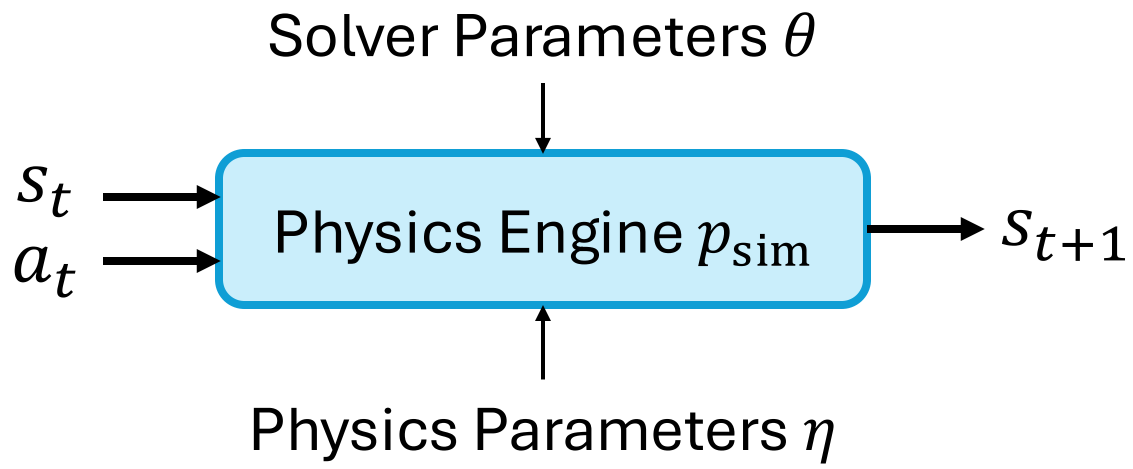


Role of Simulation with a small cost

Sim2Real Pipeline: Two major Sim2Real gaps to deal with

Contact Dynamics

How <u>**physics engine**</u> models contacts vs how our <u>**actual world**</u> models contacts

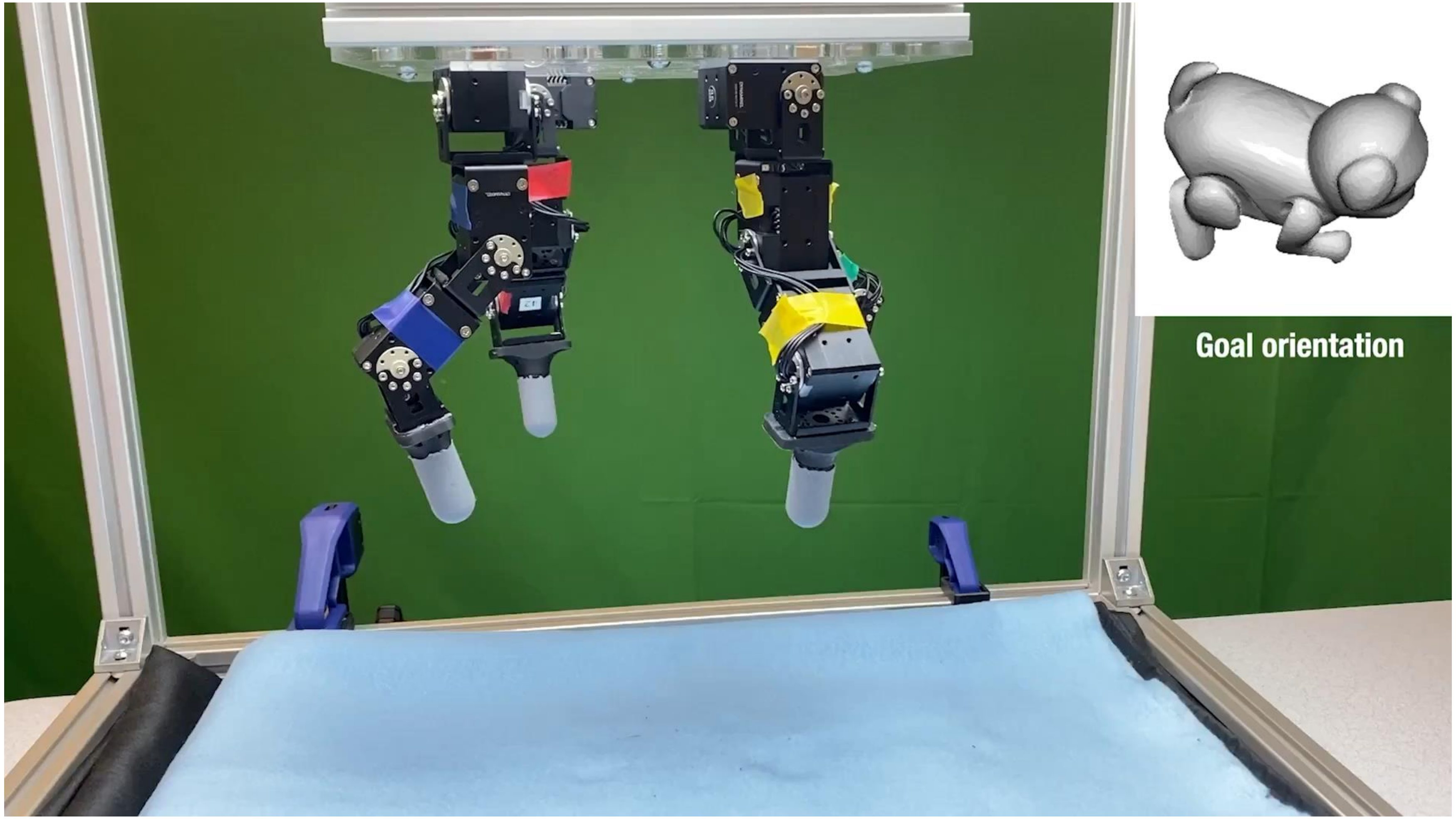


$$p_{\mathbf{sim}}(s_{t+1}|s_t, a_t; \theta, \eta) \not\approx p_{\mathbf{real}}(s_{t+1}|s_t, a_t)$$

1. **System Identification (SysID)**: Find the right θ, η that best matches p_{real}
2. **Domain Randomization (DR):**
   $$\theta \sim p_1(\theta)$$
   $$\eta \sim p_2(\eta)$$

Role of Simulation with a small cost

Sim2Real Pipeline: Two major Sim2Real gaps to deal with



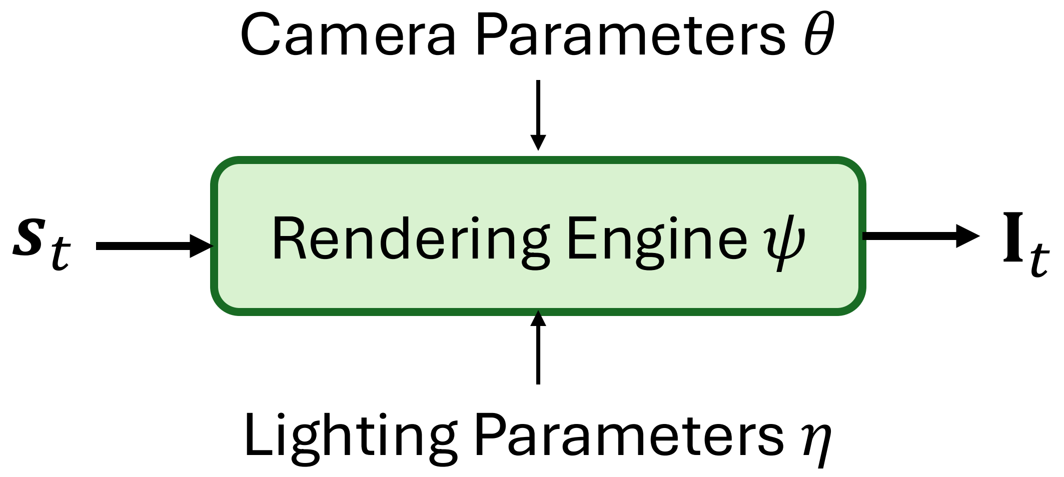
Chen, Tao, et al. "Visual dexterity: In-hand reorientation of novel and complex object shapes." *Science Robotics* 8.84 (2023): eadc9244

Role of Simulation with a small cost

Sim2Real Pipeline: Two major Sim2Real gaps to deal with

Visual Rendering

How **physics engine** renders camera vs output of **actual camera** models



$$\mathbf{I}_t^{\mathrm{sim}} = \psi_{\mathrm{renderer}}(\boldsymbol{s}_t; \theta)$$

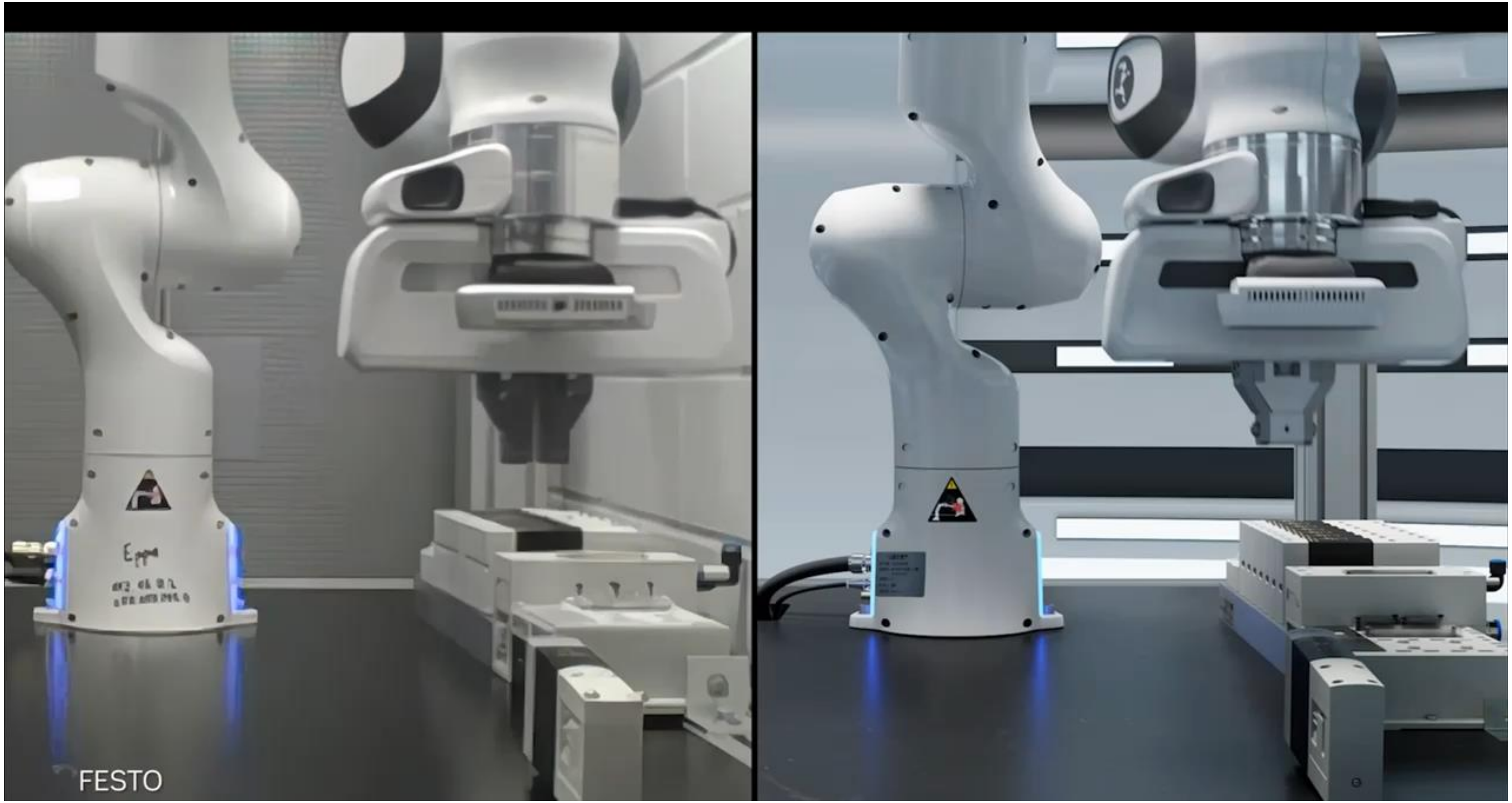
$$\mathbf{I}_t^{\mathrm{real}} = \psi_{\mathrm{real\_cam}}(\boldsymbol{s}_t)$$

1. **System Identification (SysID):** Find the right θ, η that best matches $\mathbf{I}_t^{\mathrm{real}}$
2. **Domain Randomization (DR):**

$$\theta \sim p_1(\theta)$$
$$\eta \sim p_2(\eta)$$

Role of Simulation with a small cost

Sim2Real Pipeline： Two major Sim2Real gaps to deal with



This Wednesday

Haoshu Fang

Policy Learning with
alternative datasets
without teleoperation!

Fill out a survey!