

# Pathway to Generalist Robots:

## Scaling Law, Data Flywheel, and Humanlike Embodiment

Yuke Zhu

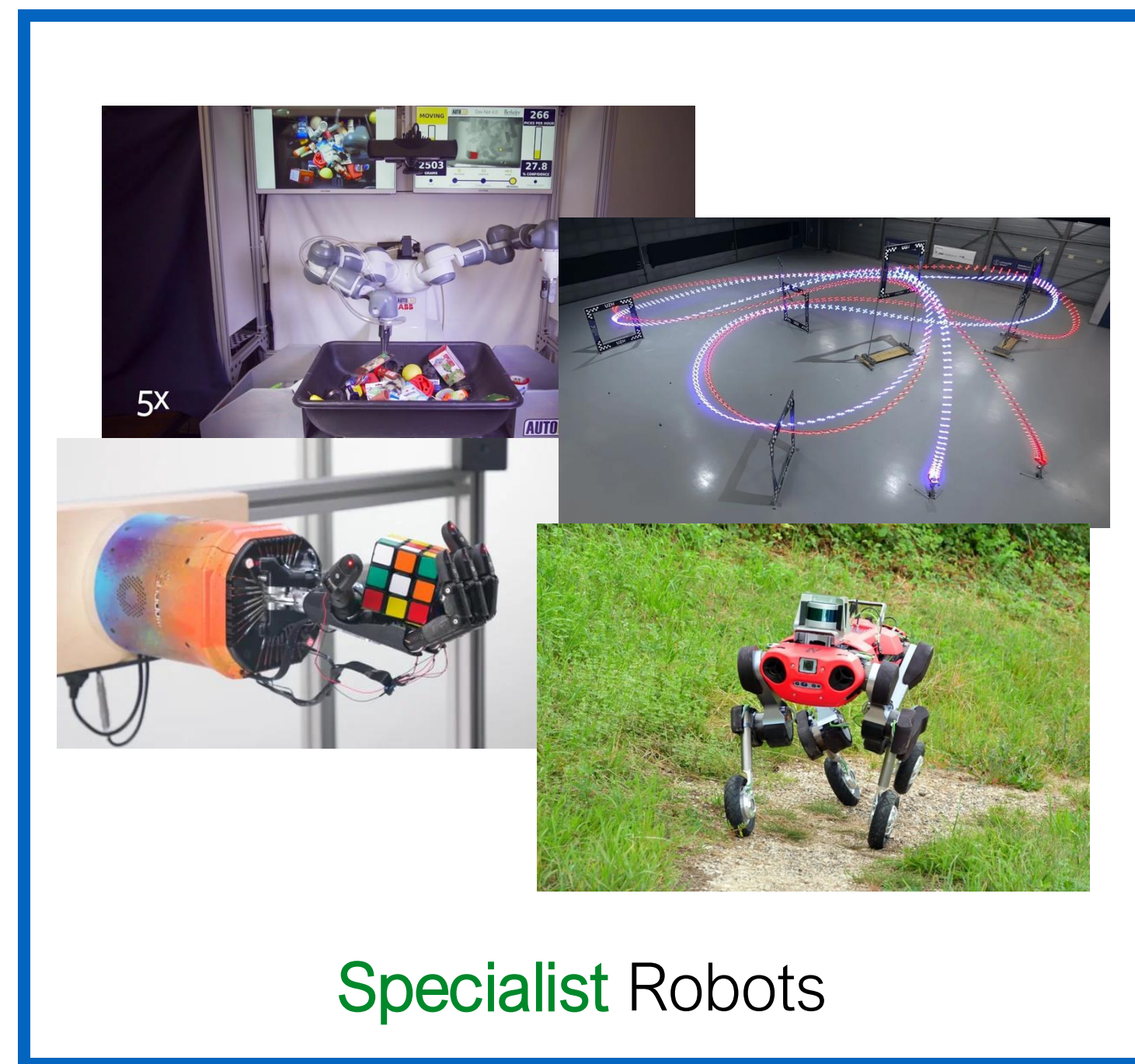
November 8<sup>th</sup>, 2023



TEXAS Robotics



# Specialist → Generalist → Specialized Generalist



Generalist Robot Models





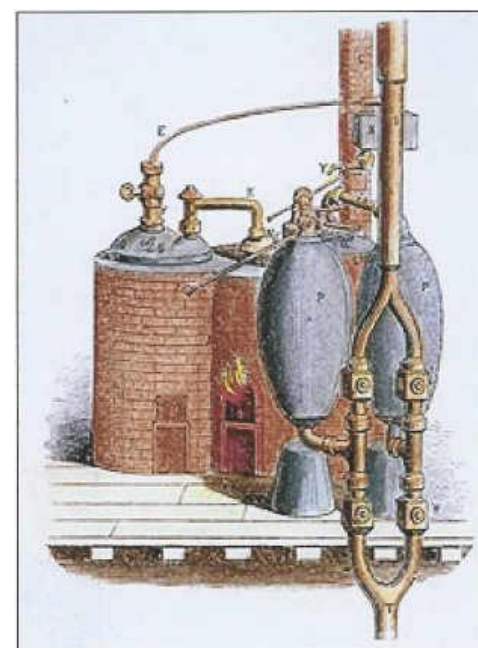
# Specialist → Generalist → Specialized Generalist



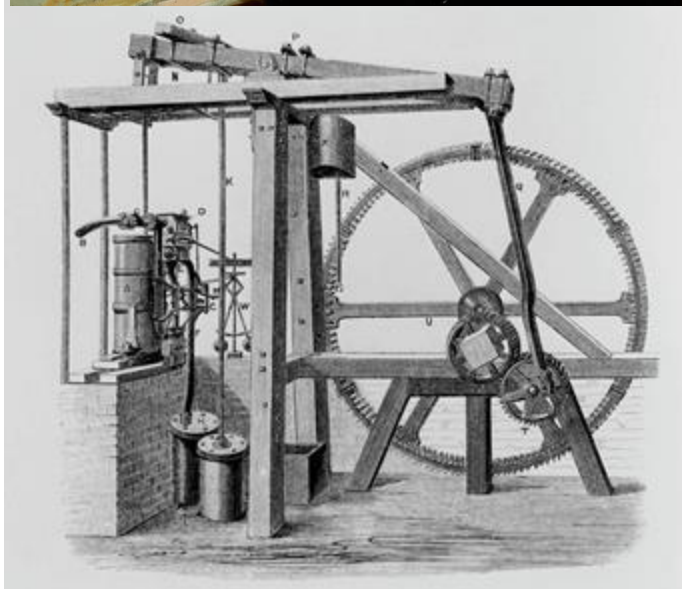
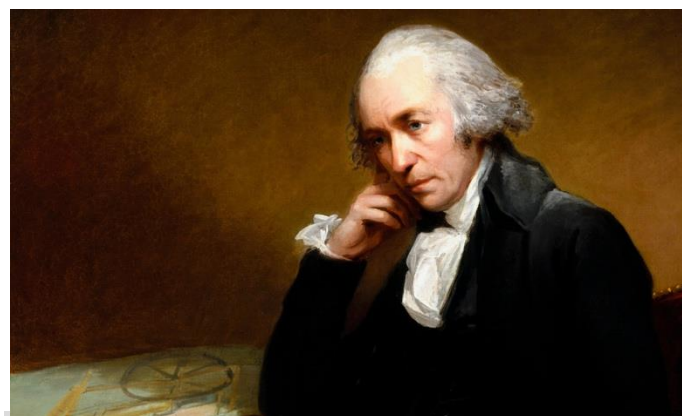
Aeolipile (Heron of Alexandria, 10–70 AD)



Rudimentary Steam Turbine (Taqi al-Din, 1551)



Savery Steam Pump (Thomas Savery, 1698)



Watt's Steam Engine (James Watt, 1764)



Internal Combustion Engine



Steam Turbine



# Specialist → Generalist → Specialized Generalist

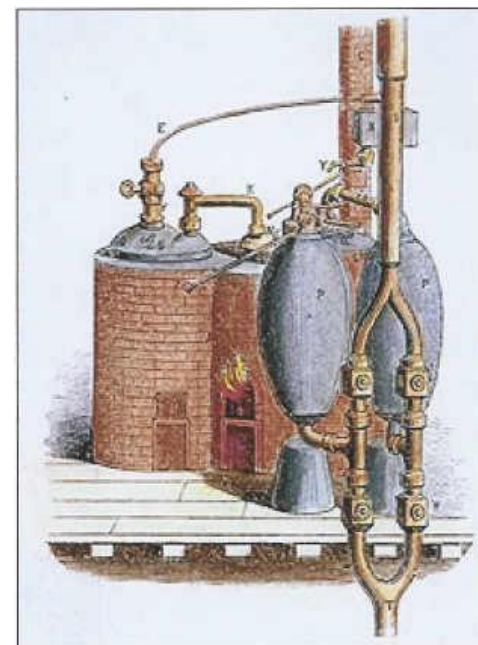
Specialist technology working in limited conditions



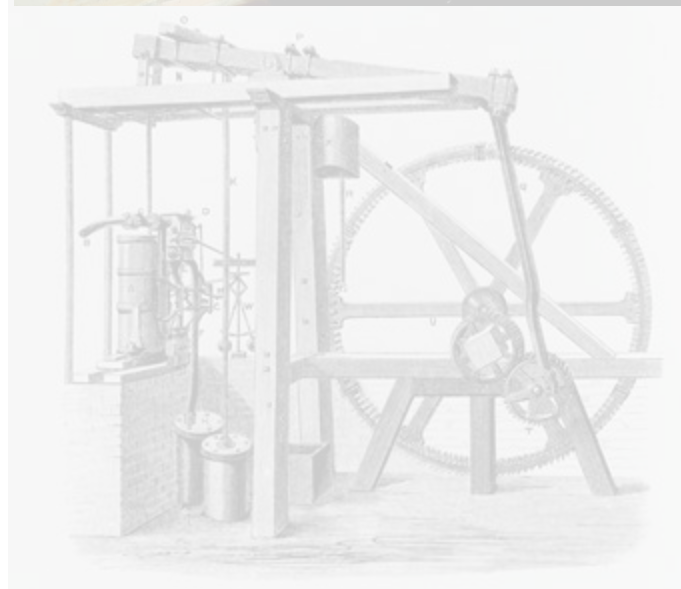
Aeolipile (Heron of Alexandria, 10–70 AD)



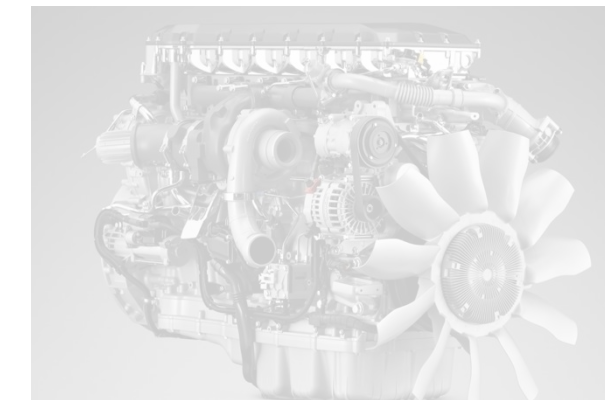
Rudimentary Steam Turbine (Taqi al-Din, 1551)



Savery Steam Pump (Thomas Savery, 1698)



Watt's Steam Engine (James Watt, 1764)



Internal Combustion Engine



Steam Turbine



# Specialist → Generalist → Specialized Generalist

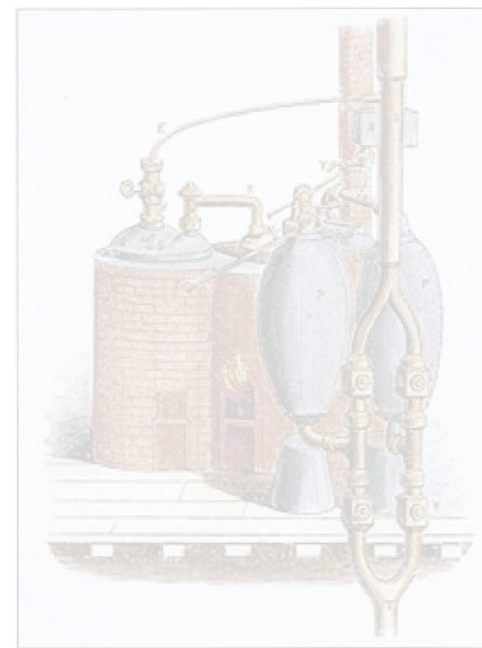
Generalist technology leading to widespread adoption



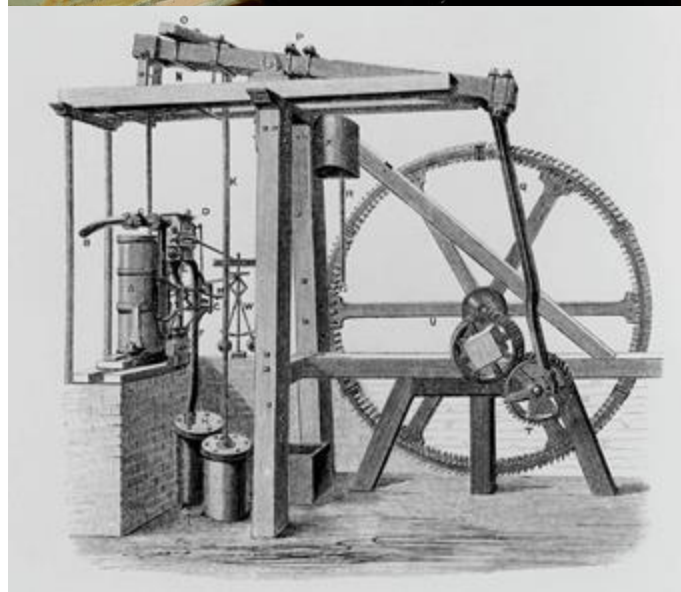
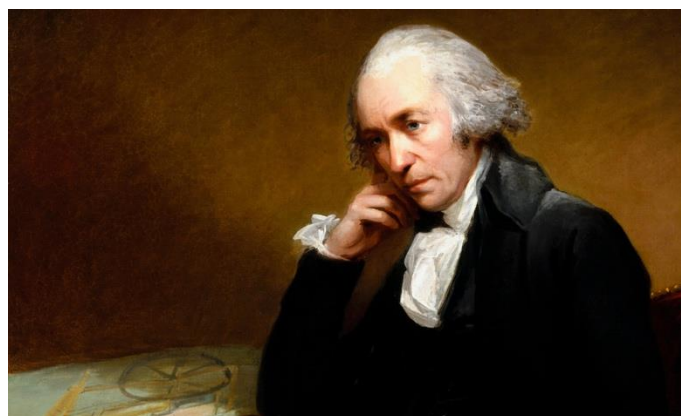
Aeolipile (Heron of Alexandria, 10–70 AD)



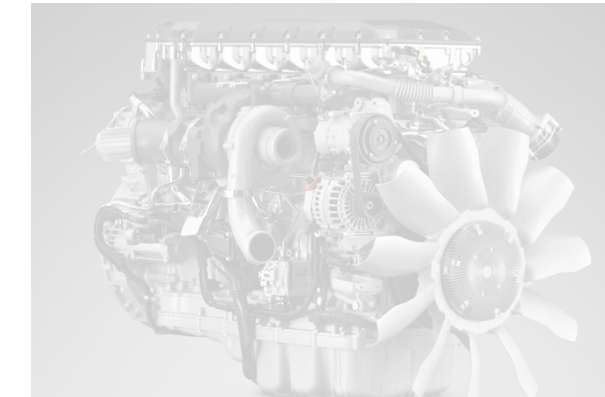
Rudimentary Steam Turbine (Taqi al-Din, 1551)



Savery Steam Pump (Thomas Savery, 1698)



Watt's Steam Engine  
(James Watt, 1764)



Internal Combustion Engine



Steam Turbine





# Specialist → Generalist → Specialized Generalist

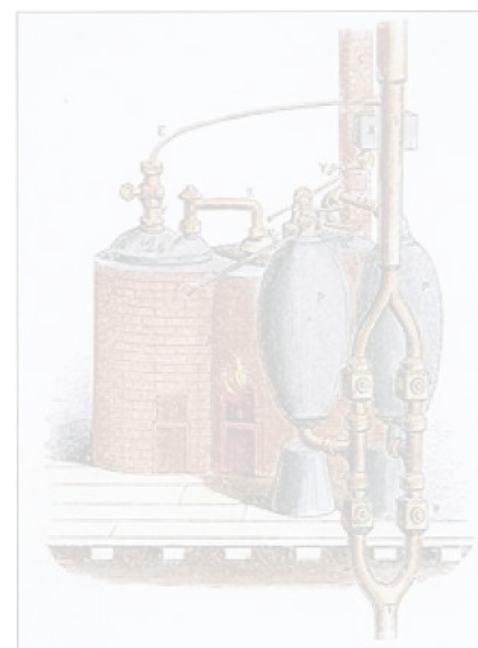
Further specialization for various problem domains



Aeolipile (Heron of Alexandria, 10–70 AD)



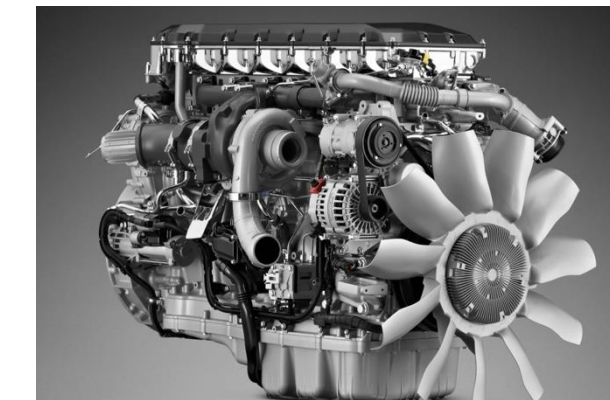
Rudimentary Steam Turbine (Taqi al-Din, 1551)



Savery Steam Pump (Thomas Savery, 1698)



Watt's Steam Engine (James Watt, 1764)



Internal Combustion Engine

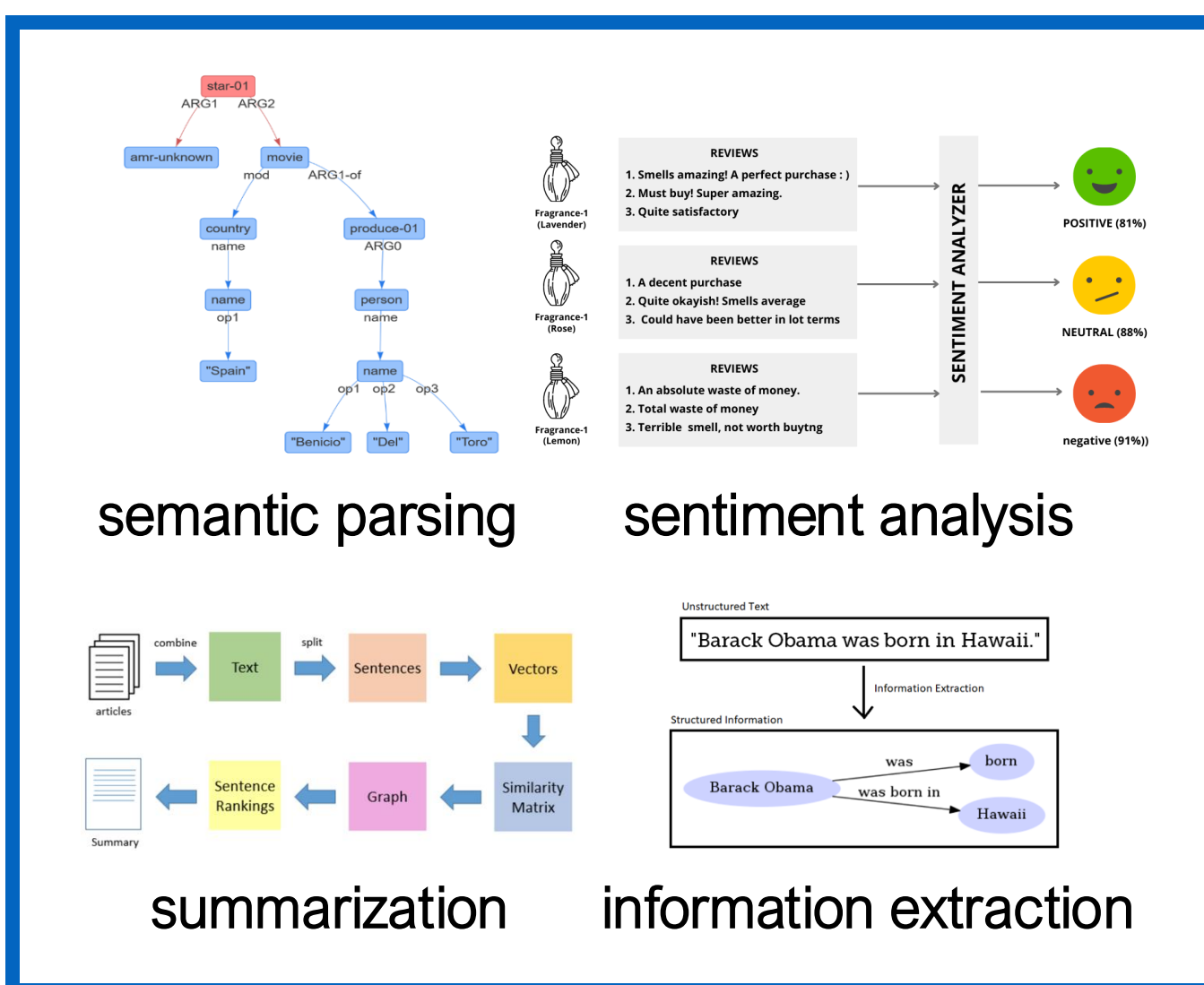
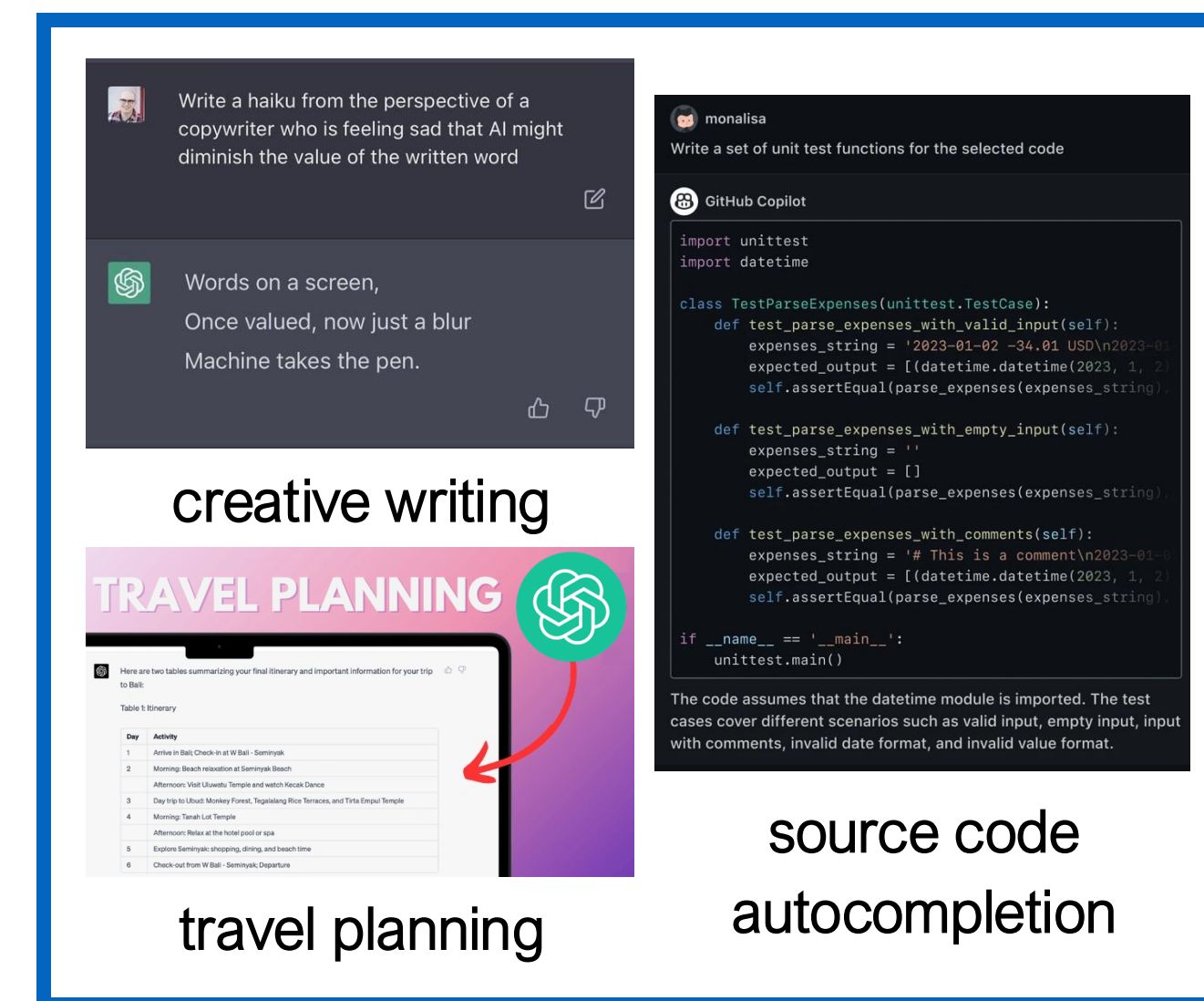
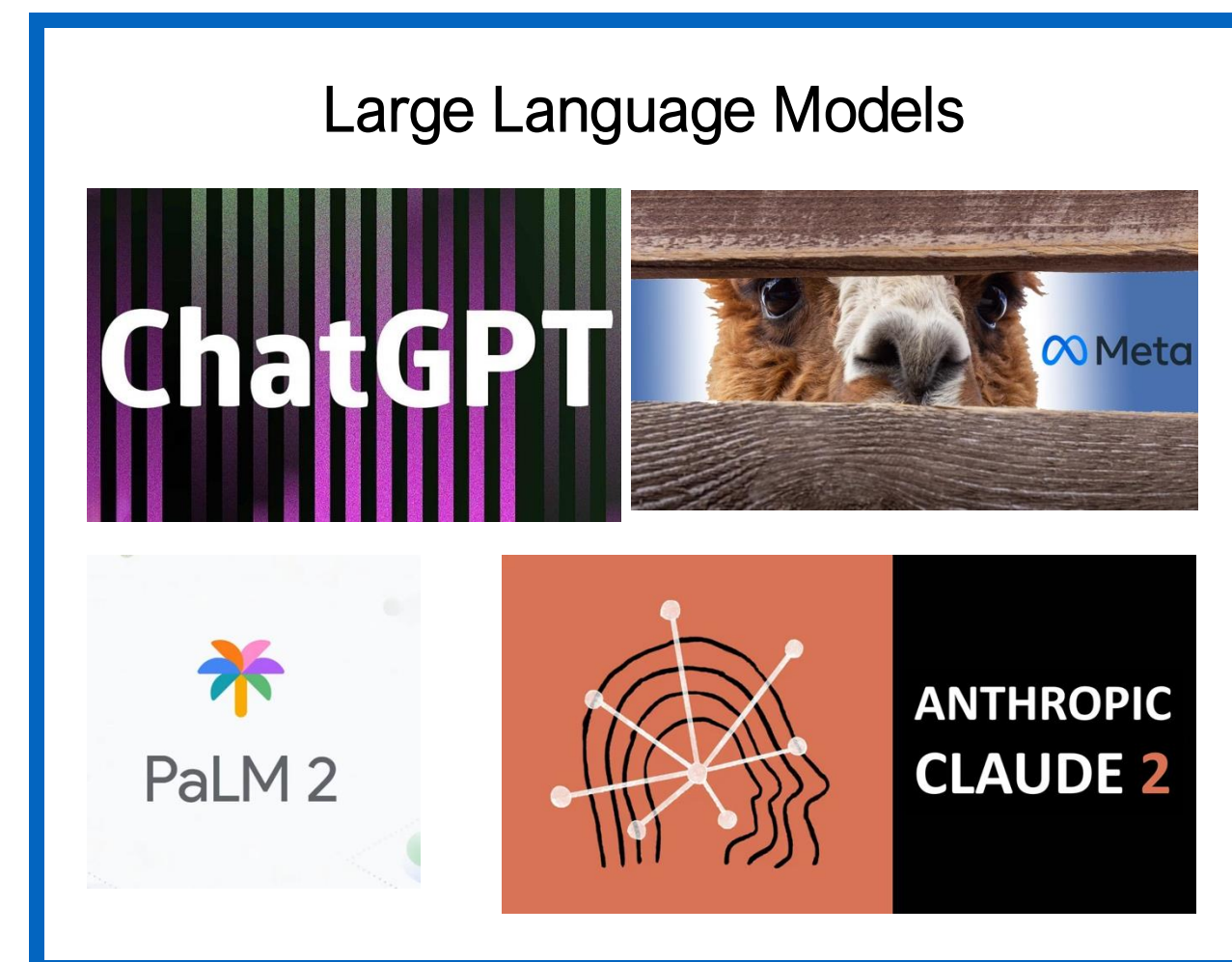


Steam Turbine

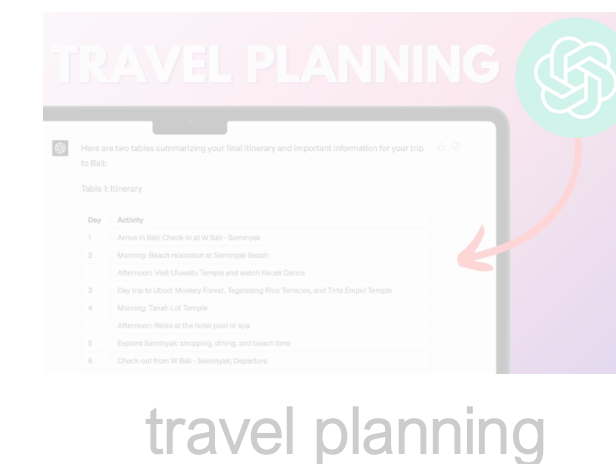
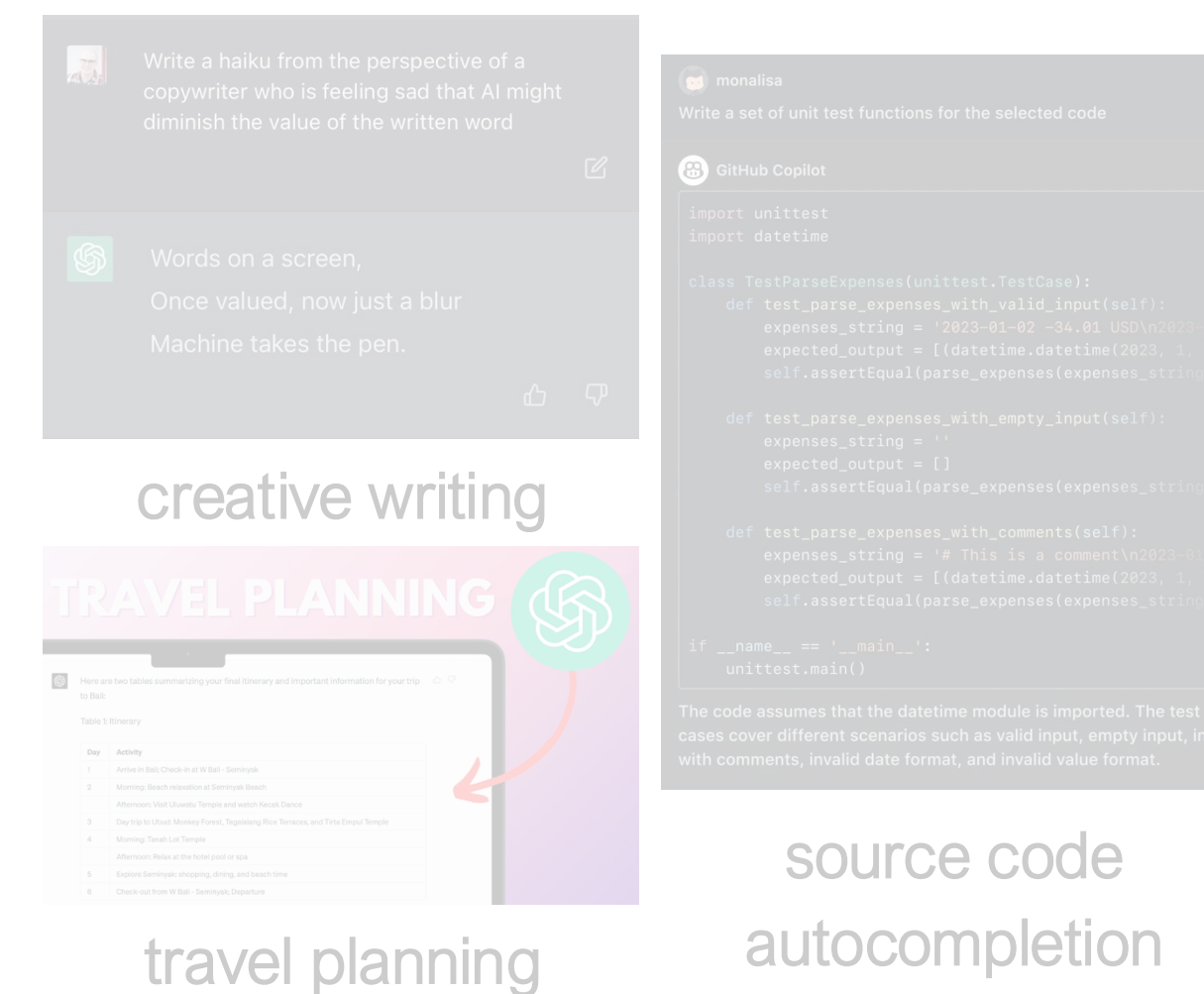
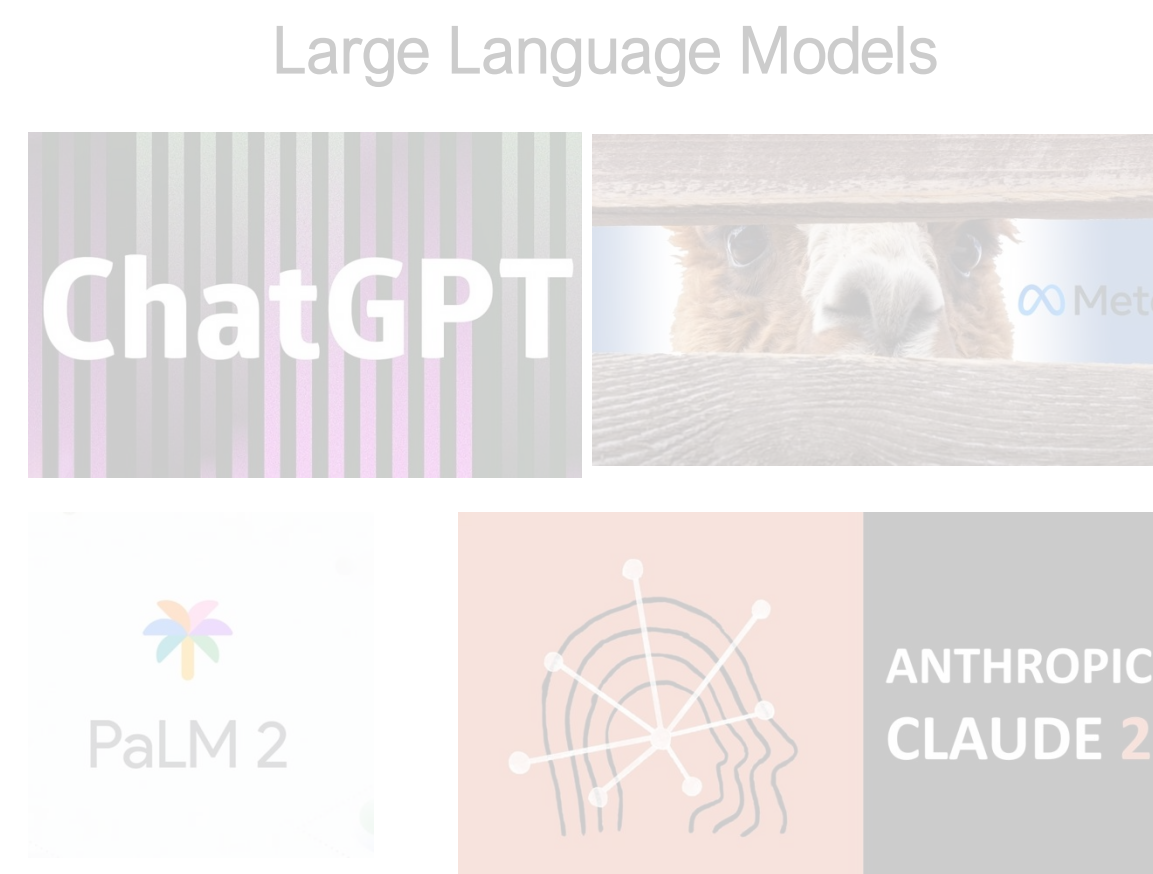
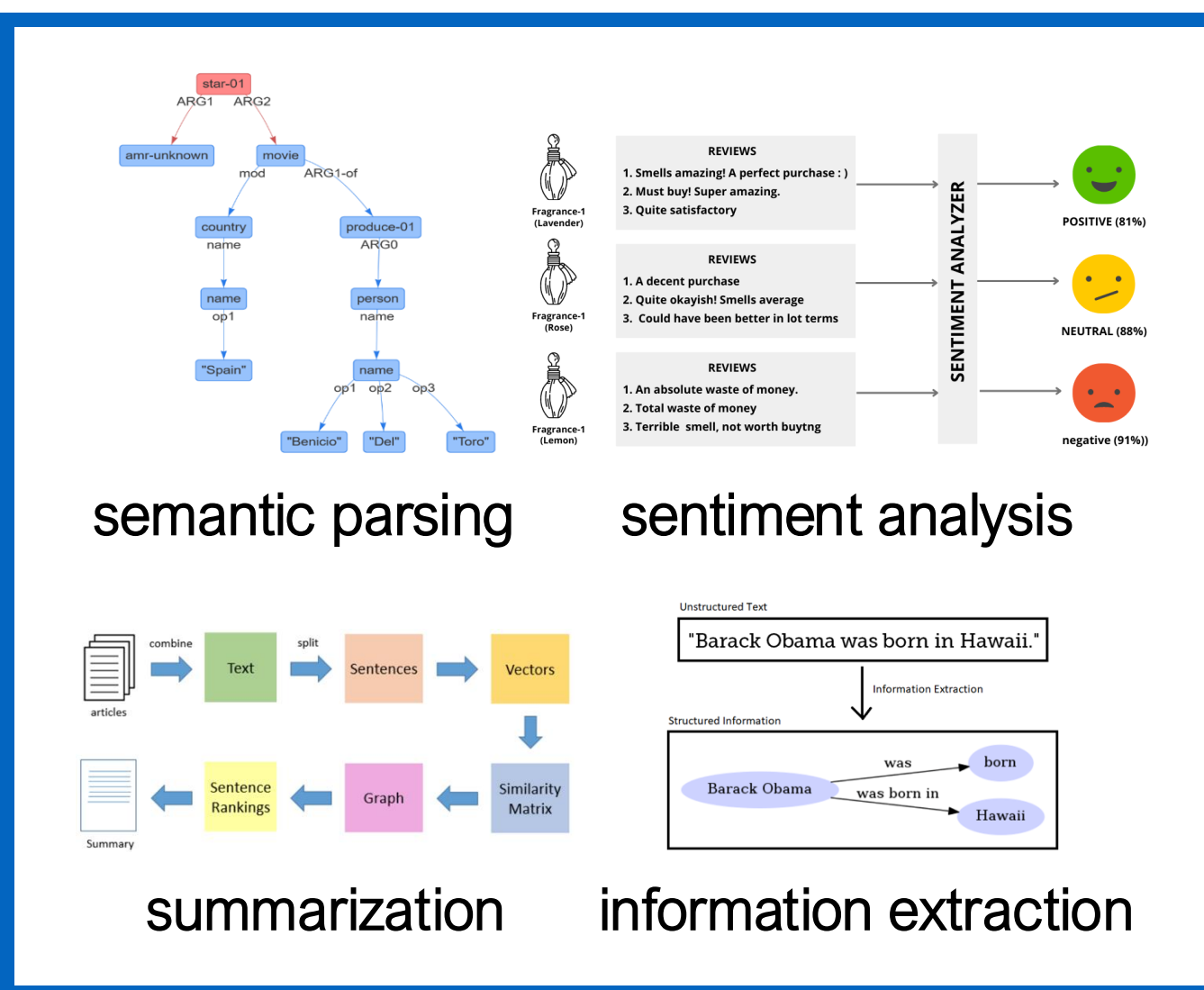




# Specialist → Generalist → Specialized Generalist

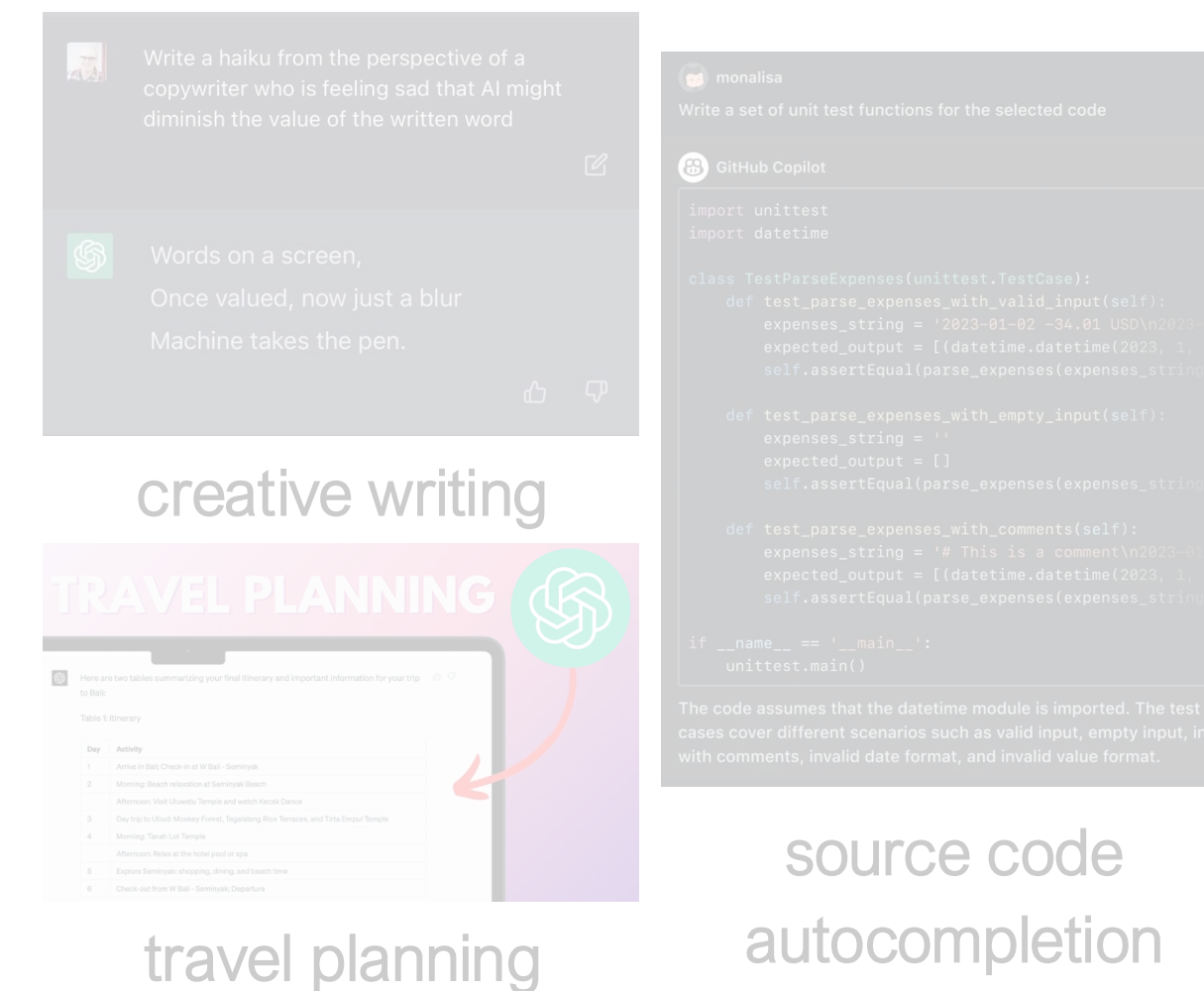
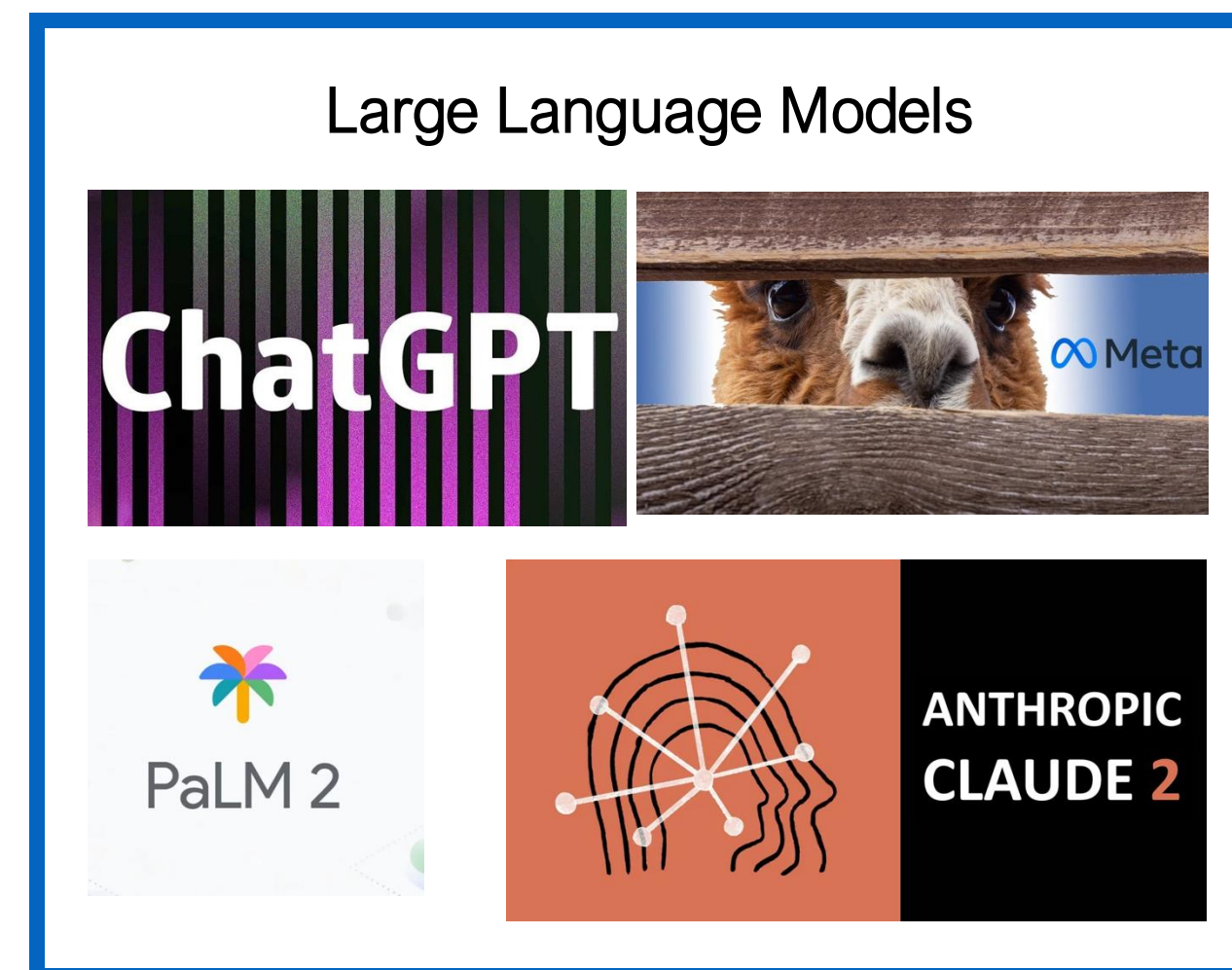
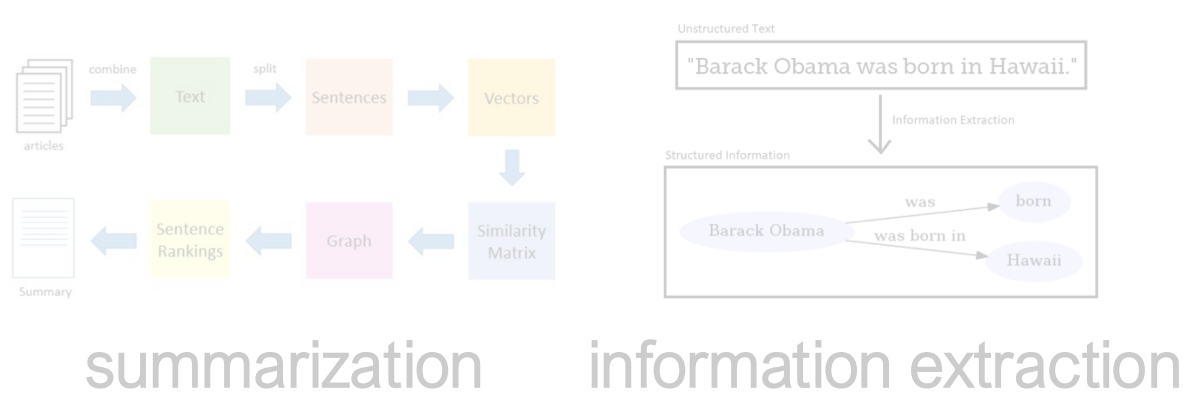
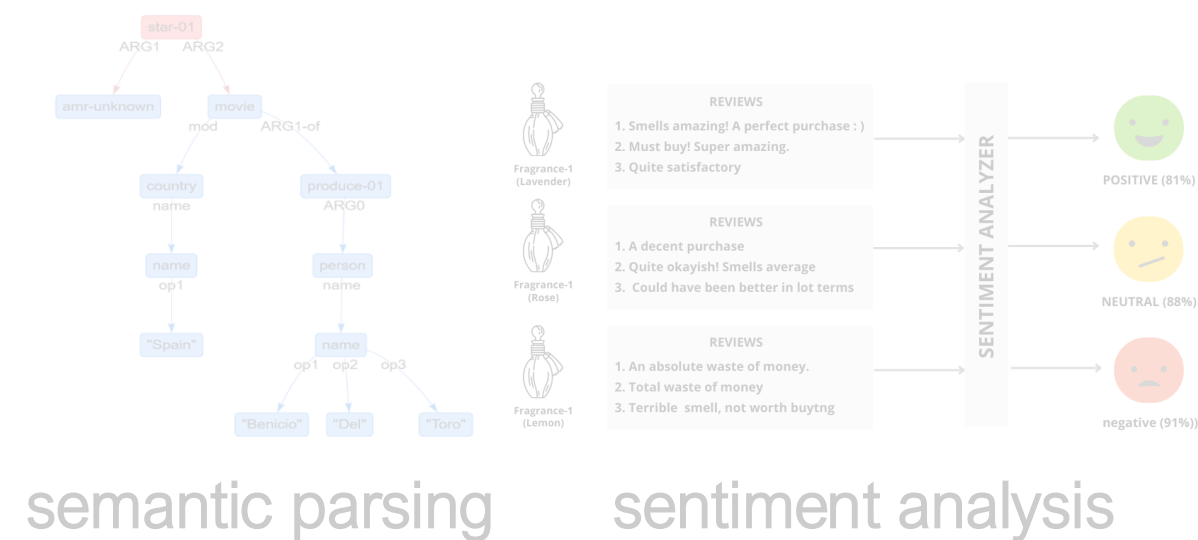


# Specialist → Generalist → Specialized Generalist



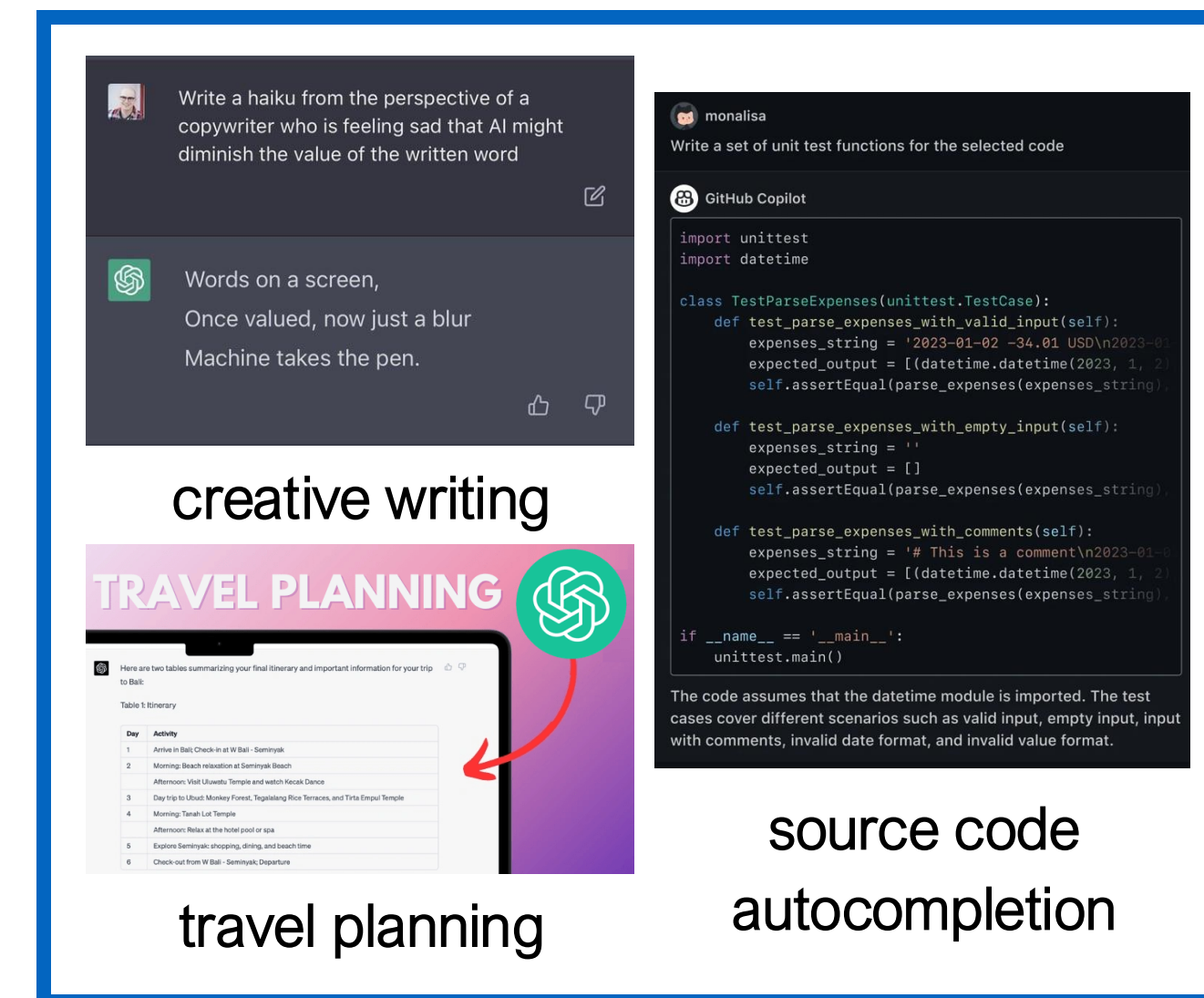
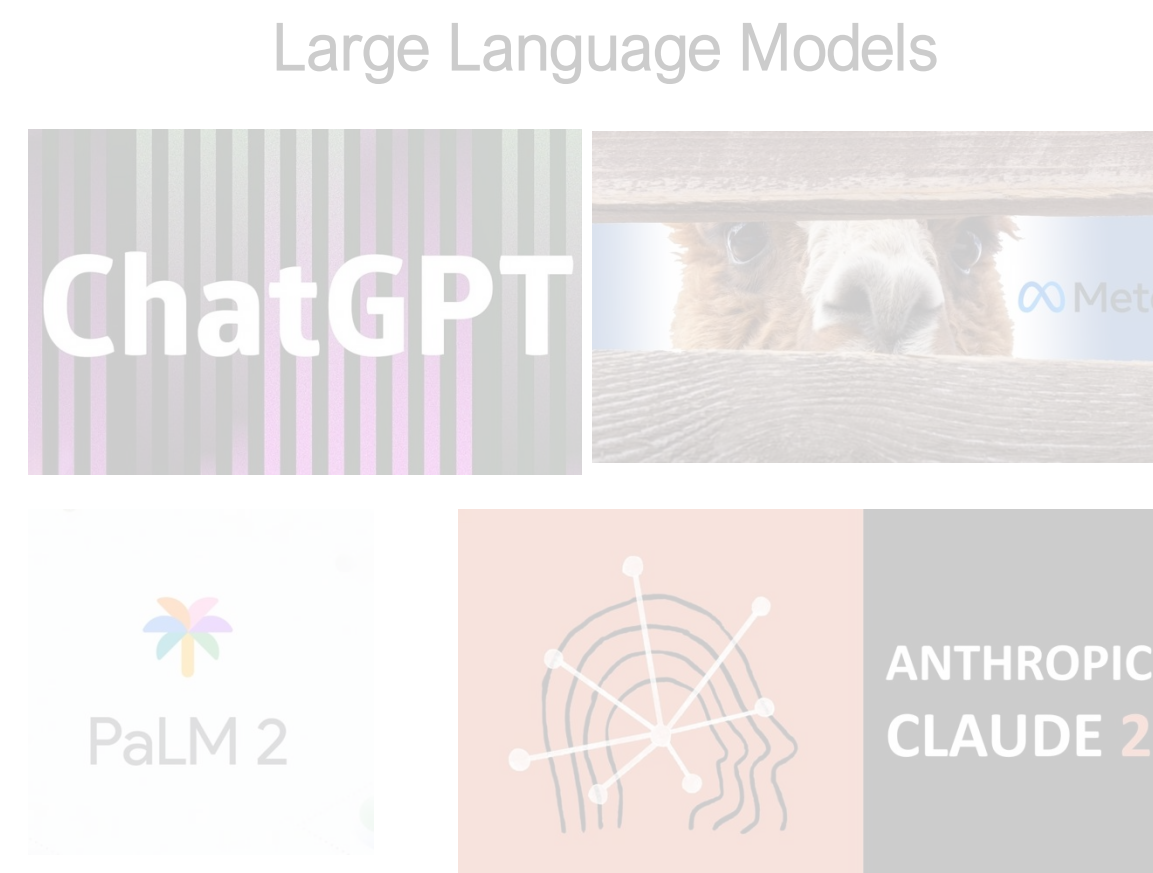
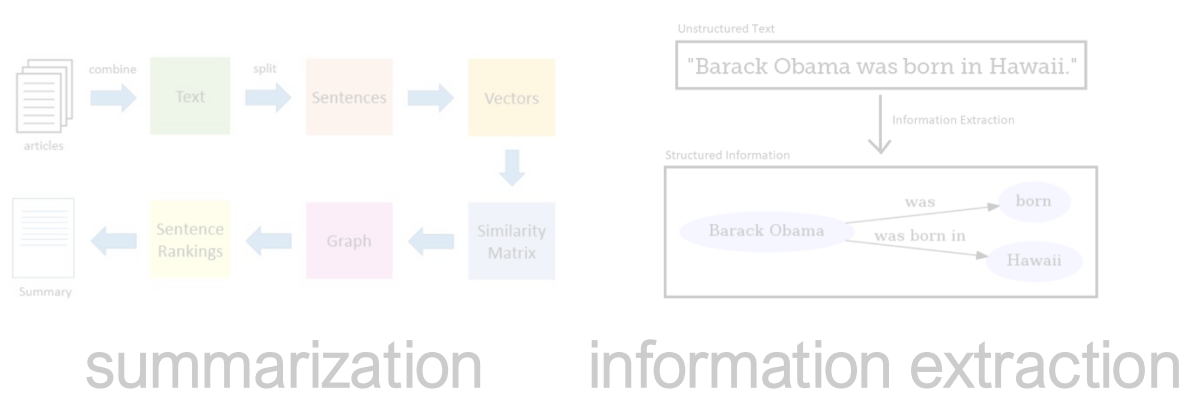
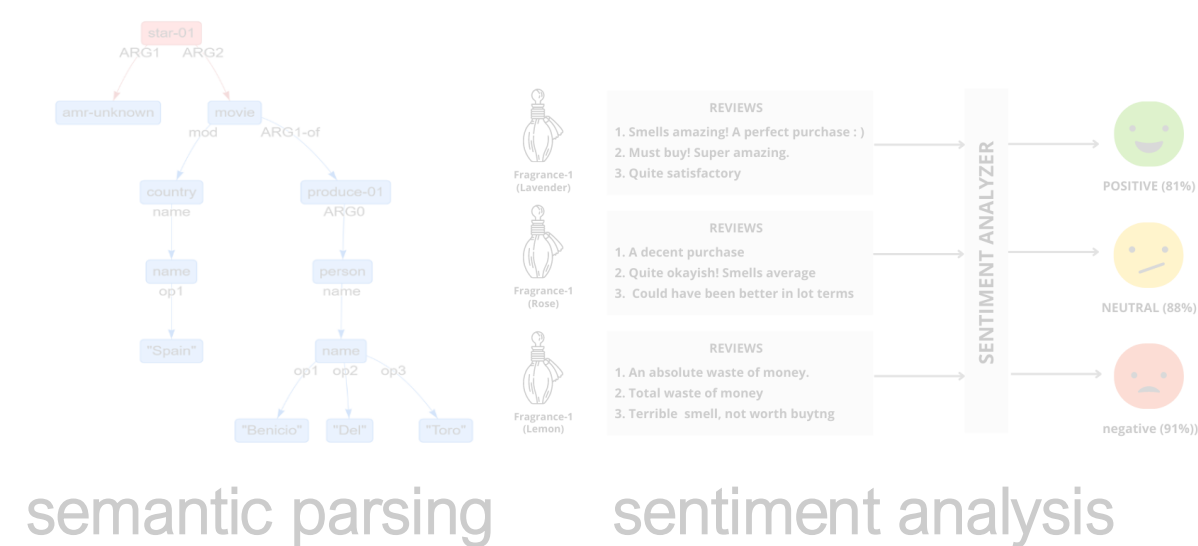


# Specialist → Generalist → Specialized Generalist





# Specialist → Generalist → Specialized Generalist





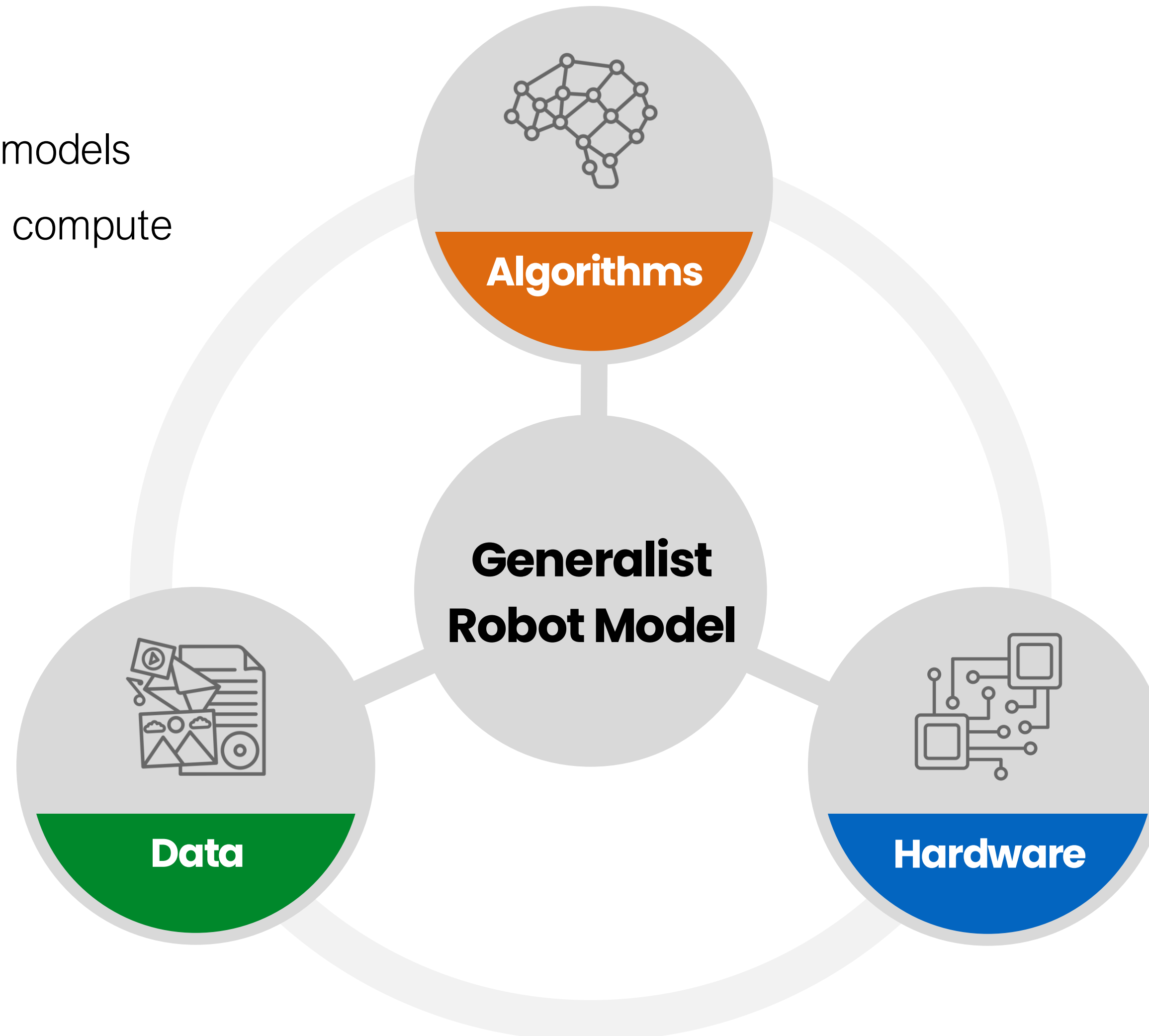
# Recipe for Building Generalist Robot Models

## Scaling Law

Powerful robot learning models  
that scale with data and compute

## Data Flywheel

New mechanism to collect  
massive training data



## Humanlike Embodiment

Humanoid robot platform for  
broad applications



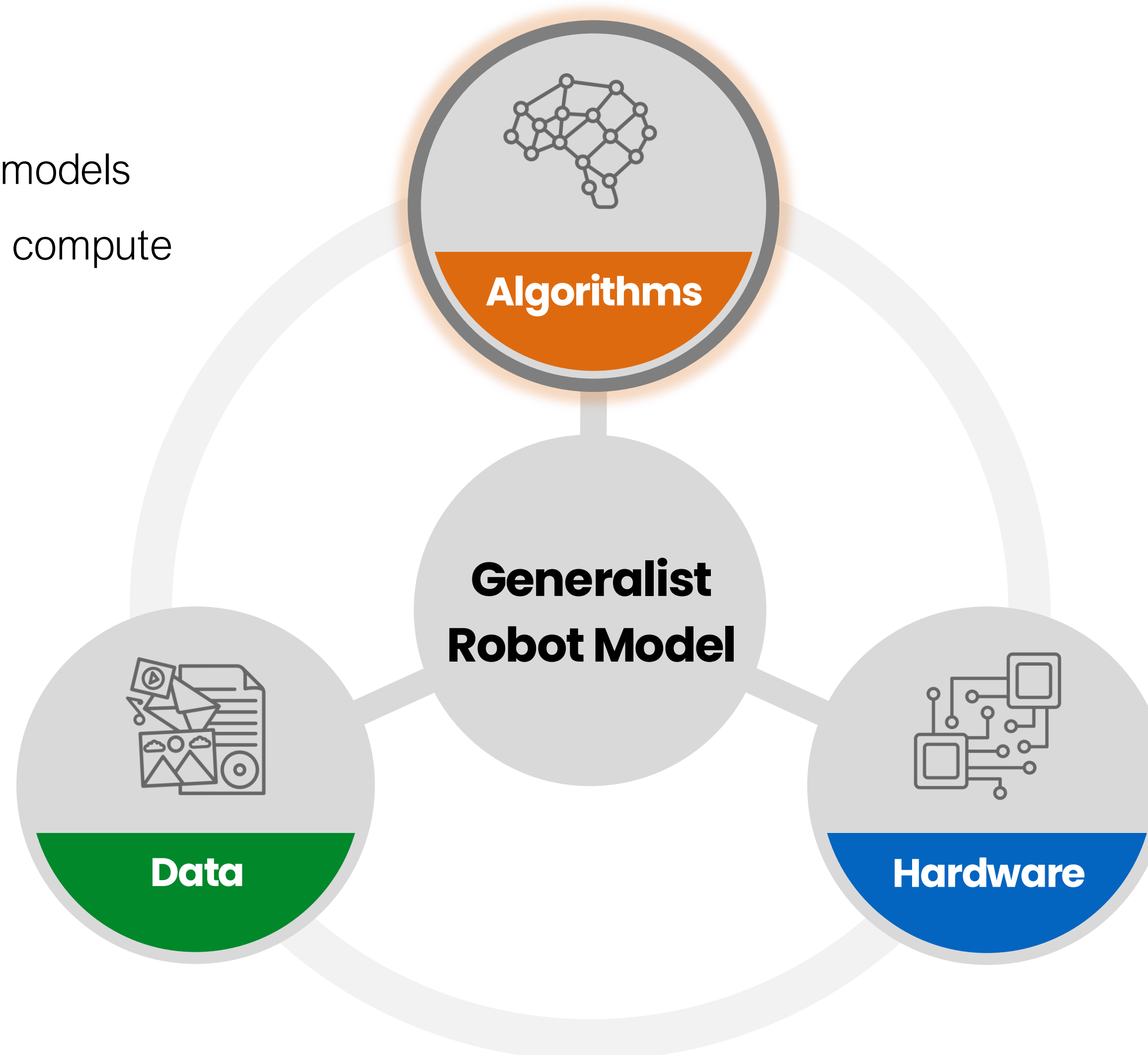
# Recipe for Building Generalist Robot Models

## Scaling Law

Powerful robot learning models  
that scale with data and compute

## Data Flywheel

New mechanism to collect  
massive training data

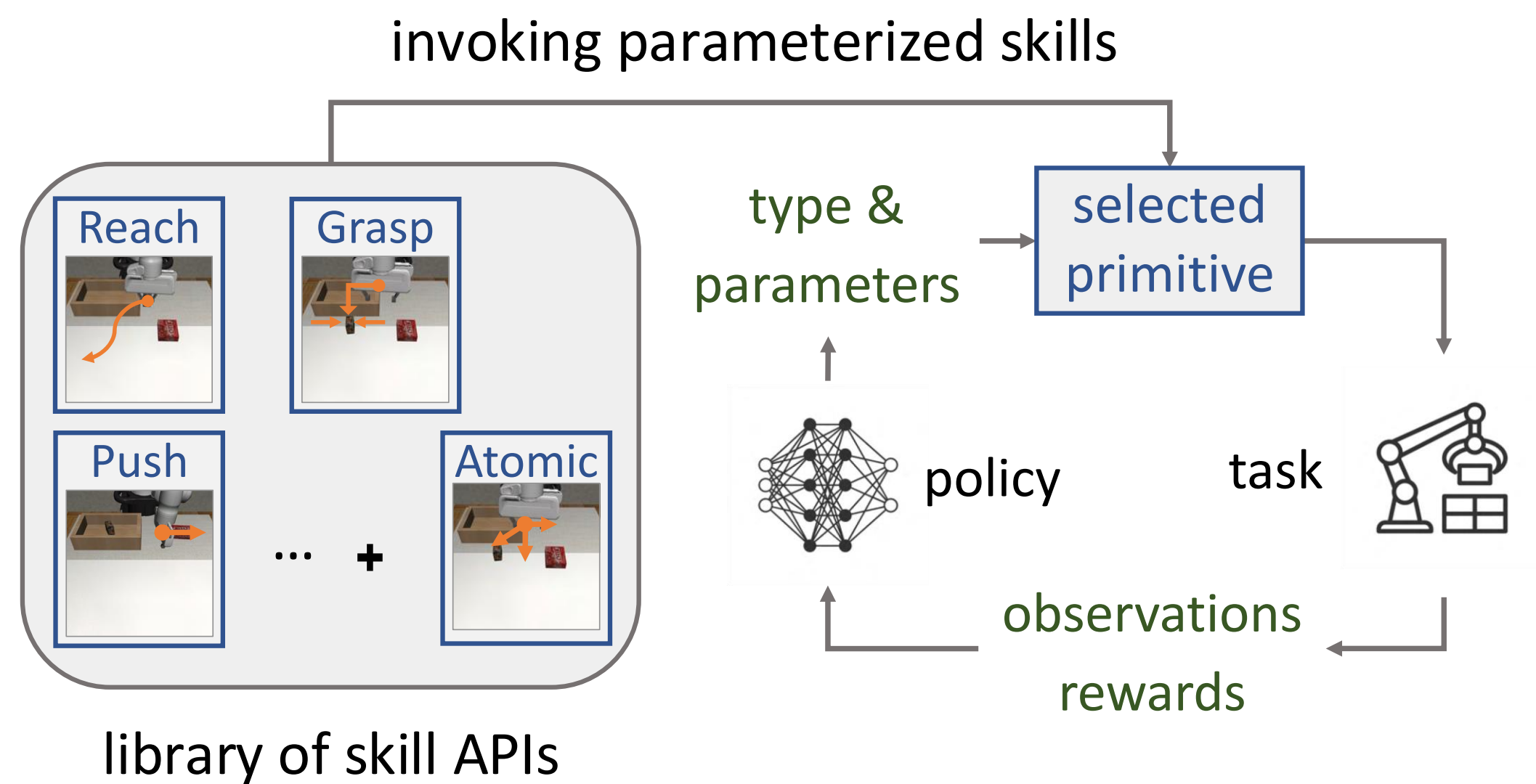


## Humanlike Embodiment

Humanoid robot platform for  
broad applications



# Key Idea: Skills as APIs and Scaling Law



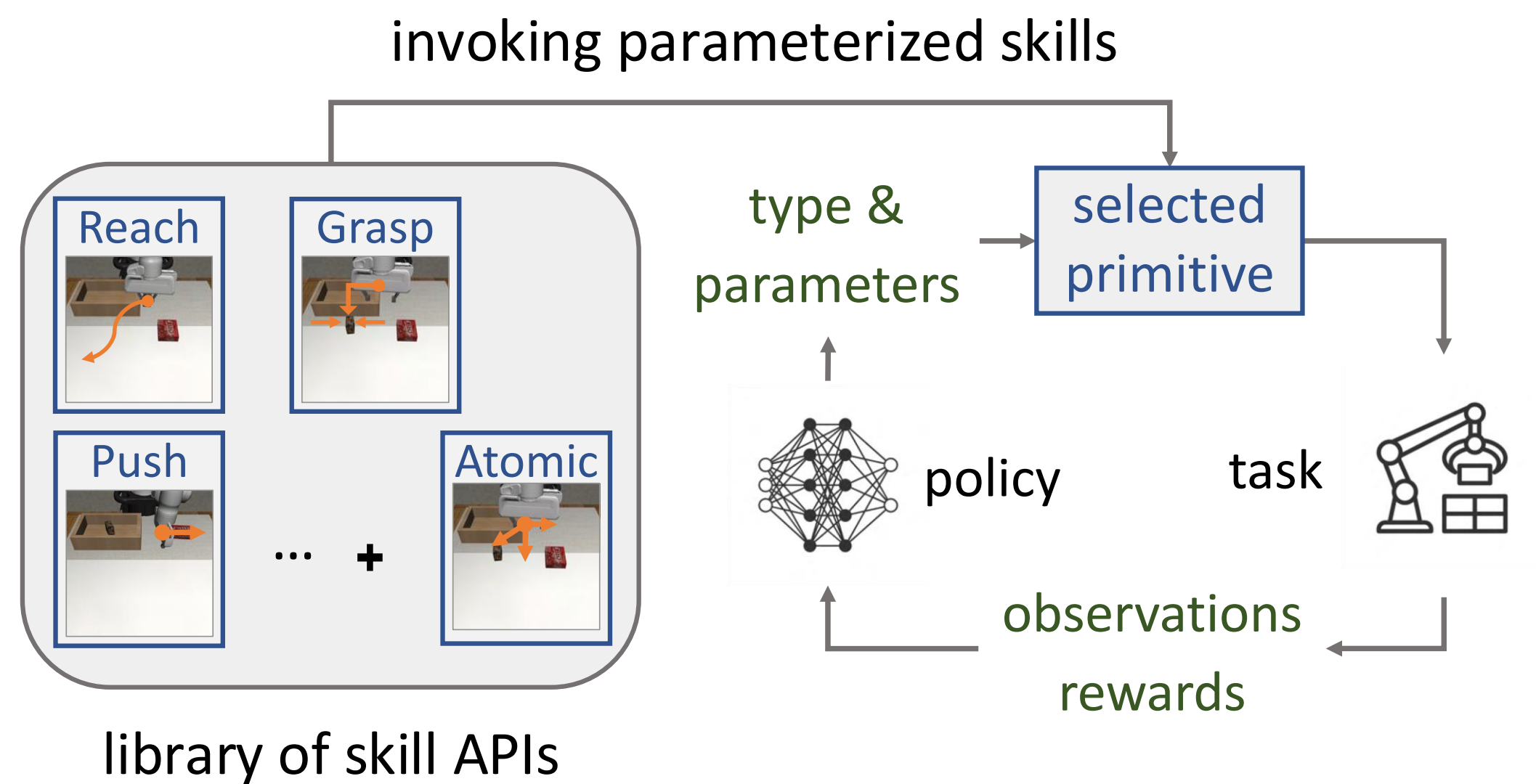
## Skills as APIs

(Nasiriany et al. ICRA 2022)



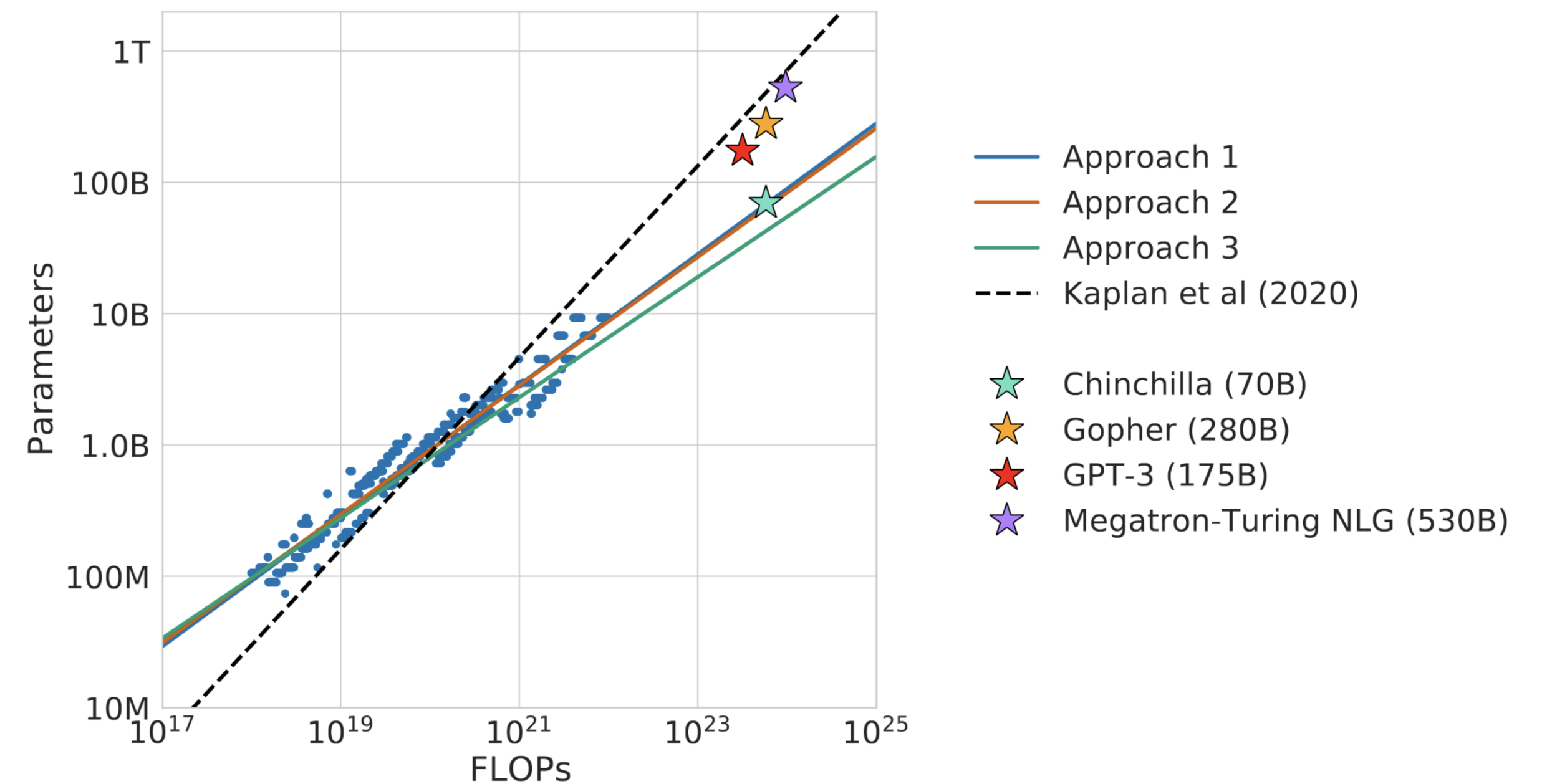
Soroush Nasiriany

# Key Idea: Skills as APIs and Scaling Law



## Skills as APIs

(Nasiriany et al. ICRA 2022)



## Scaling Law for Language Models

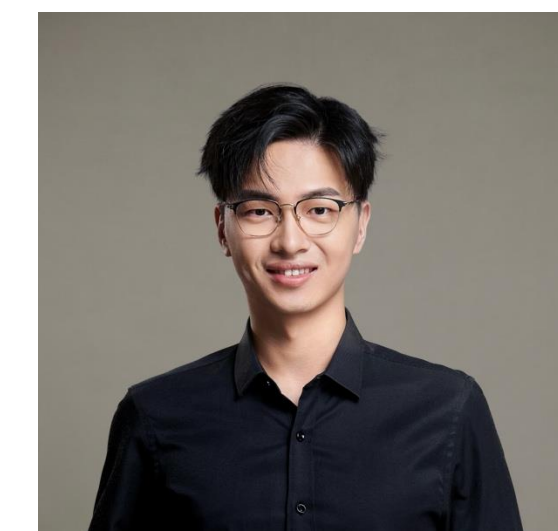
(Kaplan et al. 2020; Hoffmann et al. 2022)



# Data-Efficient Imitation Learning for Sensorimotor Skills

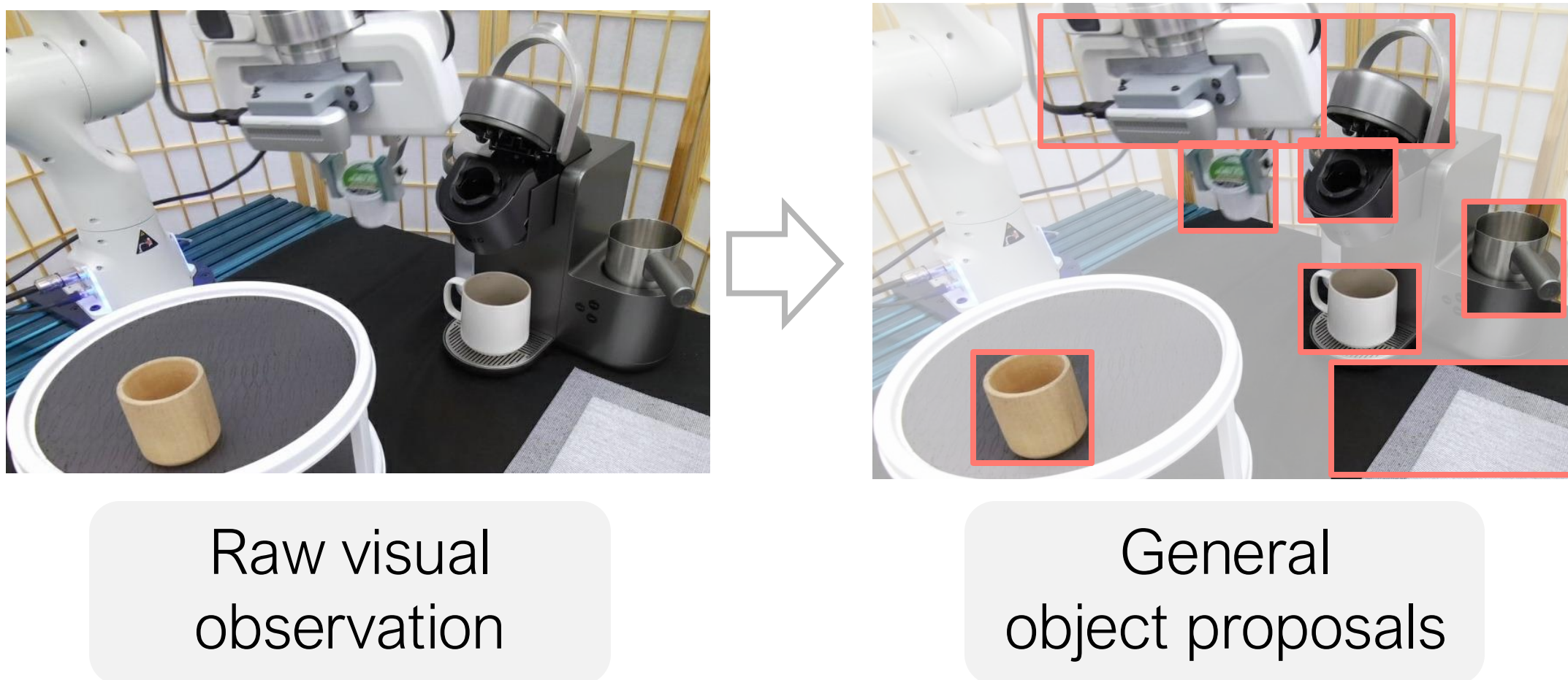


Raw visual  
observation



Yifeng Zhu

# Data-Efficient Imitation Learning for Sensorimotor Skills



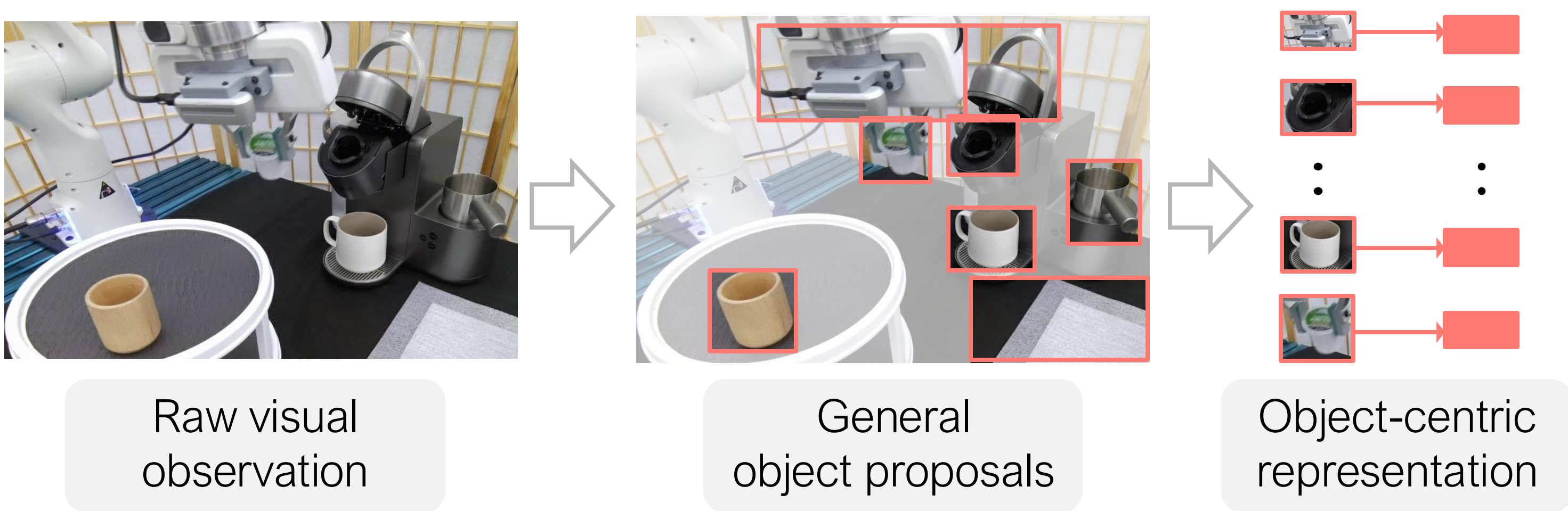
Yifeng Zhu



Large vision model trained on Internet-scale image datasets



# Data-Efficient Imitation Learning for Sensorimotor Skills

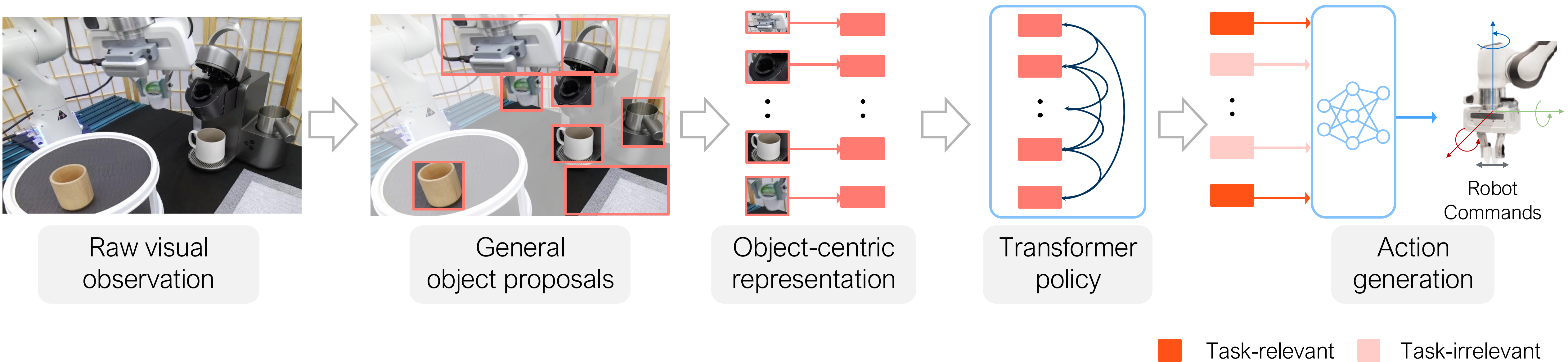


Encoding object **visual appearances**  
and their **spatial locations**



Yifeng Zhu

# Data-Efficient Imitation Learning for Sensorimotor Skills



Using transformers to select **task-relevant objects** and reason about **their relations**



Yifeng Zhu





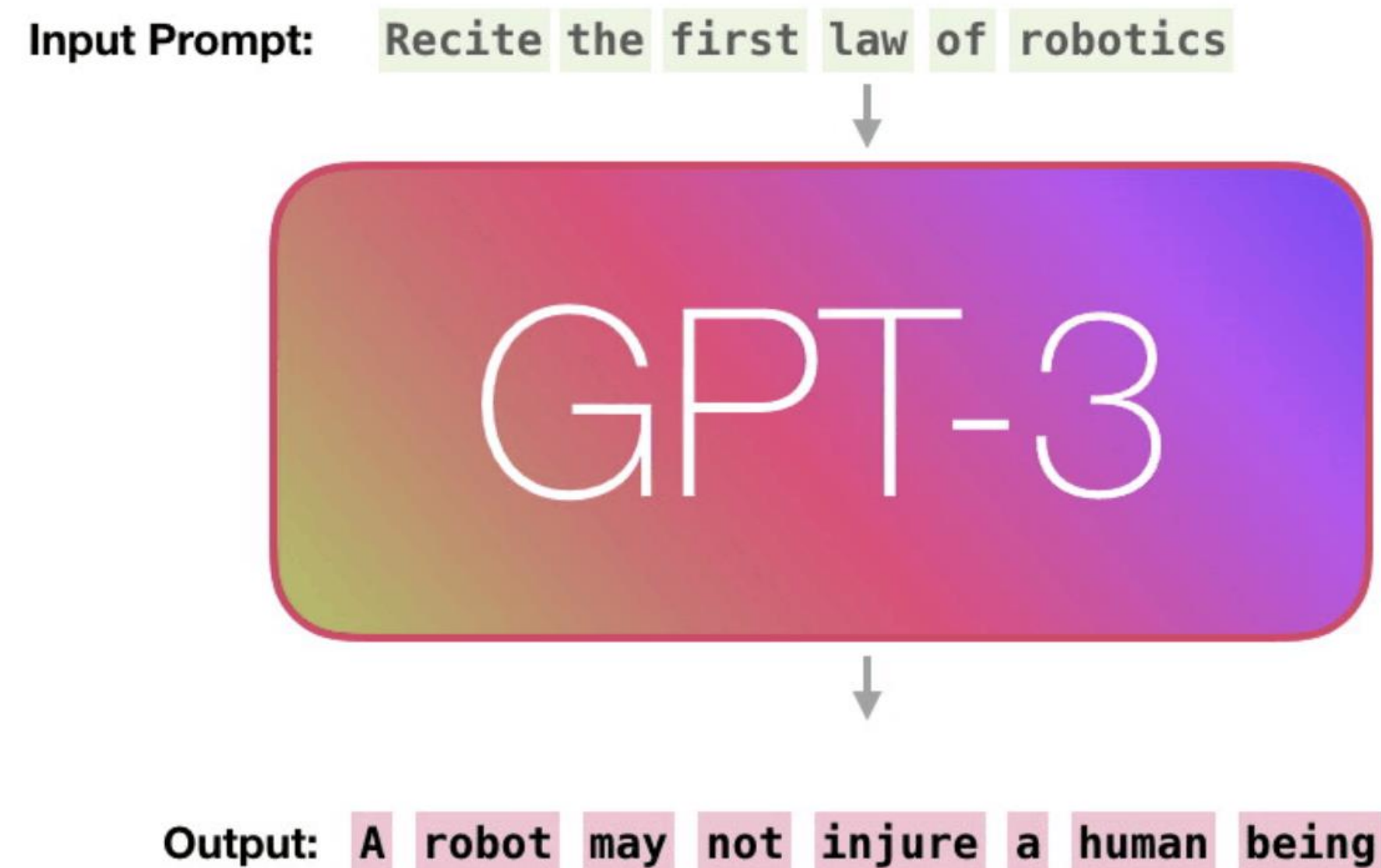
Policy Input

It learns to make coffee without additional 50 annotations on human demonstrations.

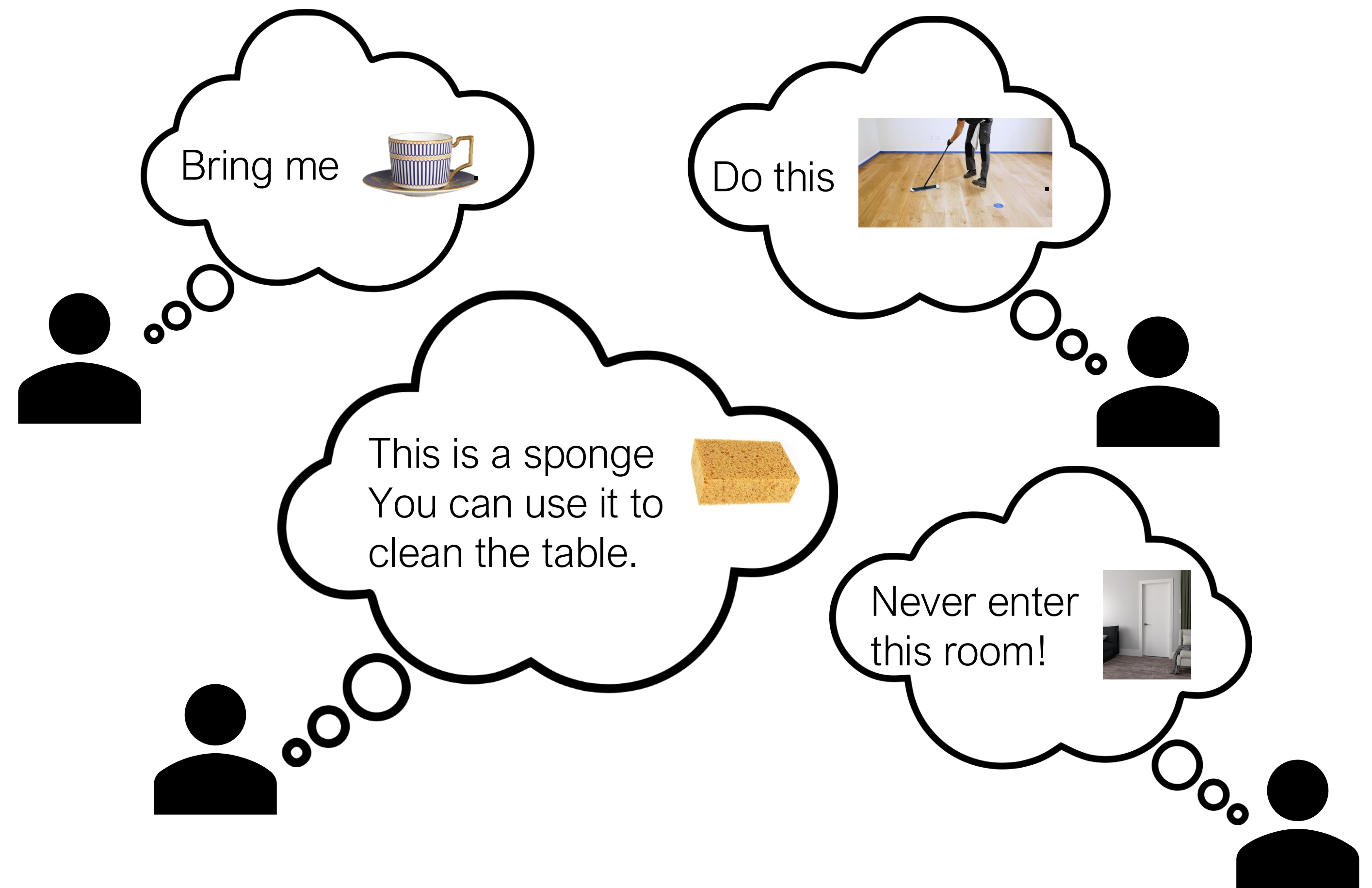


# Massively Multi-Task Robot Learning for Model Scaling

What if we can prompt a household robot to ...

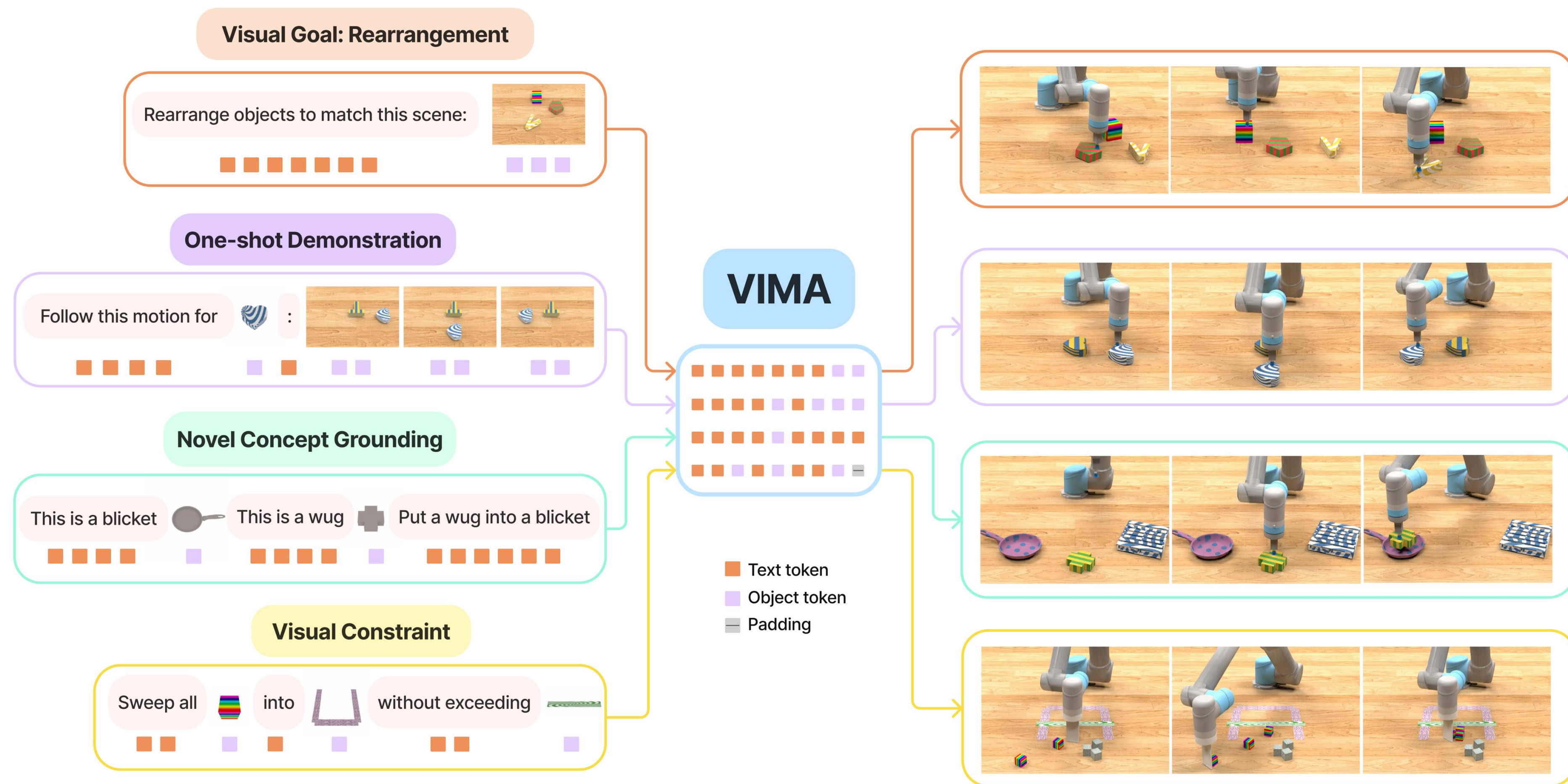


[Credit: Jay Alammar]





# Massively Multi-Task Robot Learning for Model Scaling



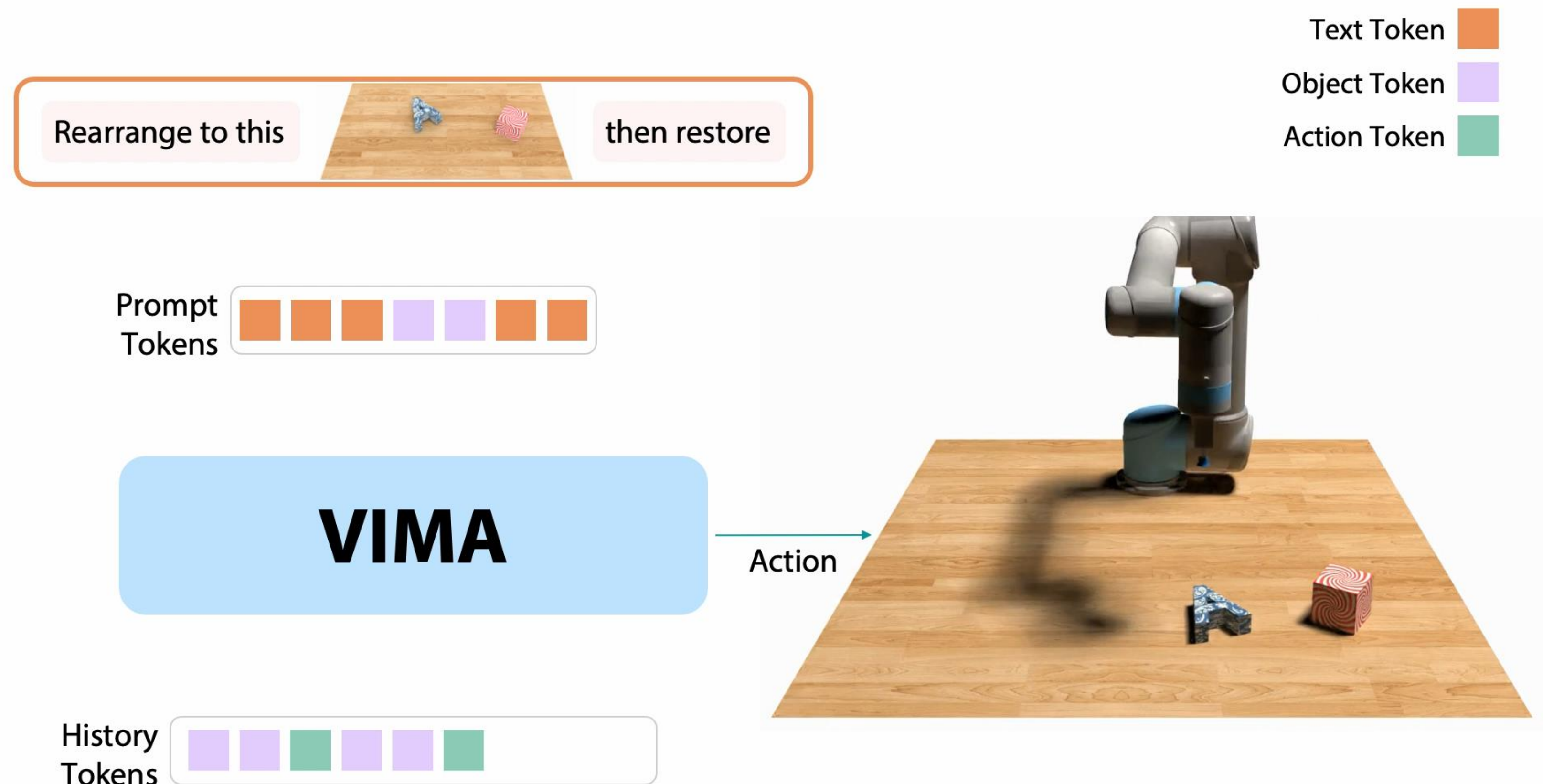
**generalist robot agent** for multi-task learning and zero-shot generalization



# Massively Multi-Task Robot Learning for Model Scaling

VIMA: Visuo-Motor Attention model

- Transformer **encoder-decoder**;
- Encodes **multimodal prompts** with a frozen language model;
- **Object-centric** representations for visual observations
- Predicts **skill APIs** given the prompt and interaction history.





# Massively Multi-Task Robot Learning for Model Scaling

Data scalability from 0.1% to full dataset

## Level 1 Object Placement

Put the  into the 



## Level 2 Novel Combination

Put the  into the 



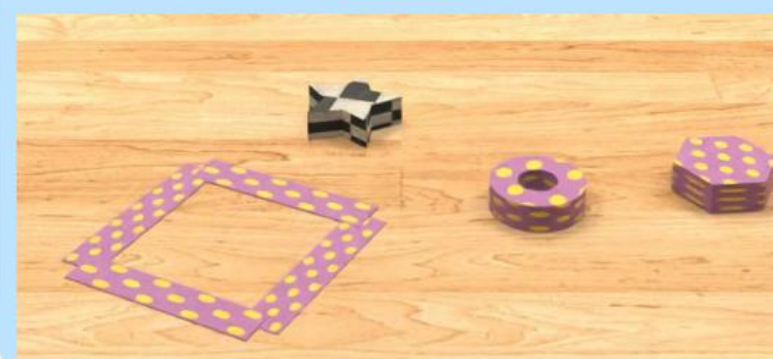
## Level 3 Novel Object

Put the  into the 

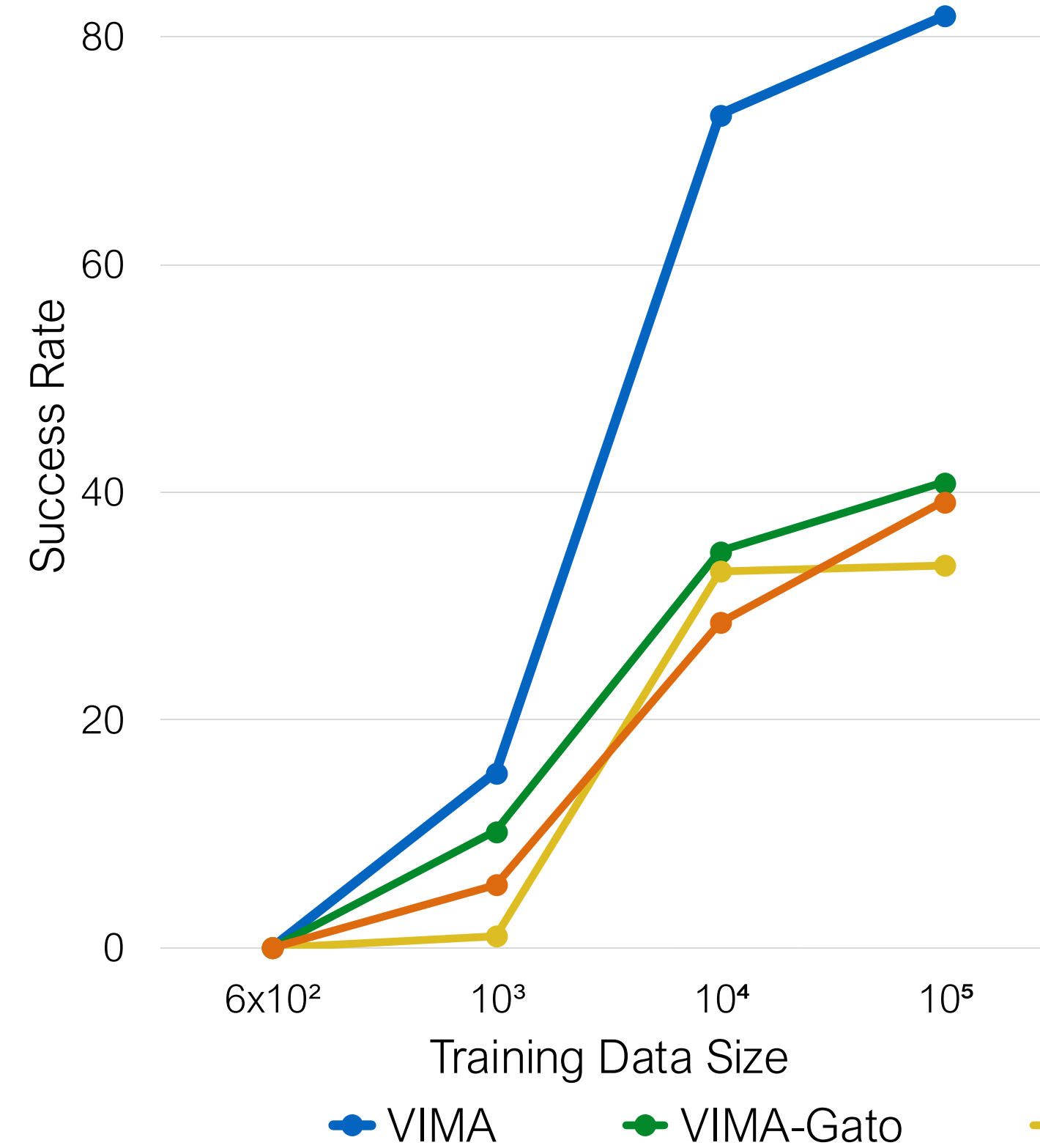


## Level 4 Novel Task

Put all objects with the same texture as  into it



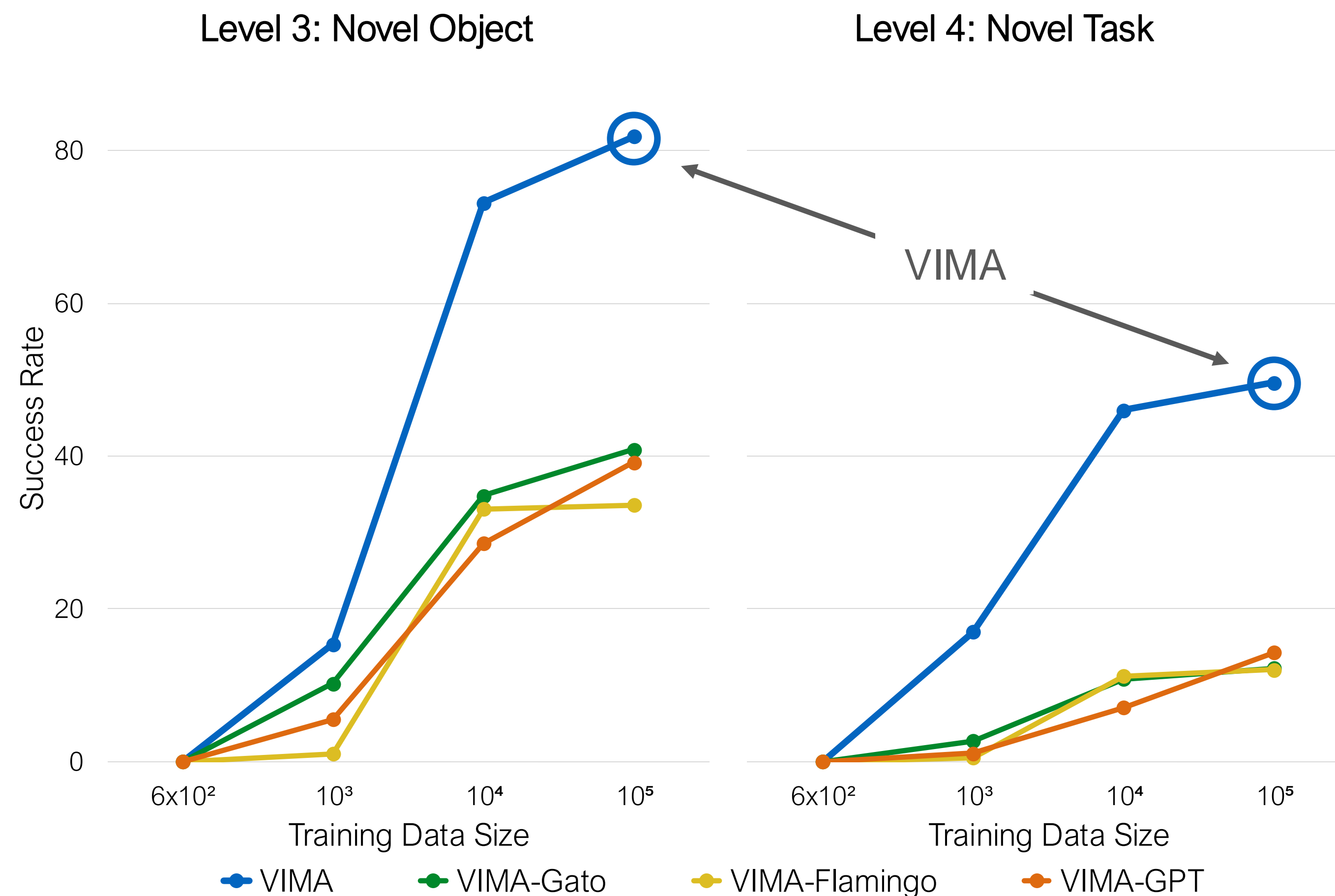
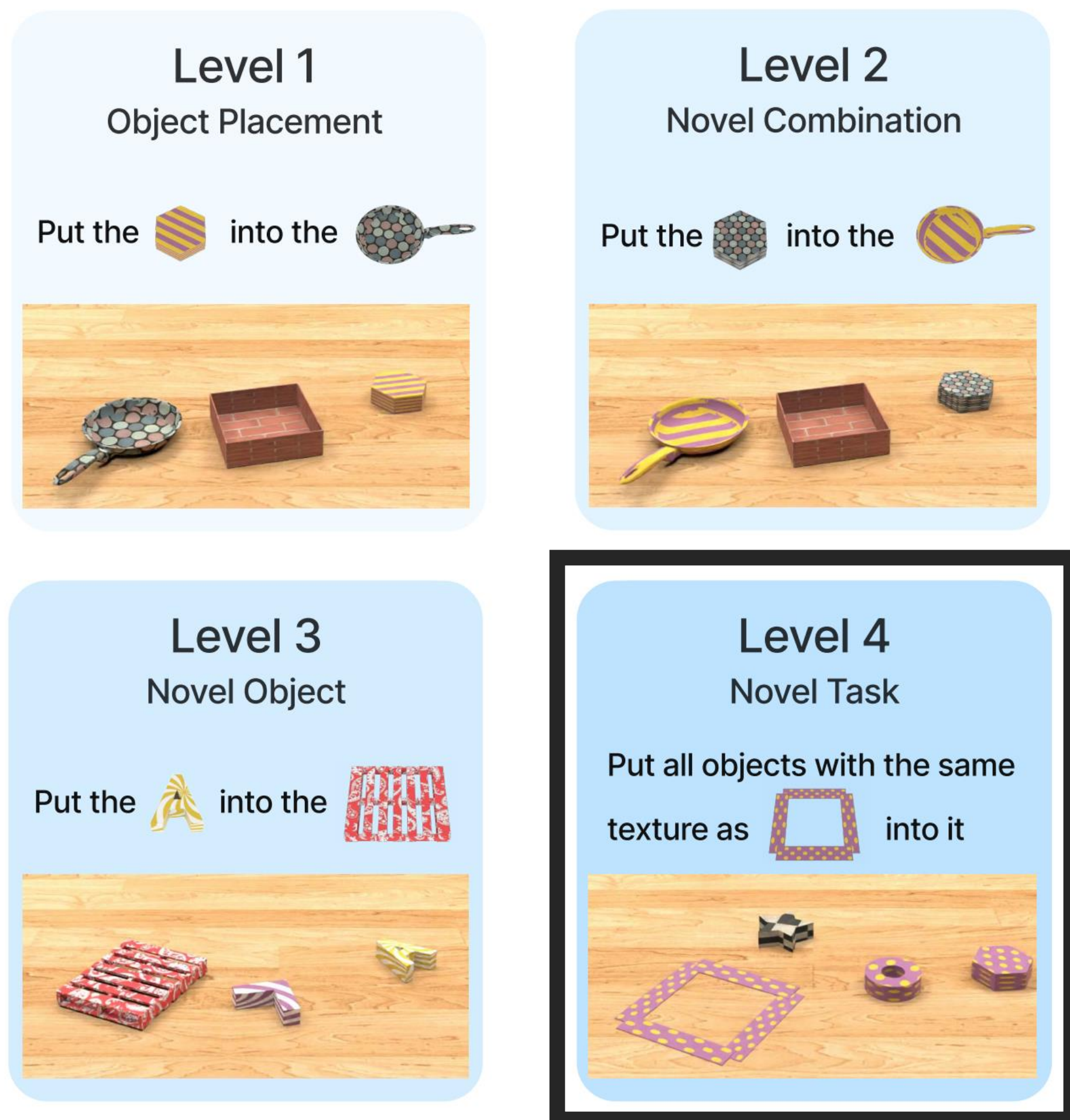
## Level 3: Novel Object





# Massively Multi-Task Robot Learning for Model Scaling

Data scalability from 0.1% to full dataset

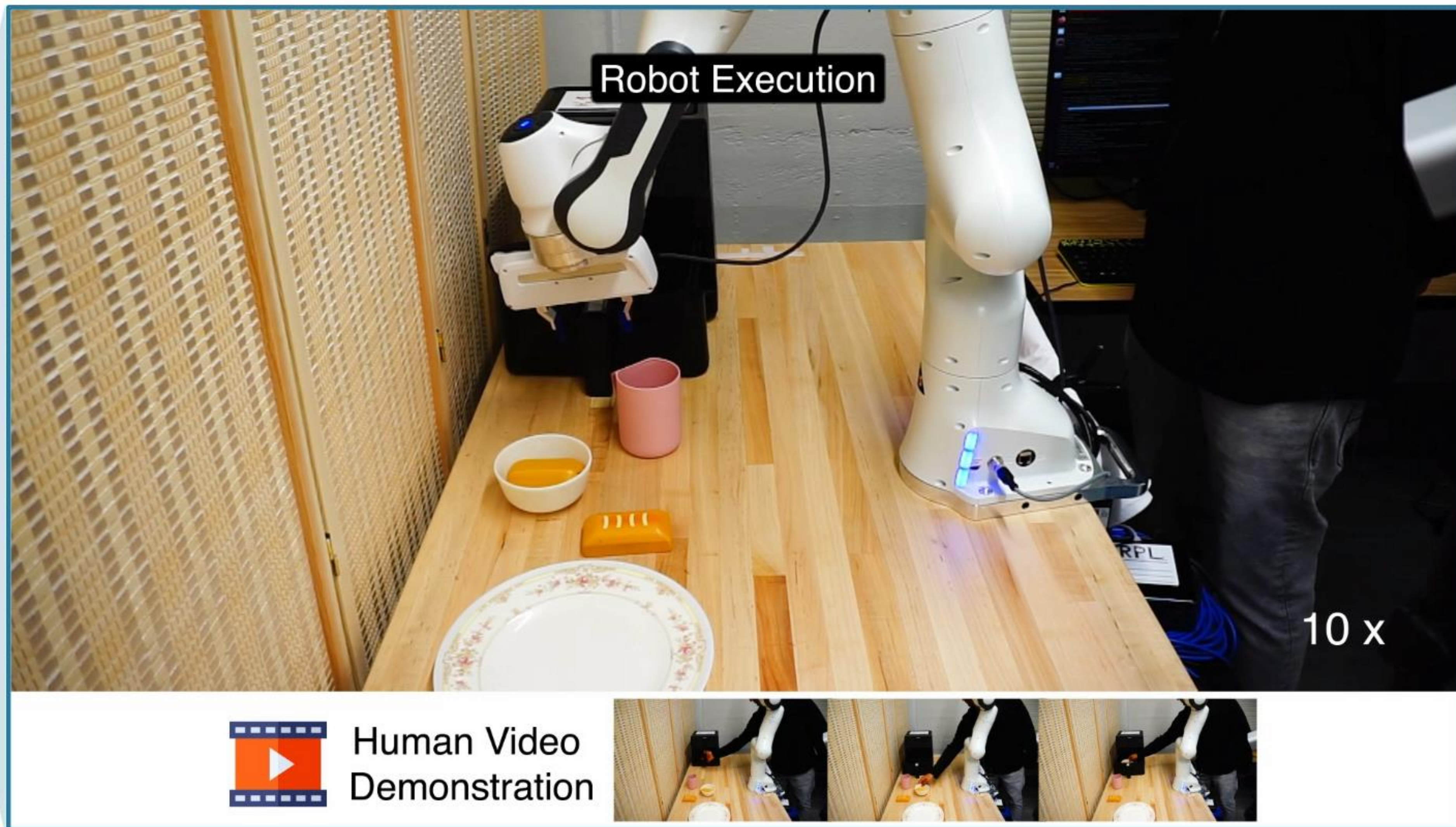




# Massively Multi-Task Robot Learning for Model Scaling



Rutav Shah





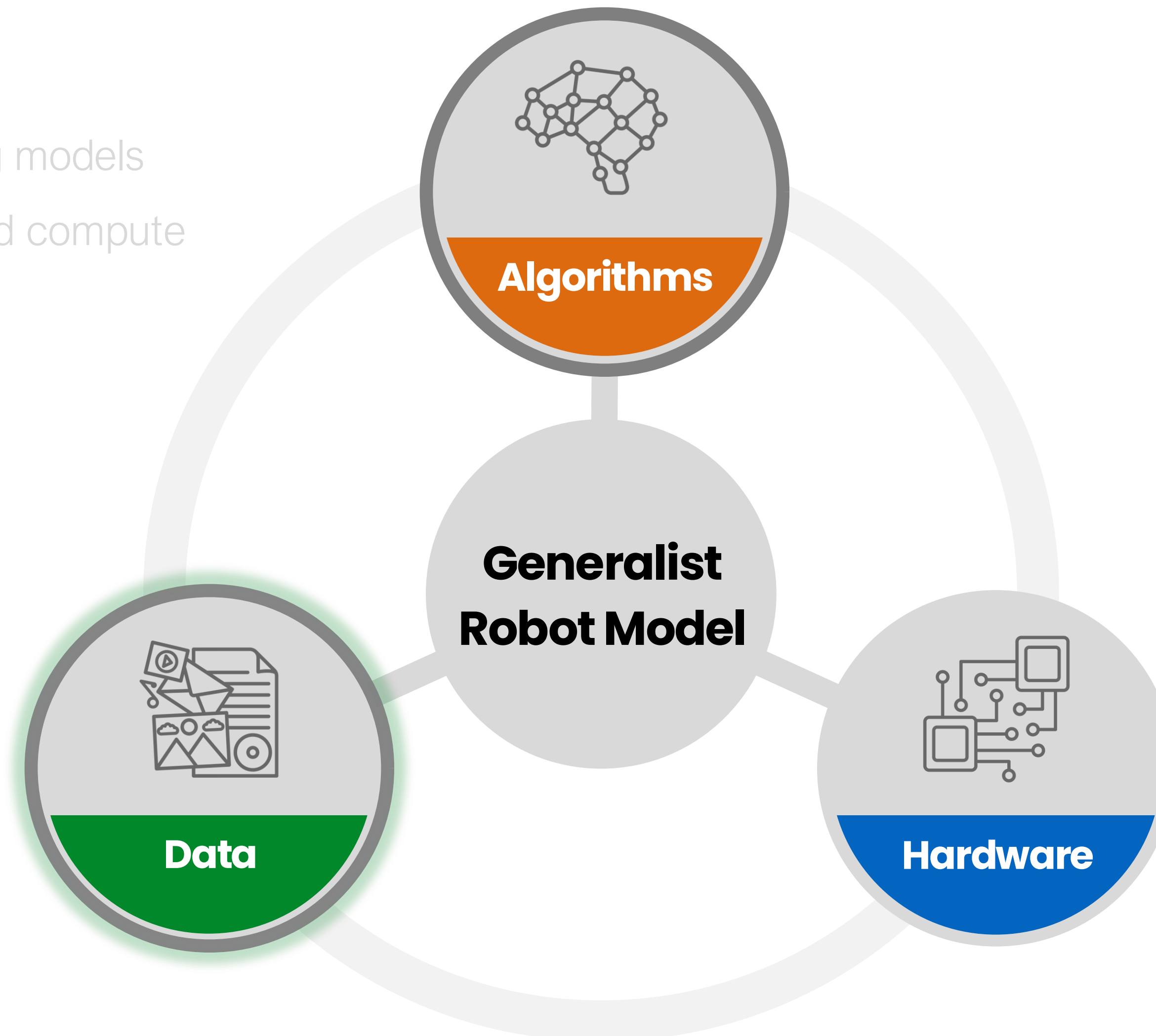
# Recipe for Building Generalist Robot Models

## Scaling Law

Powerful robot learning models  
that scale with data and compute

## Data Flywheel

New mechanism to collect  
massive training data

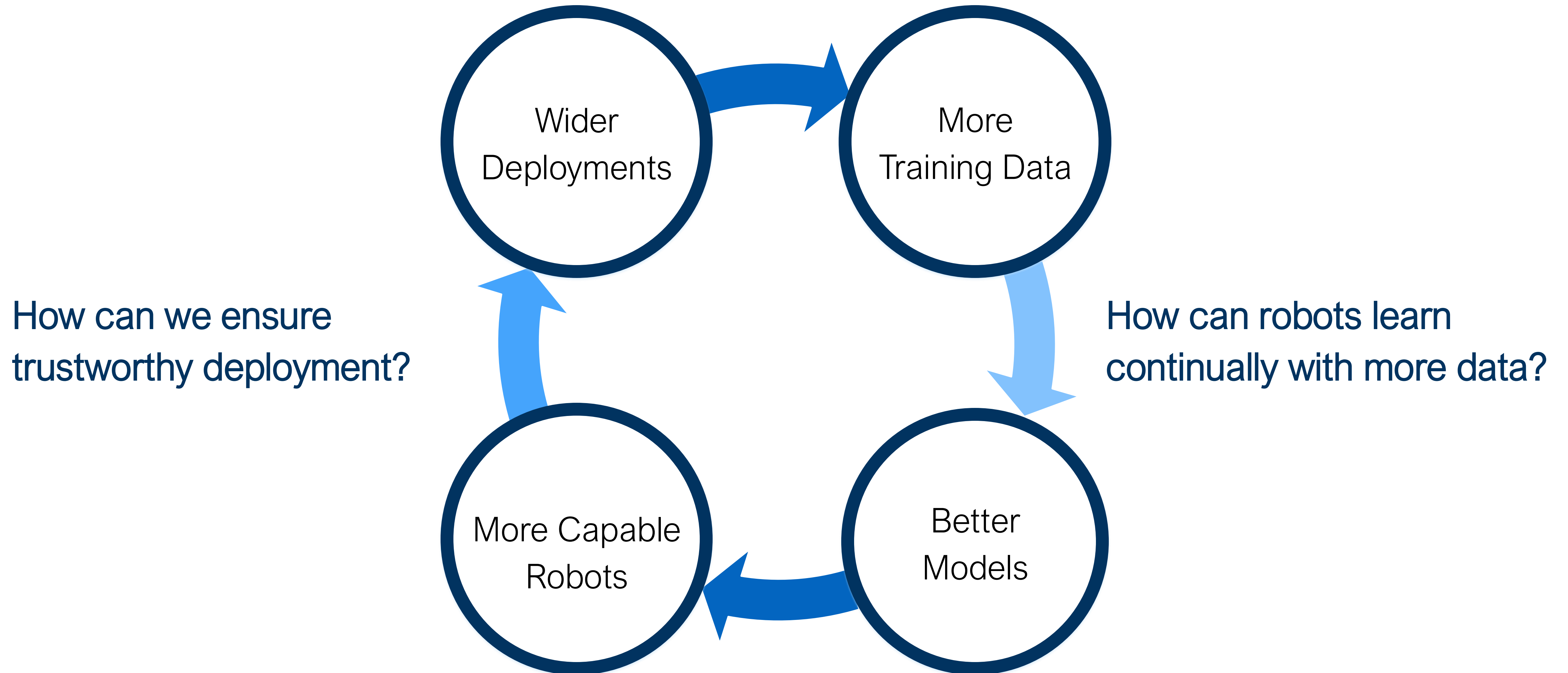


## Humanlike Embodiment

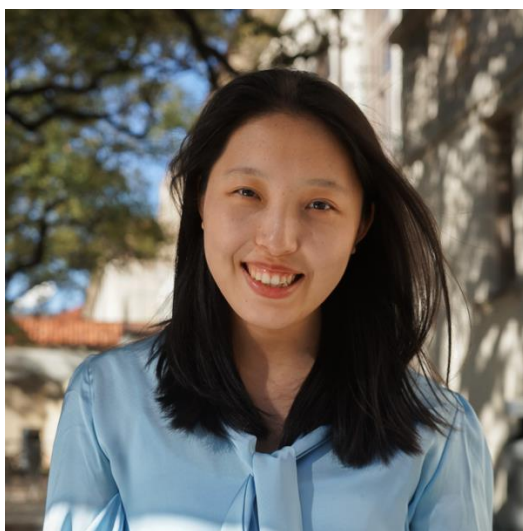
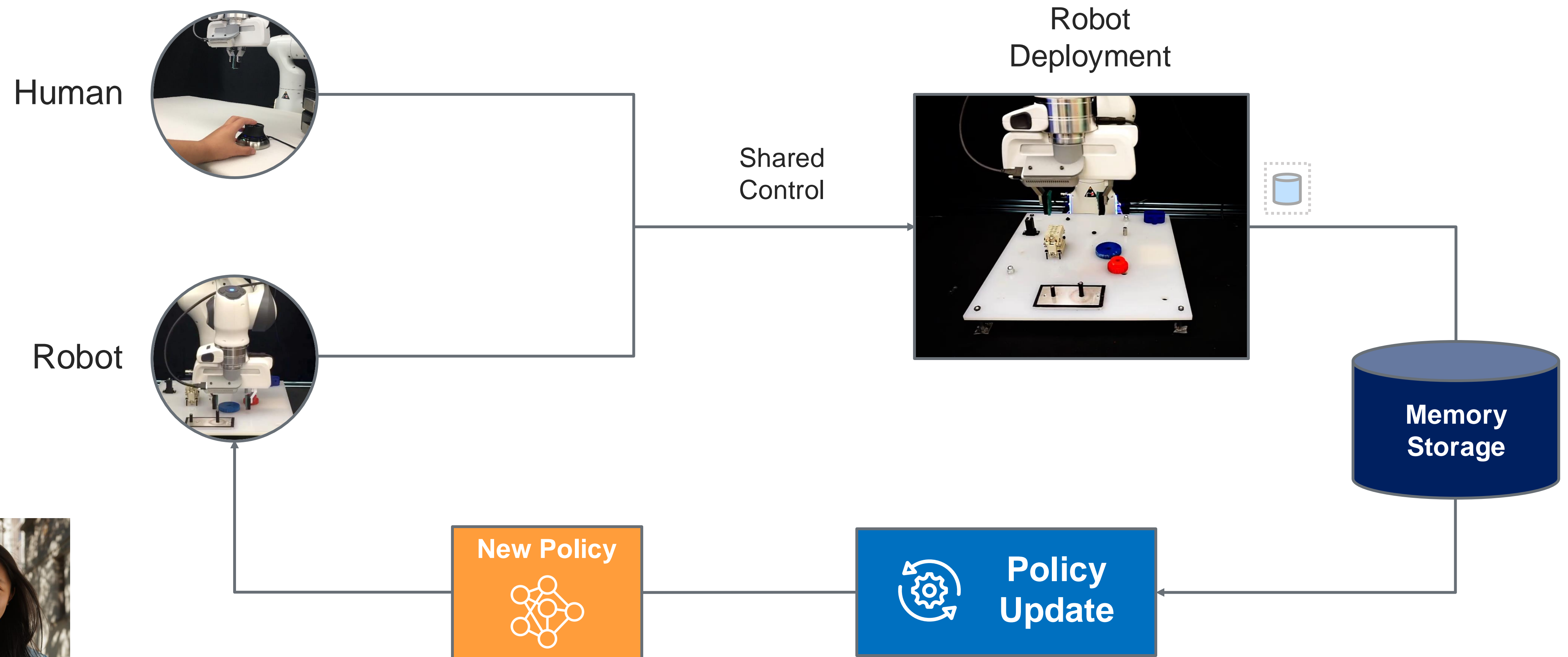
Humanoid robot platform for  
broad applications



# Key Idea: Robot Learning Data Flywheel



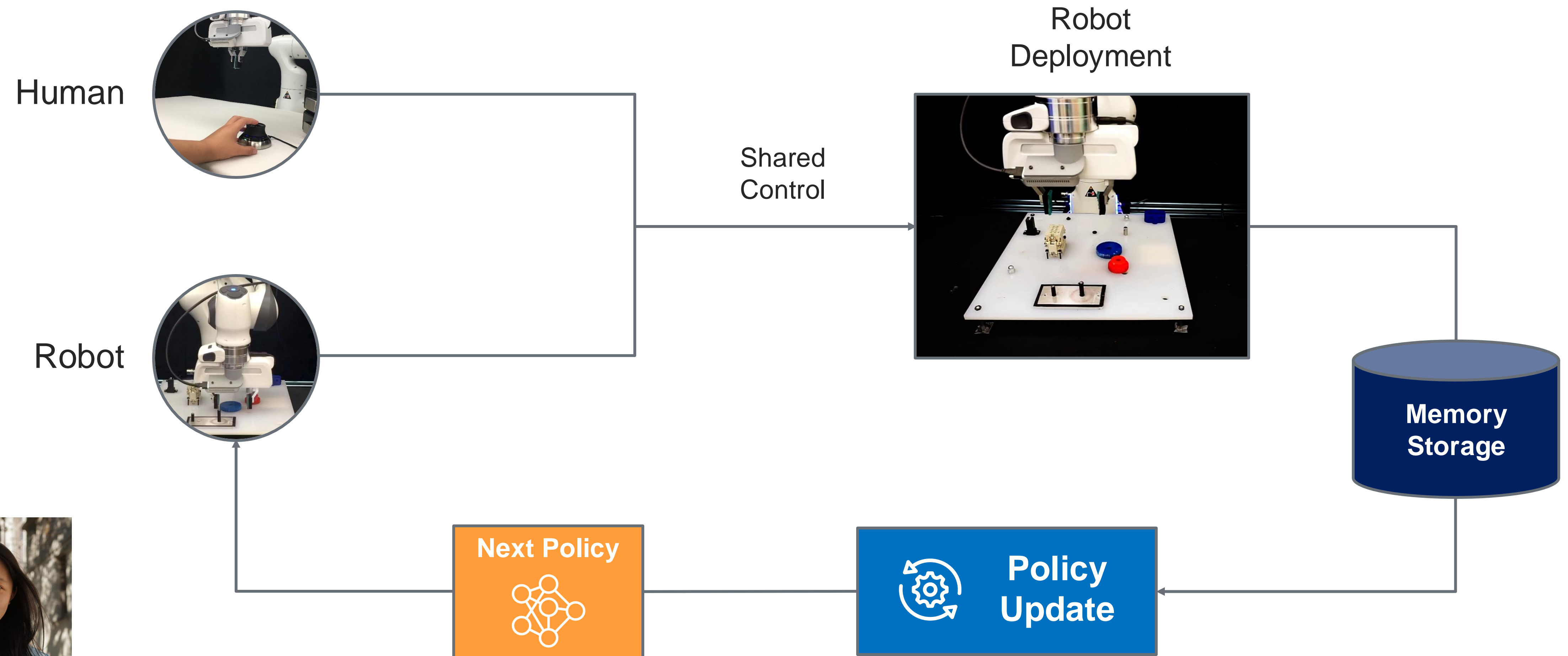
# Robot Learning on the Job: Building the Data Flywheel



Huihan Liu



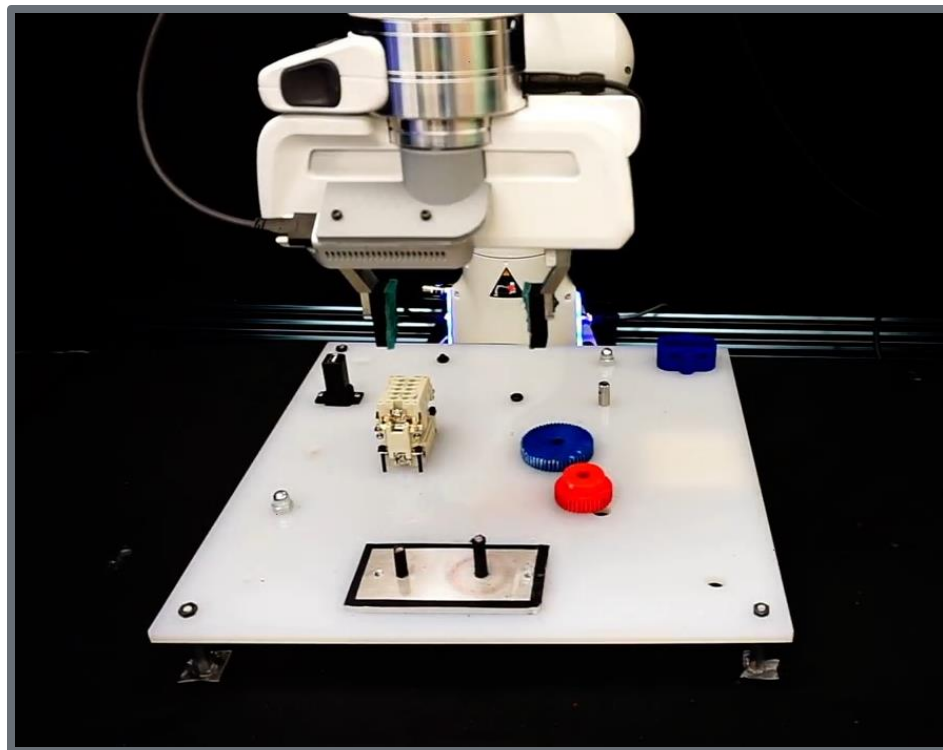
# Robot Learning on the Job: Building the Data Flywheel



Huihan Liu

# Robot Learning on the Job: Building the Data Flywheel

Robot  
Deployment

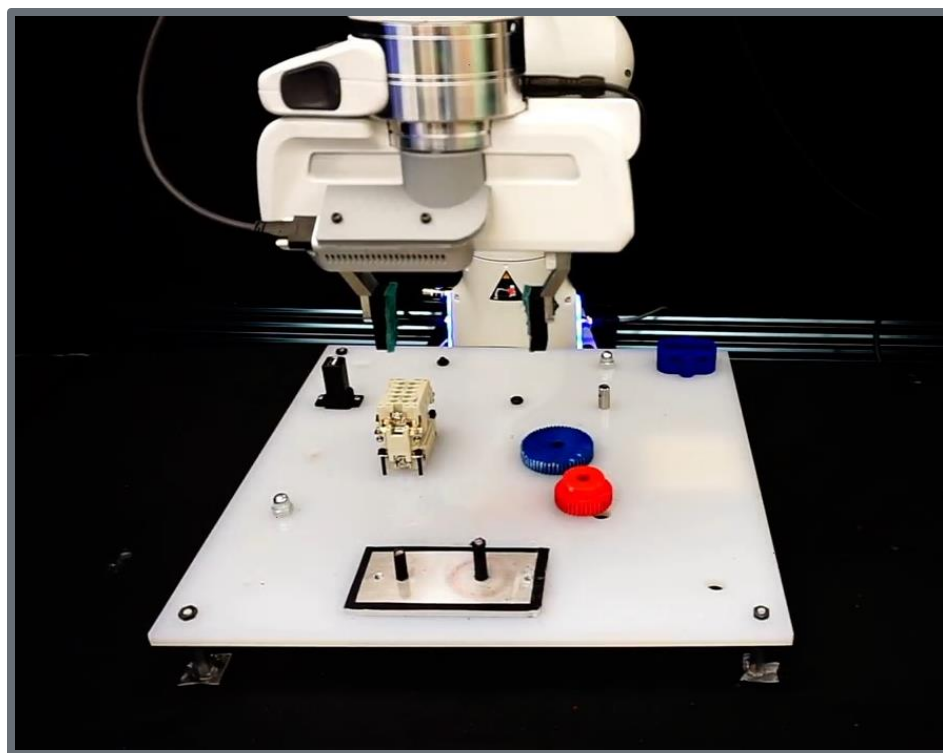


Policy  
Update

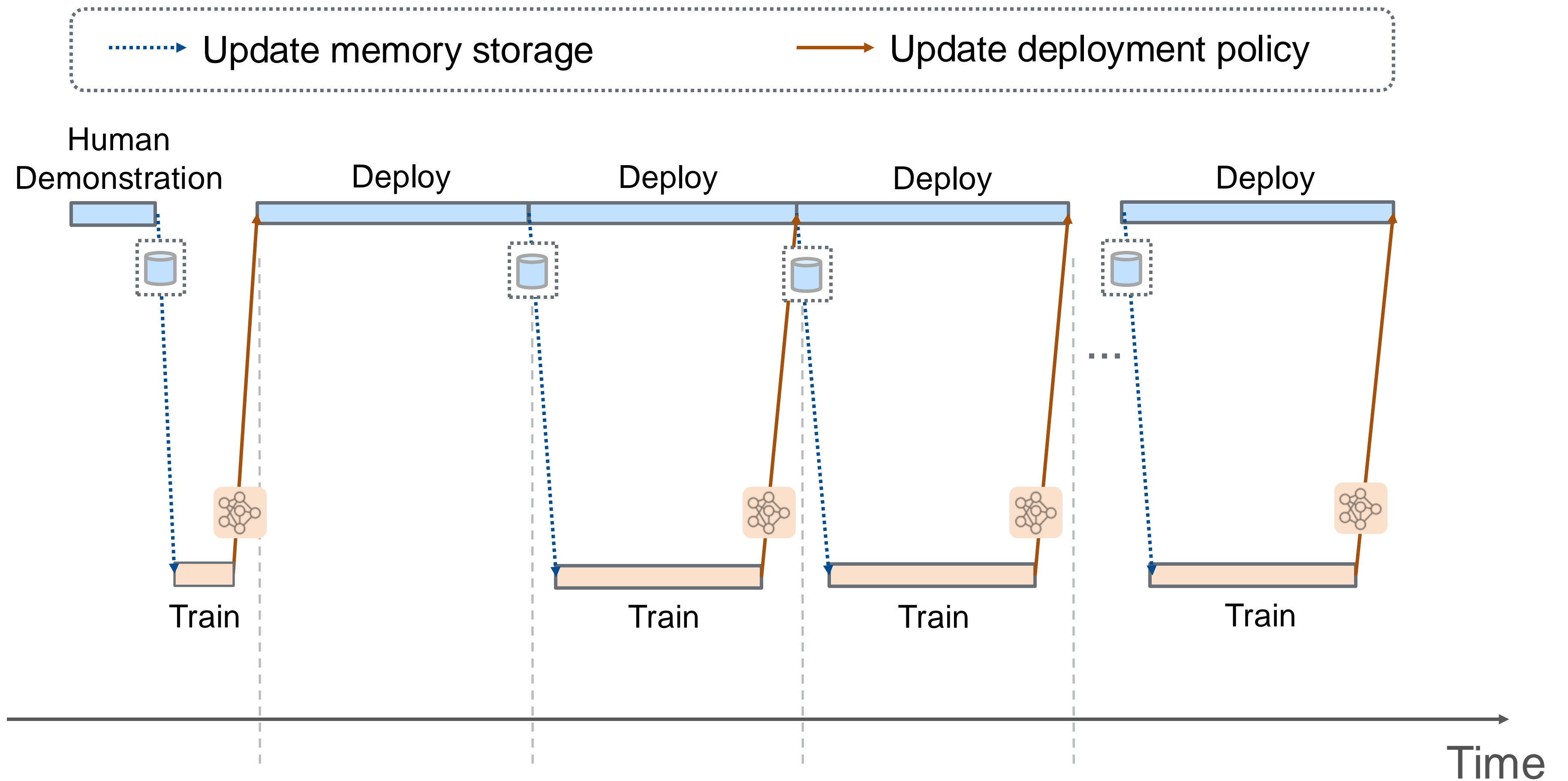


# Robot Learning on the Job: Building the Data Flywheel

Robot  
Deployment



Policy  
Update





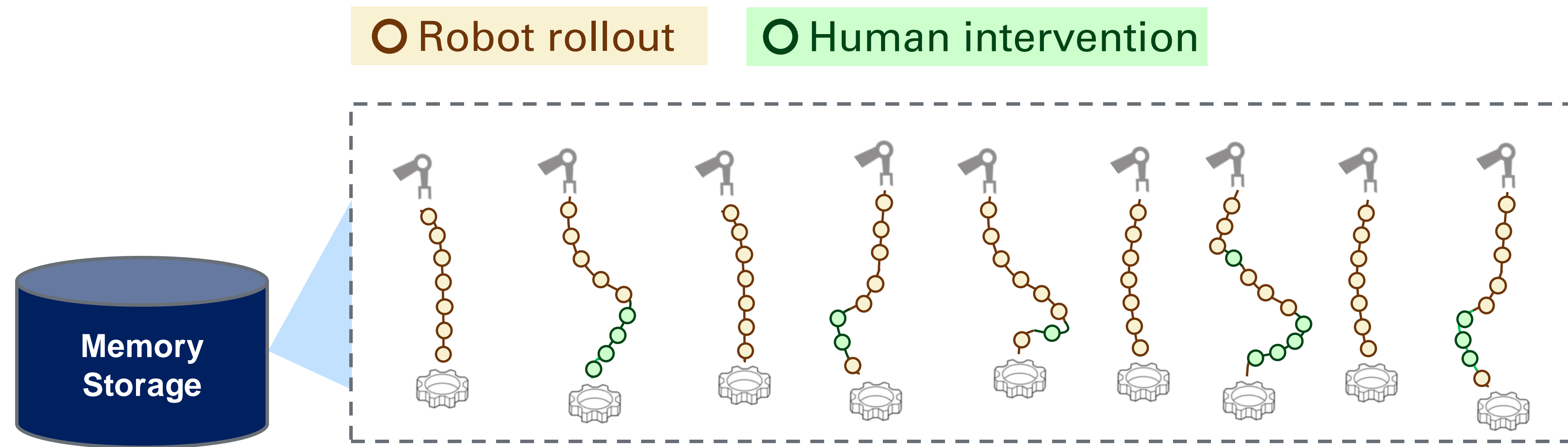
2x

Robot takes  
control again

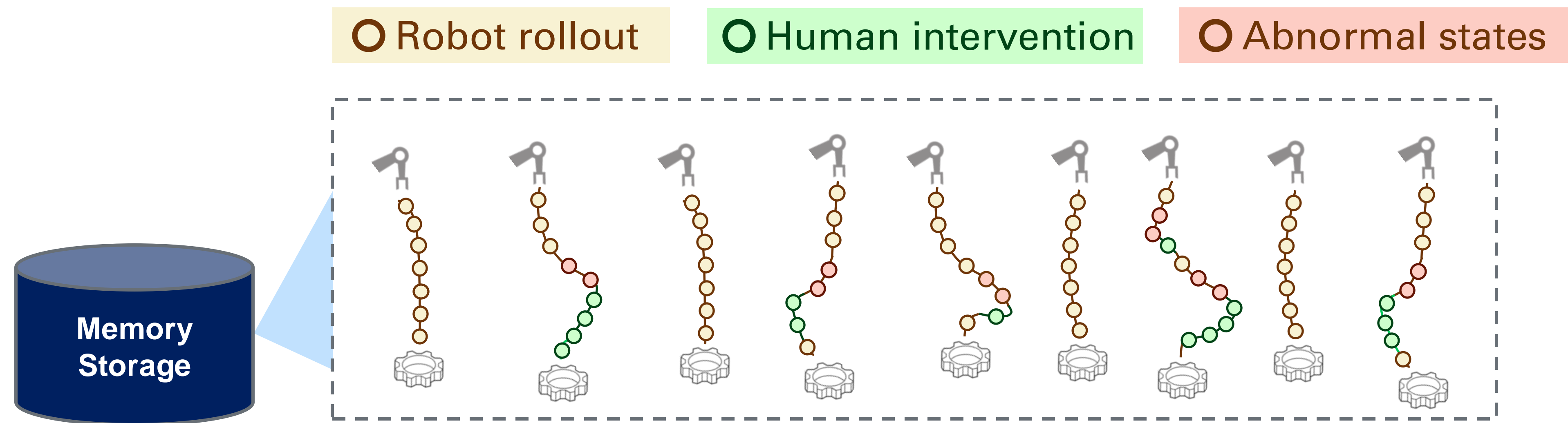
Human provides  
an intervention



# Robot Learning on the Job: Building the Data Flywheel



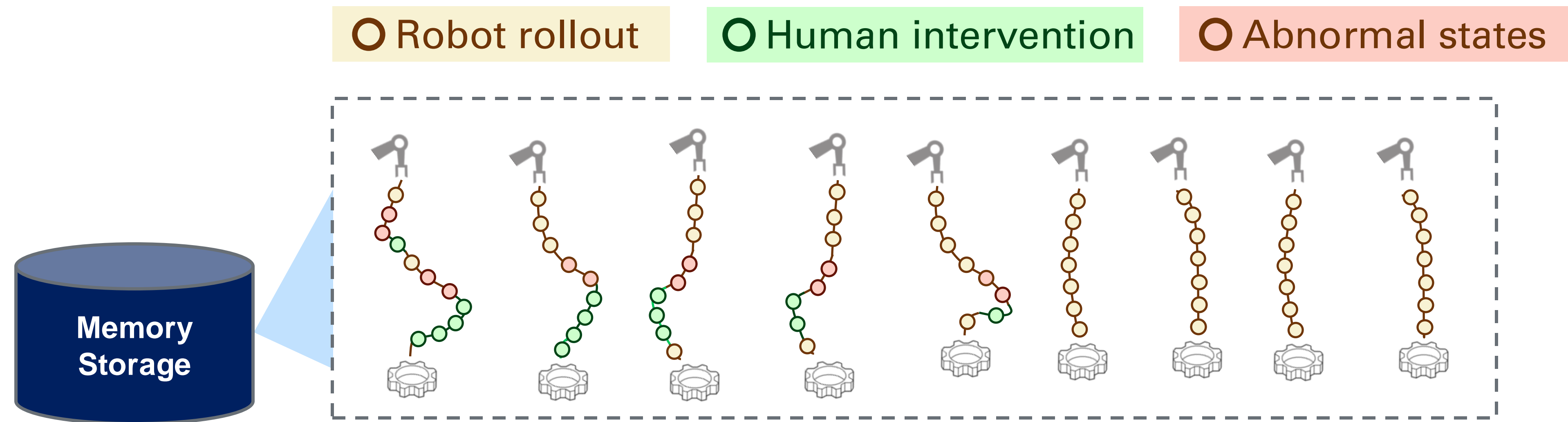
# Robot Learning on the Job: Building the Data Flywheel



Relabel **abnormal states** before interventions

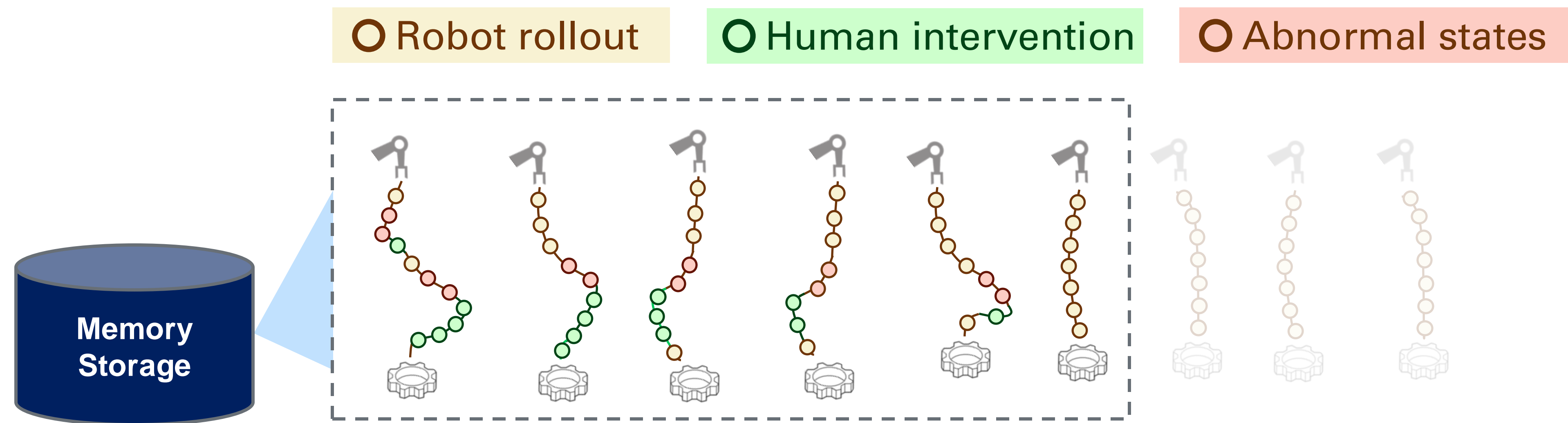


# Robot Learning on the Job: Building the Data Flywheel



Sort by a **dataset curation** strategy

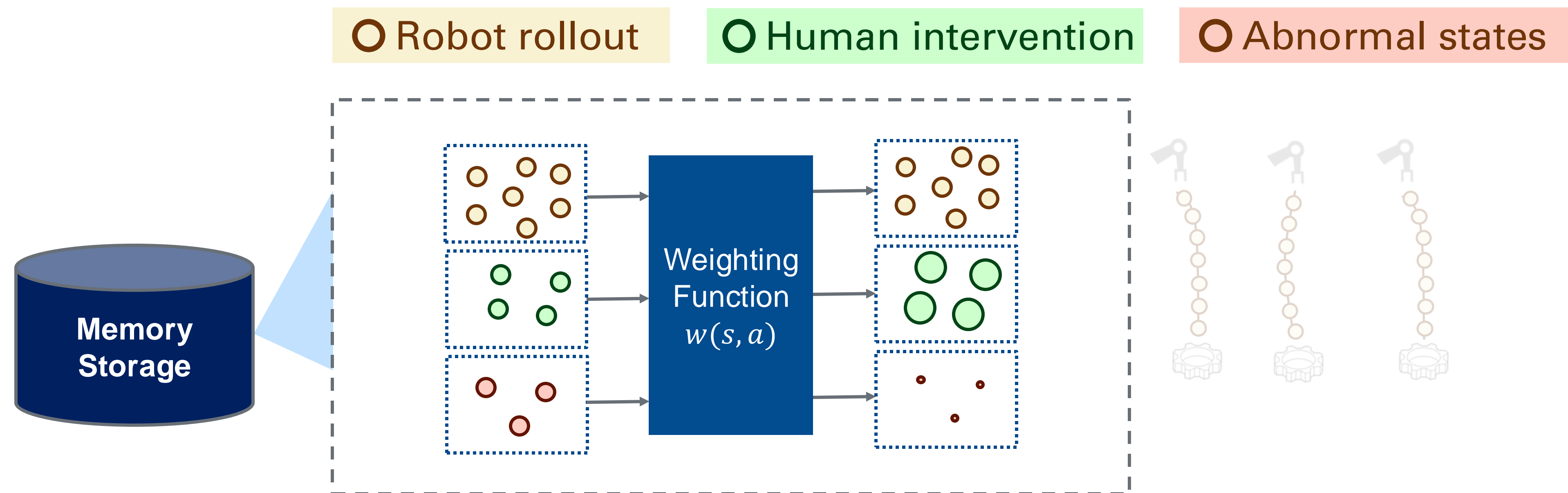
# Robot Learning on the Job: Building the Data Flywheel



Reduce memory to a **fixed size**



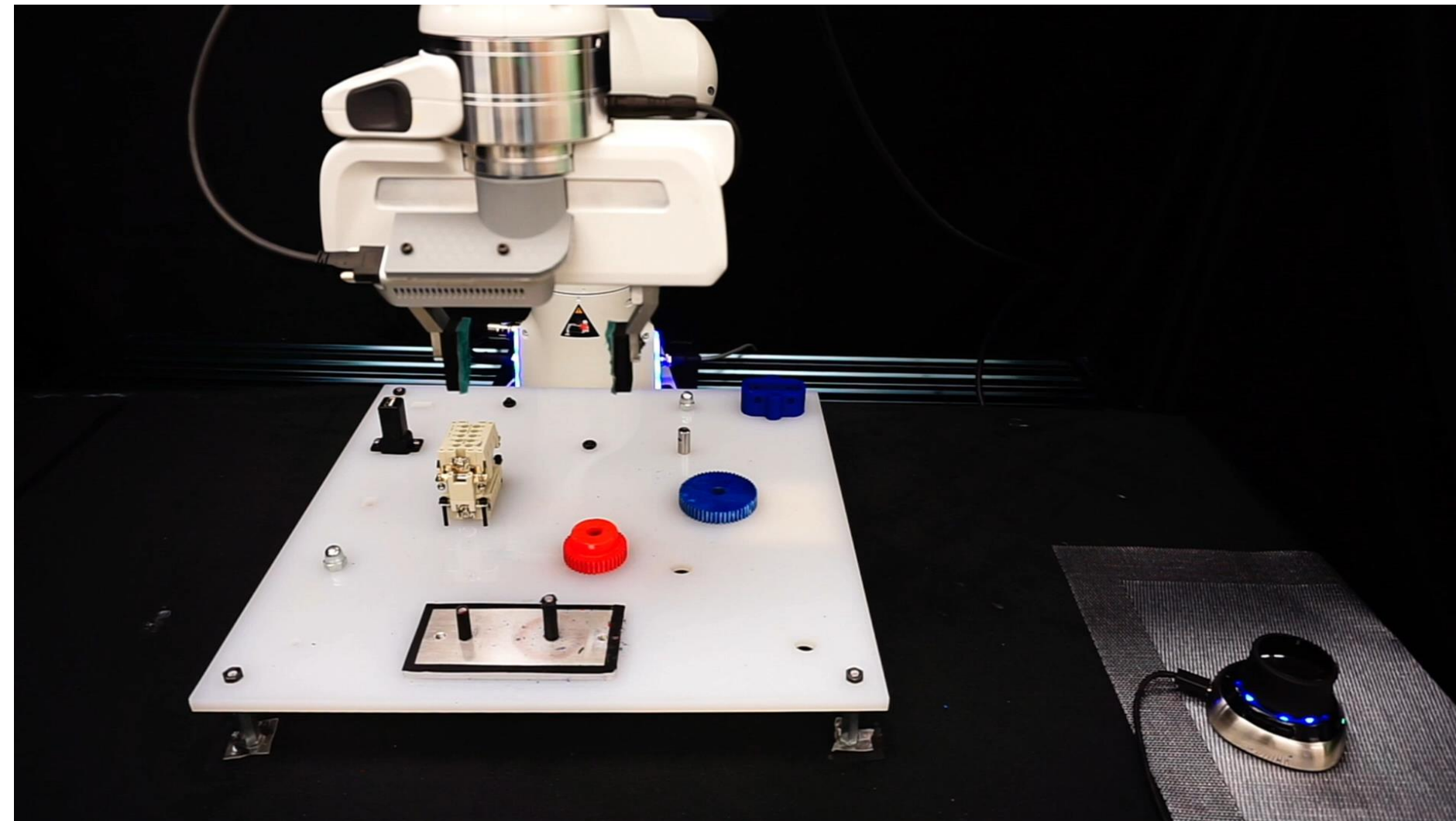
# Robot Learning on the Job: Building the Data Flywheel



Reweight samples for imitation learning

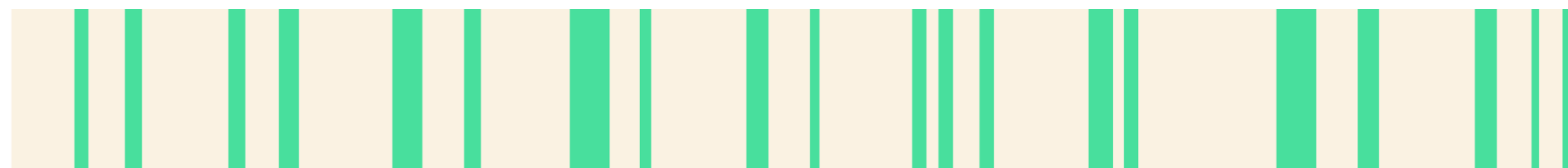
# Robot Learning on the Job: Building the Data Flywheel

Round 1 Deployment

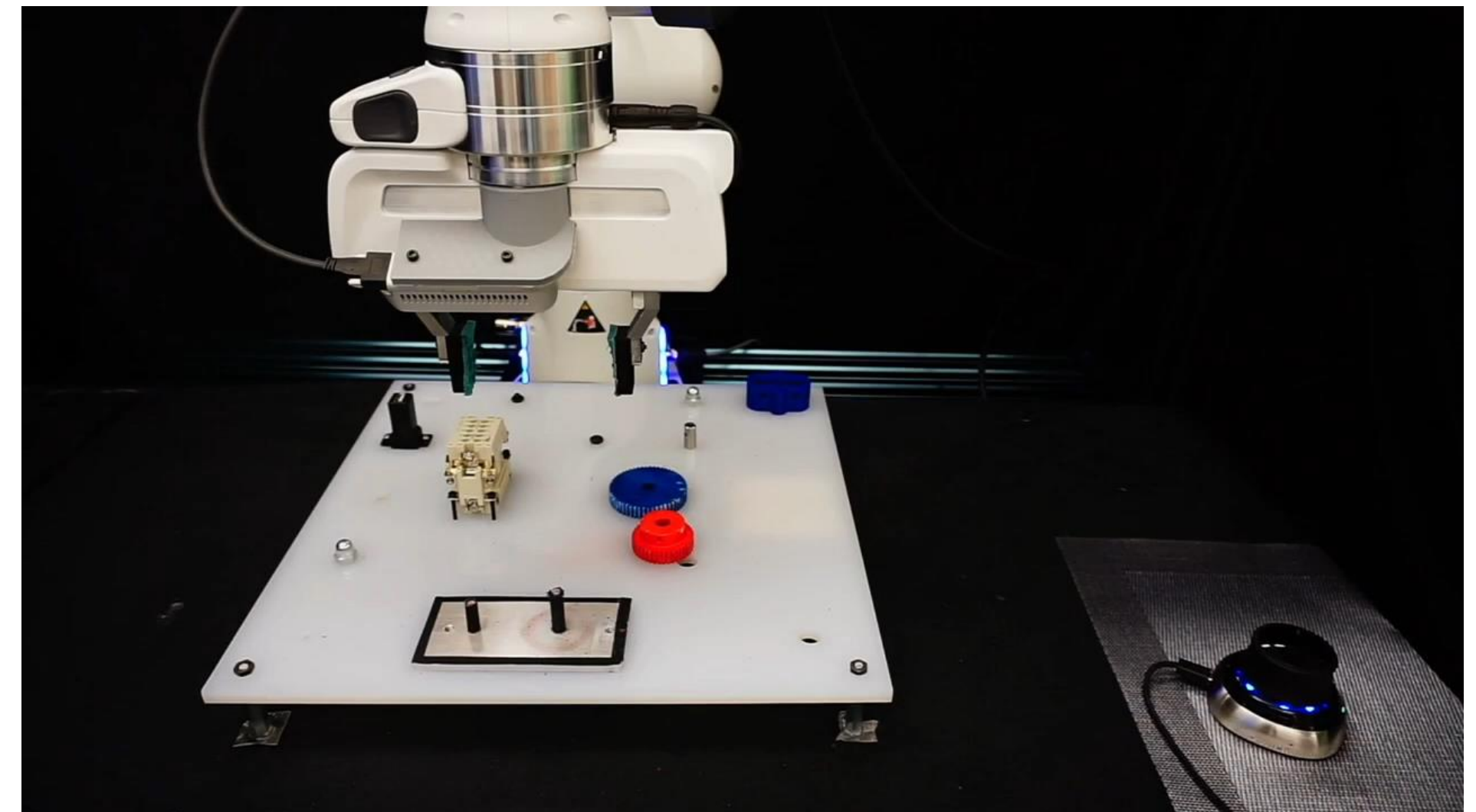


† Green masks indicate human intervention.

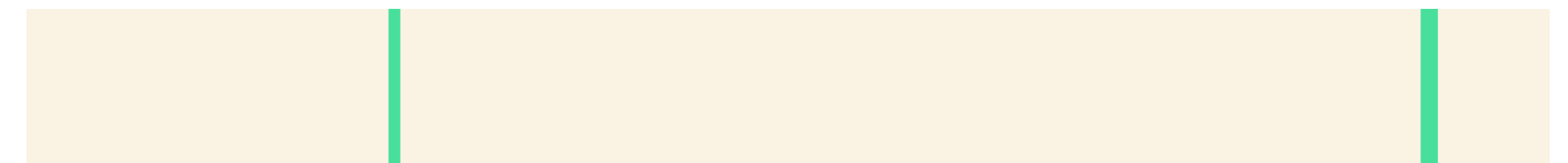
Intervention Distribution



Round 3 Deployment



Intervention Distribution

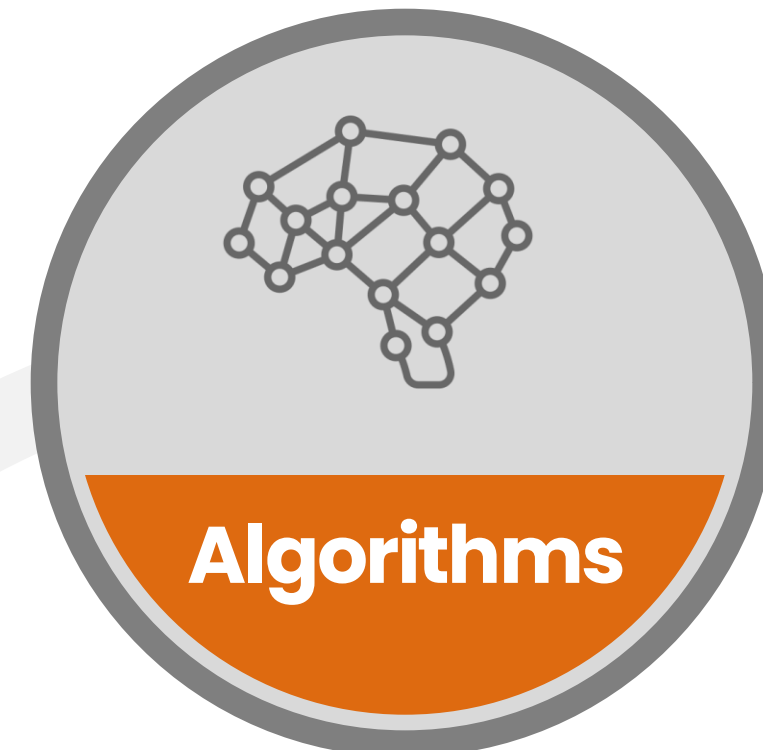




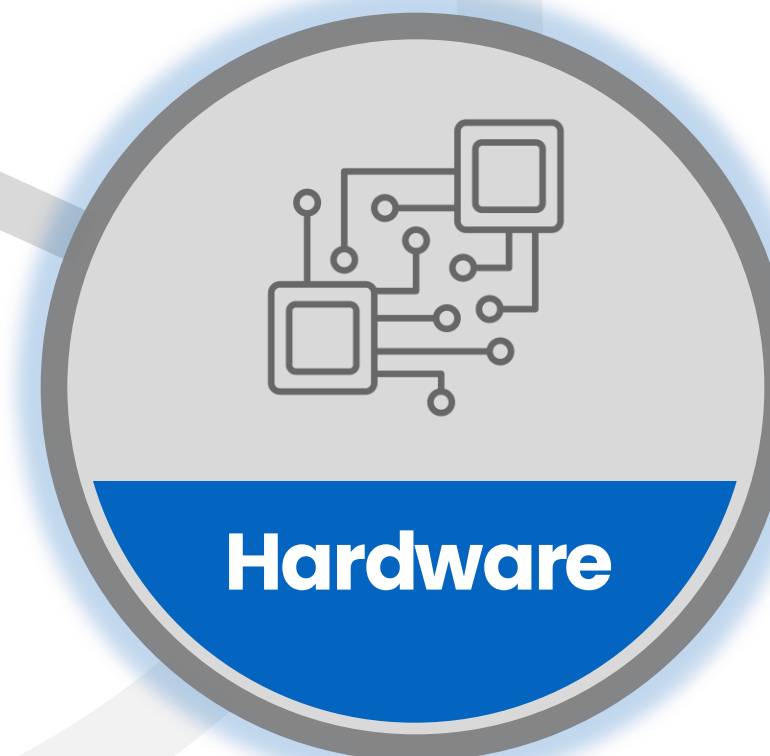
# Recipe for Building Generalist Robot Models

## Scaling Law

Powerful robot learning models  
that scale with data and compute



**Generalist  
Robot Model**



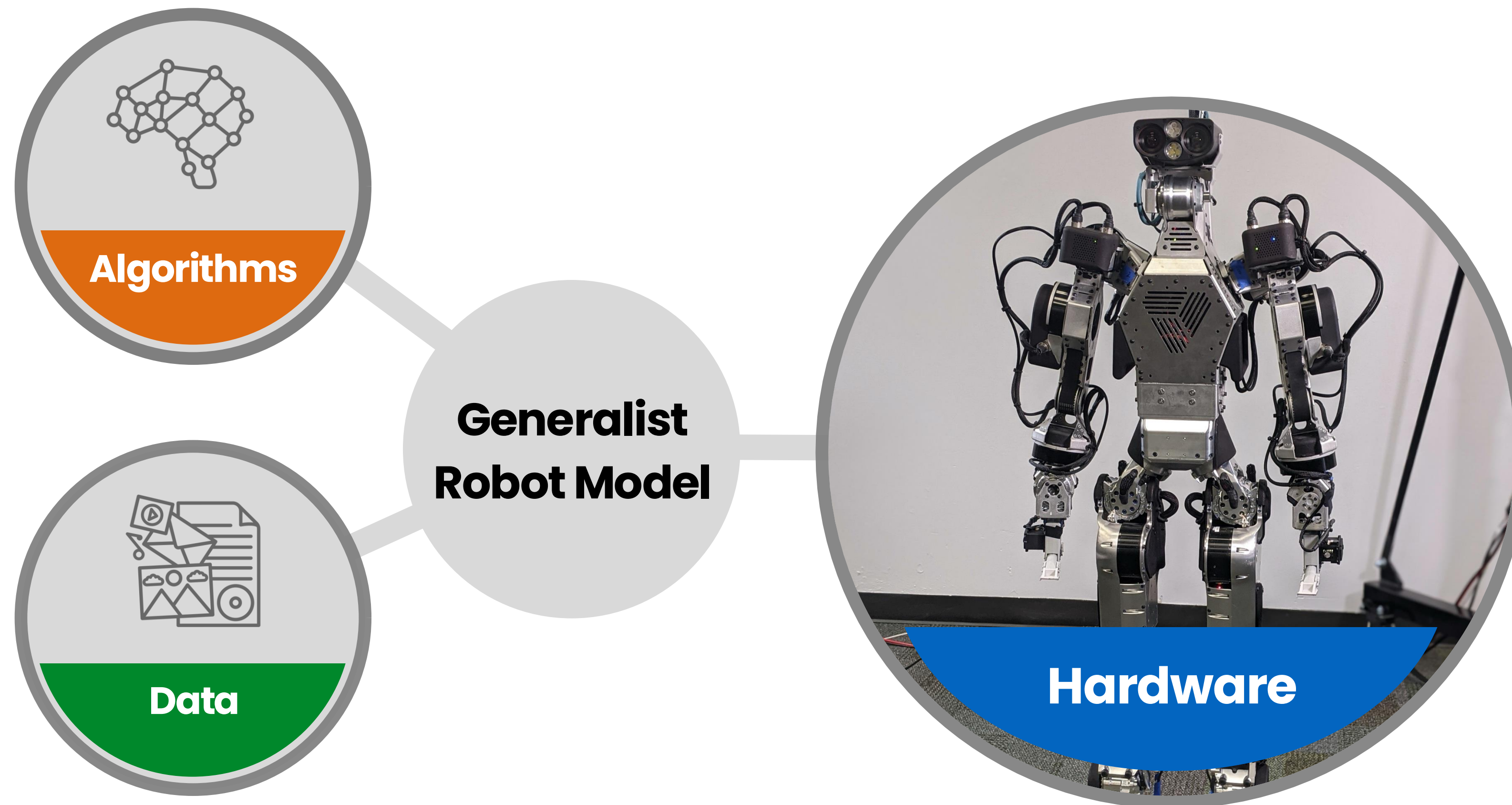
## Humanlike Embodiment

Humanoid robot platform for  
broad applications

## Data Flywheel

New mechanism to collect  
massive training data

# Humanlike Embodiment: Generalist Humanoid Robot



Draco3 Humanoid

## Team

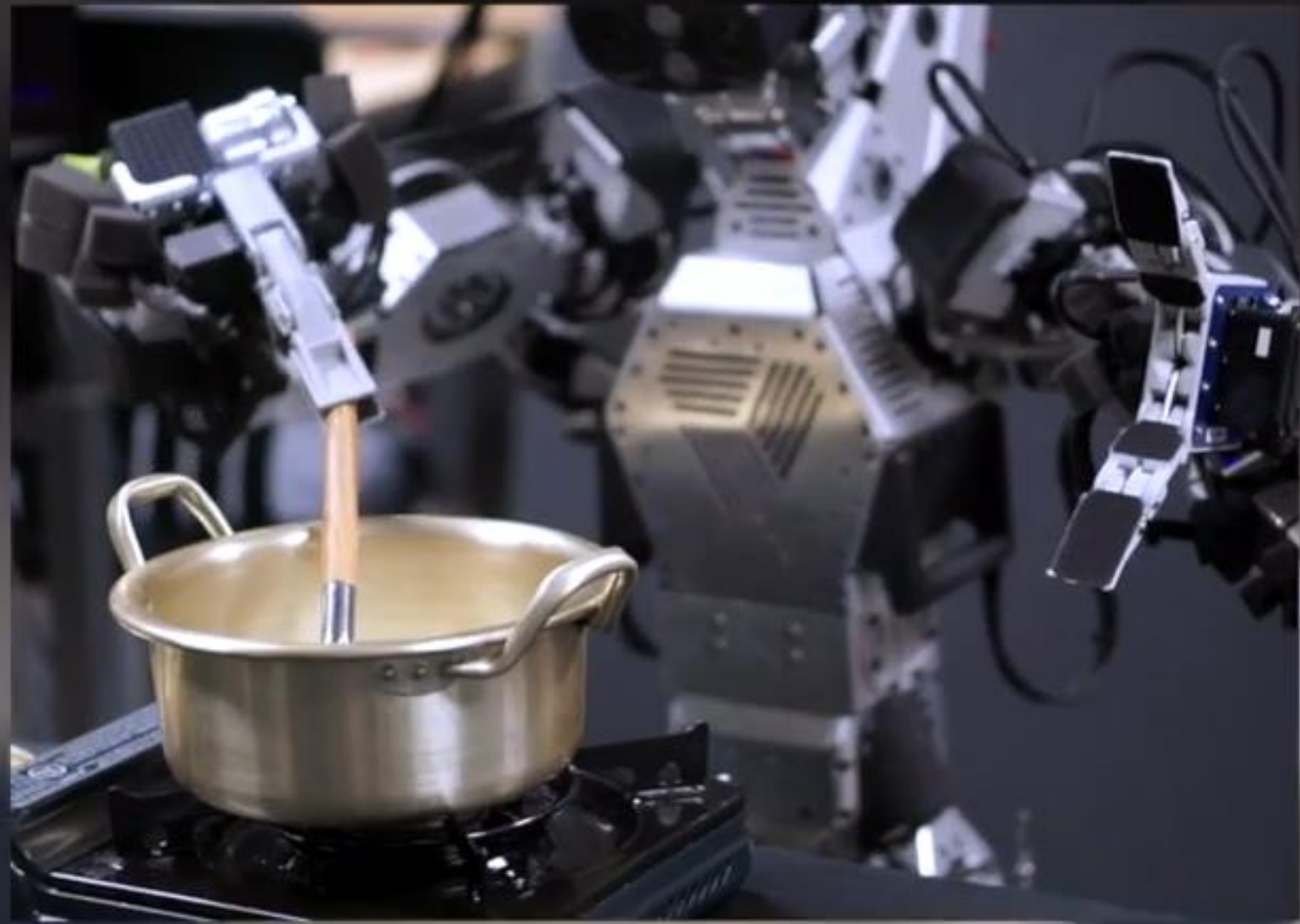


Mingyo Seo



Luis Sentis

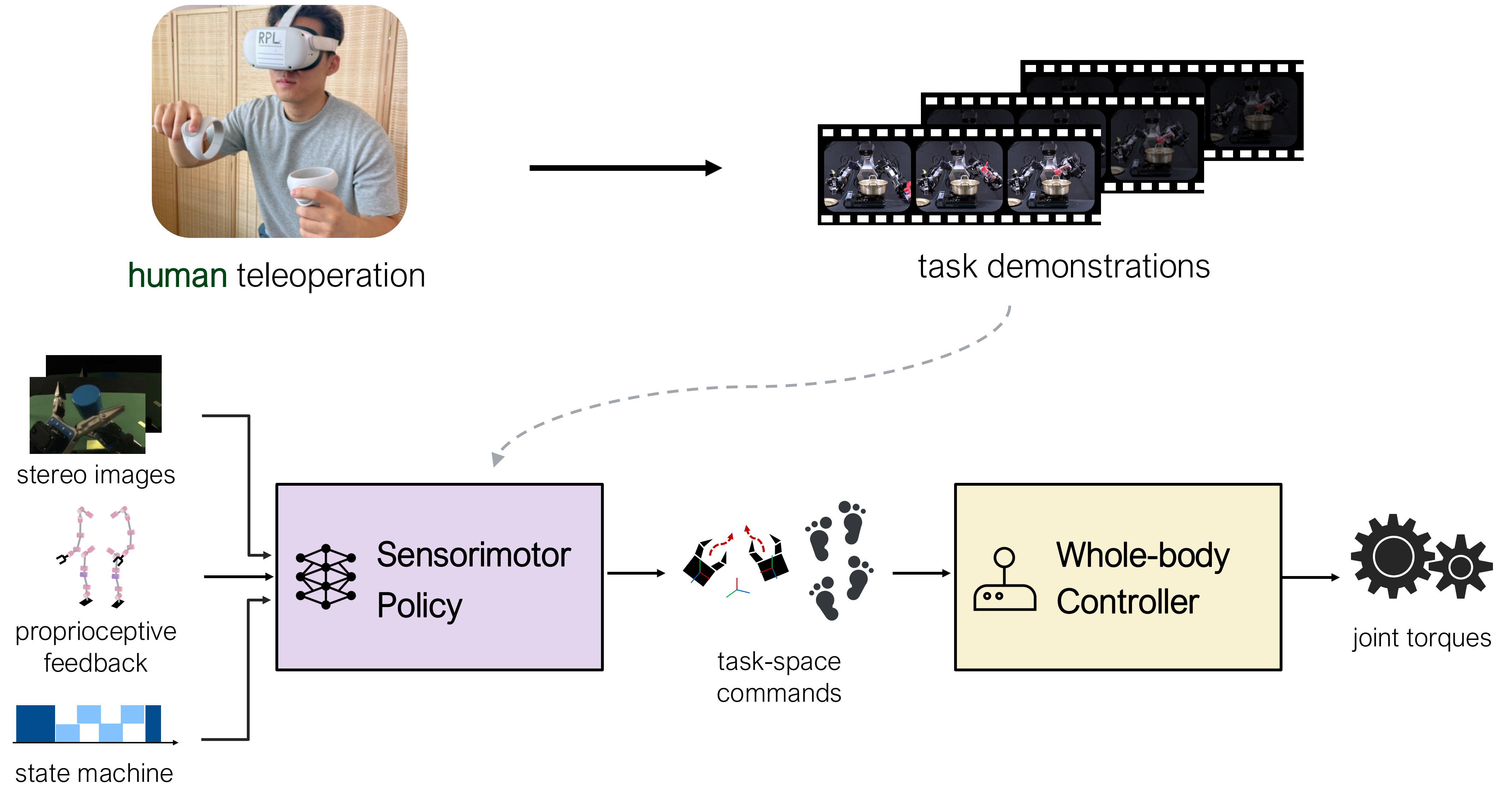




2x



# Humanlike Embodiment: Generalist Humanoid Robot



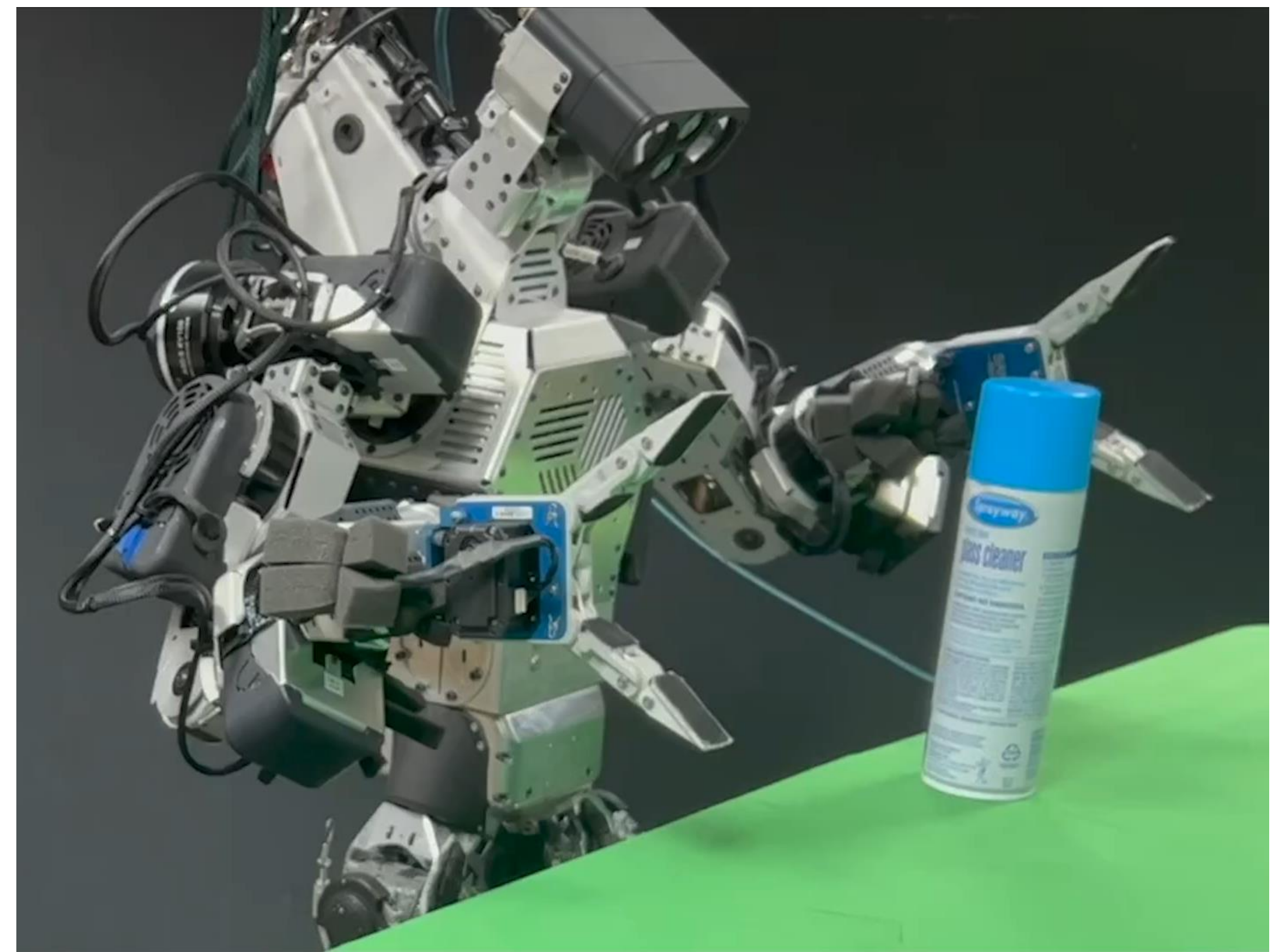


# Humanlike Embodiment: Generalist Humanoid Robot

Policy deployments in simulation and real world



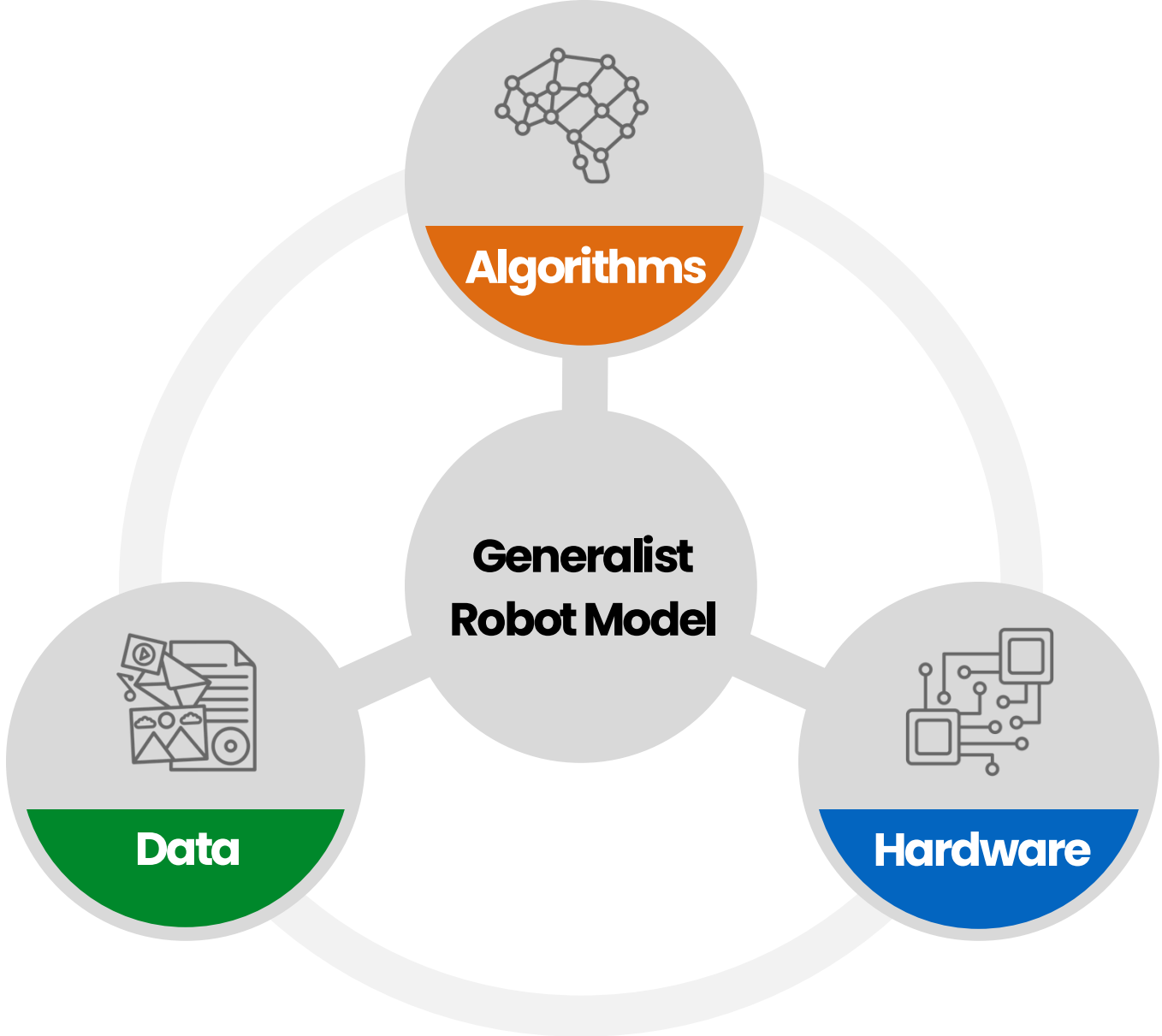
**Loco-manipulation (sim):** 92% success rate



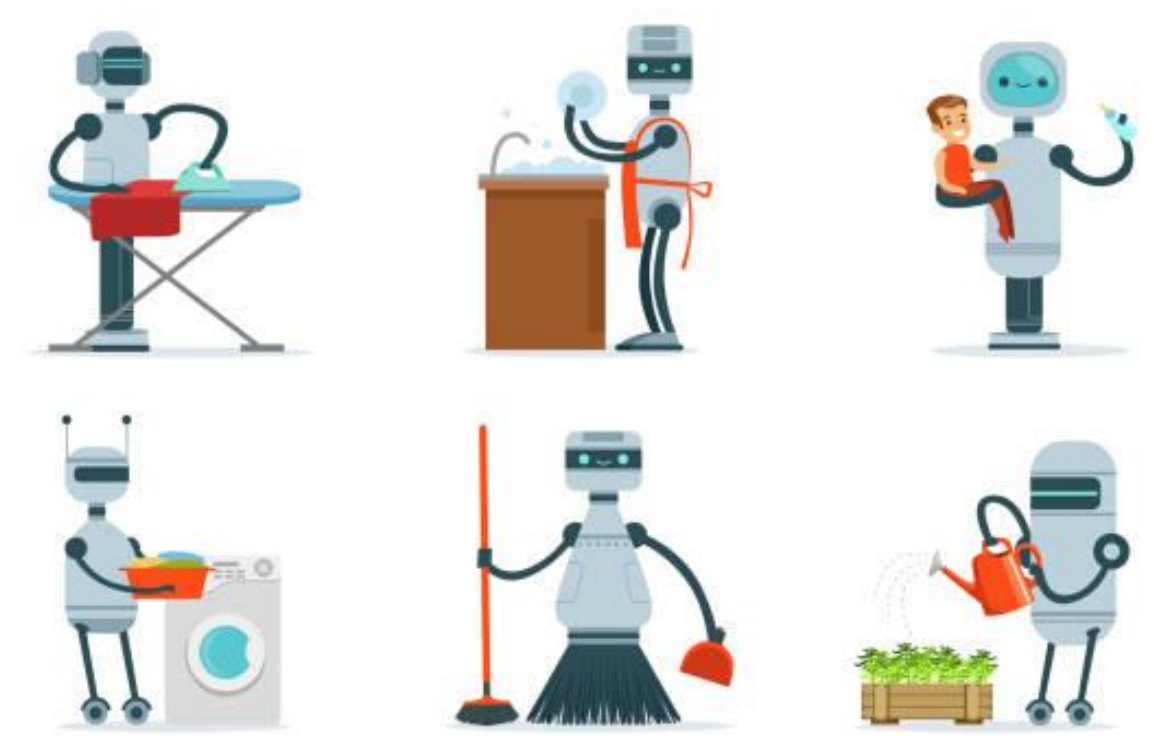
**Bimanual manipulation (real):** 90% success rate



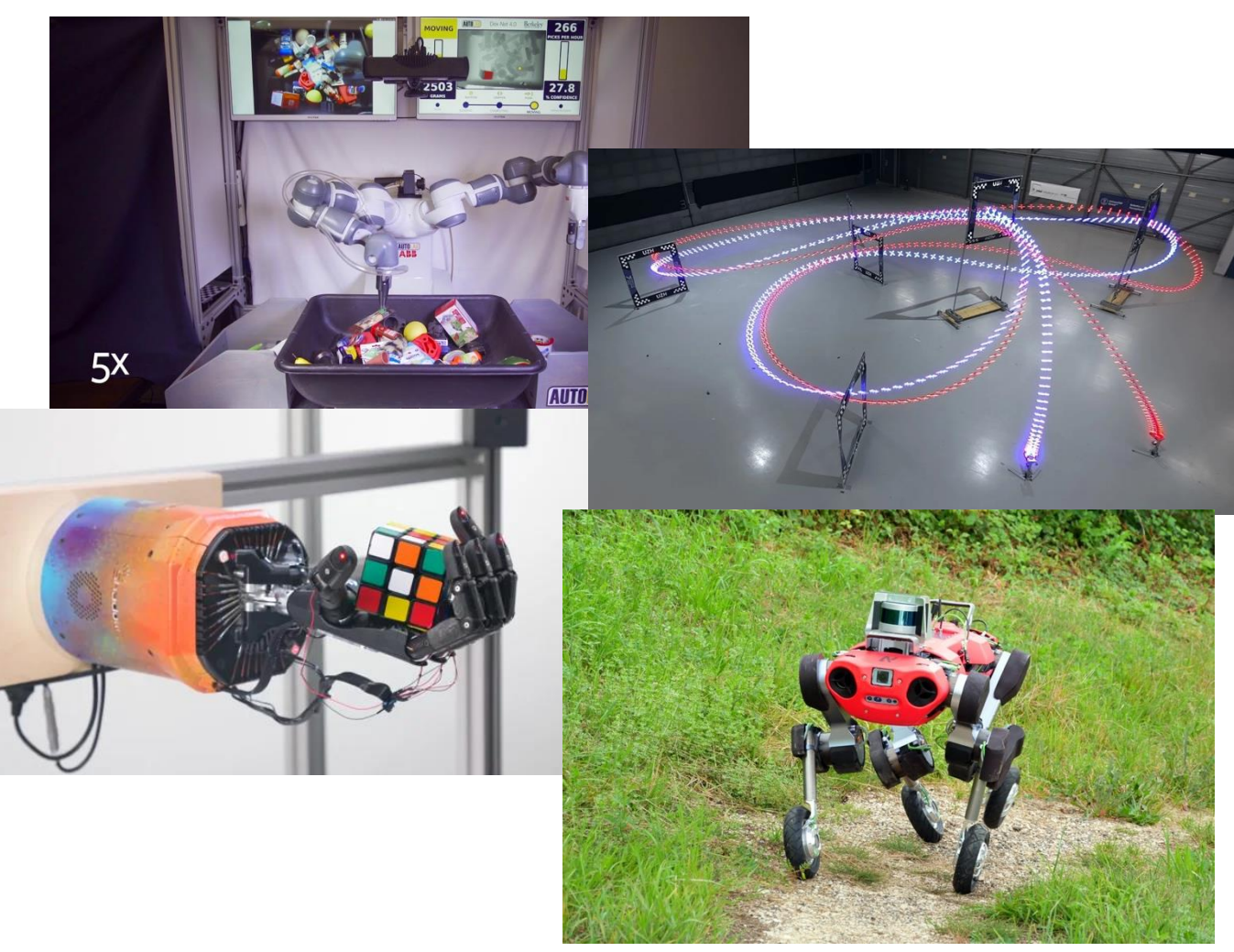
# Pathway to Generalist Robots



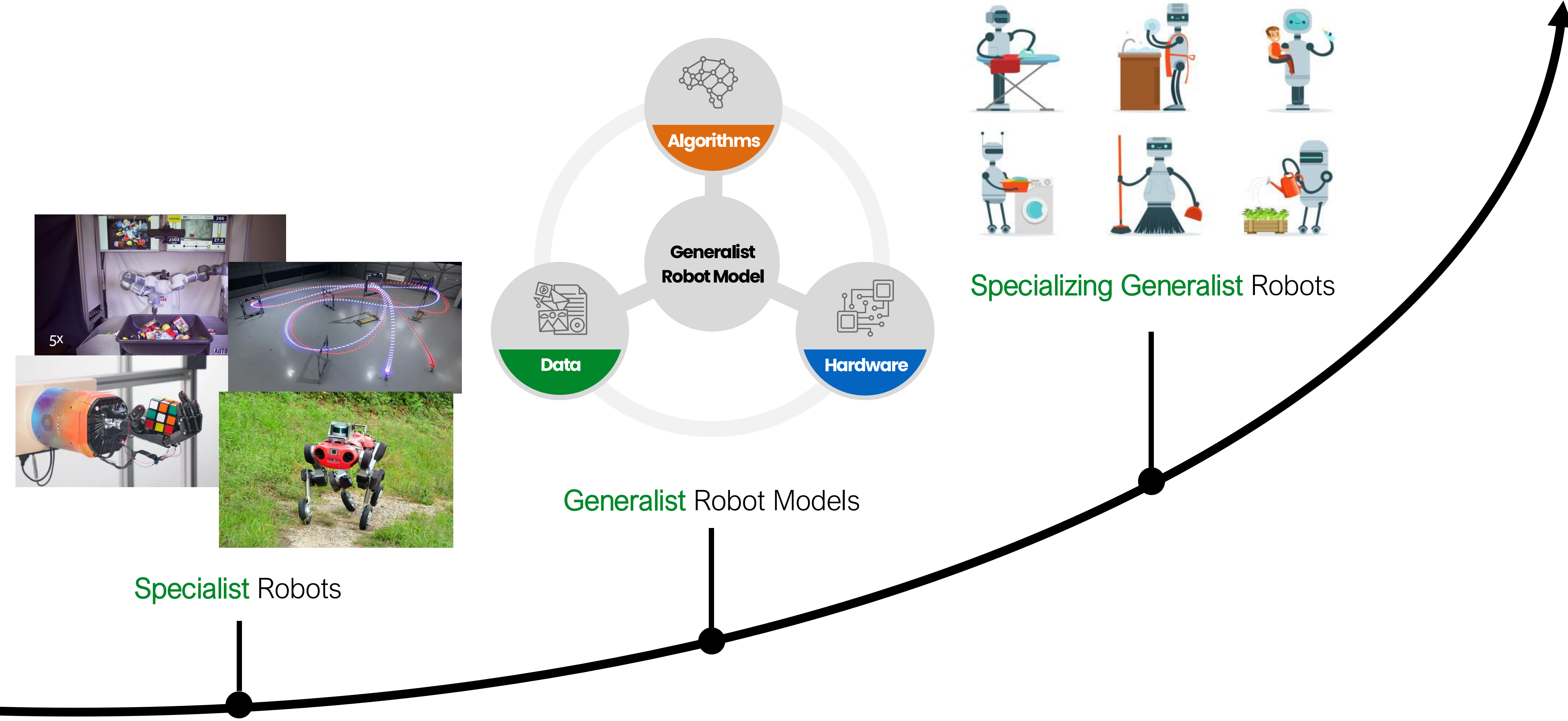
Generalist Robot Models



Specializing Generalist Robots



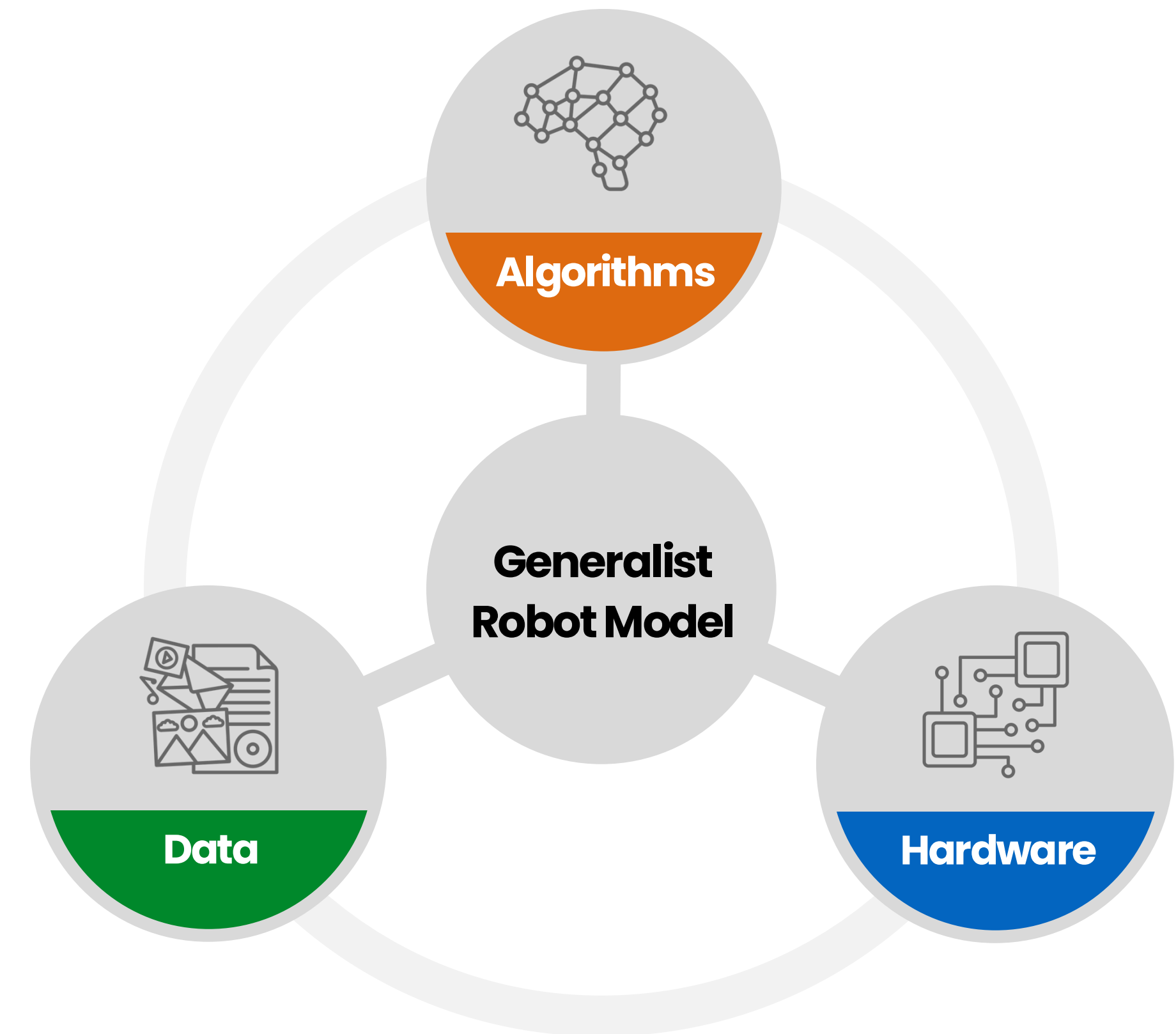
Specialist Robots





# Pathway to Generalist Robots

- **Algorithms:** Scaling up robot learning models with skill APIs and massively multi-task training
- **Data:** Building a data flywheel in real-world deployments through human-in-the-loop robot autonomy
- **Hardware:** Humanoid robots as a generalist robot platform to develop human-level physical intelligence





# UT Robot Perception & Learning Lab

**Mission statement:** building general-purpose robot autonomy in the wild

