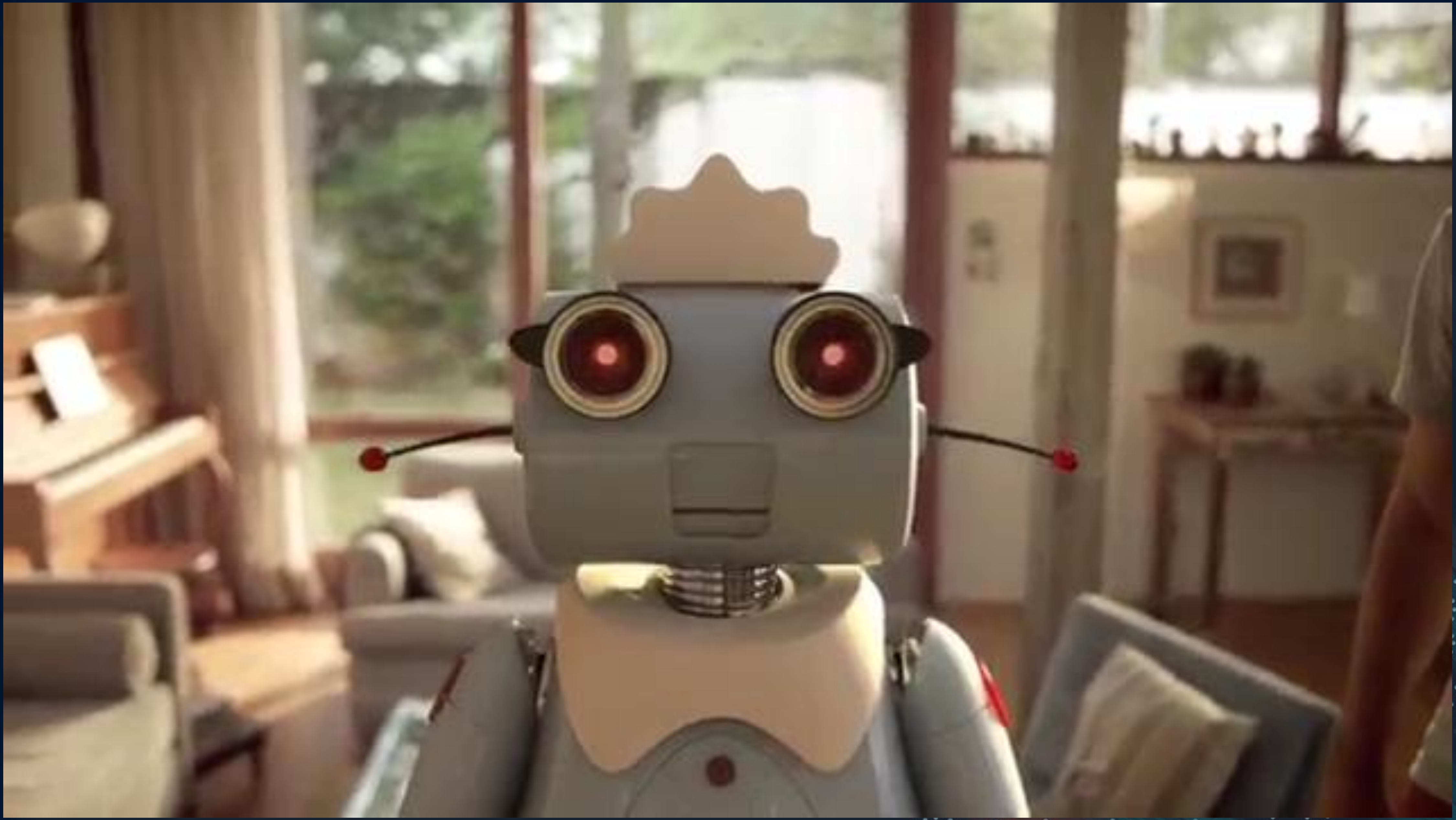
Generalizable Autonomy

Generative AI for General Purpose Robots

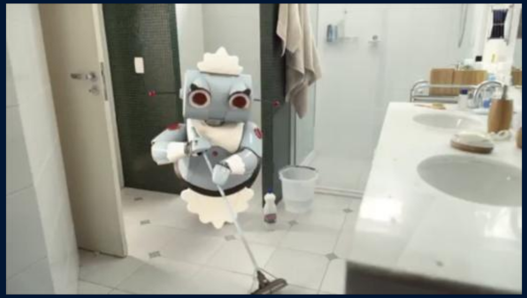
Animesh Garg

Professor of AI Robotics
Georgia Tech



Generalizable Autonomy

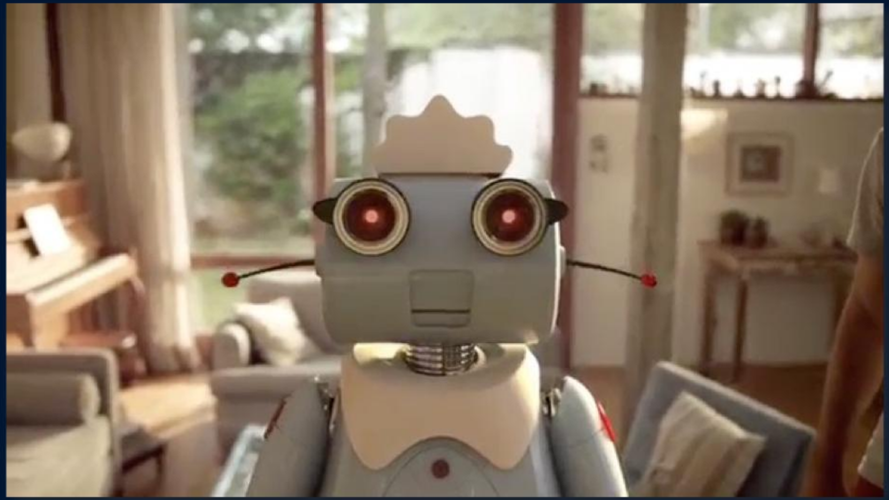
Vacuuming



Sweeping/Mopping

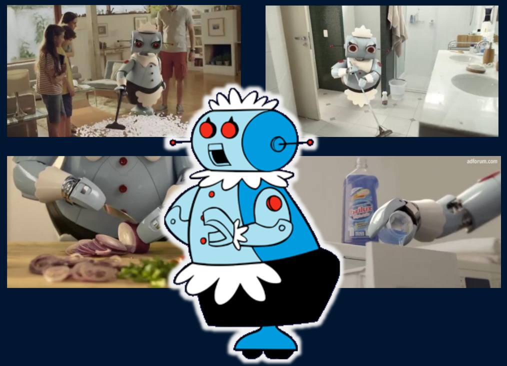
Cooking

Laundry



Generalizable Autonomy

Vision: Build Intelligent Robotic Companions towards Human Enrichment and Augmentation



Generalizable Autonomy

Vision: Build Intelligent Robotic Companions towards Human Enrichment and Augmentation

1968: Aspirational Robotic Assistant

Generalizable Autonomy

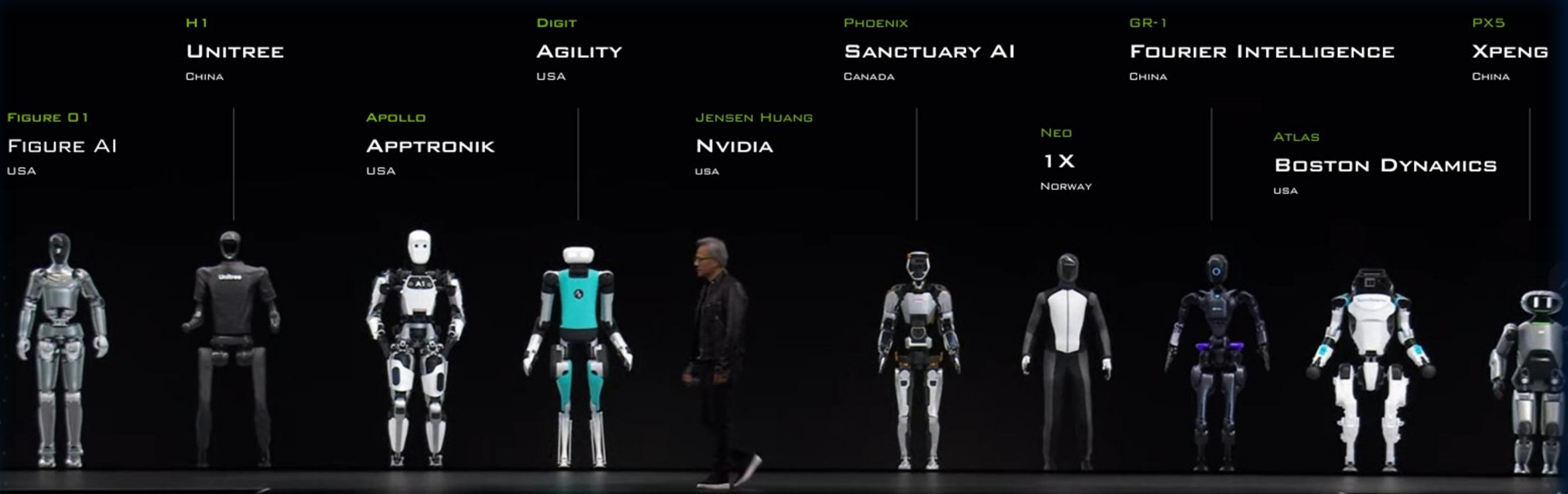
1950s
"AI"

1960s
Robots

2013
Walking

2019
Cooking

2024
Humanoids



Generalizable Autonomy

Vision: Build Intelligent Humanoid Platform

Strategy: Foundation Models with Large Scale Data

Structure

What do we need for sequential decision making in a physical setting?

Data

C4: How to Create, Collect, Clean, & Curate large-scale data for Robot Learning?

Generalizable Autonomy

Structure

What do we need for sequential decision making in a physical setting?

Data

How to Create, Collect, Clean, & Curate large-scale data for Robot Learning?

The Computing Stack

Digital AI

App

General-Purpose Applications
Ease of Use

OS

Platform-Agnostic OS
Modular Utilities

Driver

Hardware-Specific Drivers
Optimize Performance

The Computing Stack

Digital AI

App — **General-Purpose Applications**
Ease of Use

OS — **Platform-Agnostic OS**
Modular Utilities

Driver — **Hardware-Specific Drivers**
Optimize Performance

Physical AI

App — **Natural Interaction Interface**
Ease of Use

OS — **Platform-Agnostic Planning**
Modular Tool-Use

Driver — **Hardware-Specific Skills**
Optimize for Morphology

The Computing Stack

Physical AI

Natural Interaction Interface
Ease of Use

App

Platform-Agnostic Planning
Modular Tool-Use

OS

Hardware-Specific Skills
Optimize for Morphology

Driver

The Computing Stack

Physical AI

Internet Data
Language, Image, Video
$, Very Diverse

Synthetic Data
Simulation
$$, Engineered Designs

Real World Data
Teleoperation
$$$$, Limited Diversity

Natural Interaction Interface Ease of Use	App
Platform-Agnostic Planning Modular Tool-Use	OS
Hardware-Specific Skills Optimize for Morphology	Driver

The Computing Stack

Physical AI

LLM for Reasoning

Semantic Procedures

Planning as Program Synthesis

Multimodal Feedback Control

Natural Interaction Interface
Ease of Use

Platform-Agnostic Planning
Modular Tool-Use

Hardware-Specific Skills
Optimize for Morphology

Structure: What and How

”WHAT”

Multimodal Foundation Model

Does High-level Reasoning

Slow Inference

”HOW”

Generic Observation-to-control

Low-level Reasoning

Fast Inference

Input

“Tidy up the Kitchen”

→ “Take” ”Jug”

→ “Open” ”Fridge”

→ “Put” ”Jug” in ”Fridge”

Goal Generation

Goal-conditioned

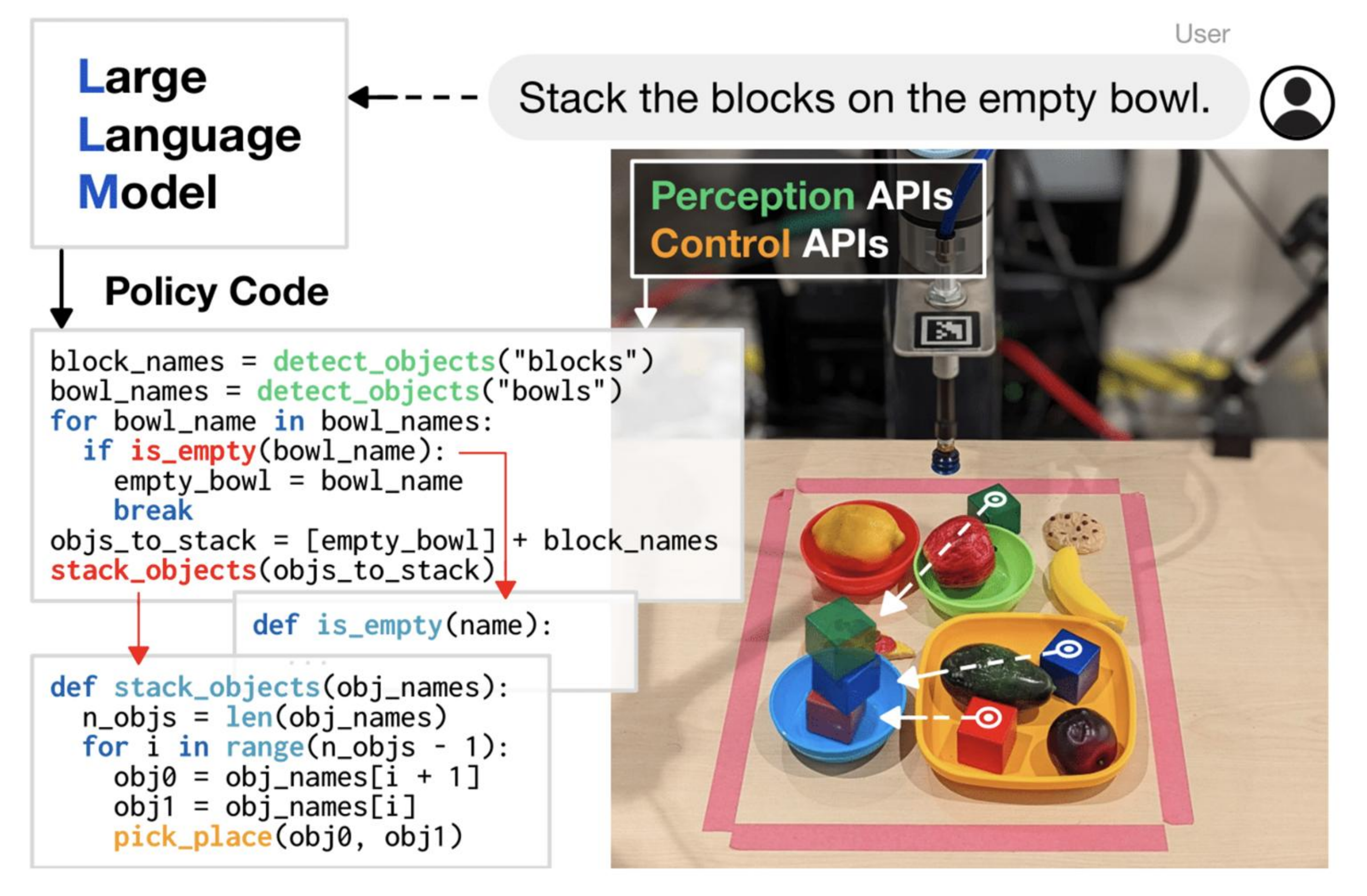
Reactive controller

Solvable online for

different agents

Program Synthesis

LLMs tell robots what to do!



Code as Policies:
Language Model Programs for Embodied Control

Jacky Liang Wenlong Huang Fei Xia Peng Xu Karol Hausman Brian Ichter Pete Florence Andy Zeng

Robotics at Google

ProgPrompt: Generating Situated Robot Task Plans using Large Language Models

ICRA 2023

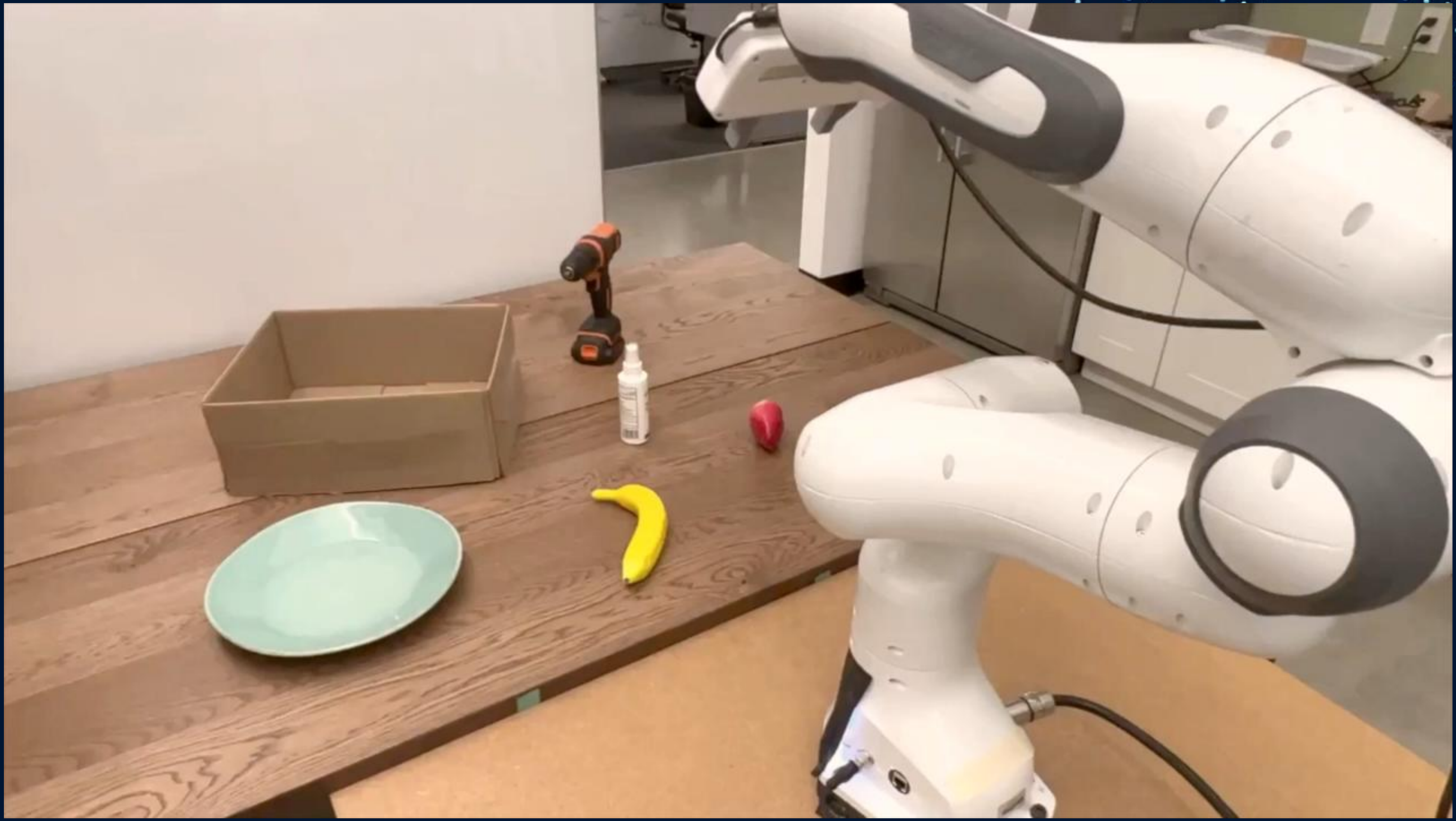
Ishika Singh[1], Valts Blukis[2], Arsalan Mousavian[2], Ankit Goyal[2], Danfei Xu[2], Jonathan Tremblay[2], Dieter Fox[2], Jesse Thomason[1], Animesh Garg[2]

Do As I Can, Not As I Say:
Grounding Language in Robotic Affordances

Michael Ahn* Anthony Brohan* Noah Brown* Yevgen Chebotar* Omar Cortes* Byron David* Chelsea Finn* Chuyuan Fu* Keerthana Gopalakrishnan* Karol Hausman* Alex Herzog* Daniel Ho* Jasmine Hsu* Julian Ibarz* Brian Ichter* Alex Irpan* Eric Jang* Rosario Jauregui Ruano* Kyle Jeffrey* Sally Jesmonth* Nikhil Joshi* Ryan Julian* Dmitry Kalashnikov* Yuheng Kuang* Kuang-Huei Lee* Sergey Levine* Yao Lu* Linda Luu* Carolina Parada* Peter Pastor* Jornell Quiambao* Kanishka Rao* Jarek Rettinghouse* Diego Reyes* Pierre Sermanet* Nicolas Sievers* Clayton Tan* Alexander Toshev* Vincent Vanhoucke* Fei Xia* Ted Xiao* Peng Xu* Sichun Xu* Mengyuan Yan* Andy Zeng*

Robotics at Google Everyday Robots

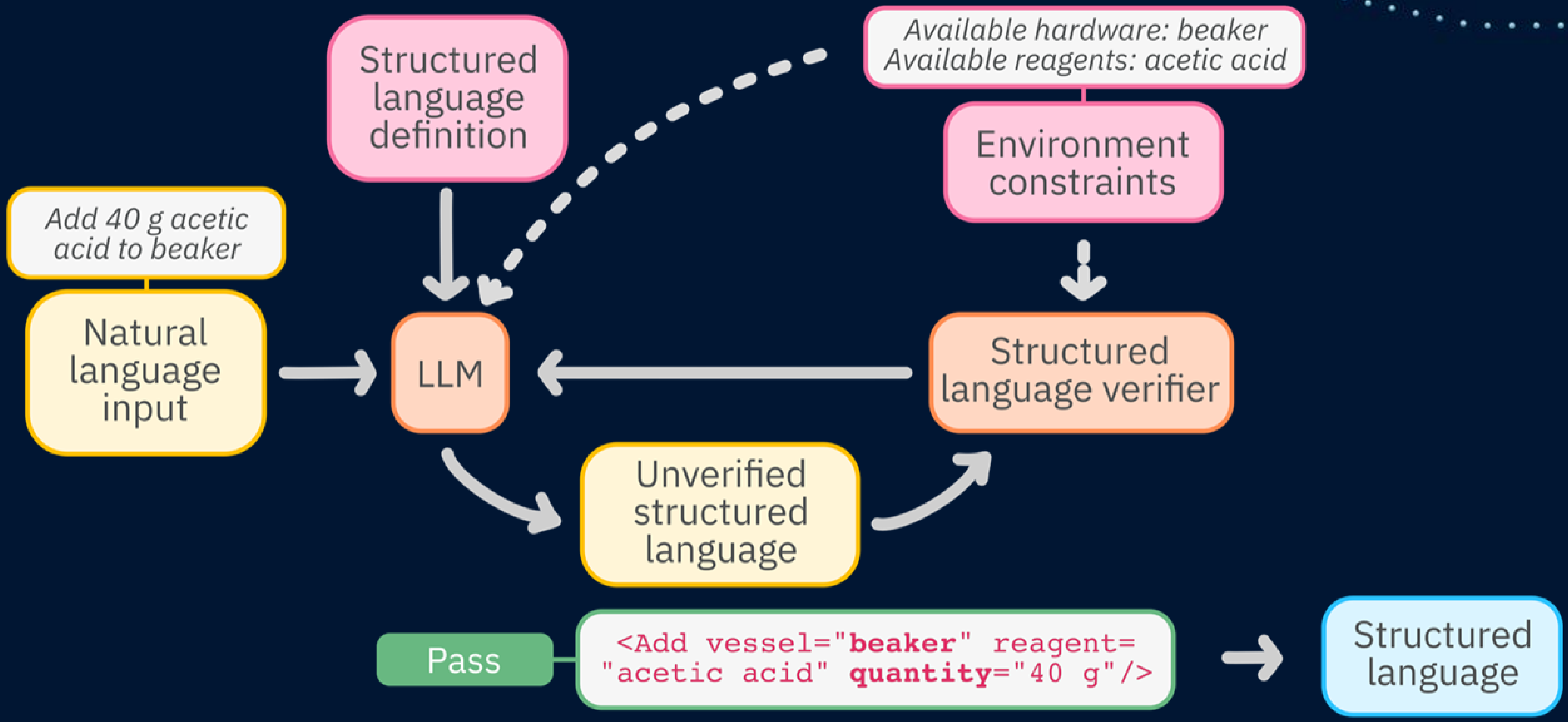
LLMs for Planning

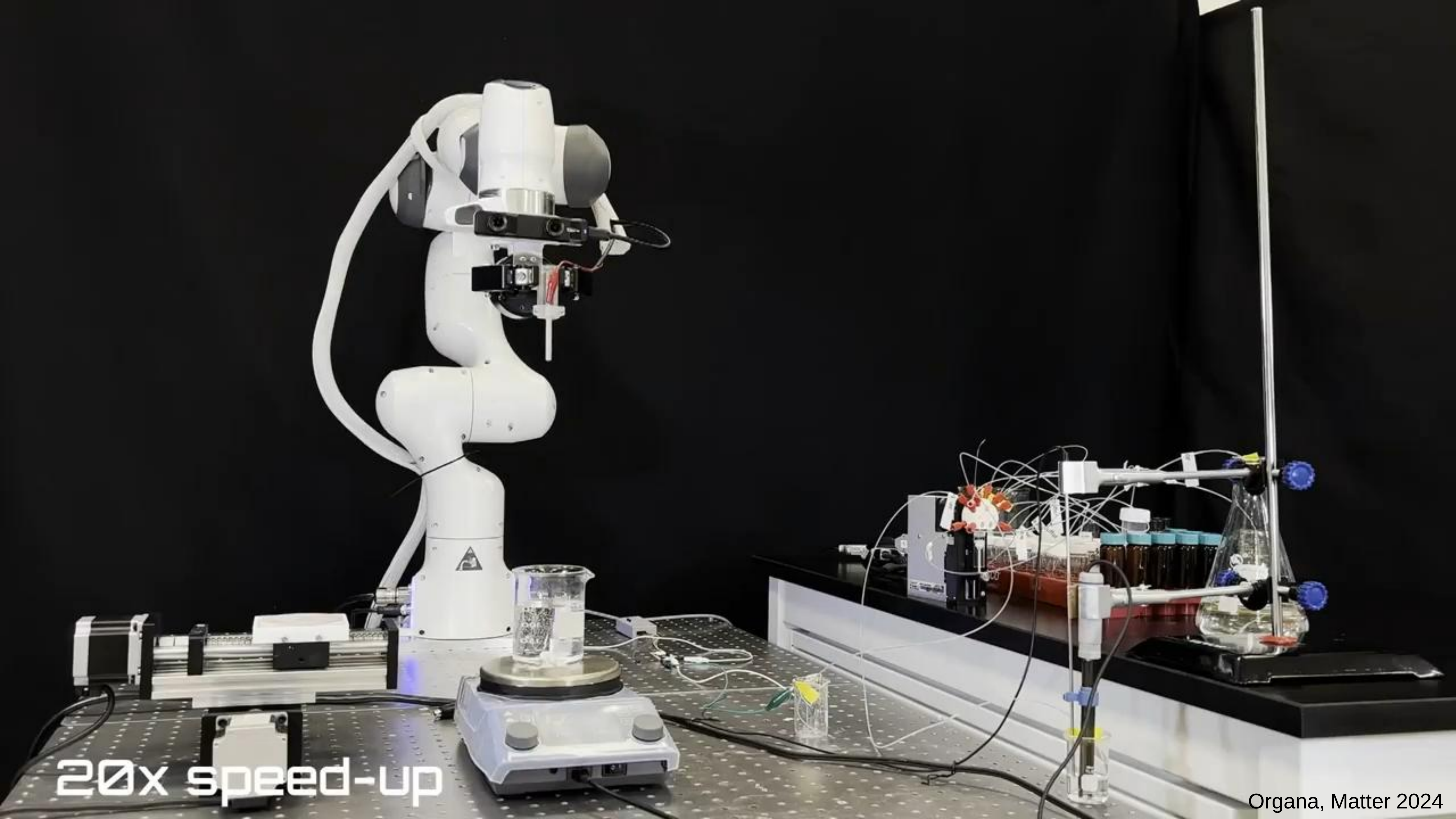


Progprompt, 2023

LLMs for Planning

CLAIRify





Organa, Matter 2024

CLIMB

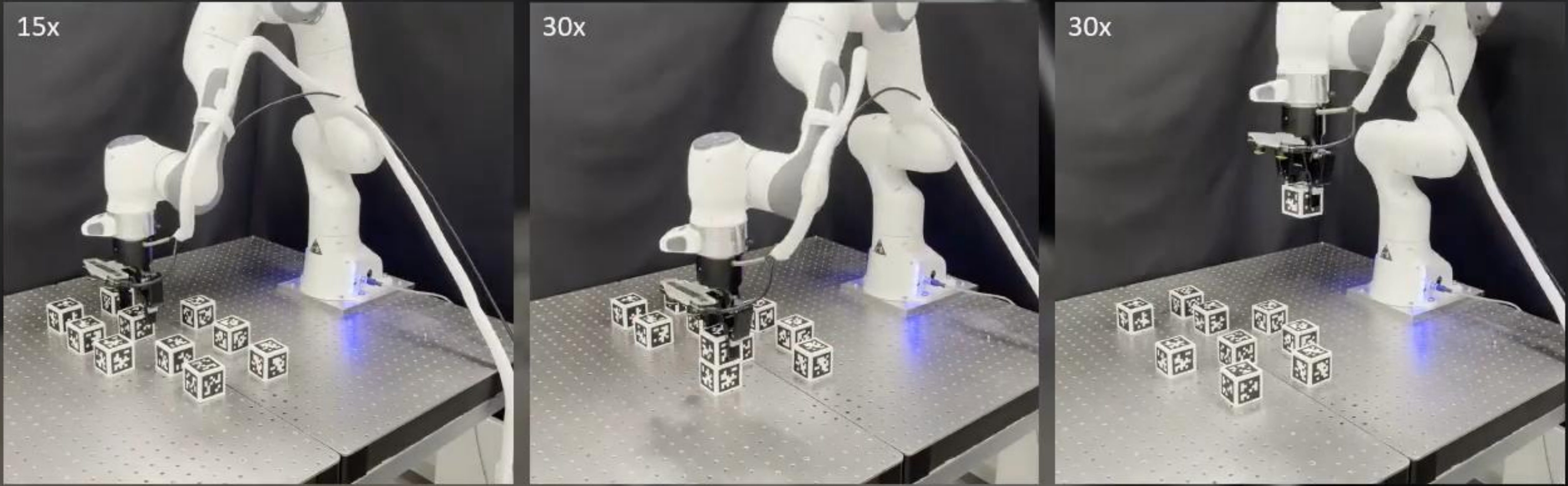
Language-Guided Continual Learning for Task Planning with Iterative Model Building

Walker Byrnes[1,2], Miroslav Bogdanovic[3], Avi Balakirsky[4], Stephen Balakirsky[2], Animesh Garg[1,3,5]

[1]Georgia Institute of Technology, [2]Georgia Tech Research Institute, [3]University of Toronto, [4]Ohio State University, [5]Nvidia

Basic Stack 3 | Basic Stack 4 | Basic Stack 6 | Line EW | Line NS | Arrange L | Pyramid 3

Note: All basic stack problems were successful on first rollout.



The Computing Stack
Physical AI

Large Behavior Models

Large Scale Imitation Learning

Learned Task Planning and replanning Behavior

Fine-Tune Generalists for better Specialists using RL

Natural Interaction Interface
Ease of Use

Platform-Agnostic Planning
Modular Tool-Use

Hardware-Specific Skills
Optimize for Morphology

Structure: What and How

”WHAT”
Multimodal Foundation Model
Does High-level Reasoning
Slow Inference

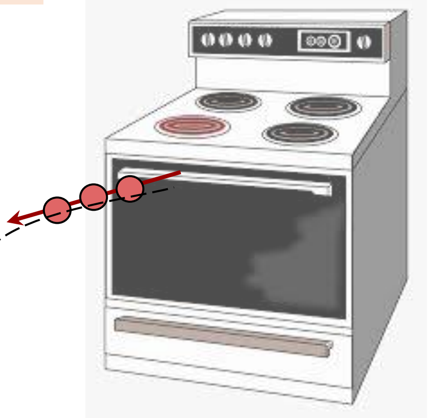
”HOW”
Generic Observation-to-control
Low-level Reasoning
Fast Inference

Input
“Tidy up the Kitchen”

→ “Take” ”Jug”
→ “Open” ”Fridge”
→ “Put” ”Jug” in ”Fridge”

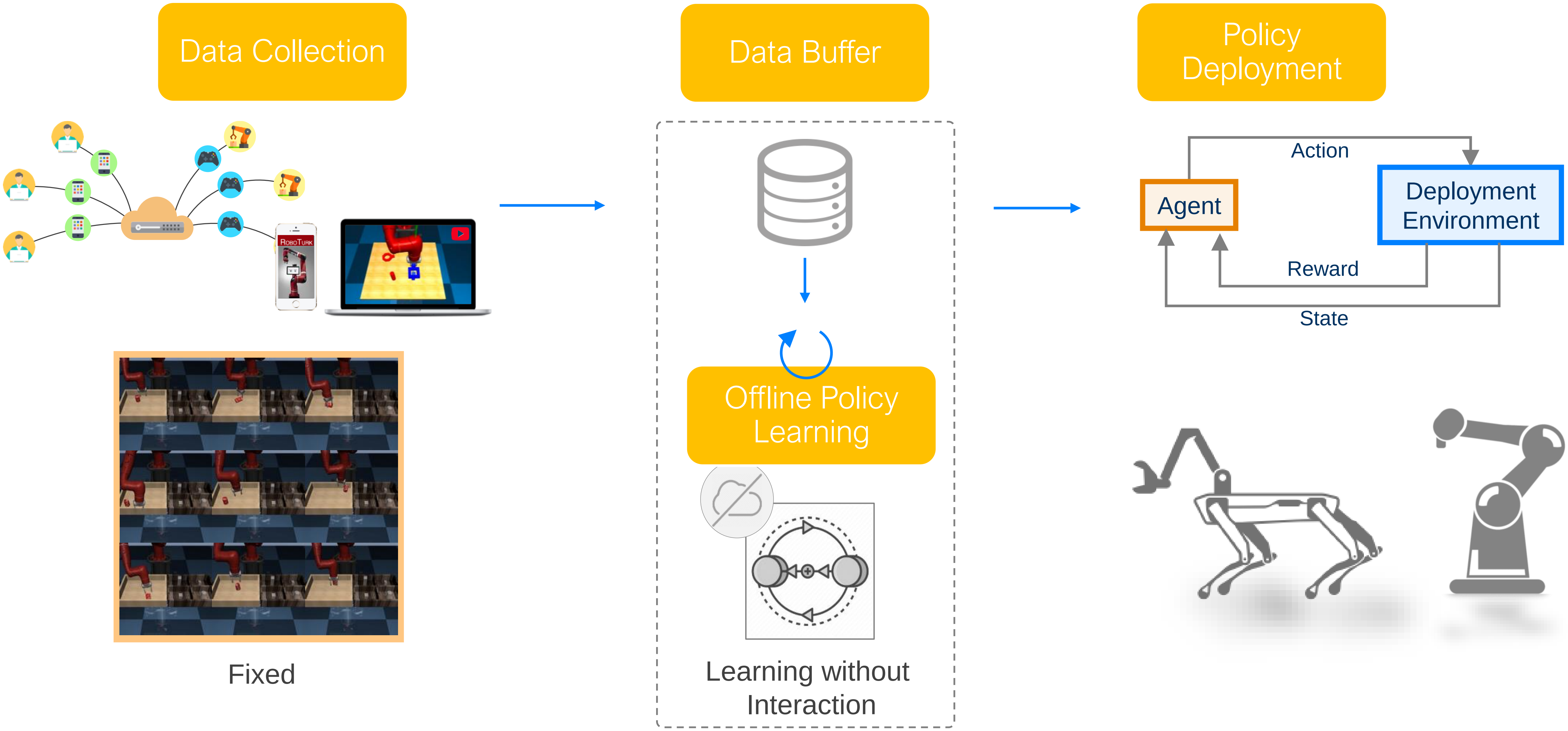
Goal Generation

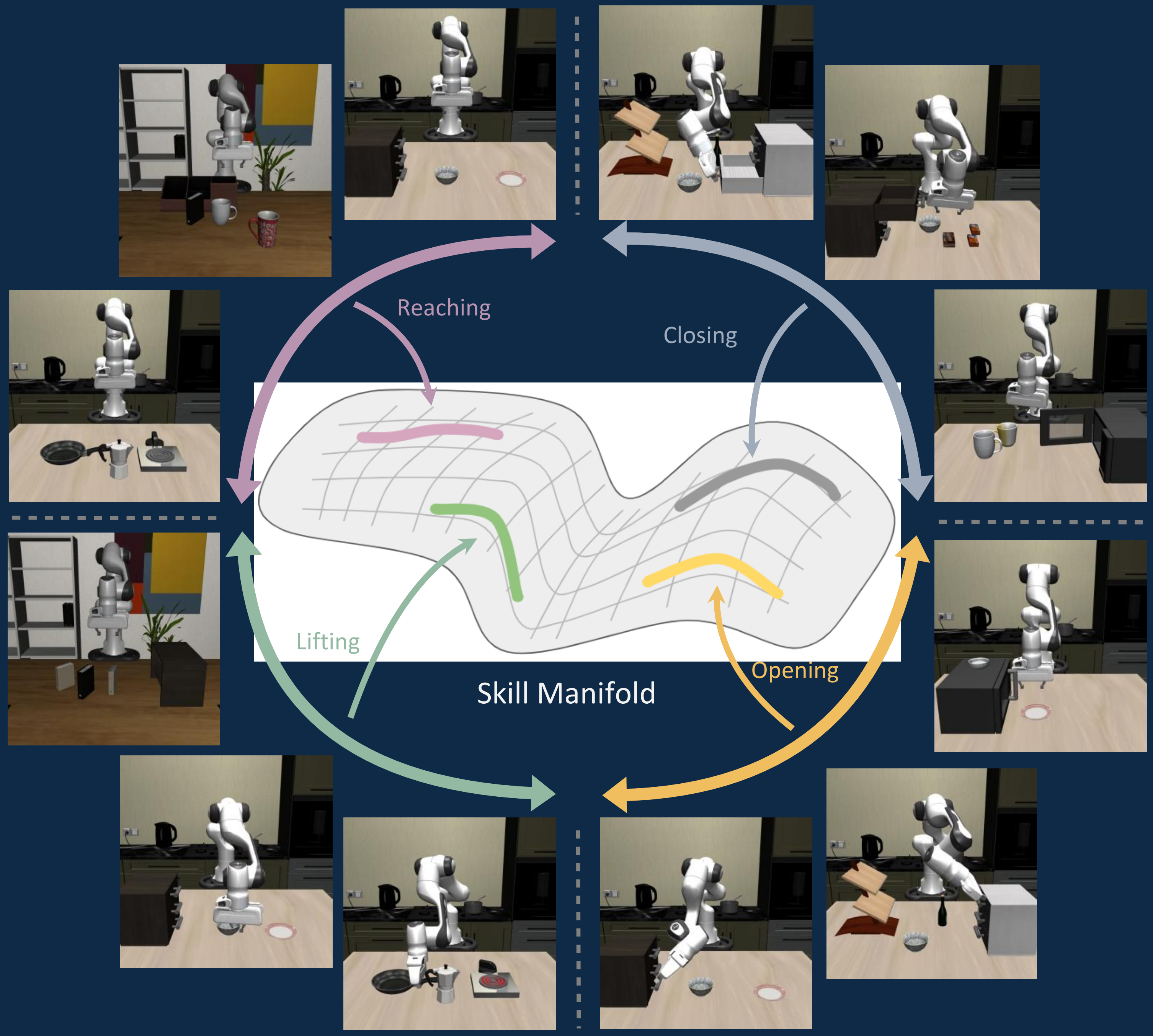
Goal-conditioned
Reactive controller



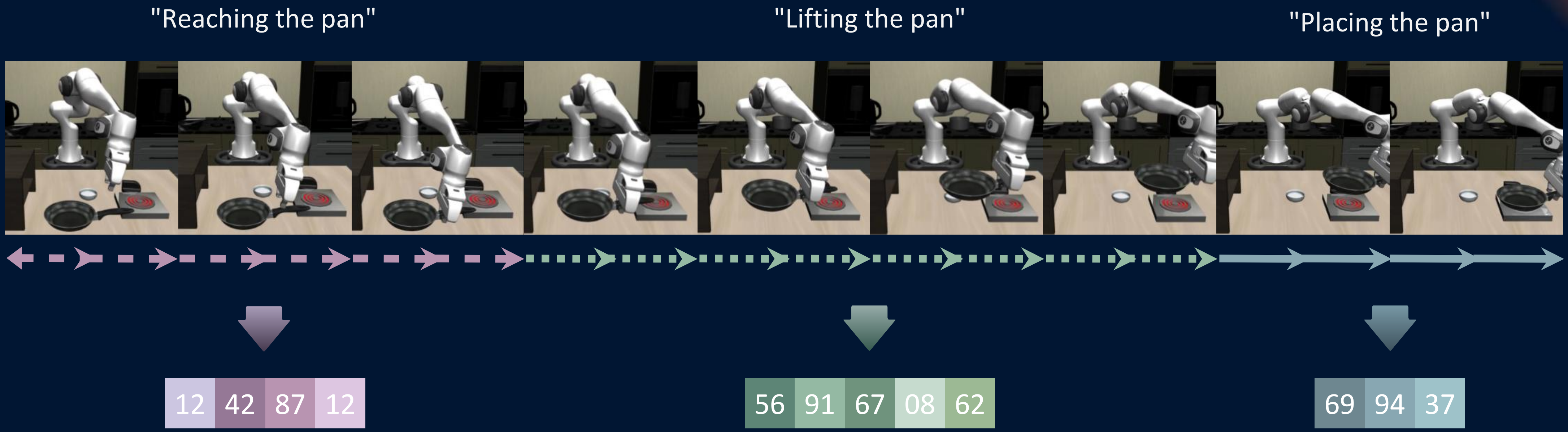
Solvable online for
different agents

Policy Learning from Offline Datasets



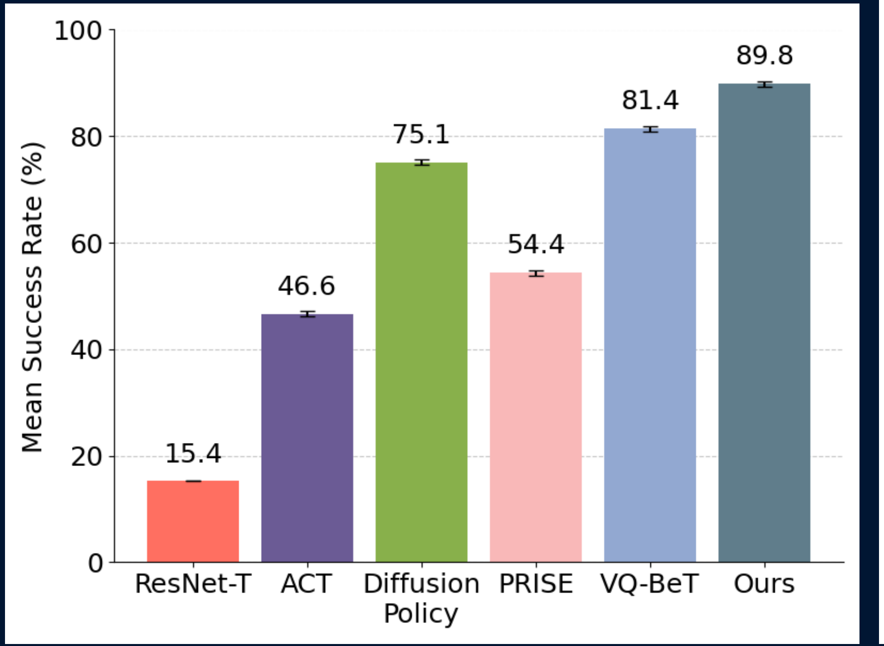

Reaching
Closing
Lifting
Opening
Skill Manifold

Multi-task Learning: QUEST



Multi-task Learning: QUEST

Multitask-IL LIBERO-90



Multitask IL: Relative improvement of 10.3% over next best baseline

How to learn many things (better than data)

PWM: Policy Learning with Large World Models

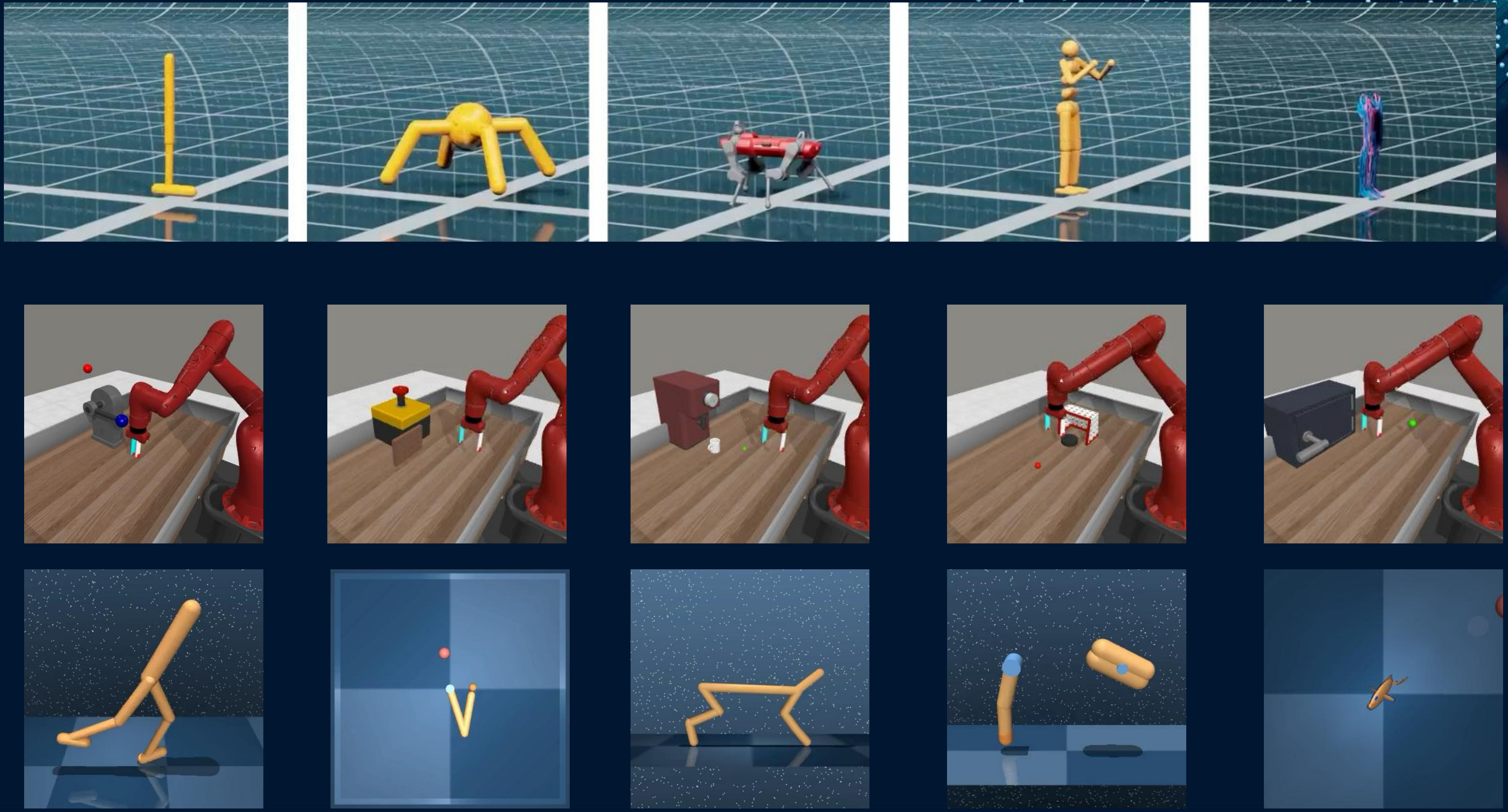
Large World Models

Multi-task

Efficient Policy Optimization

1. Regularized large models enable efficient policy learning
2. Use First-order optimization to train policies in <10m per task

PWM learns over 80 tasks



The Computing Stack

Physical AI

Motion Generation Models

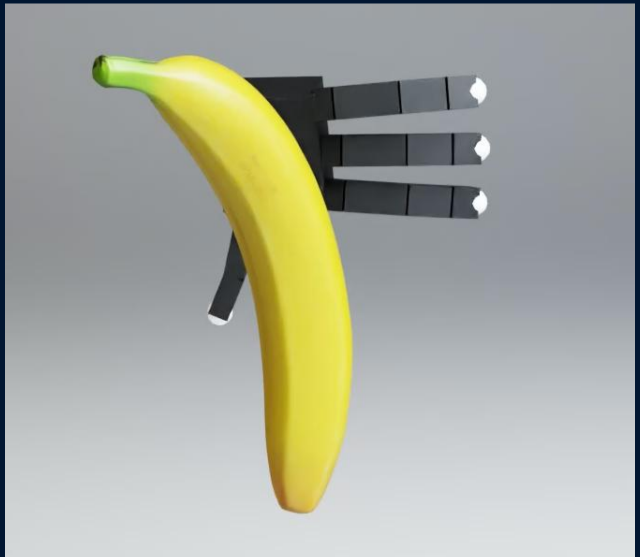
Fine-Tune Generalists for better Specialists.

Reinforcement Learning for Locomotion, WBC + Dexterity

Self-supervised learning without rewards

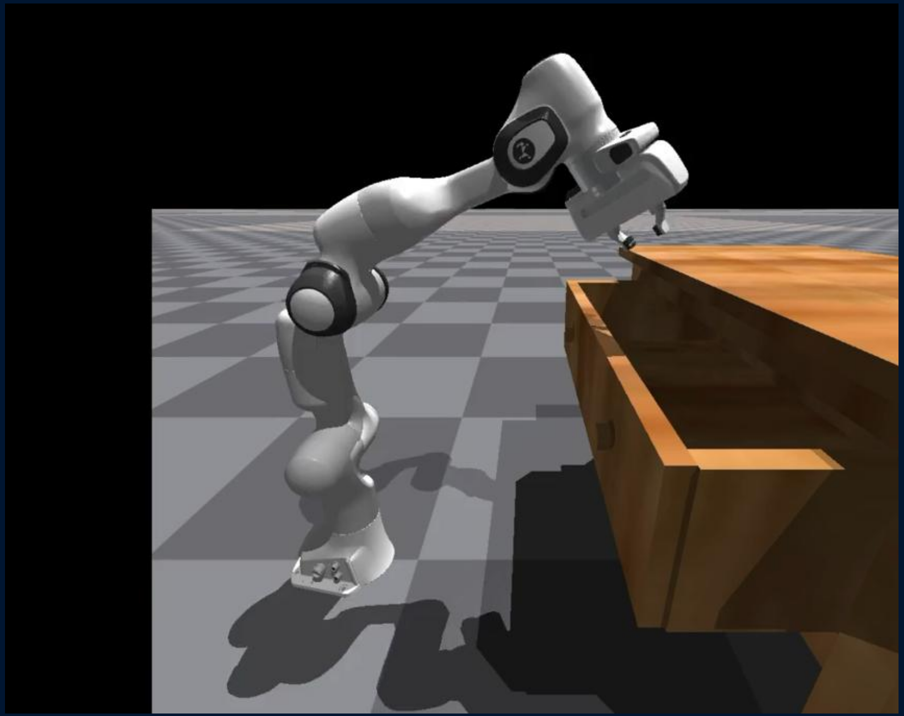
Natural Interaction Interface Ease of Use	App
Platform-Agnostic Planning Modular Tool-Use	OS
Hardware-Specific Skills Optimize for Morphology	Driver

Dexterity Primitives



- Open World Functional Grasping
- Contact-Rich Interaction
- Robust Pick and Place
- Language Conditioned Interface for Planners

Self-Supervised Learning

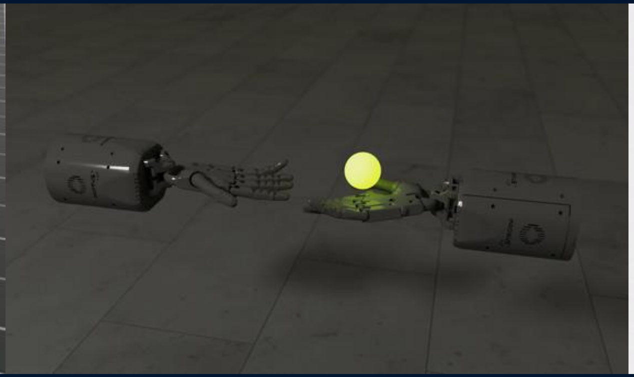
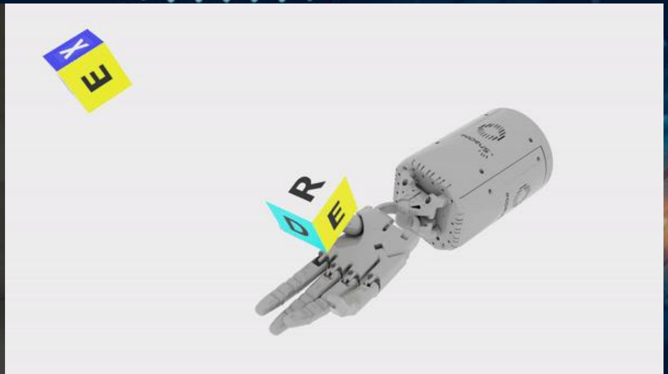


Learning to do “what can be done”
Learning from Self-Supervised Play

Transferring to Real world
Learning without data or rewards

Locomanipulation



- Robust Walking
- Robust Pick and Place

- Multifinger In-hand manipulation
- Dexterous Bimanual Manipulation


x1.5
IROS 2022 (under review)

Generalizable Autonomy

Structure

What do we need for sequential decision making in a physical setting?

Data

How to Create, Collect, Clean, & Curate large-scale data for Robot Learning?

Generalizable Autonomy

Structure

What do we need for sequential decision making in a physical setting?

Data

How to Create, Collect, Clean, & Curate large-scale data for Robot Learning?

Large-scale Datasets → Accelerated Progress

IMAGENET

LAION-5B

C4

StarCoder

The Pile

Open Images Dataset

WIKIPEDIA
The Free Encyclopedia

RedPajama

COMMON CRAWL

refinedweb

Microsoft COCO
Common Objects in Context

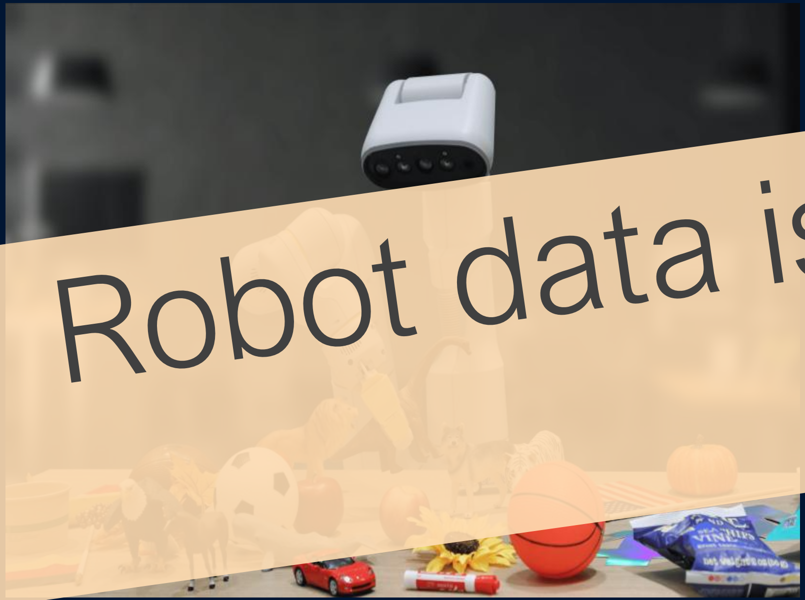
PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

places
THE SCENE RECOGNITION DATABASE

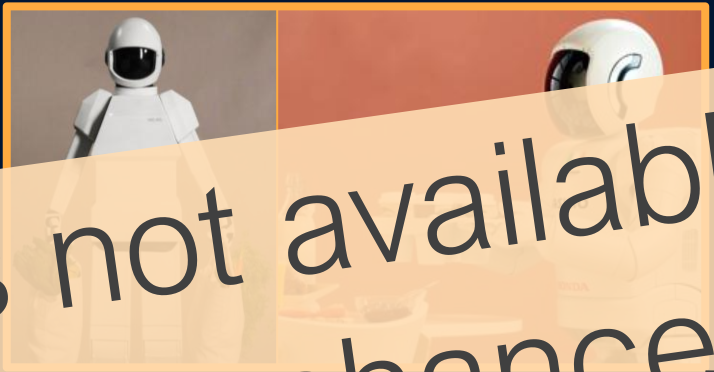
Russell et al. (2008), Deng et al. (2009), Everingham et al. (2010), Socher et al. (2013), Zhou et al. (2014),
Lin et al. (2015), Rajpurkar et al. (2016), Krishna et al. (2016), Iyer et al. (2017), Williams et al. (2018), Wang et al. (2018)

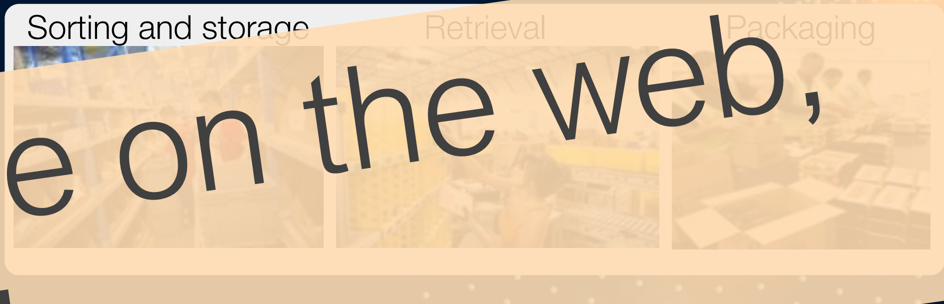
Large Scale Data mined from the web!

Generalization Autonomy: Data



RT-2, Google 2023





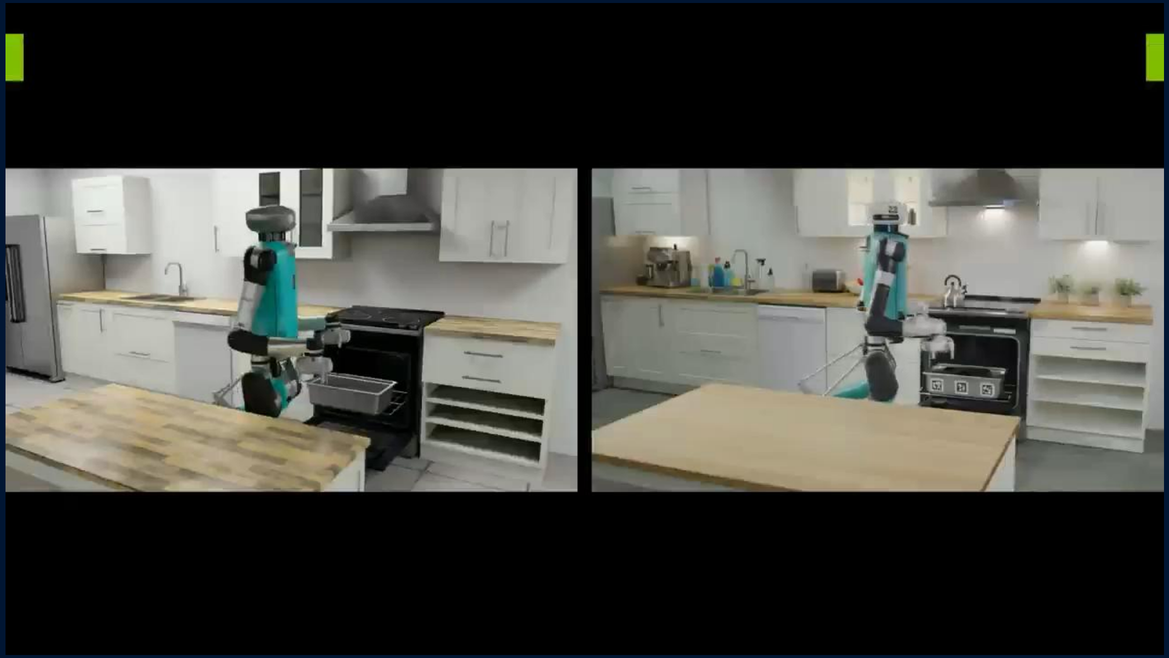
Robot data is not available on the web, by chance!

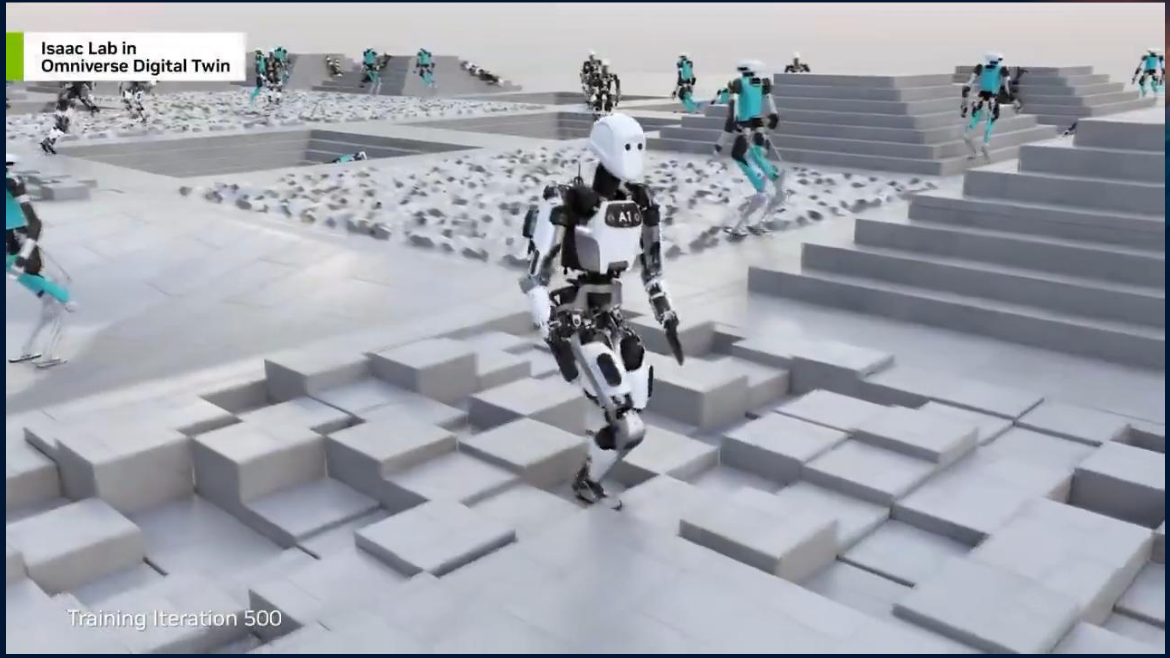
Too many problems to create datasets for each!

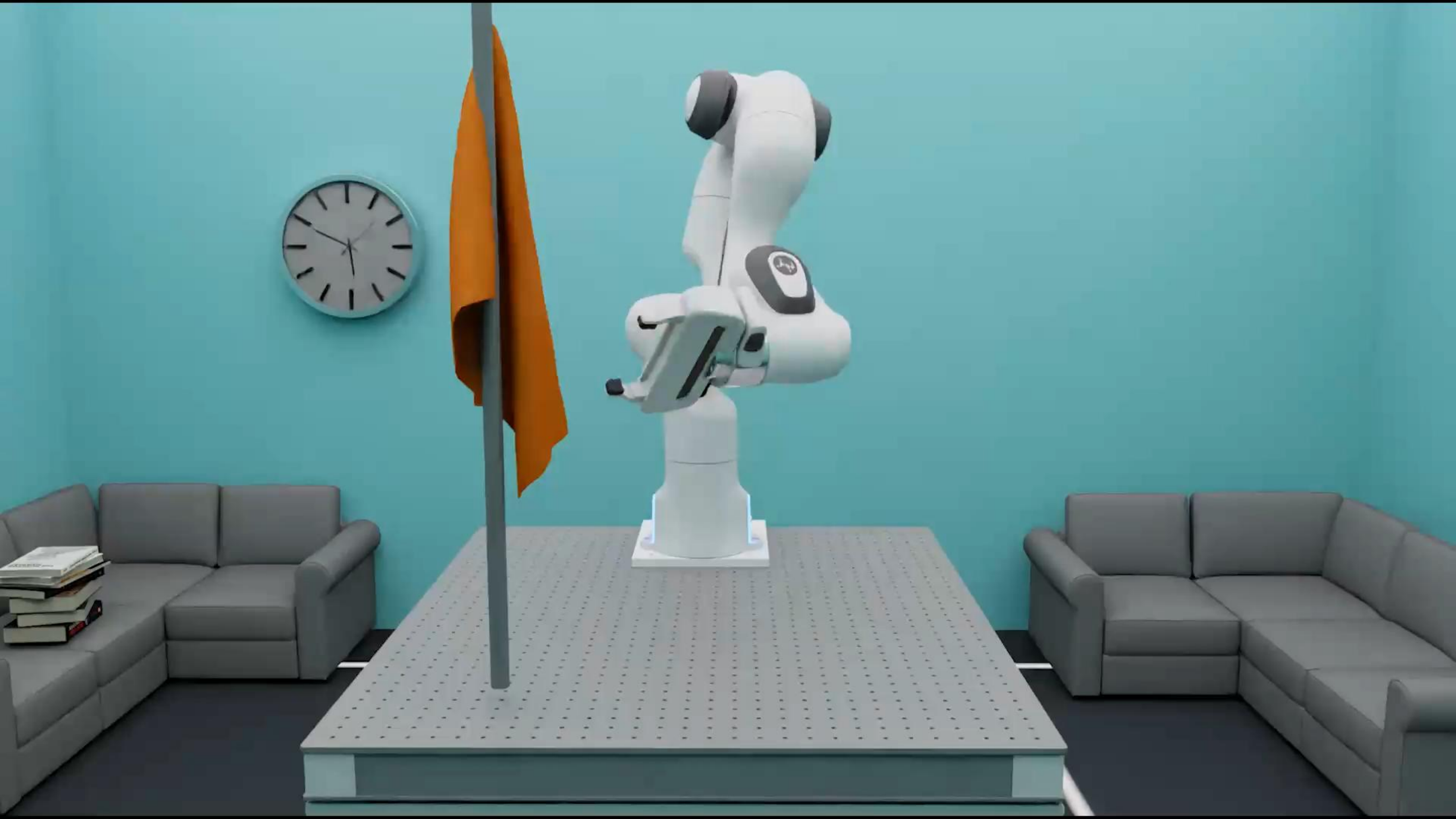
Isaac Lab

Unified Framework for Robot Learning

Isaac Lab built on Orbit (BSD OSS)

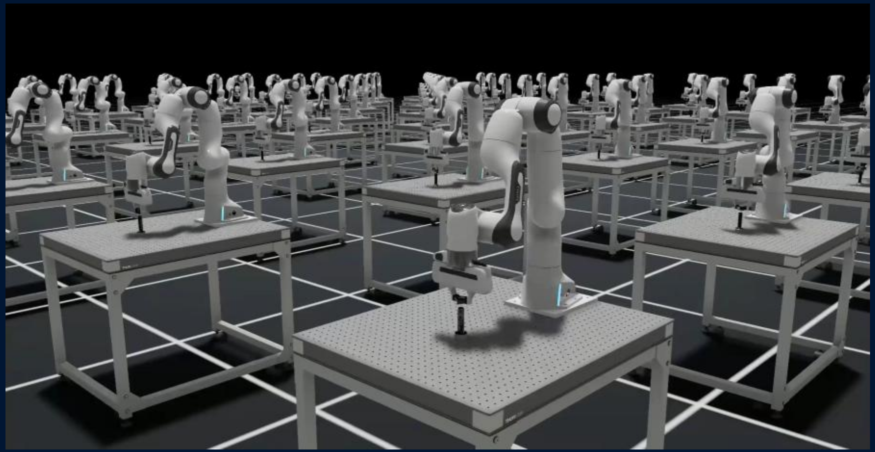


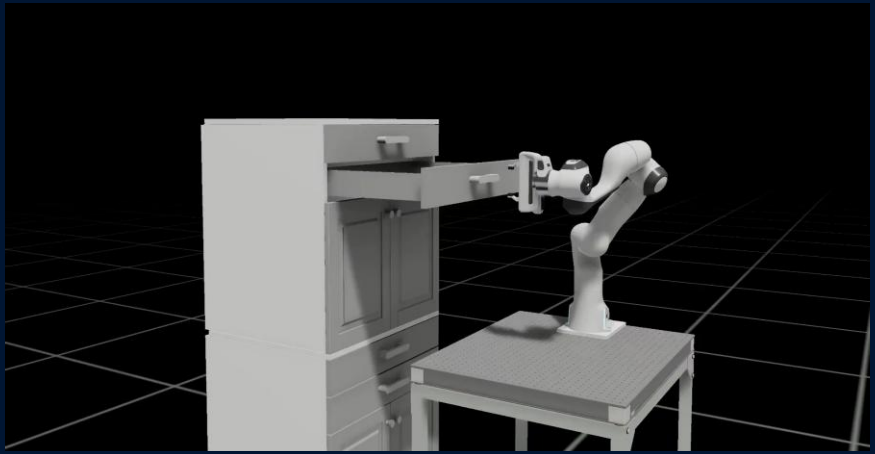




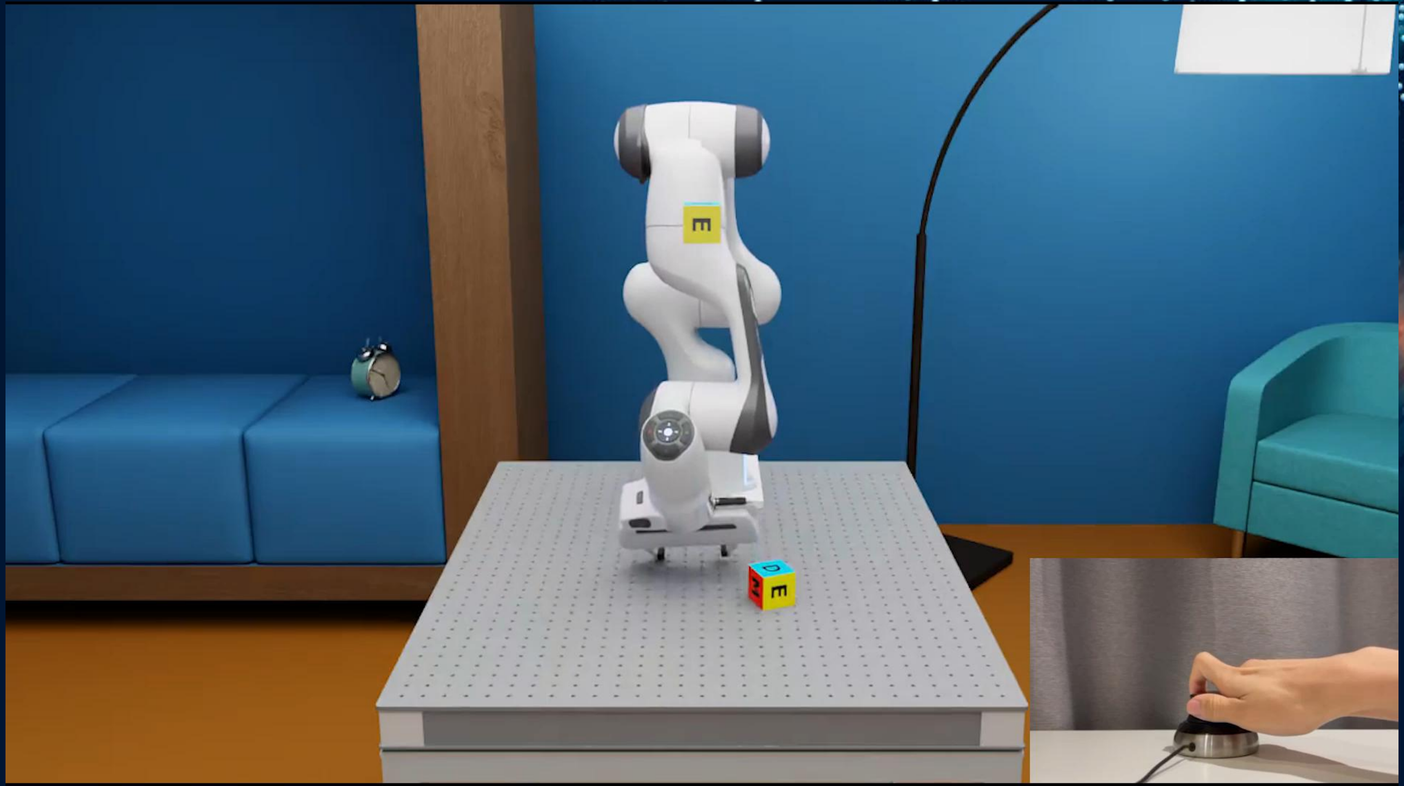
Isaac Lab: Content

Isaac Lab: Workflow



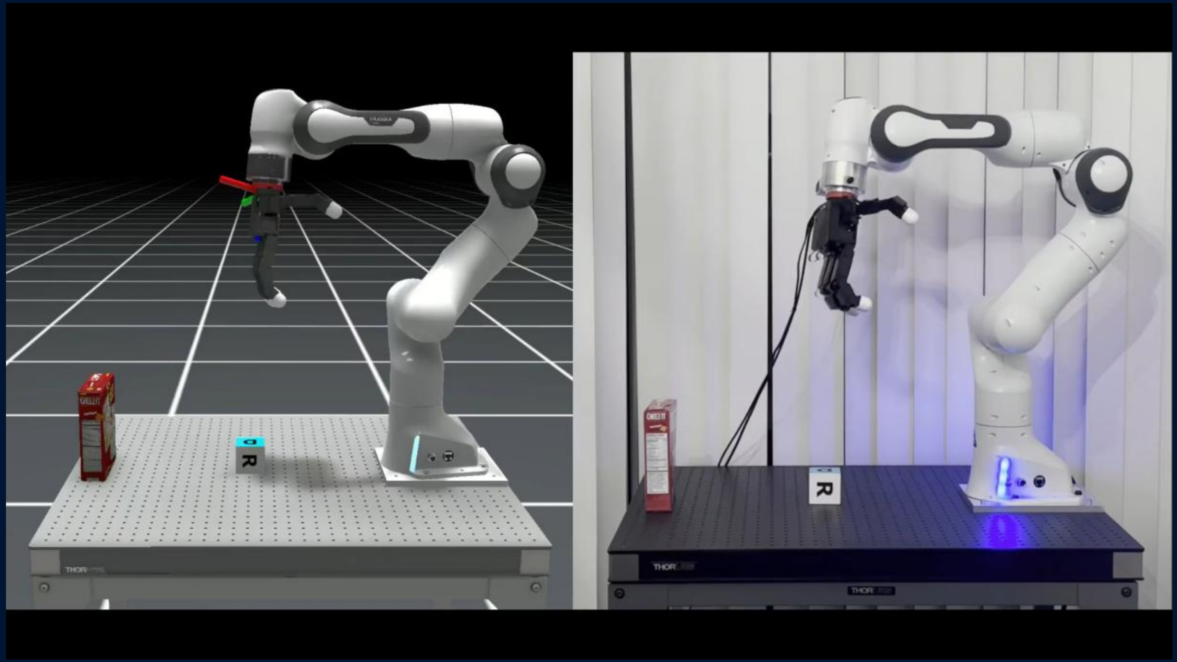


RL with Accelerated Simulation

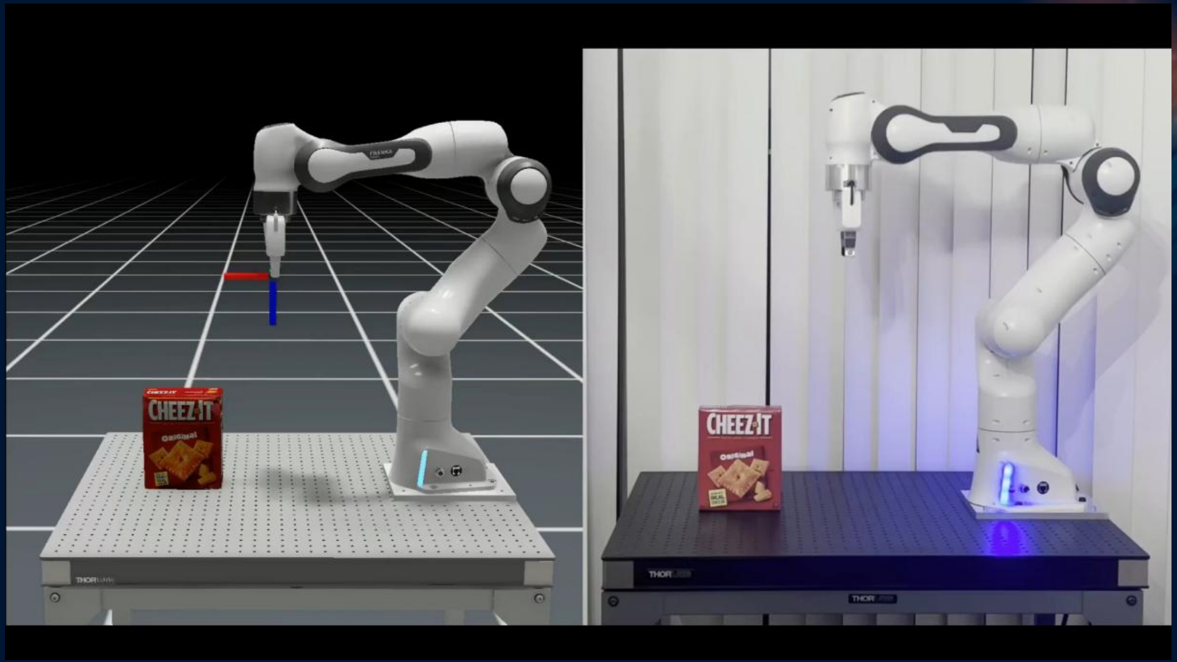


Teleoperated Data Collection and Learning

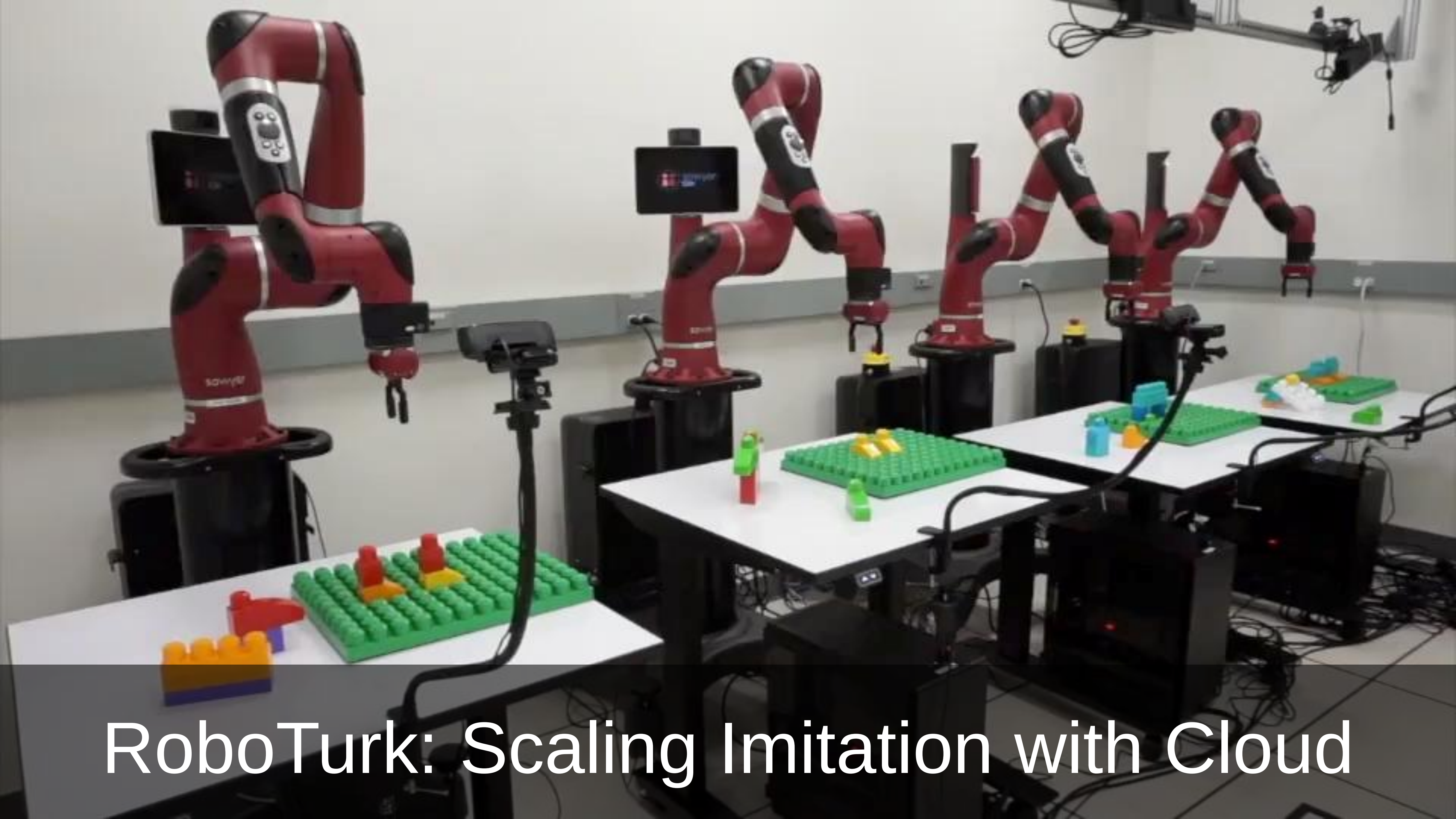
Isaac Lab: Workflow



Franka EMIKA arm with allegro hand



Franka EMIKA Arm with parallel-jaw gripper



RoboTurk: Scaling Imitation with Cloud

RoboTurk: Dexterous Data Collection

Operator specifies full 6-DoF motion of the arm by moving their phone.

Roboturk Scalability

Multiple Simultaneous Teleoperation Connections

CoRL 2018, IROS 2019

The Computing Stack

Physical AI

Internet Data
Language, Image, Video
$, Very Diverse

Synthetic Data
Simulation
$$, Engineered Designs

Real World Data
Teleoperation
$$$$, Limited Diversity

Natural Interaction Interface
Ease of Use

Platform-Agnostic Planning
Modular Tool-Use

Hardware-Specific Skills
Optimize for Morphology

Generalizable Autonomy

Structure

Data

Generative AI to Enable Robotics

Innovations in better Models and larger datasets

Generalizable Autonomy

Generative AI for General Purpose Robots

Animesh Garg

Professor of AI Robotics
Georgia Tech