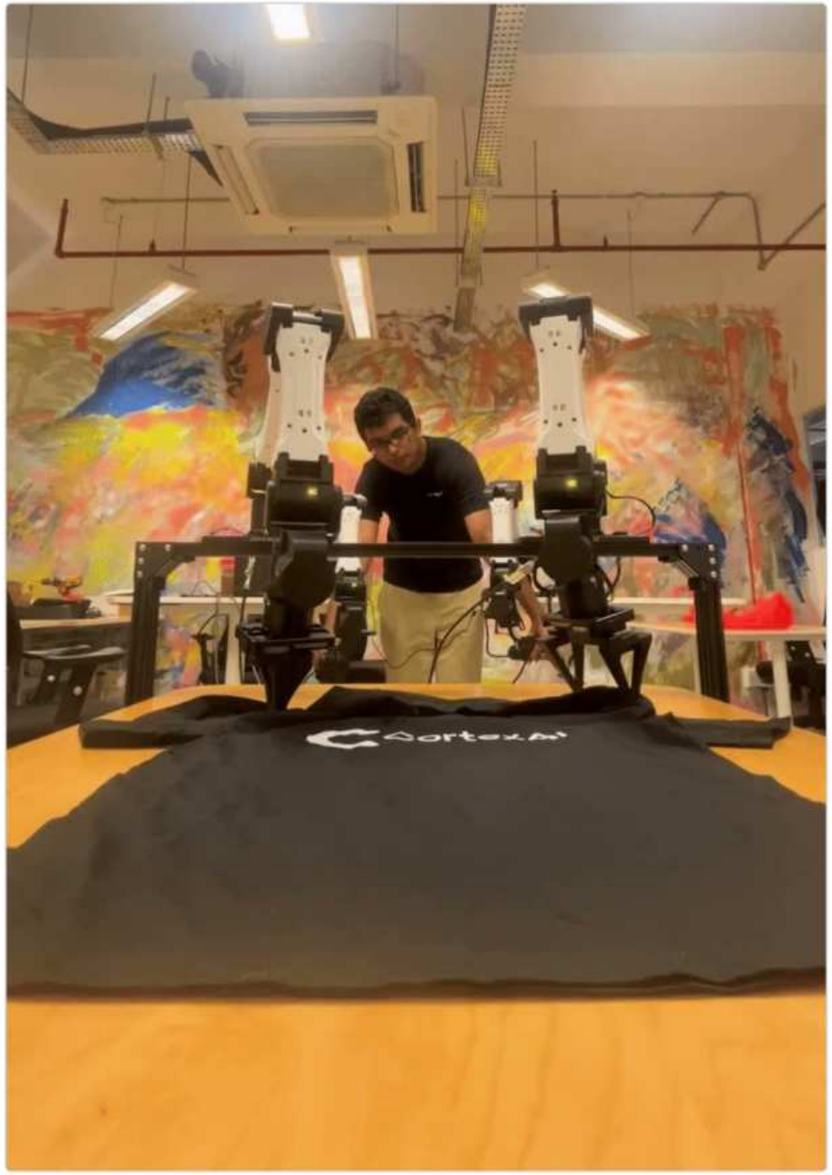
29 September 2026

When AI Gets a Body

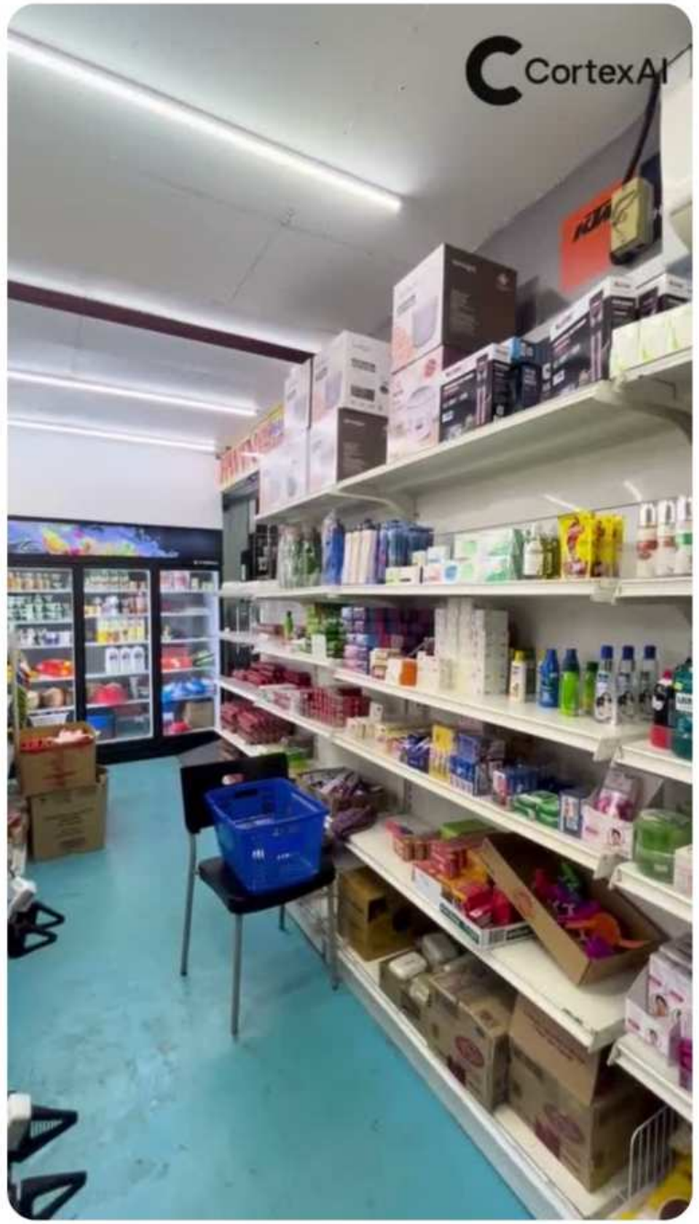
Intelligent Robots are Here

Suveen Ellawela

Founding Engineer @ Cortex AI



Making coffee
in the office



Restocking
products in a
convenience store

https://cortexrobot.ai/robots

AI is being used every day

AI agents

Grokbot

Muse

Codex

Claude

Cursor

And more…

+

Everyday tools

Gmail

Calendar

Notion

Salesforce

Google Docs

Google Slides

And more…

What happens when the work is physical?

Fold laundry

Load dishwasher

Laboratory work

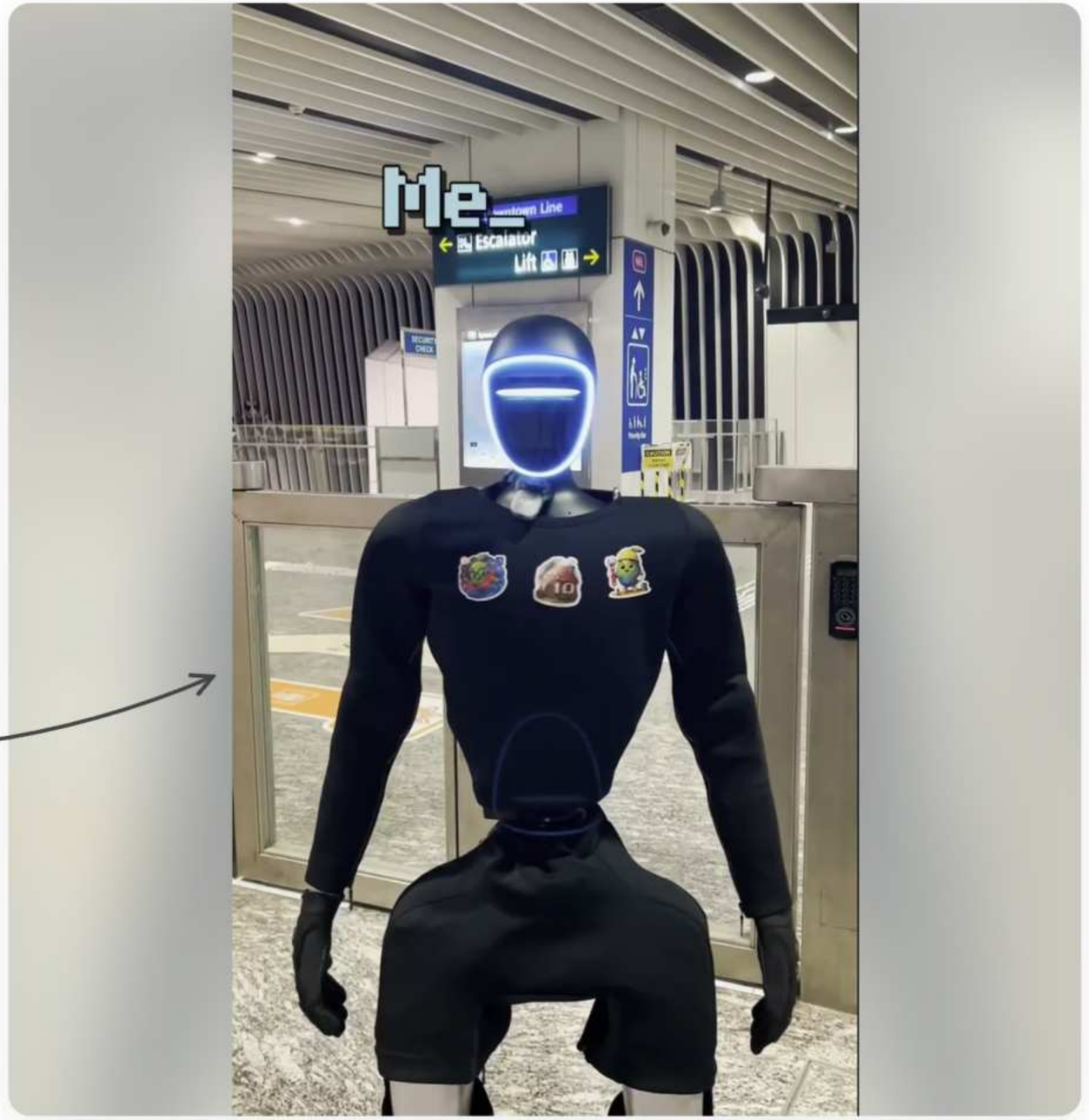
Robots in Structured Environments

Custom-built robots | Pre-programmed

But the real world is not structured

Robots are here in Singapore!

a humanoid robot
in an MRT station



> It can answer questions, but it cannot open your bottle's lid if you ask it to.

https://www.youtube.com/shorts/pEE2UGC-ygM

Moravec's paradox

Play strong chess

Hard for most people.

Easy for AI.

Pick up a mug

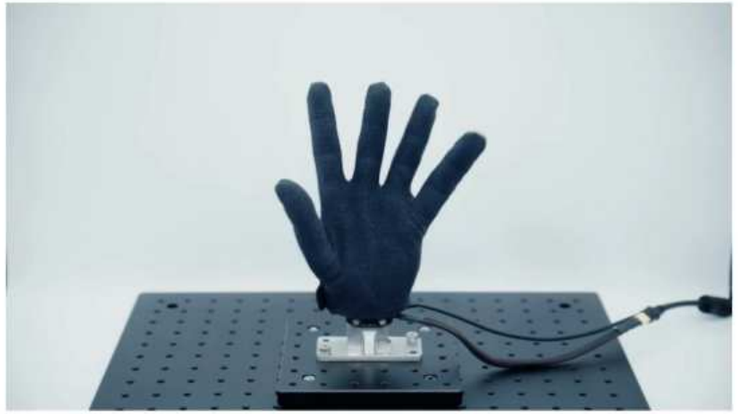
Routine for people.

Still challenging for robots under variation.

How can we achieve general-purpose robotics?
And how close are we to truly Intelligent Robots?

Robots can already move like us, or even more.

Unitree



Wuji Hand

Boston Dynamics · Atlas

https://www.youtube.com/watch?v=y7ojRmPxqNg ·

https://www.youtube.com/watch?v=LXVV-oErD8s ·

https://www.youtube.com/watch?v=xKK5ze3FukQ

But robots capable of doing tasks have been here for a while.

2007

Stanford PR1

Teleoperated

A person supplied the decisions.

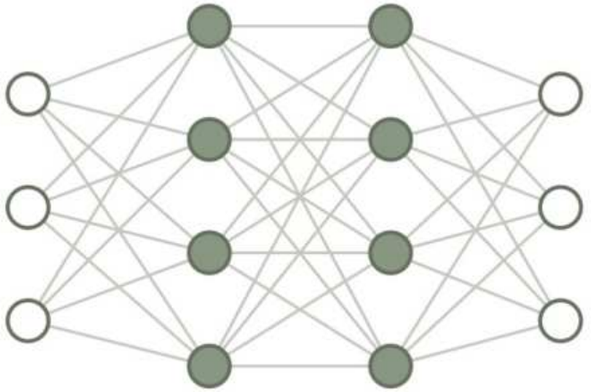
https://www.youtube.com/watch?v=jJ4XtyMoxIA

The missing part for robotics is the brain.

There's a lesson from language models

More varied data

+

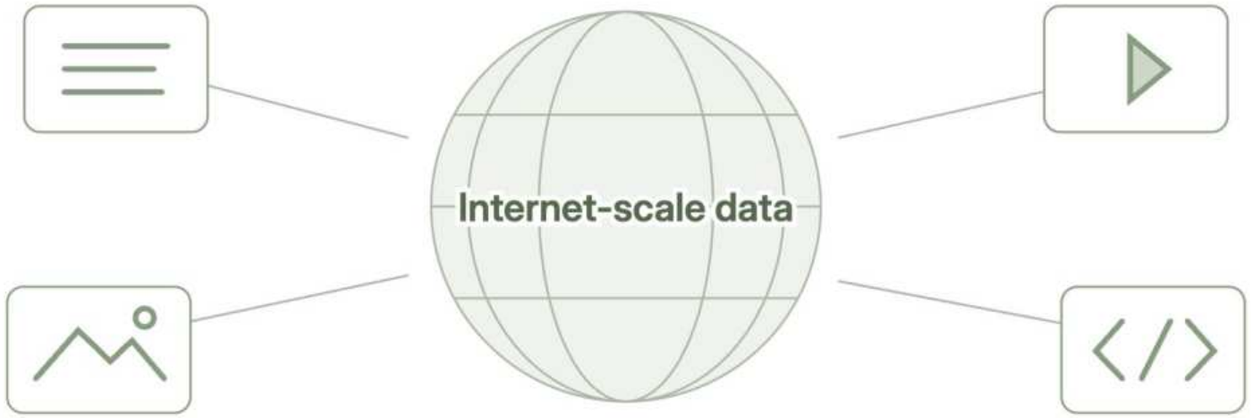


Larger models

→

Broader robot skills?

LLMs learned from internet-scale data.



Internet text what LLMs train on

"How do I fold a fitted sheet? Tuck each corner into the next, then lay it flat and…"

× trillions of tokens

Robot-action data what robots need

cameras overhead wrist

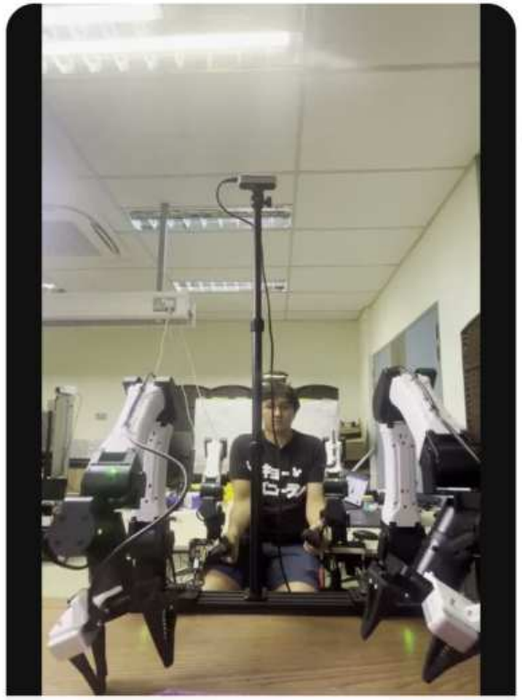
joints 0.12 -0.84 1.03 0.00 1.57 -0.21

gripper 0.80 → 0.35

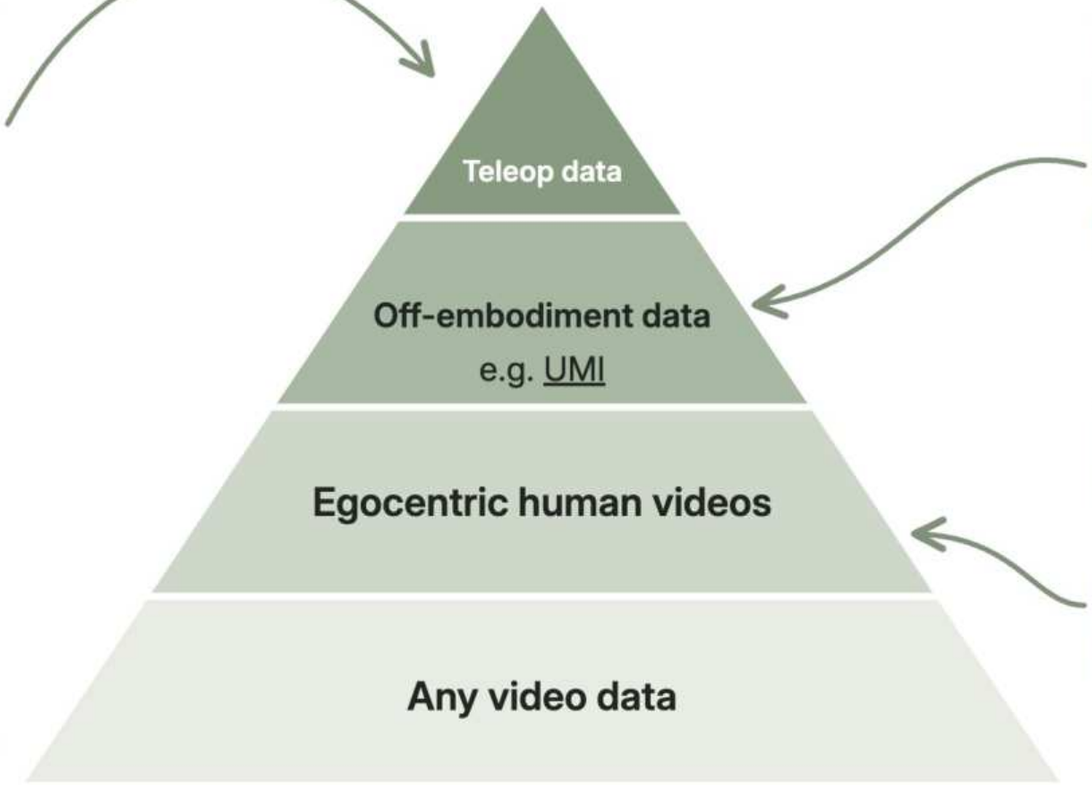
one timestep · many per second · on a real robot

But robotics doesn't have an internet of **robot-action data** to train on.

Robotics Data Pyramid



Teleop data

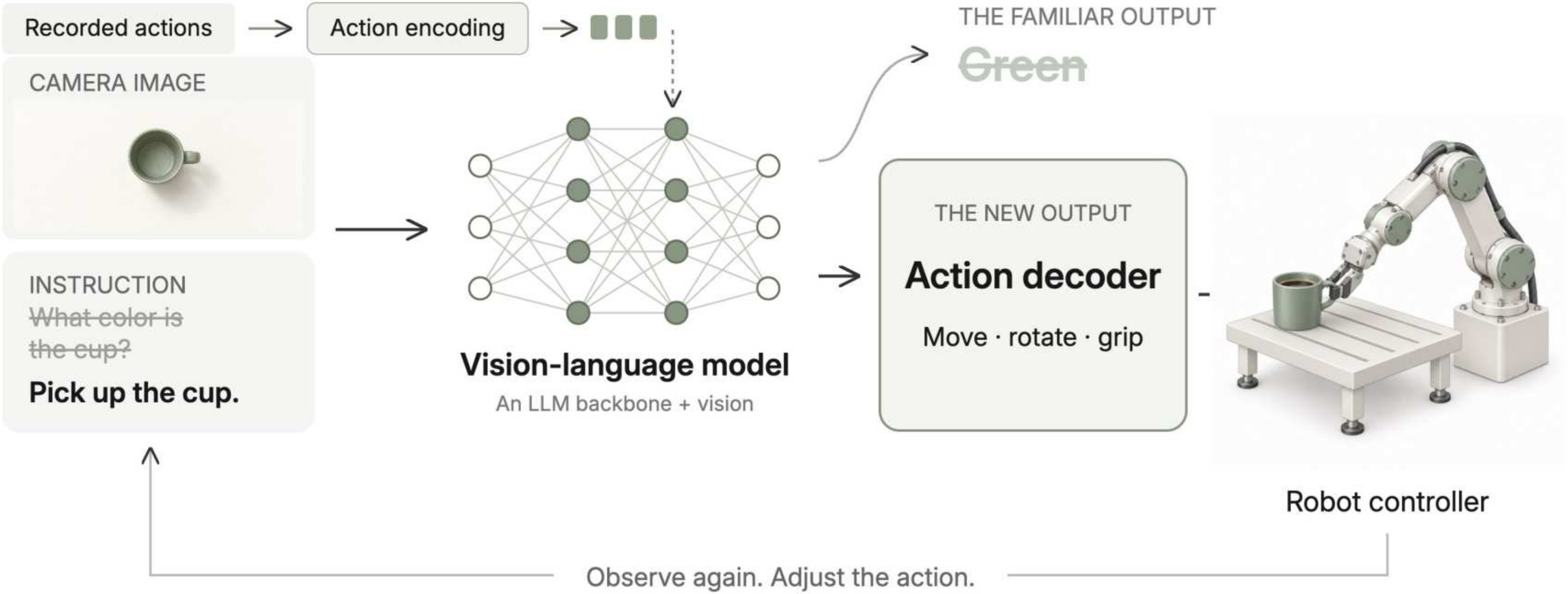


UMI (Universal Manipulation Interface) data

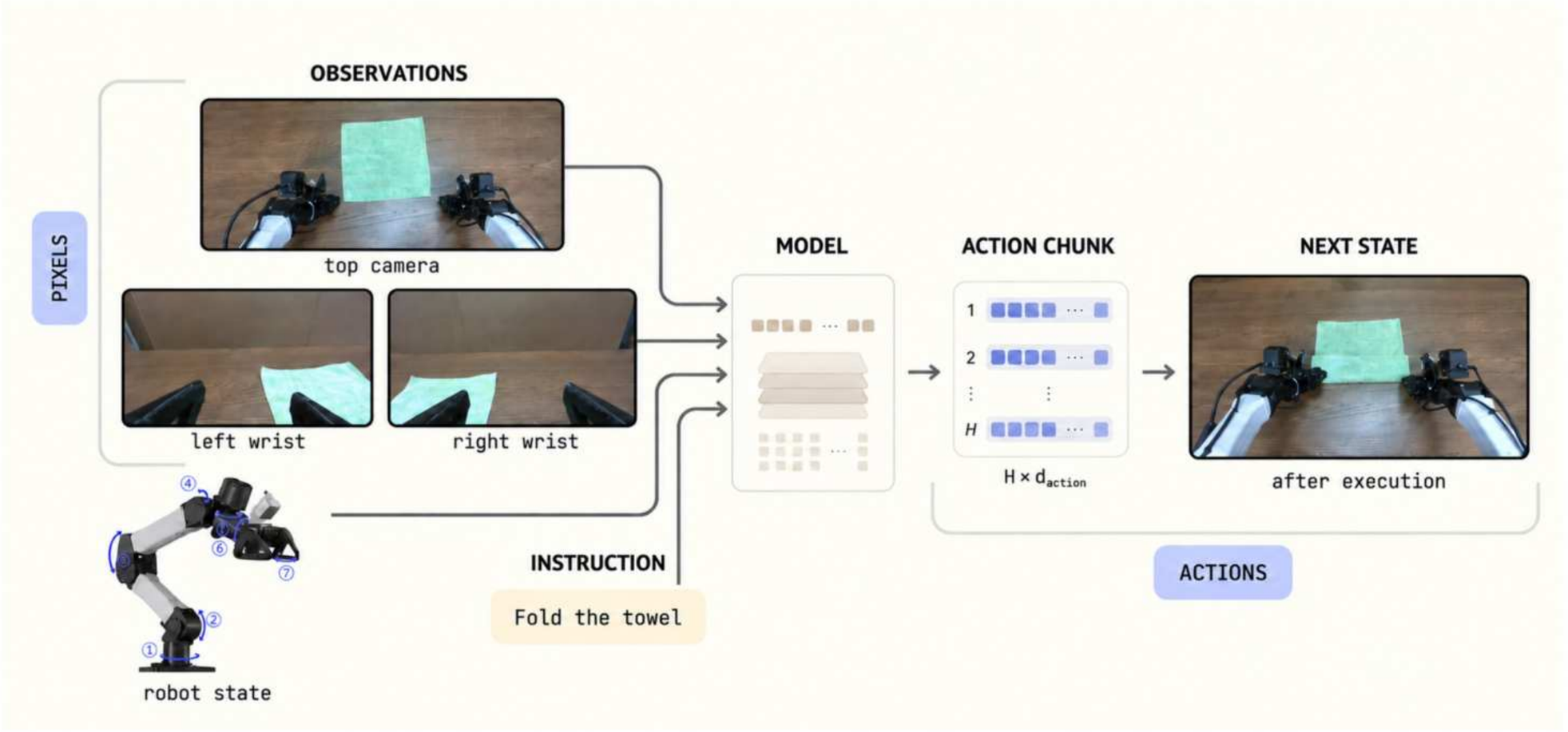
Egocentric human videos

UMI – https://www.youtube.com/watch?v=gXtlW1HIlTo

From understanding to doing



Vision language action models: from pixels to actions



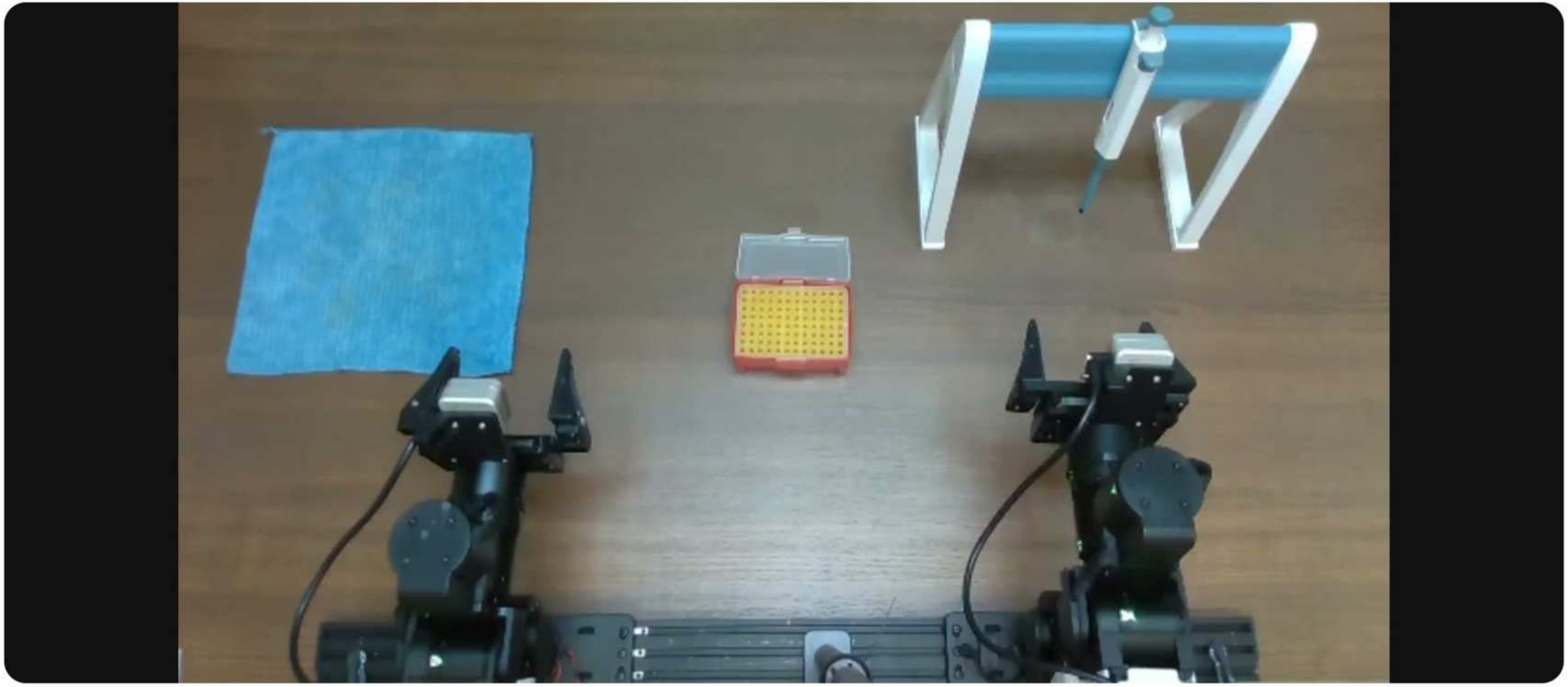
Examples OpenVLA · π_0 · RT-2

RT-2: Vision-Language-Action Models

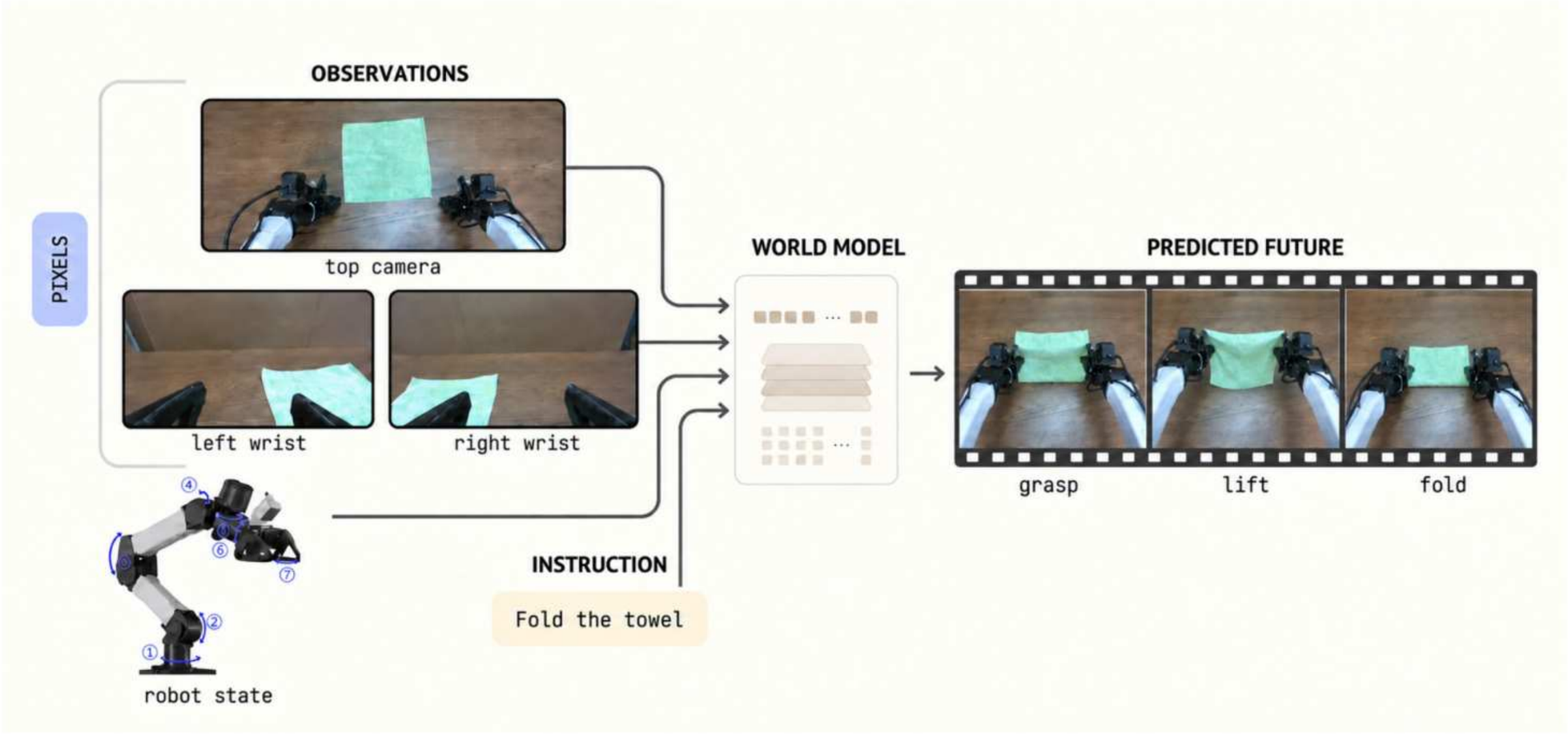
Transfer Web Knowledge to Robotic Control

https://deepmind.google/discover/blog/rt-2-new-model-translates-vision-and-language-into-action/

MolmoAct2 in action



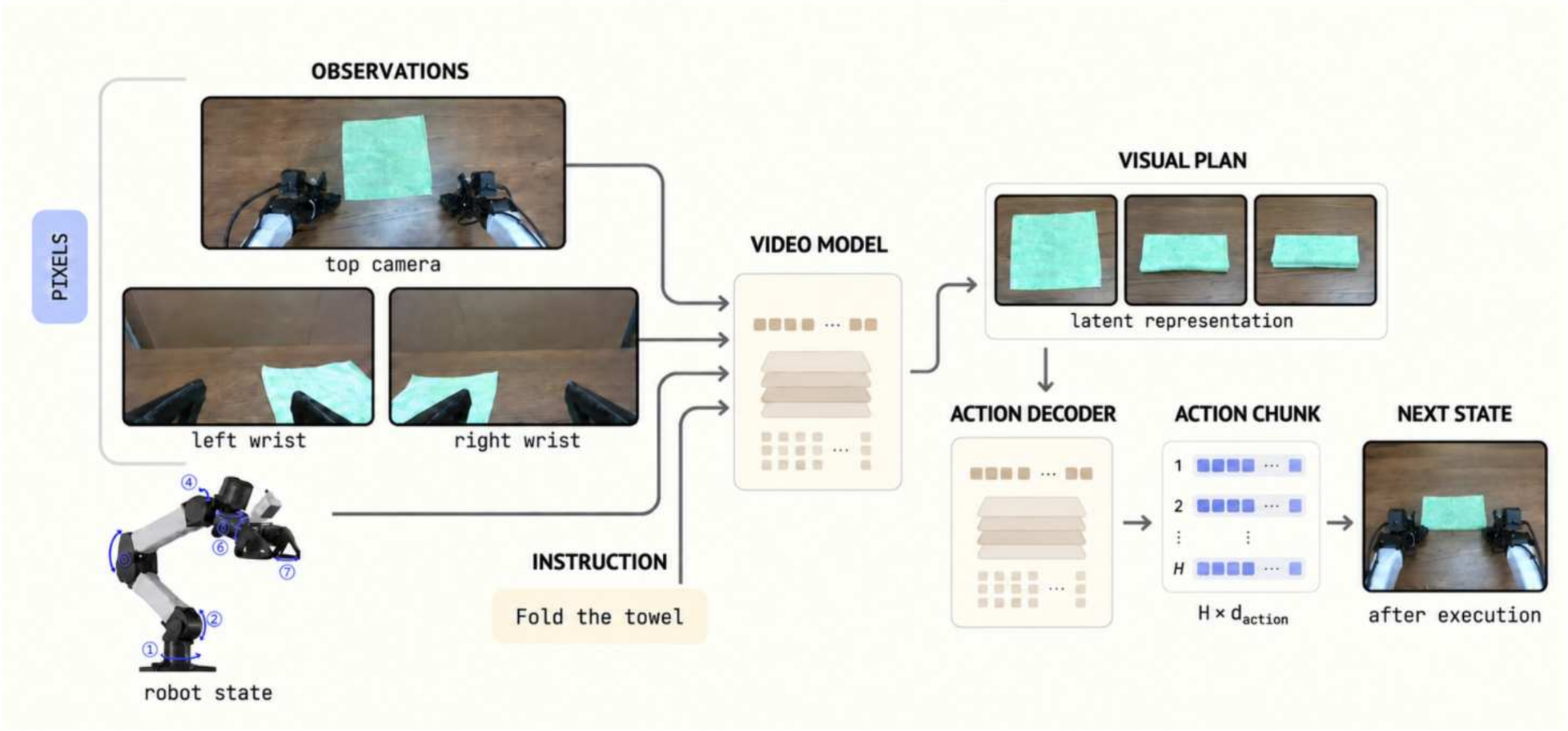
World models: predict what happens next



Predicting a future ≠ producing robot commands

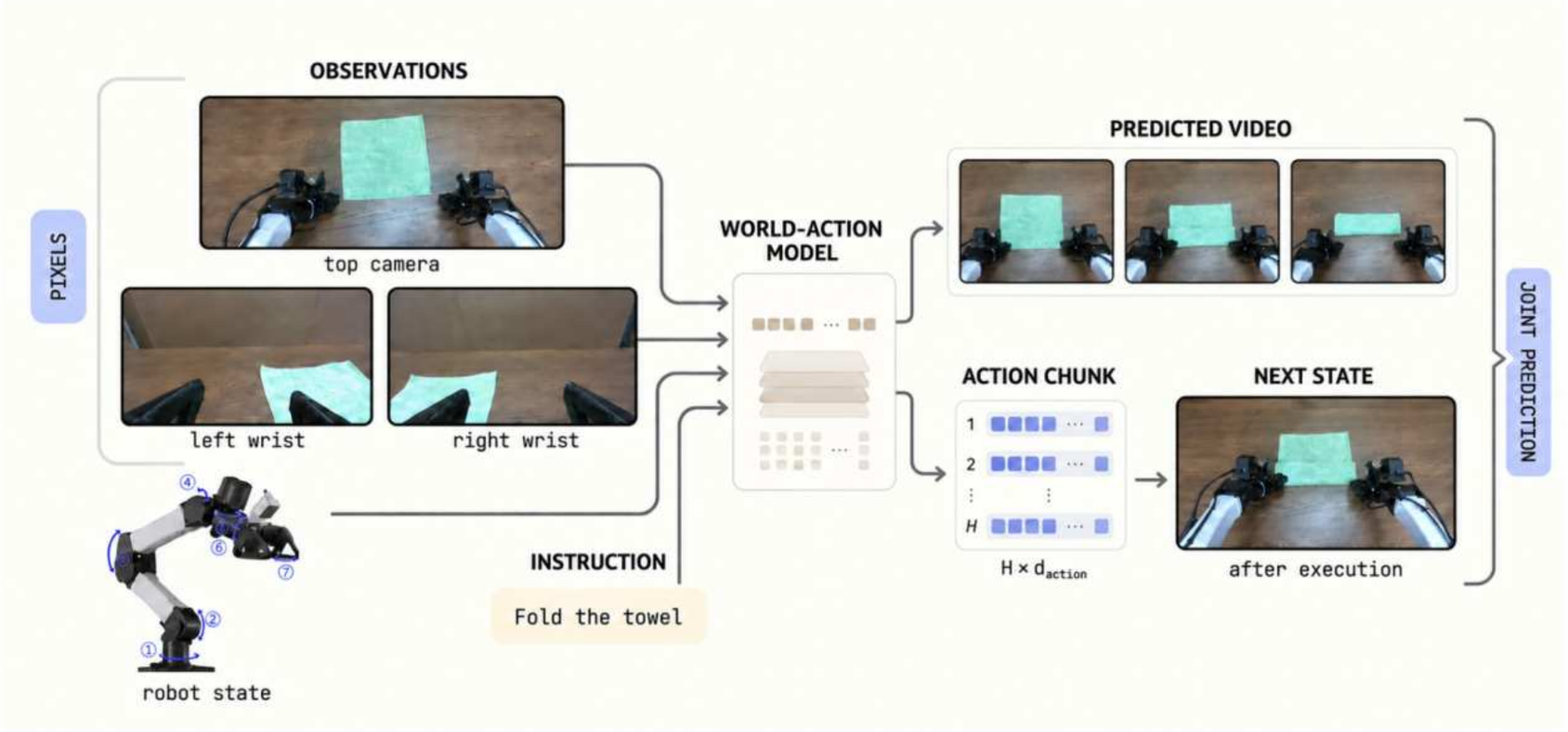
Examples Genie 2 · Cosmos Predict

Video-action models: turn plans into actions



Examples mimic-video

World-action models: predict both together

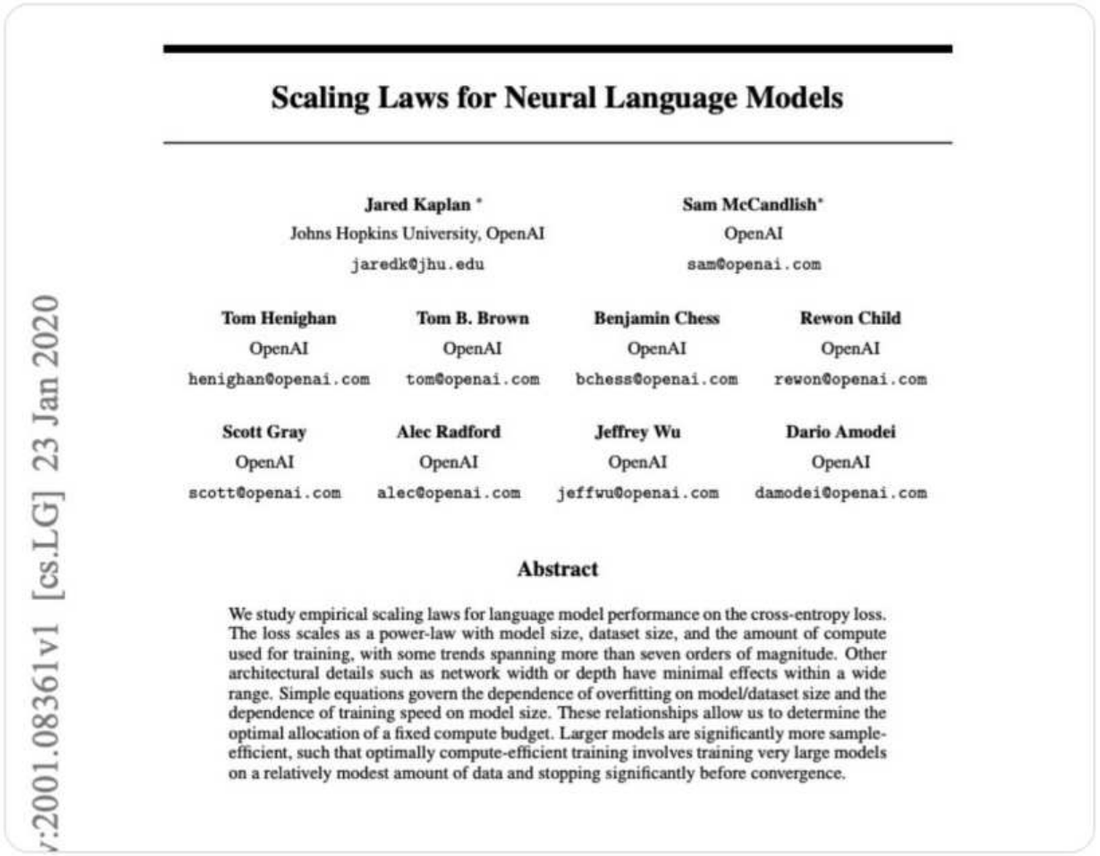


Recent work suggests these models handle new situations better than VLAs: "Do World Action Models Generalize Better Than VLAs?" (Zhang et al.)

Examples DreamZero · VideoVLA

Scaling laws

Language models · Kaplan et al., 2020



Scaling Laws for Neural Language Models

Jared Kaplan *
Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*
OpenAI
sam@openai.com

Tom Henighan
OpenAI
henighan@openai.com

Tom B. Brown
OpenAI
tom@openai.com

Benjamin Chess
OpenAI
bchess@openai.com

Rewon Child
OpenAI
rewon@openai.com

Scott Gray
OpenAI
scott@openai.com

Alec Radford
OpenAI
alec@openai.com

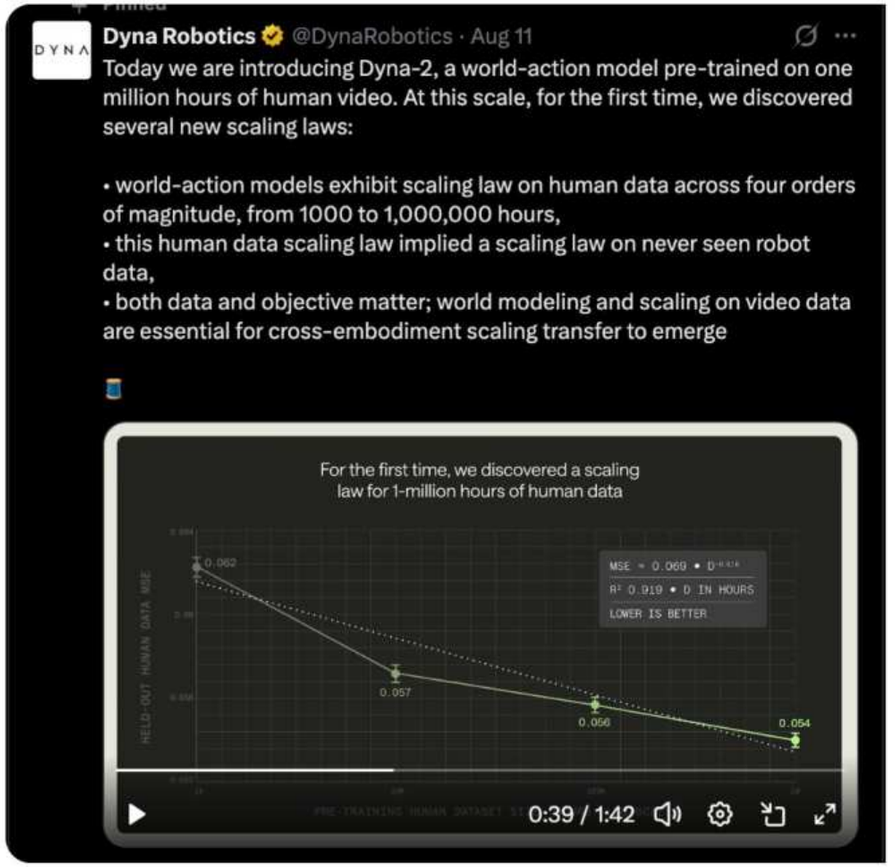
Jeffrey Wu
OpenAI
jeffwu@openai.com

Dario Amodei
OpenAI
damodei@openai.com

Abstract

We study empirical scaling laws for language model performance on the cross-entropy loss. The loss scales as a power-law with model size, dataset size, and the amount of compute used for training, with some trends spanning more than seven orders of magnitude. Other architectural details such as network width or depth have minimal effects within a wide range. Simple equations govern the dependence of overfitting on model/dataset size and the dependence of training speed on model size. These relationships allow us to determine the optimal allocation of a fixed compute budget. Larger models are significantly more sample-efficient, such that optimally compute-efficient training involves training very large models on a relatively modest amount of data and stopping significantly before convergence.

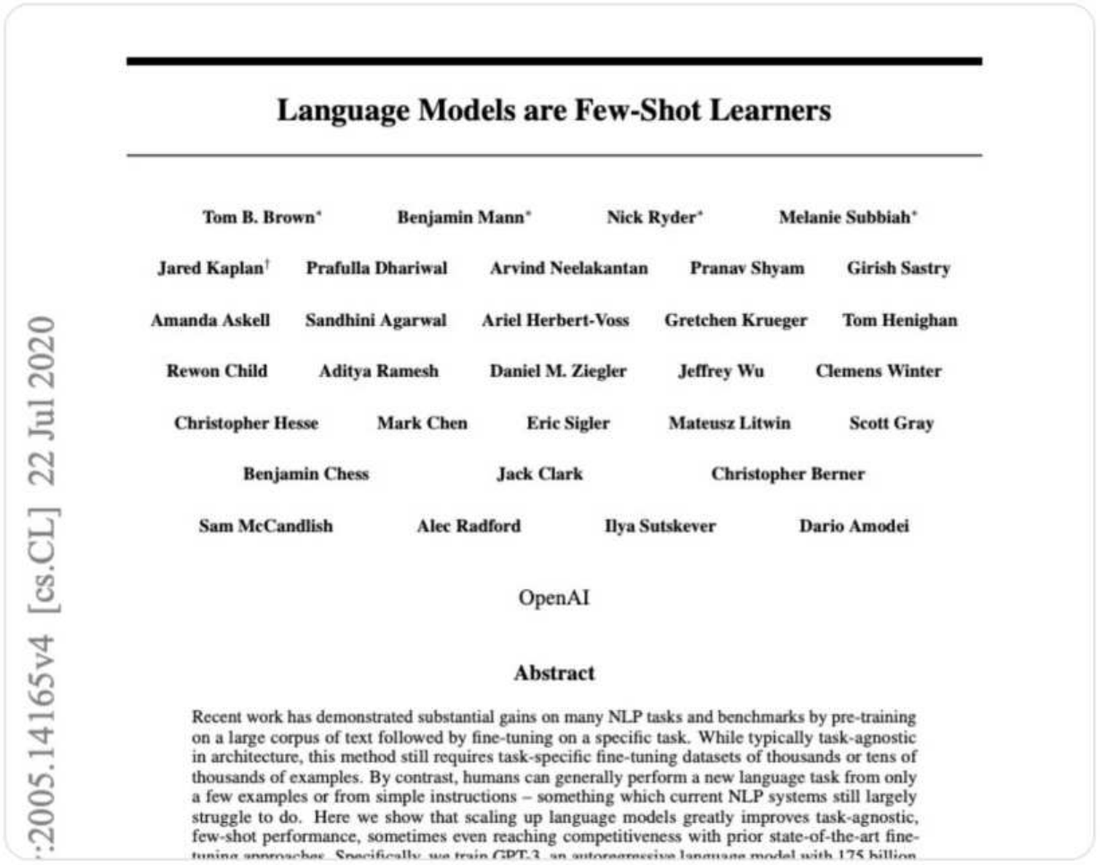
Robotics · Dyna-2, 2026



https://arxiv.org/abs/2001.08361 · https://www.dyna.co/dyna-2

In-context learning

Language models · Brown et al., 2020



arXiv:2005.14165v4 [cs.CL] 22 Jul 2020

Language Models are Few-Shot Learners

Tom B. Brown* Benjamin Mann* Nick Ryder* Melanie Subbiah*

Jared Kaplan† Prafulla Dhariwal Arvind Neelakantan Pranav Shyam Girish Sastry

Amanda Askell Sandhini Agarwal Ariel Herbert-Voss Gretchen Krueger Tom Henighan

Rewon Child Aditya Ramesh Daniel M. Ziegler Jeffrey Wu Clemens Winter

Christopher Hesse Mark Chen Eric Sigler Mateusz Litwin Scott Gray

Benjamin Chess Jack Clark Christopher Berner

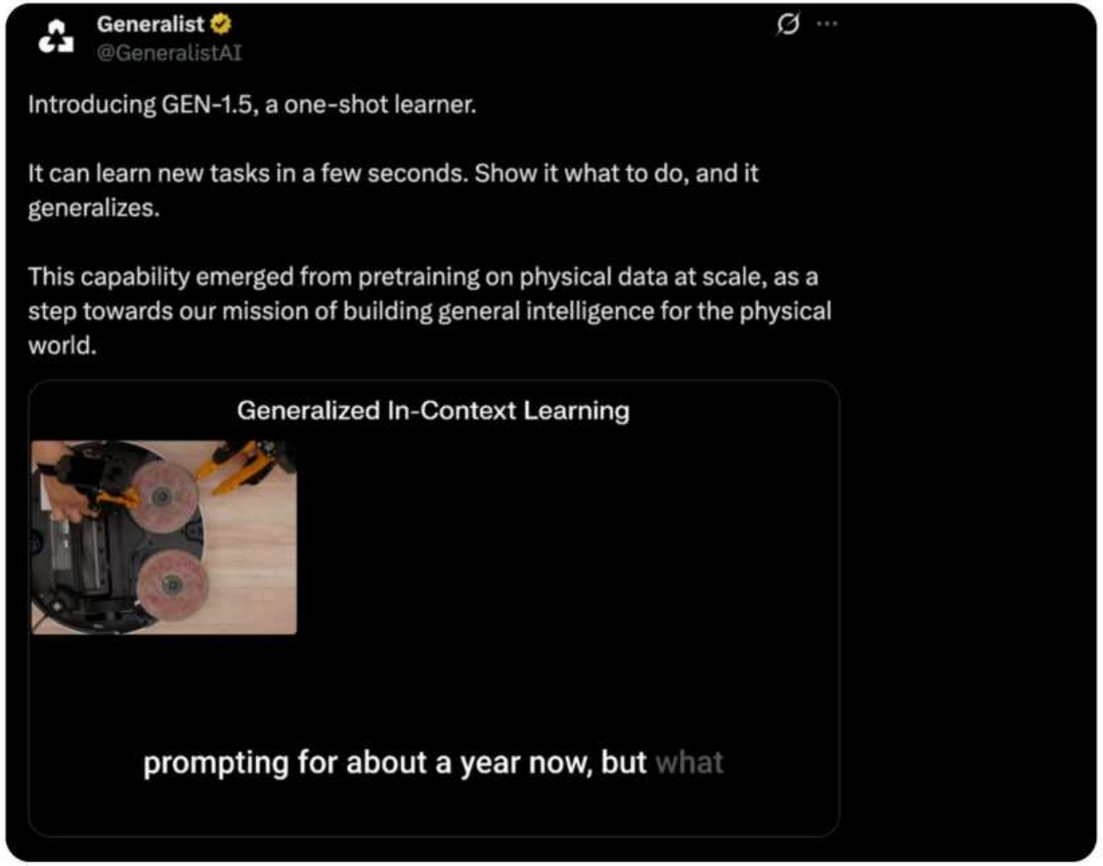
Sam McCandlish Alec Radford Ilya Sutskever Dario Amodei

OpenAI

Abstract

Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. By contrast, humans can generally perform a new language task from only a few examples or from simple instructions – something which current NLP systems still largely struggle to do. Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion

Robotics · Generalist GEN-1.5, 2026



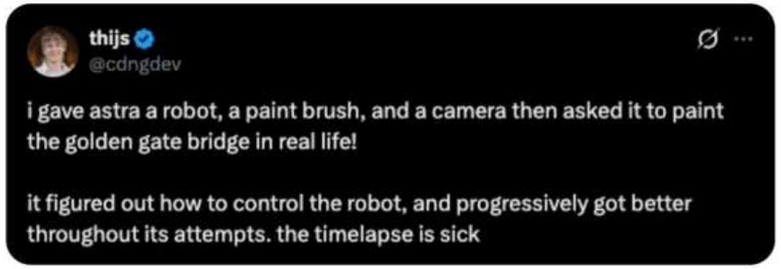
https://arxiv.org/abs/2005.14165 · https://generalistai.com/blog/gen-1.5

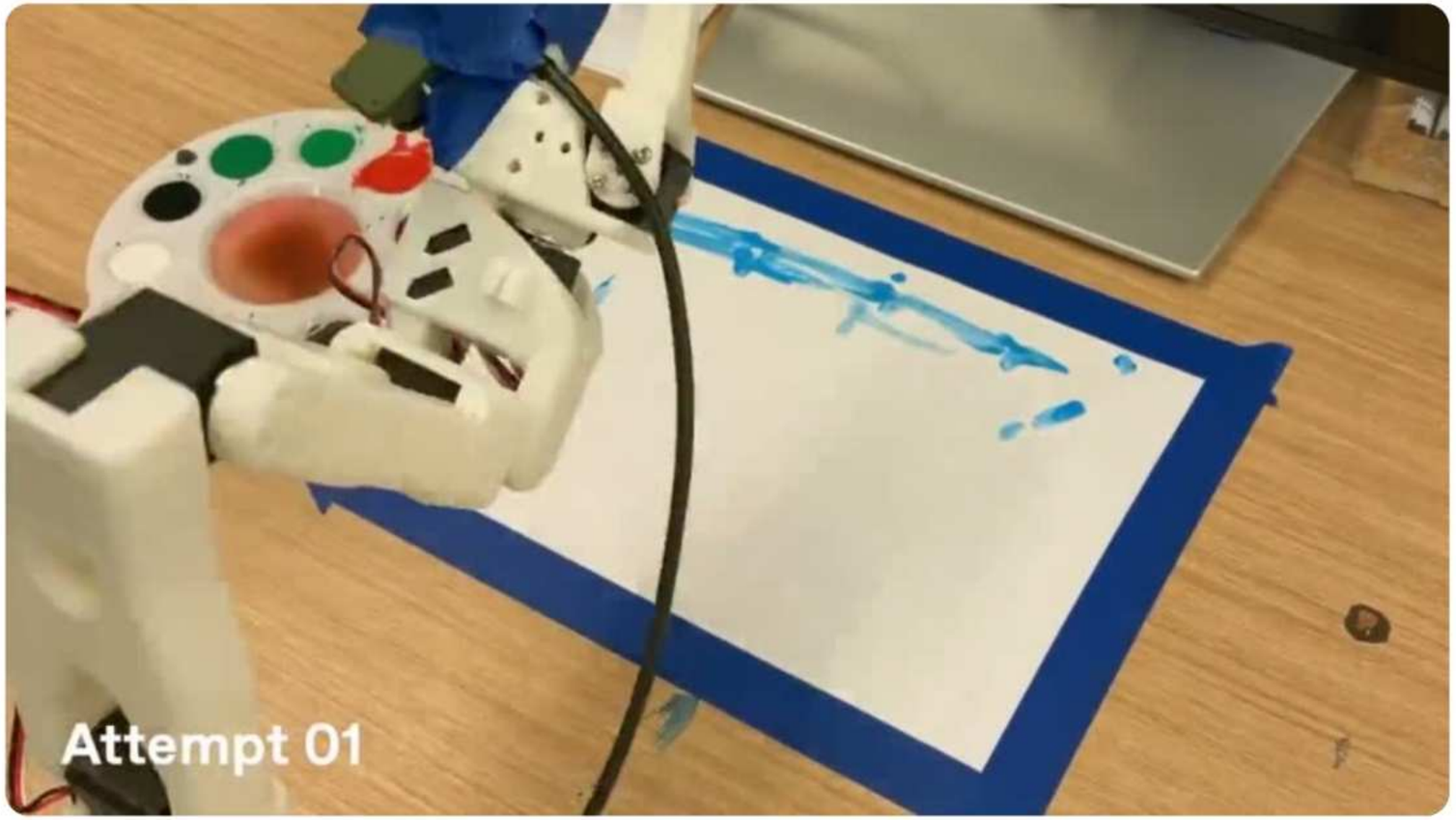
Intelligent robots that can do specific tasks 99.1% of the time are here!

Folding and stacking clothes at home

https://www.sunday.ai/blog/act-2-preview

Now gpt-6-astra can paint using a robot arm!





https://x.com/cdngdev/status/2097339677128982873

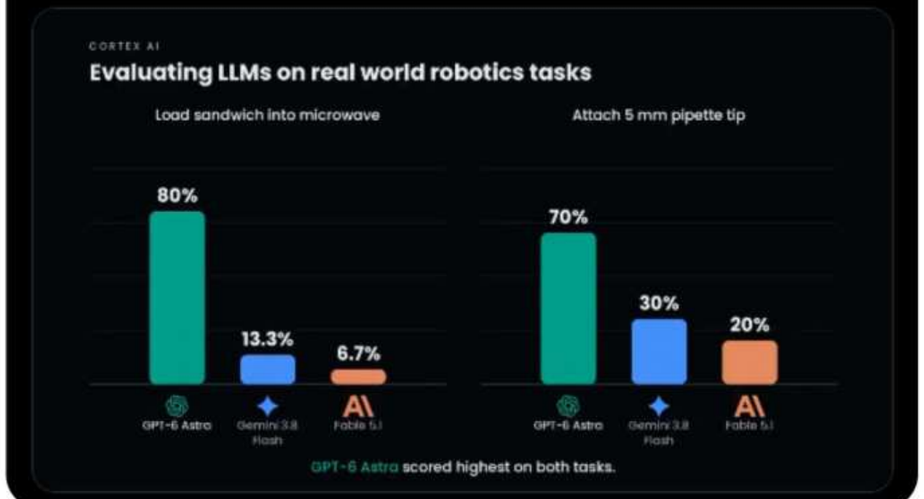
We tested frontier AI models on real-world robotics tasks

CortexAI
@cortexairobot

We tested GPT-6 Astra, Gemini 3.8 Flash, and Fable 5.1 on two real-world robot tasks from our MolmoAct2 real-world evaluations:
1. Fitting a 5 mm pipette tip
2. Loading a microwave

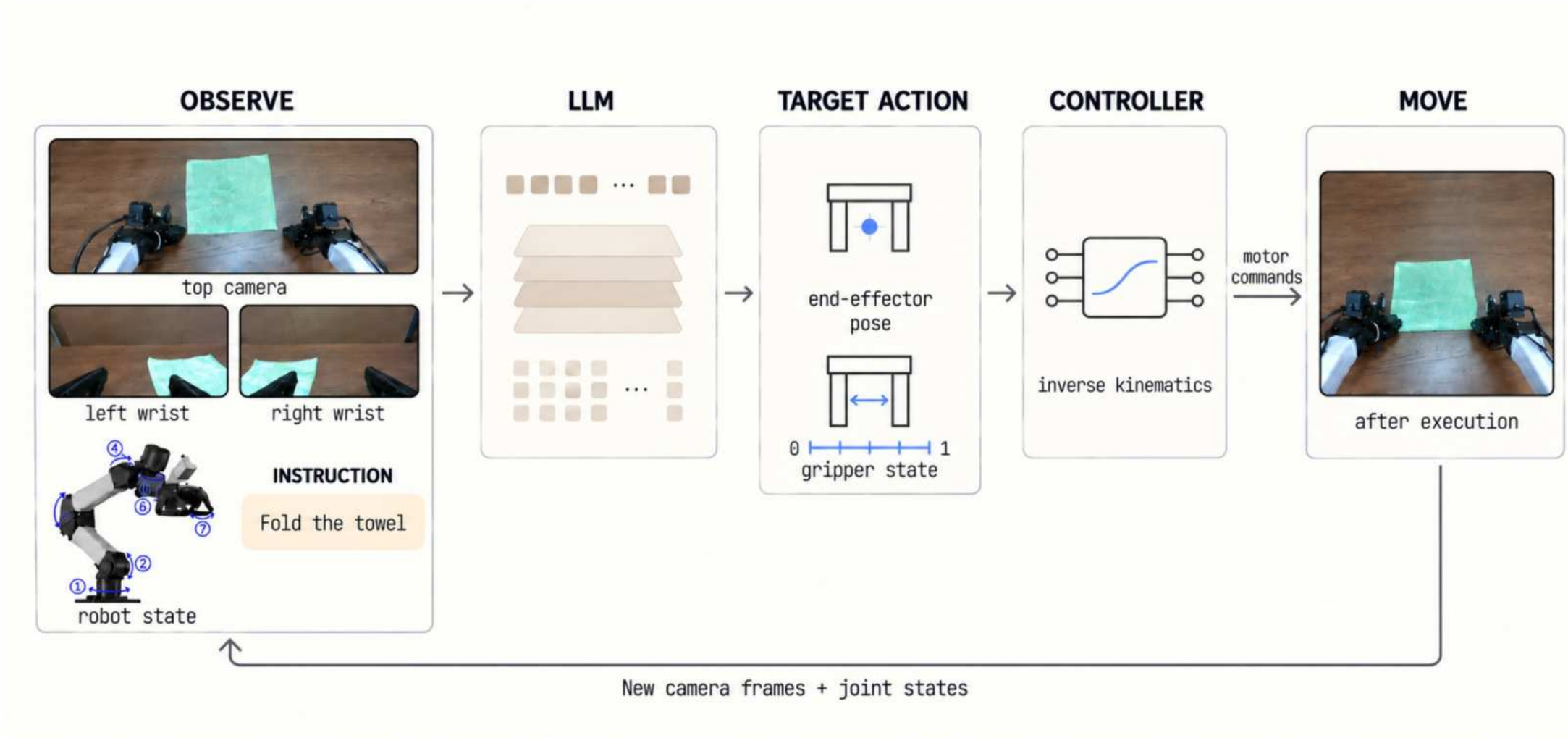
Astra scored highest on both. 🥇

Here are some few surprising things that Astra did to come out on top. 🧵

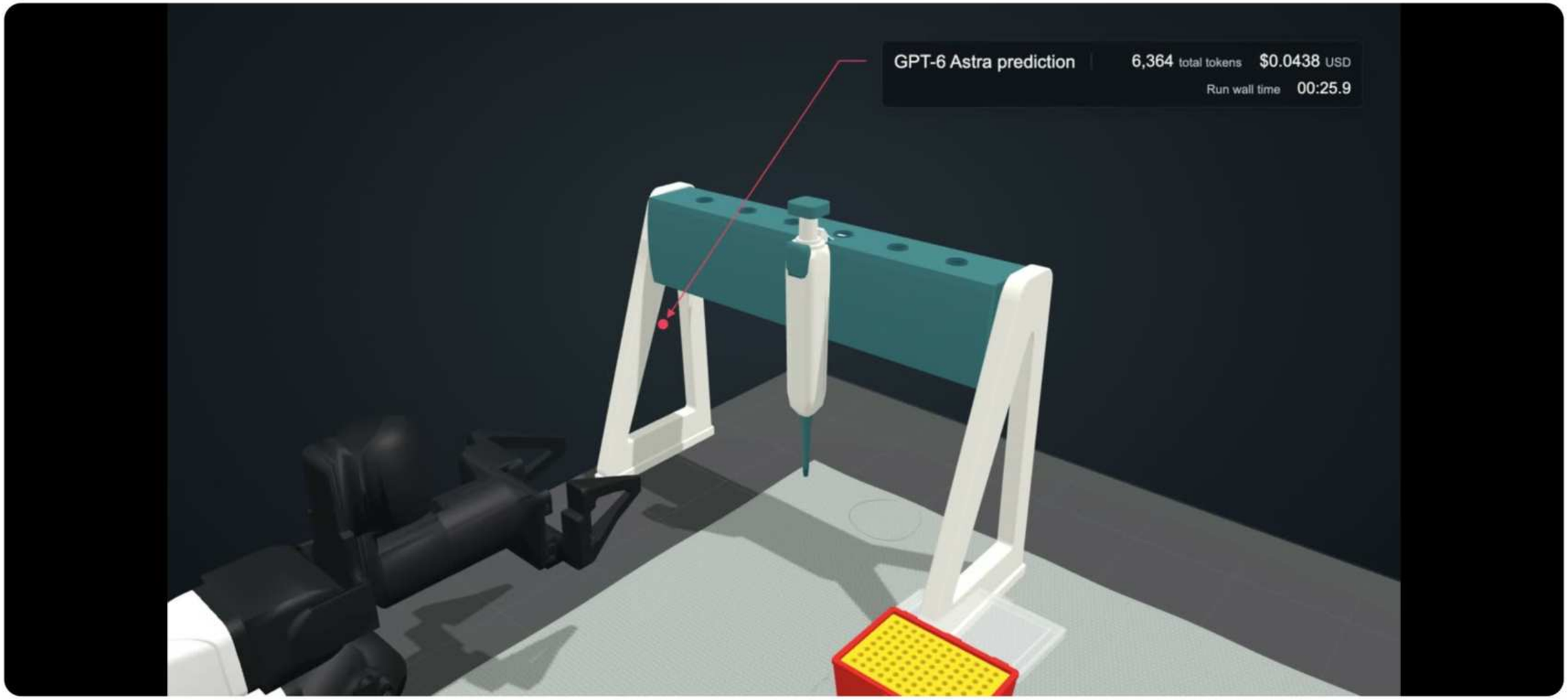


https://x.com/cortexairobot/status/2100628691114201489

How do LLMs control robots?



The control loop in action: attaching a pipette tip



LLMs' reasoning skills shine through.

unprecedented
behavior
from VLA models.

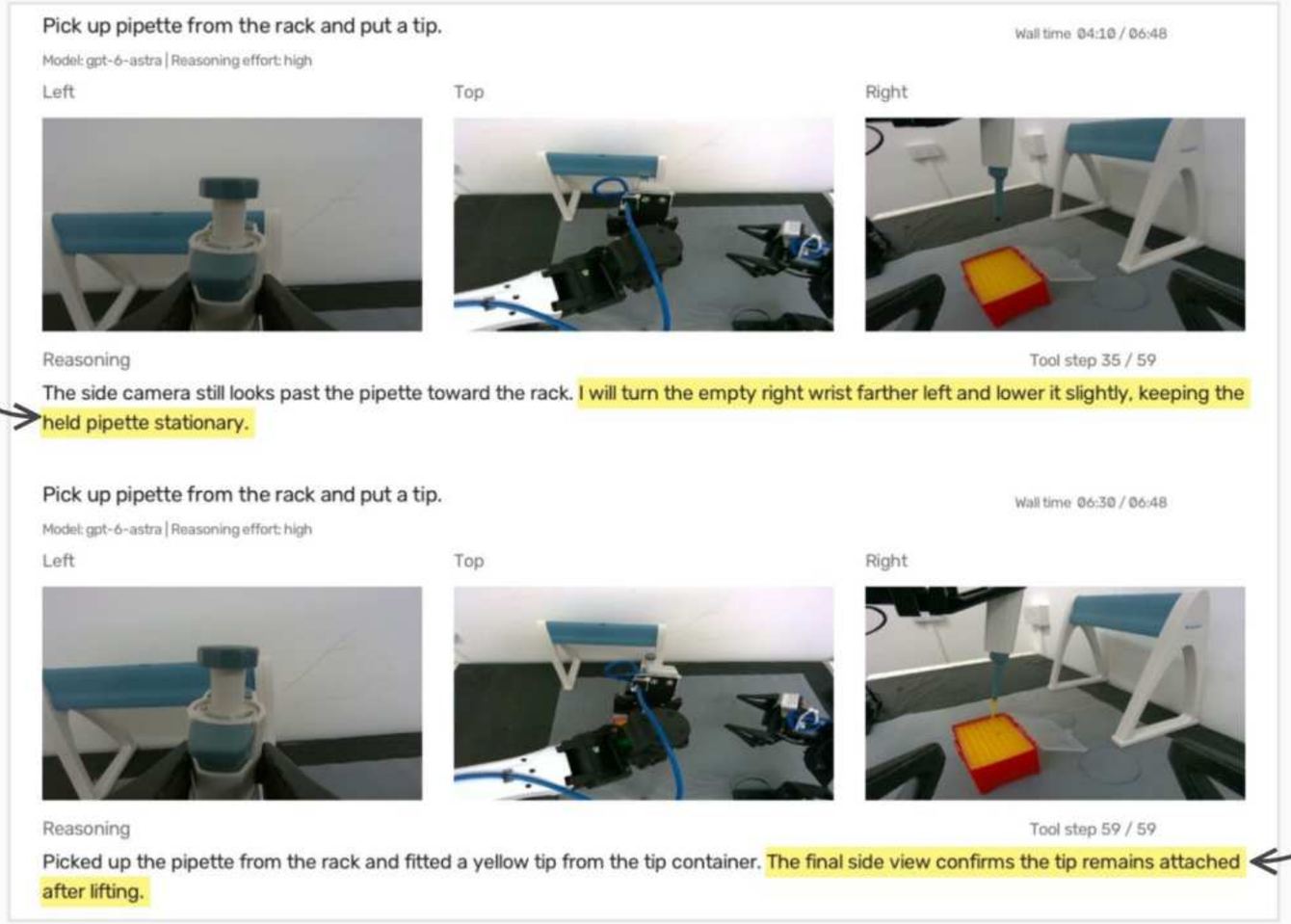


Figure 5: Astra uses its free arm's camera for a better view.

acts as a built-in
reward model.

https://cortexrobot.ai/research/llm-real-world-robotics-evals/

Intelligent robots will continue to get more intelligent.

01 Robot bodies are ready.

Atlas · Wuji Hand · Unitree

02 Robotics is following the LLM playbook.

Scaling laws · in-context learning

03 Frontier LLMs can already reason about and control real robots.

GPT-6 Astra

FOOD FOR THOUGHT

GPT-6 Astra: the GPT moment for robotics?

My picture of AGI in robotics

On a personal note, my picture of AGI is a physical robot that

sits next to me and my friends, ← SAFE HUMAN-ROBOT INTERACTION

walks around the room, ← MOBILITY

deals the cards, ← DEXTEROUS MANIPULATION

reads people's faces and what they say, ← MULTIMODAL PERCEPTION

and then wins a game of Avalon (a social deduction game) by

playing Mordred (one of the villains) flawlessly. ← MEMORY, REASONING, THEORY OF MIND

Thank you

@SuveenE
suveenellawela.com