

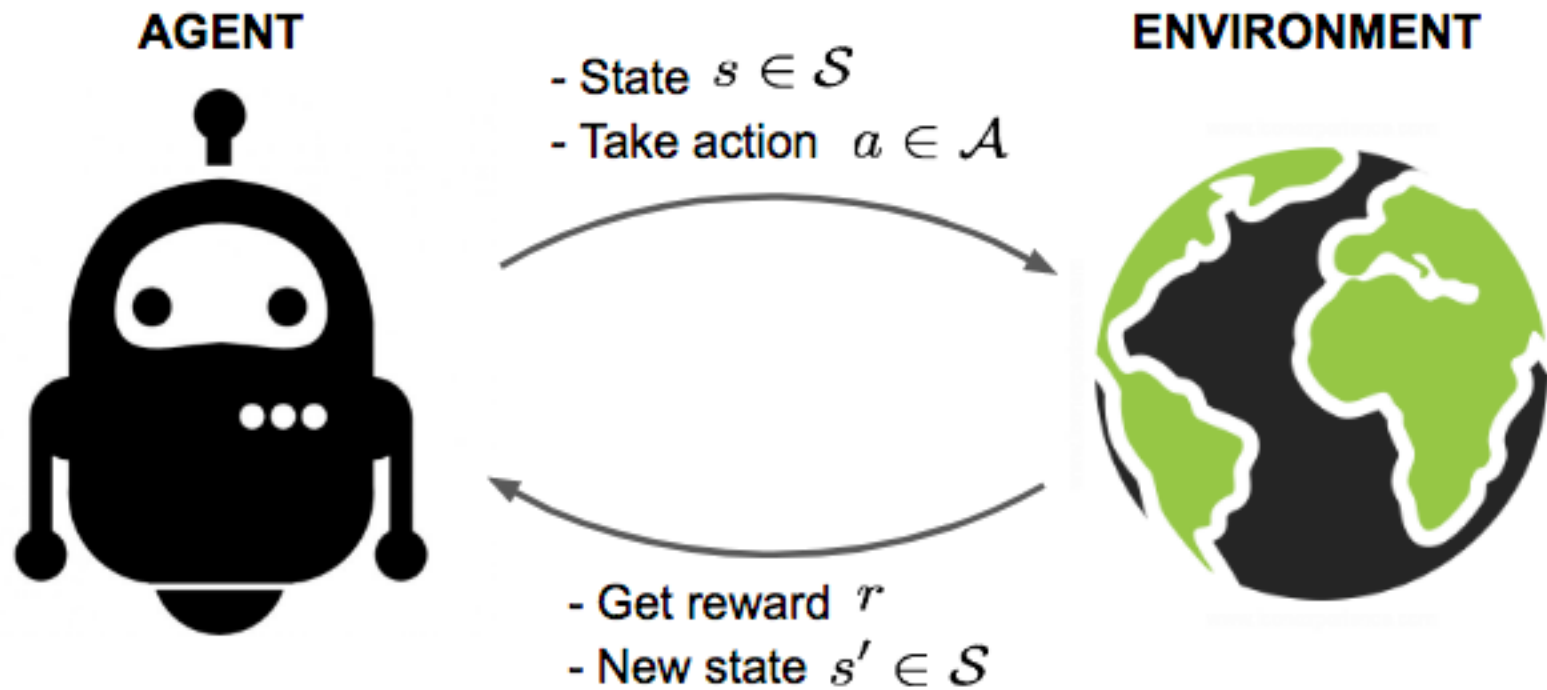
Deep Learning and Robotics

William Chen

Data C182 Guest Lecture

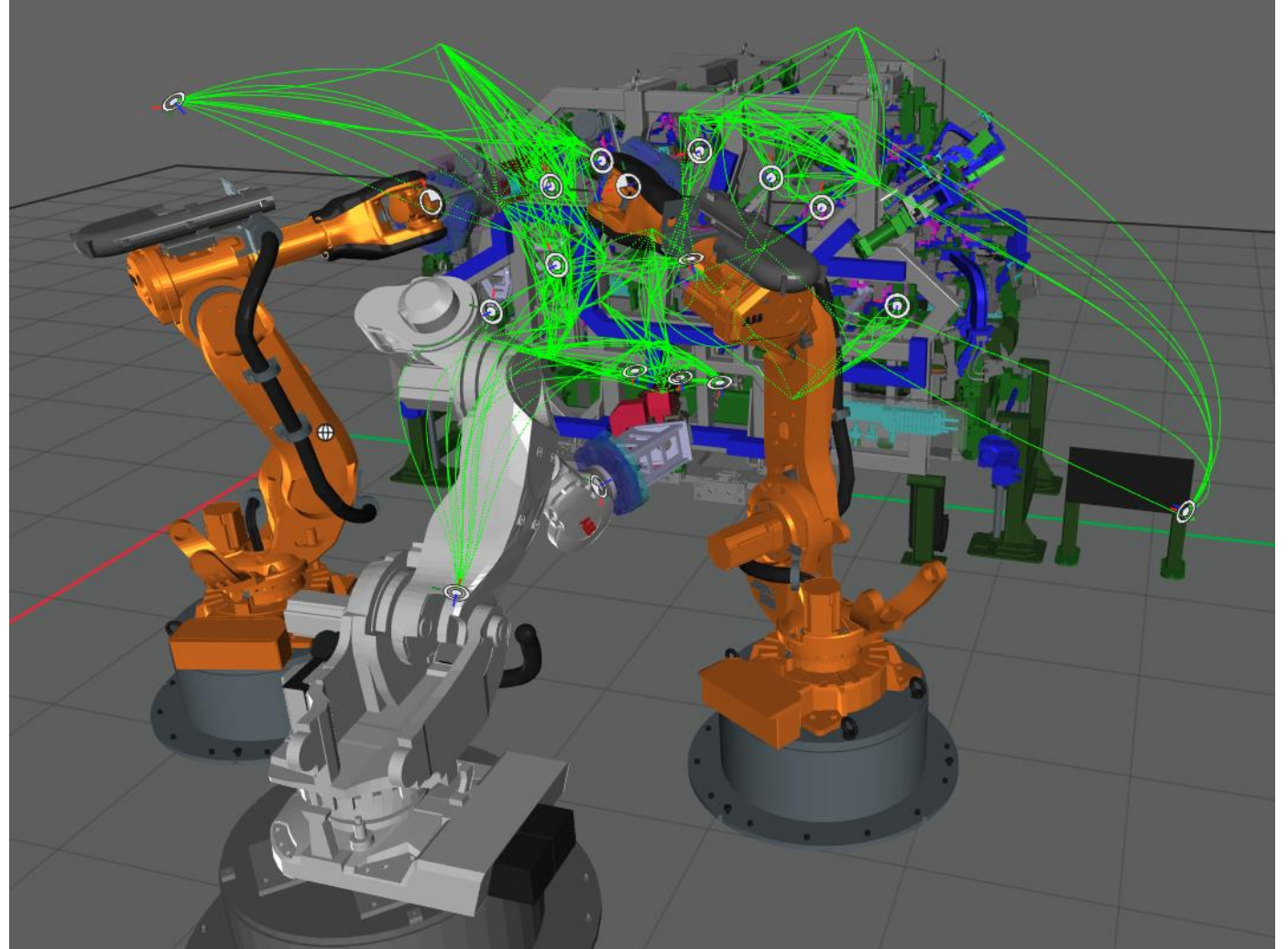
The Robotics Problem Statement

Given current and past sensor observations of the world, produce and take actions to affect the world in some way. Repeat.



Classical Robotics

- In classical robotics: Hand design an algorithm that models the world and picks actions
- Typically formulated as optimization
- Requires human ingenuity
→ Quite difficult to do!

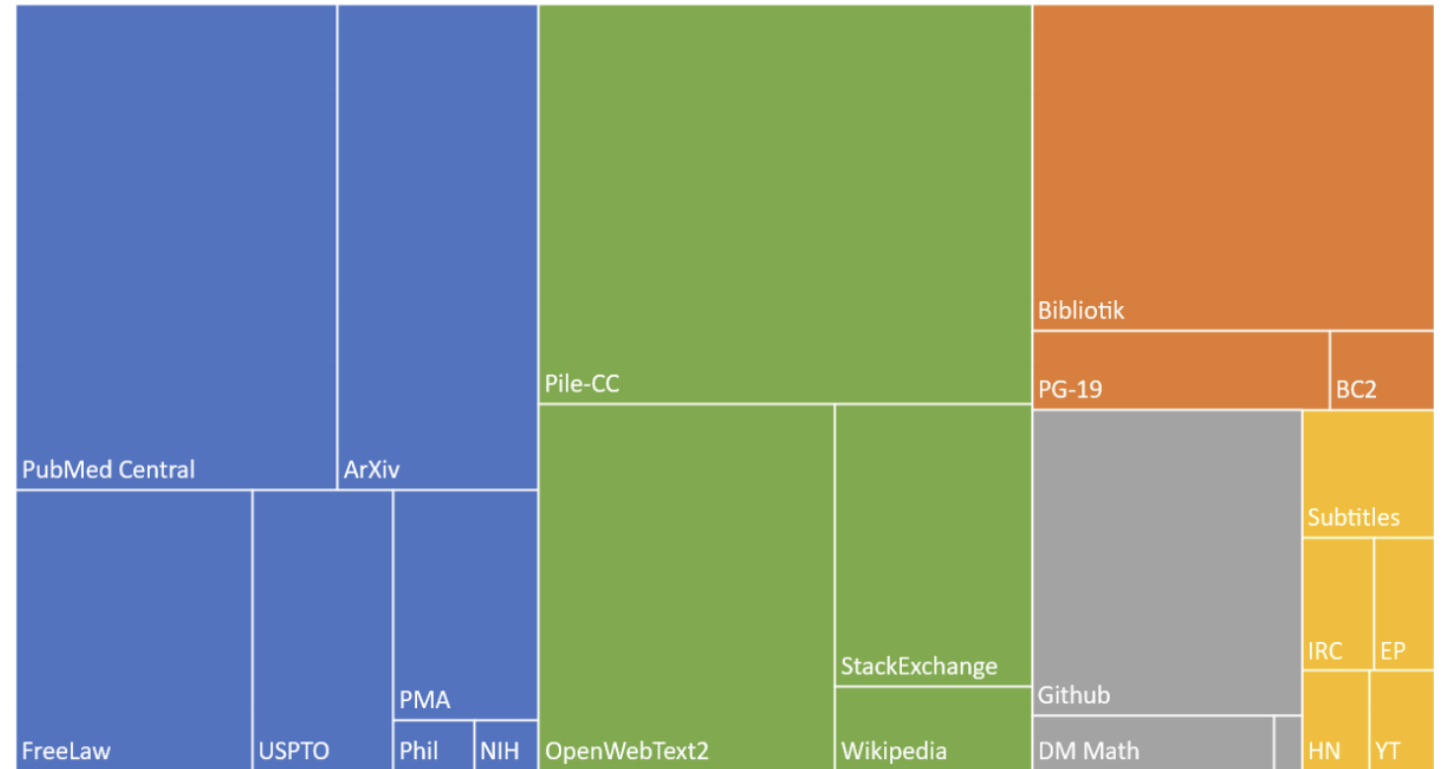


Successes of Deep Learning



Composition of the Pile by Category

■ Academic ■ Internet ■ Prose ■ Dialogue ■ Misc



Deep learning extracts complex patterns from (lots of) data.

Imitation Learning

Learn robot behaviors from demonstrations.

Formulate as supervised learning: given observation (or sequence of observations), learn a model that maps it to robot actions.

$$a = \pi_{\theta}(s); \quad \text{minimize}_{\theta} \mathcal{L}(a, a^*)$$

Imitation Learning

QT-Opt
pick anything

TOTO
pour

sweep the green cloth to the left side of the table

Push T

stack cups

place the black bowl in the dish rack

pick red block

Jaco Play **ALOHA** **Taco Play**

1M Episodes from **311 Scenes**
34 Research Labs across **21 Institutions**

22 Embodiments

527 Skills

pour stack route

60 Datasets

1,798 Attributes • **5,228 Objects** • **23,486 Spatial Relations**

Cable Routing

pick green chip bag from counter

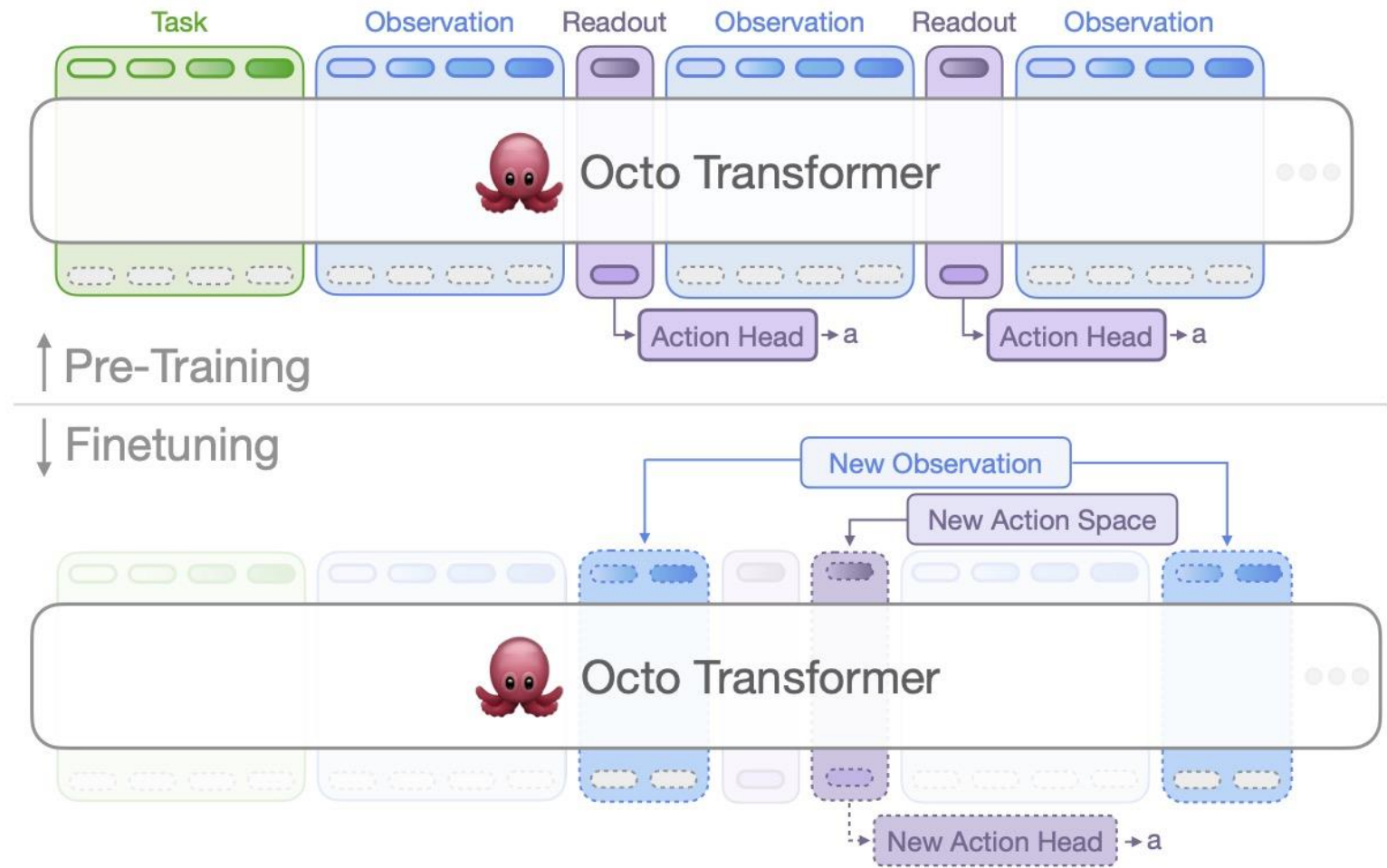
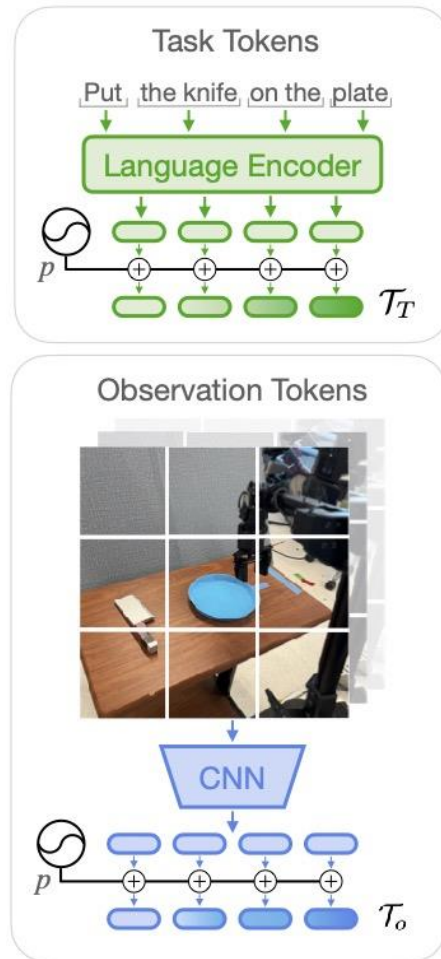
RT-1

set the bowl to the right side of the table

Bridge

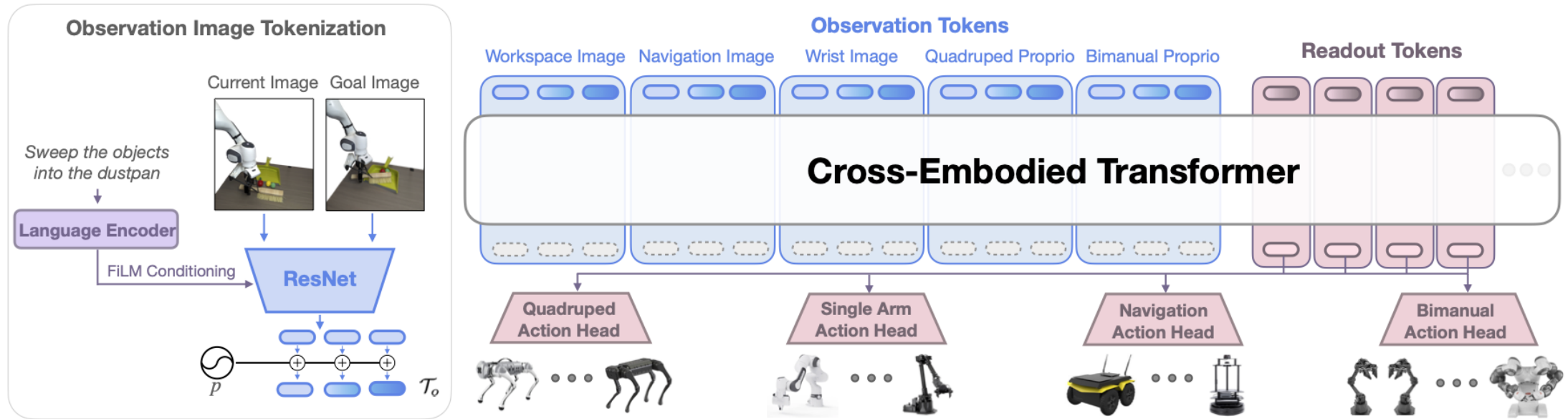
Door Opening

Imitation Learning



Train a big model to map from instructions and observations to actions on demo data from various robot arms

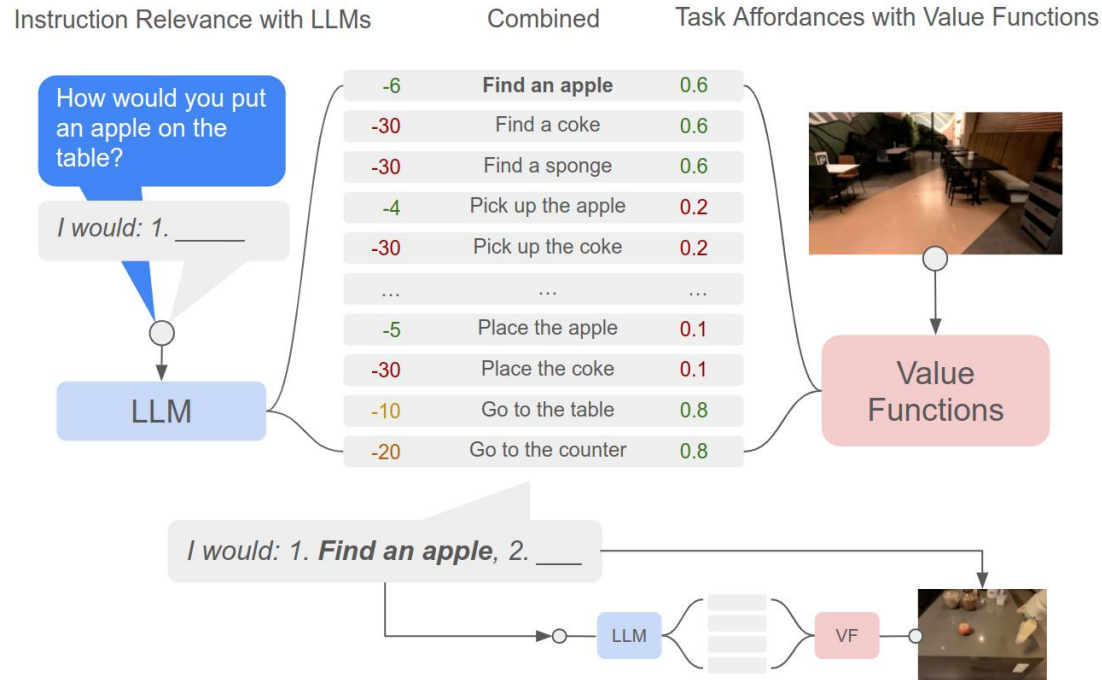
Imitation Learning



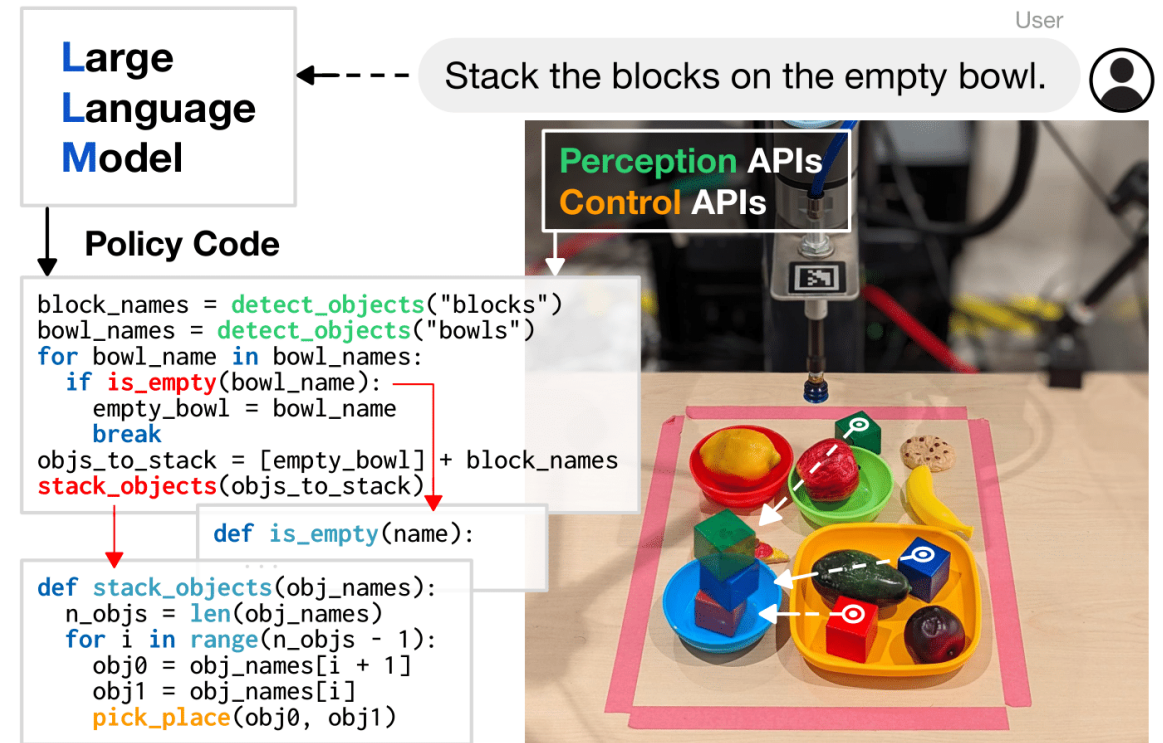
Train a big model to map observations to actions on data from a wide variety of robots (single arms, double arms, wheeled robots, quadrupeds, ...)

→ Get a model that can control all sorts of robots!

Using Large Pretrained Models

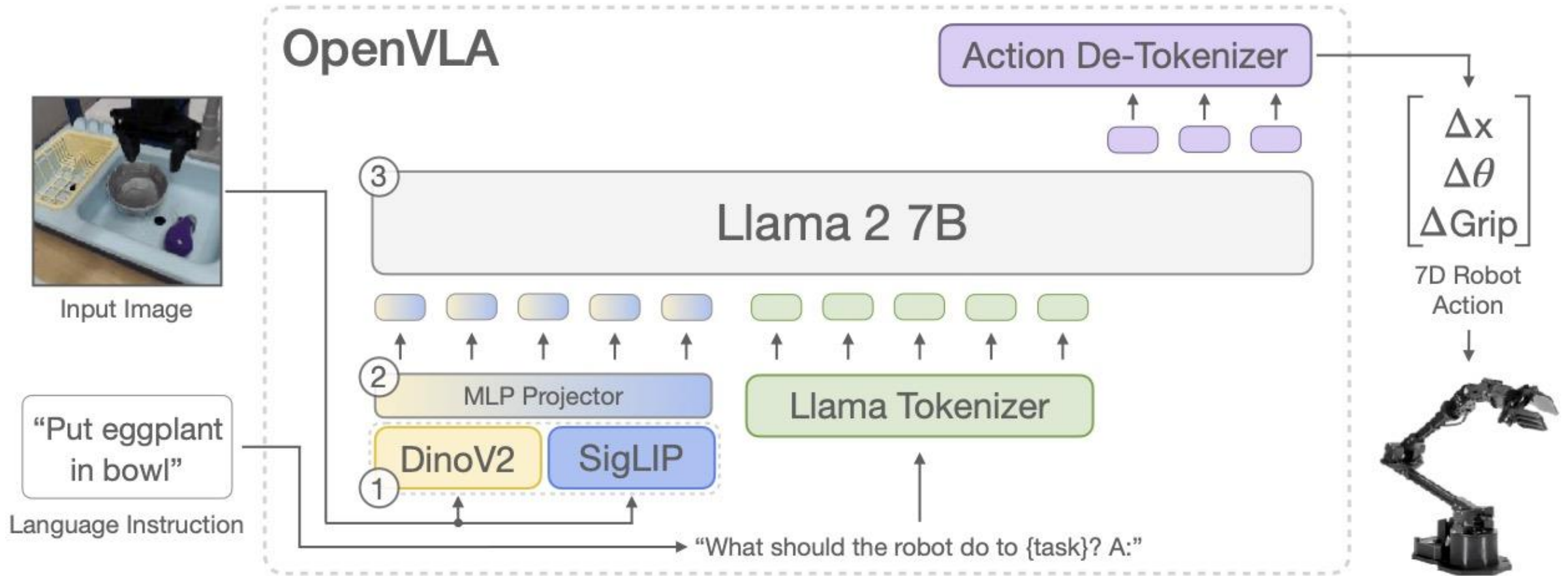


Use language models to pick reasonable subtasks



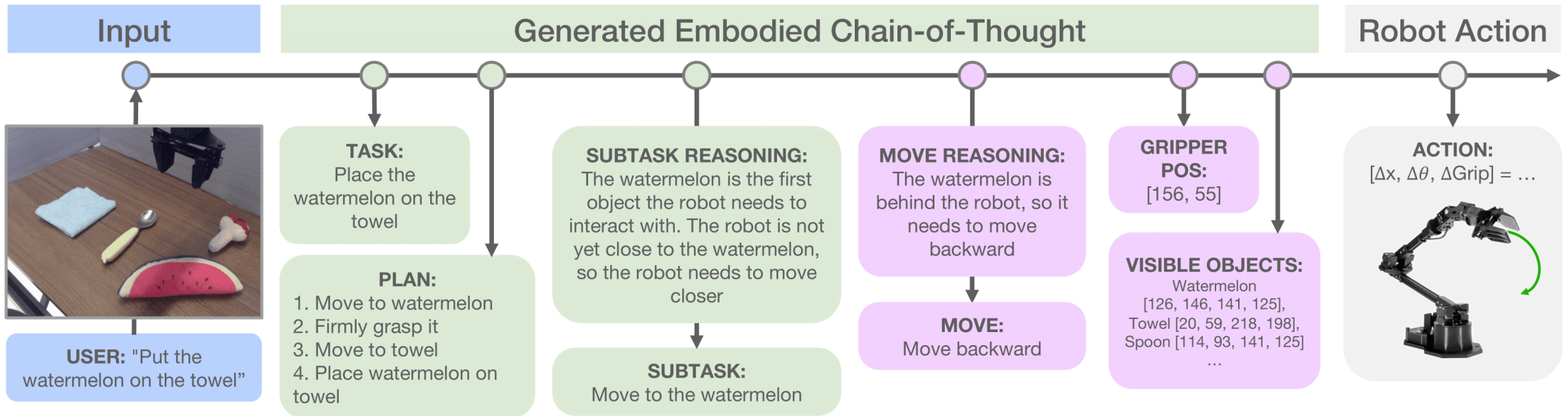
Use language models to write code that a robot can execute

Vision-Language-Action Models



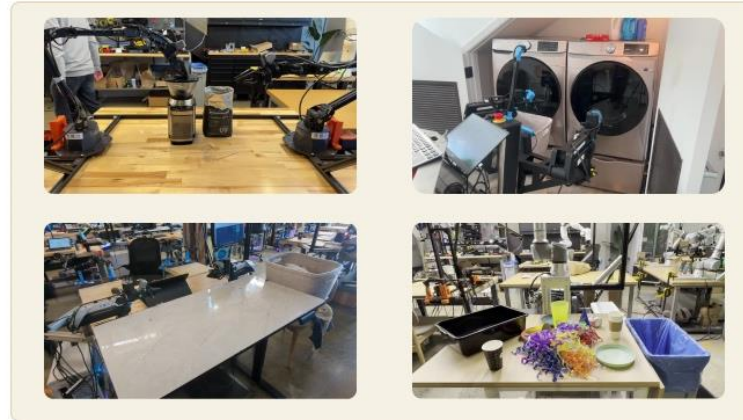
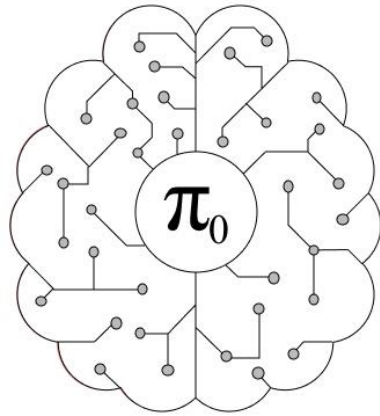
Start with a pretrained vision-language model and fine-tune it to produce robot actions in response to an instruction and observation with imitation learning

Robot Reasoning



Do the same thing as before, but generate *reasoning* before picking the robot action

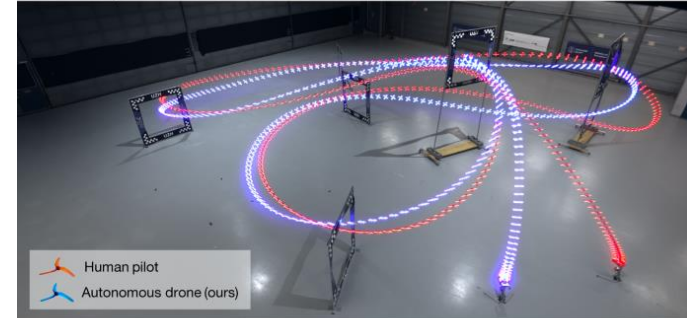
Where Do We Go From Here?



Further scaling up imitation learning

Or something else?

a Drone racing: human versus autonomous



b Head-to-head competition



c Human champions



Reinforcement learning