Foundation Models for Robotics

Dorsa Sadigh

Stanford University

iliad

intelligent and interactive autonomous systems

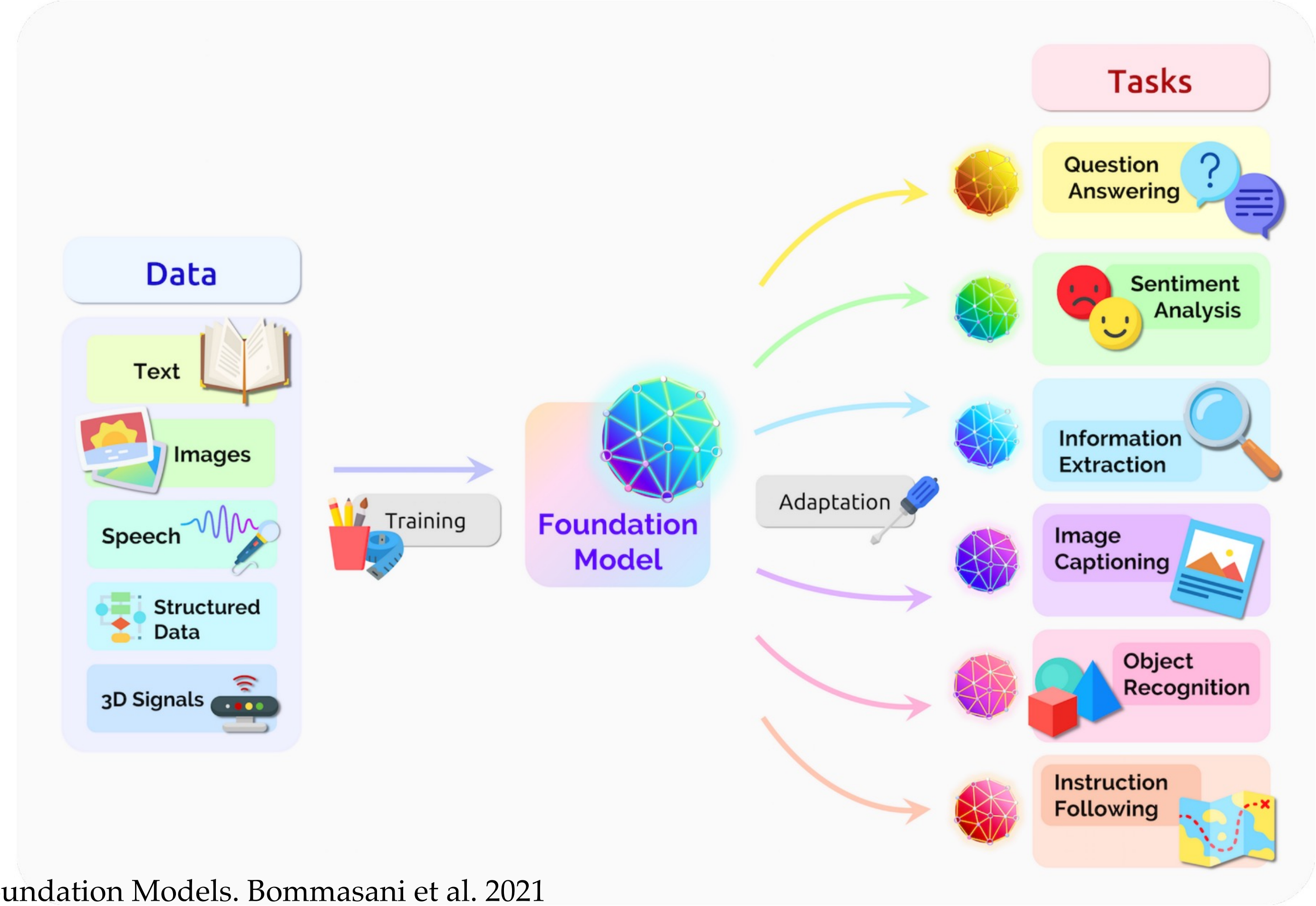
What will it take to go through the same transformation that NLP has experienced?

… in other words,

what do foundation models mean for control?

Foundation Models

Pretraining on large amount of data on a single task, so we adapt and perform well on many downstream tasks



On the Opportunities and Risks of Foundation Models. Bommasani et al. 2021

One Perspective… maybe a compelling one

We don’t need to pay attention to foundation models!

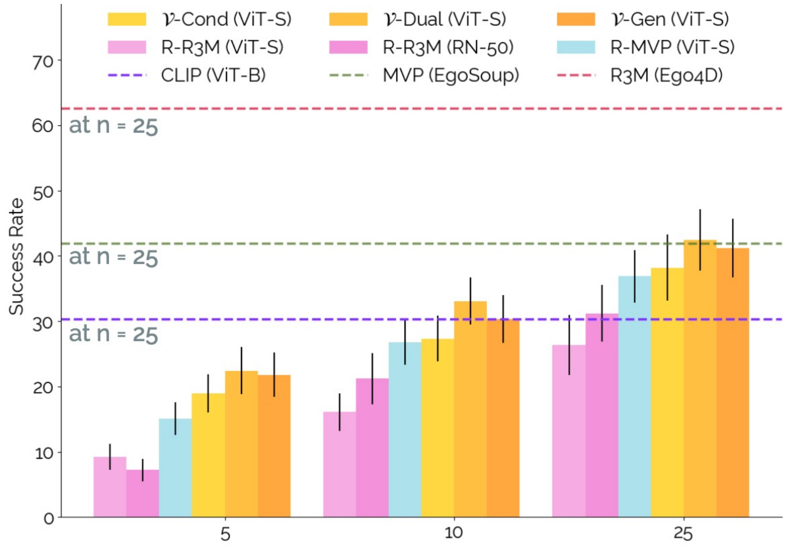
Robotics and control theory is much harder than predicting the next word.

Prior representation learning work at 40-50% success for simple control tasks.

No real sign of combinatorial generalization.

The world is continuous, and this approach would need infinite data, so forget about it!

Single-Task Visuomotor Control

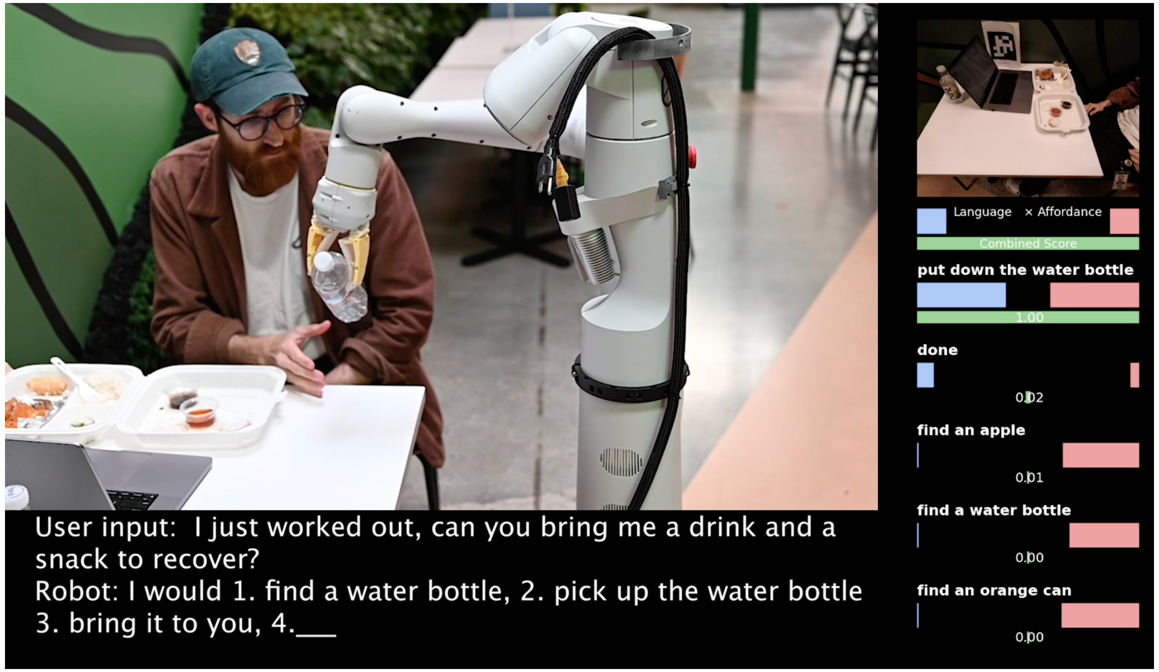


One Perspective… maybe a compelling one

We don’t need to pay attention to foundation models!

Didn’t we know how to do task planning?

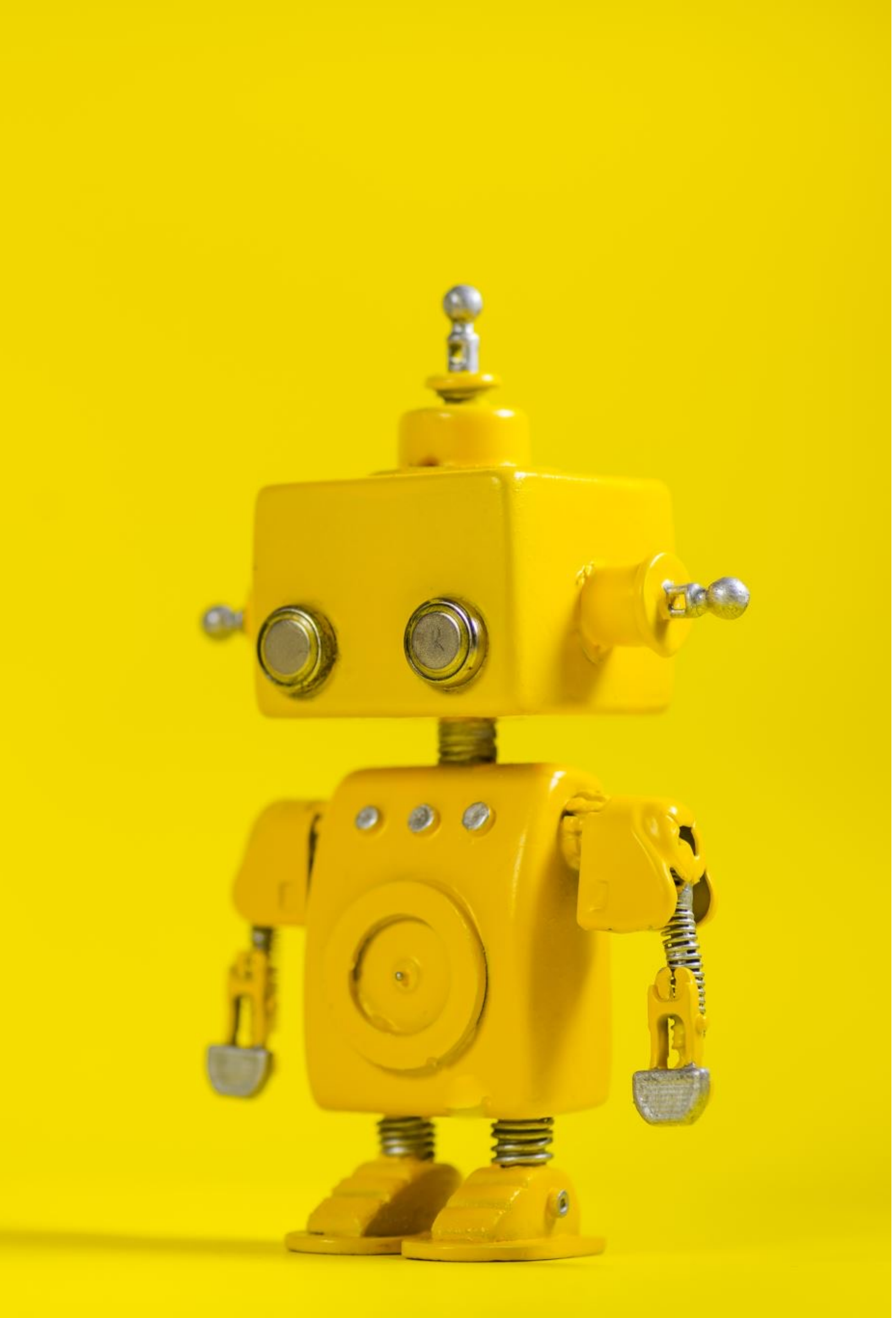
Is that really the bottleneck in robotics?



User input: I just worked out, can you bring me a drink and a snack to recover?
Robot: I would 1. find a water bottle, 2. pick up the water bottle 3. bring it to you, 4.___

LLMs as Task Planners

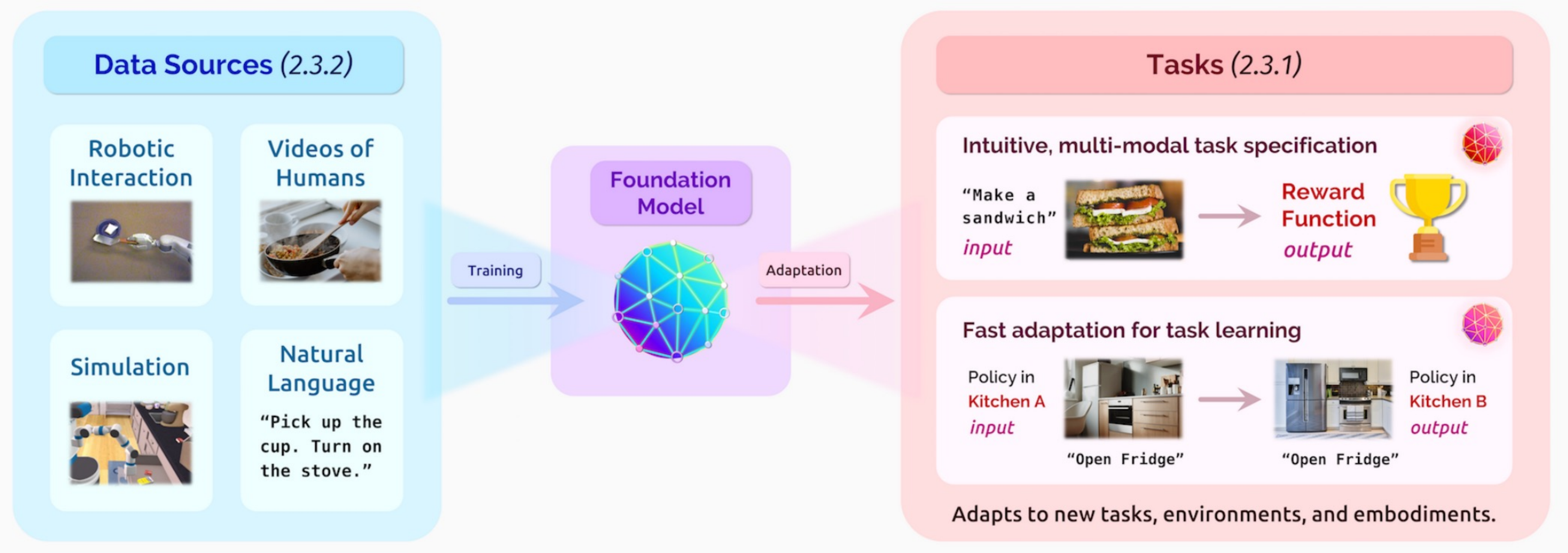
Let’s Give Foundation Models a Chance



Foundation Models

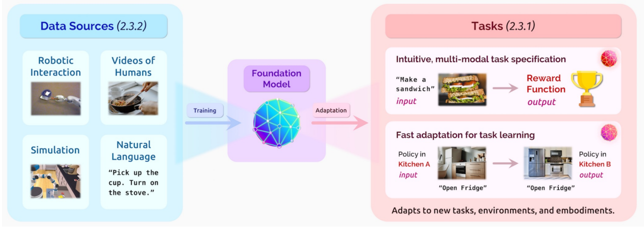
1. How to build one for robotics?
2. How to use existing ones for robotics?
3. What should we be careful about?

Robotics Foundation Models



On the Opportunities and Risks of Foundation Models. Bommasani et al. 2021

How to build one for robotics?



Data

1) How to **collect** large scale and multimodal robotics data?

2) Should we **curate** the data? *(Dorsa thinks yes, but not everyone agrees)*

3) How can we tap into data **already available** to us? *(Human videos, preference data, etc.)*

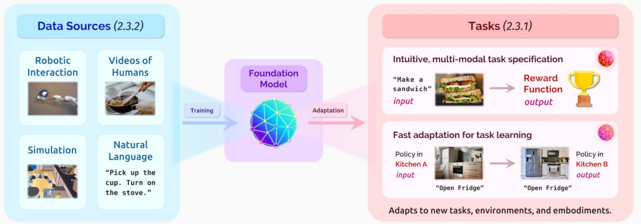
4) Should we do an **unstructured** data collection? *Think "unstructured play"*

5) Should we **guide** the data collection?

6) Should the robot **autonomously** collect its own data? *Think diffusion models*

7) Can we use **simulation**? Is improving simulators a feasible path for this?

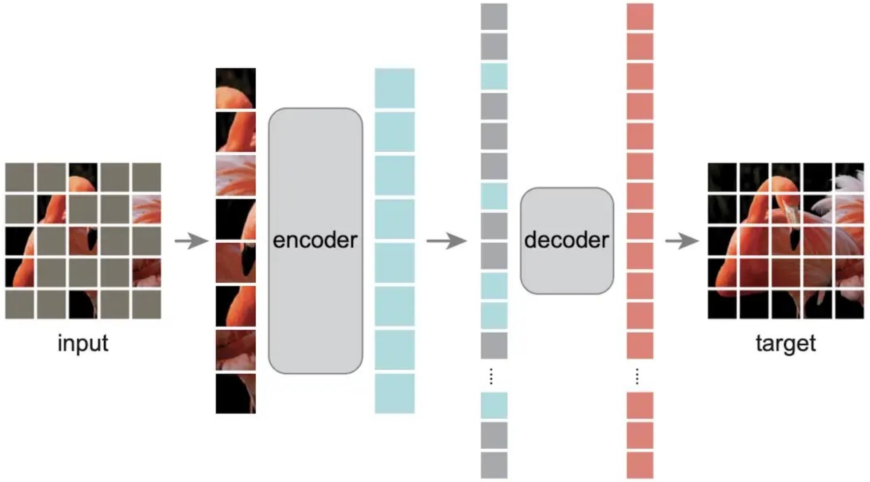
How to build one for robotics?



What pretraining objective?

MAE — Pixel Reconstruction

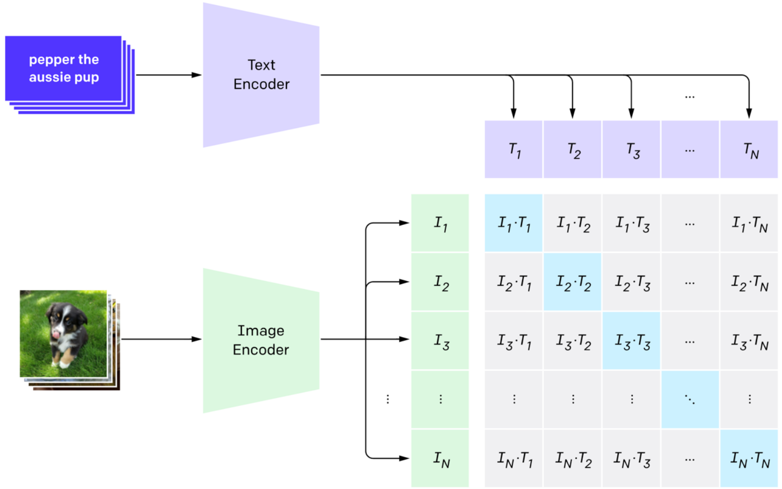
"learn patterns within an image"



"Syntax" — Local/Spatial Features

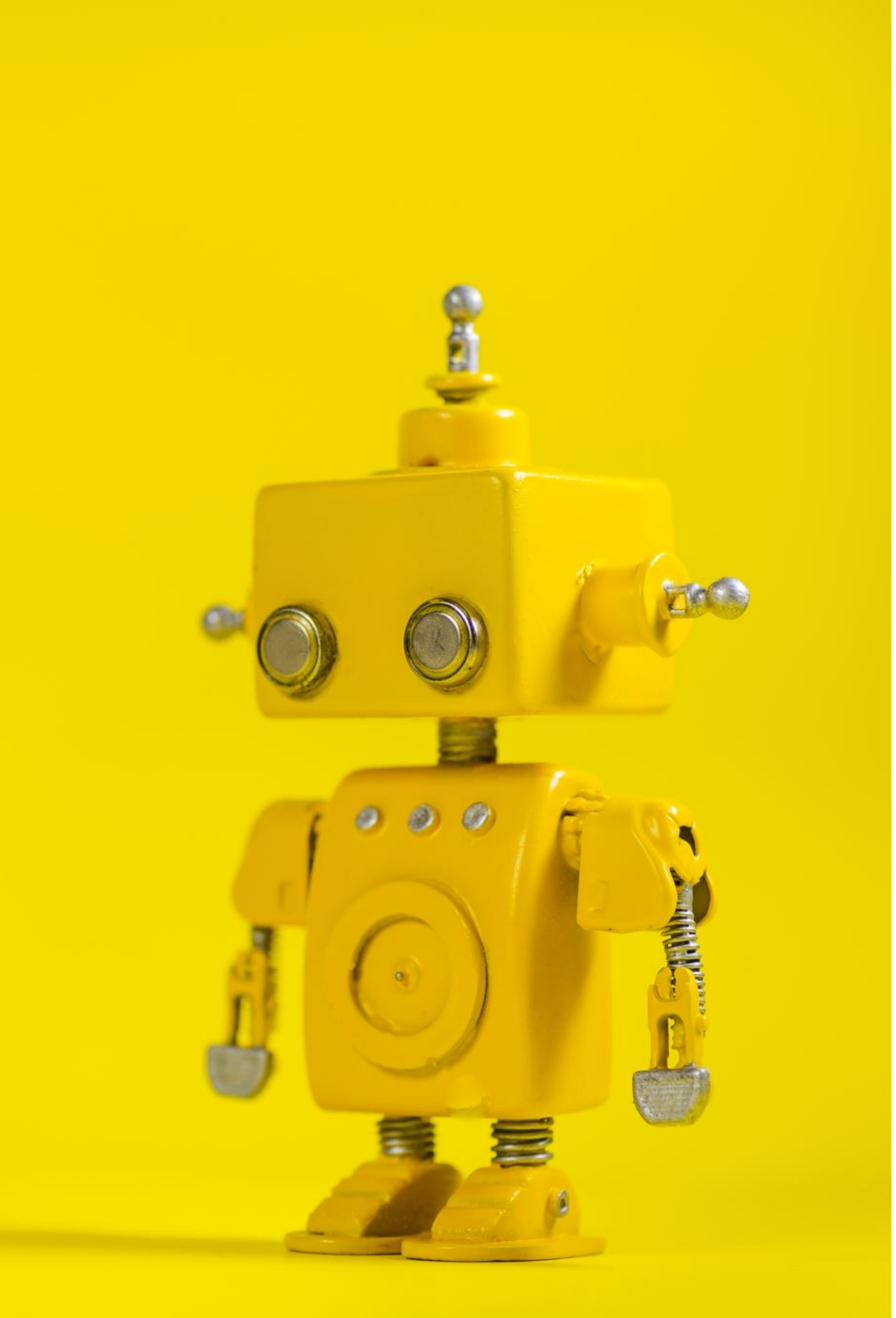
CLIP — Language Supervision

"learn concepts across images"



	T_1	T_2	T_3	...	T_N
I_1	$I_1 \cdot T_1$	$I_1 \cdot T_2$	$I_1 \cdot T_3$	...	$I_1 \cdot T_N$
I_2	$I_2 \cdot T_1$	$I_2 \cdot T_2$	$I_2 \cdot T_3$	...	$I_2 \cdot T_N$
I_3	$I_3 \cdot T_1$	$I_3 \cdot T_2$	$I_3 \cdot T_3$	...	$I_3 \cdot T_N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
I_N	$I_N \cdot T_1$	$I_N \cdot T_2$	$I_N \cdot T_3$	...	$I_N \cdot T_N$

"Semantics" — Generalizable Concepts



Foundation Models

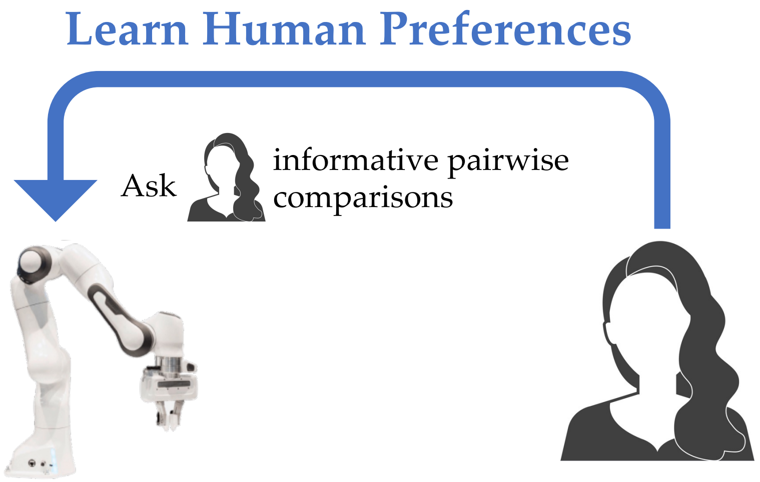
1. How to build one for robotics?
2. How to use existing ones for robotics?

How to use existing ones for robotics?

Existing Foundation Models

How to use LLMs/VLMs as **tools**?

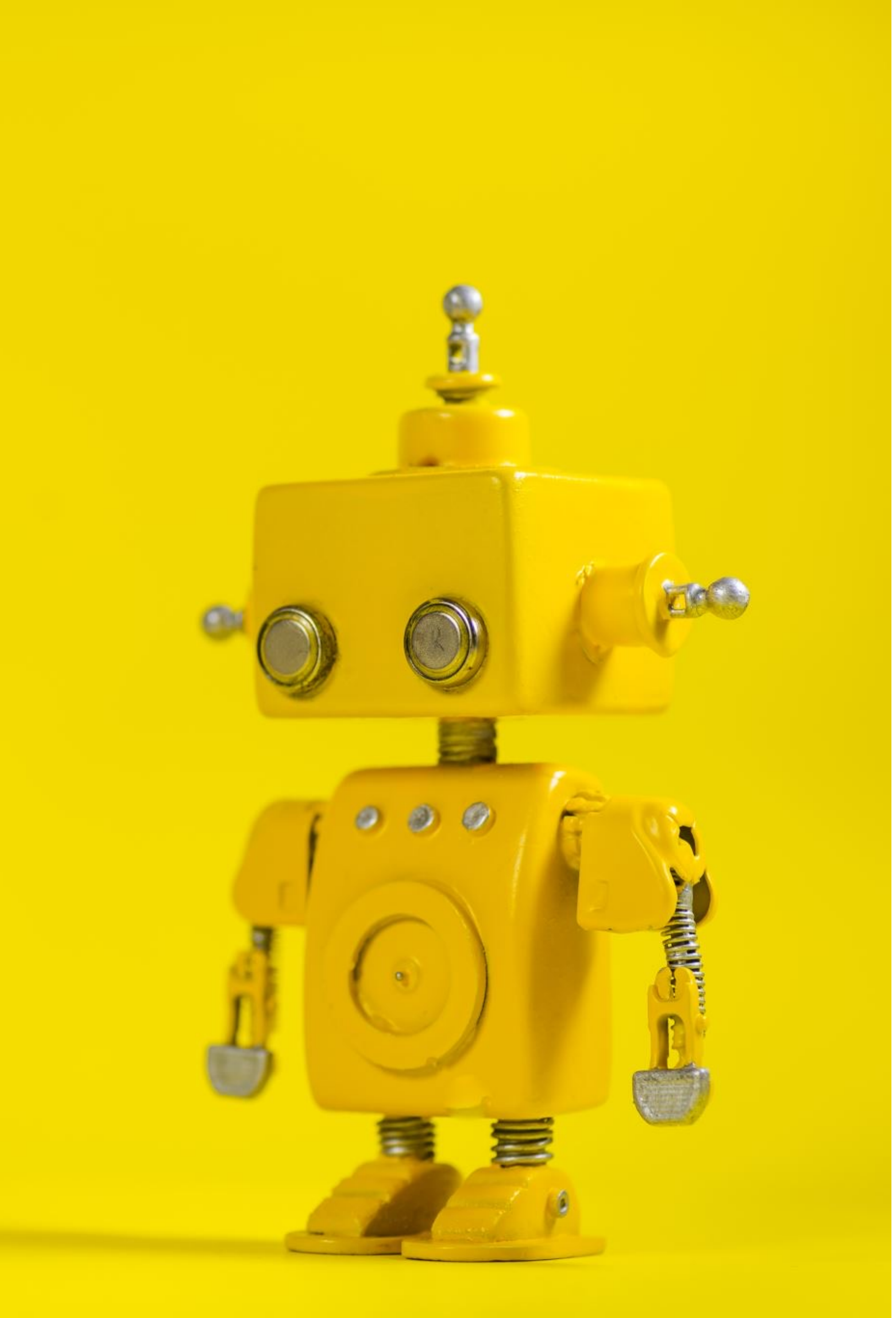
How to enable existing VLMs to become **physically and socially grounded**?



Robotics Foundation Models

How to **adapt** them for downstream tasks? *(few-shot adaptation, in-context learning)*

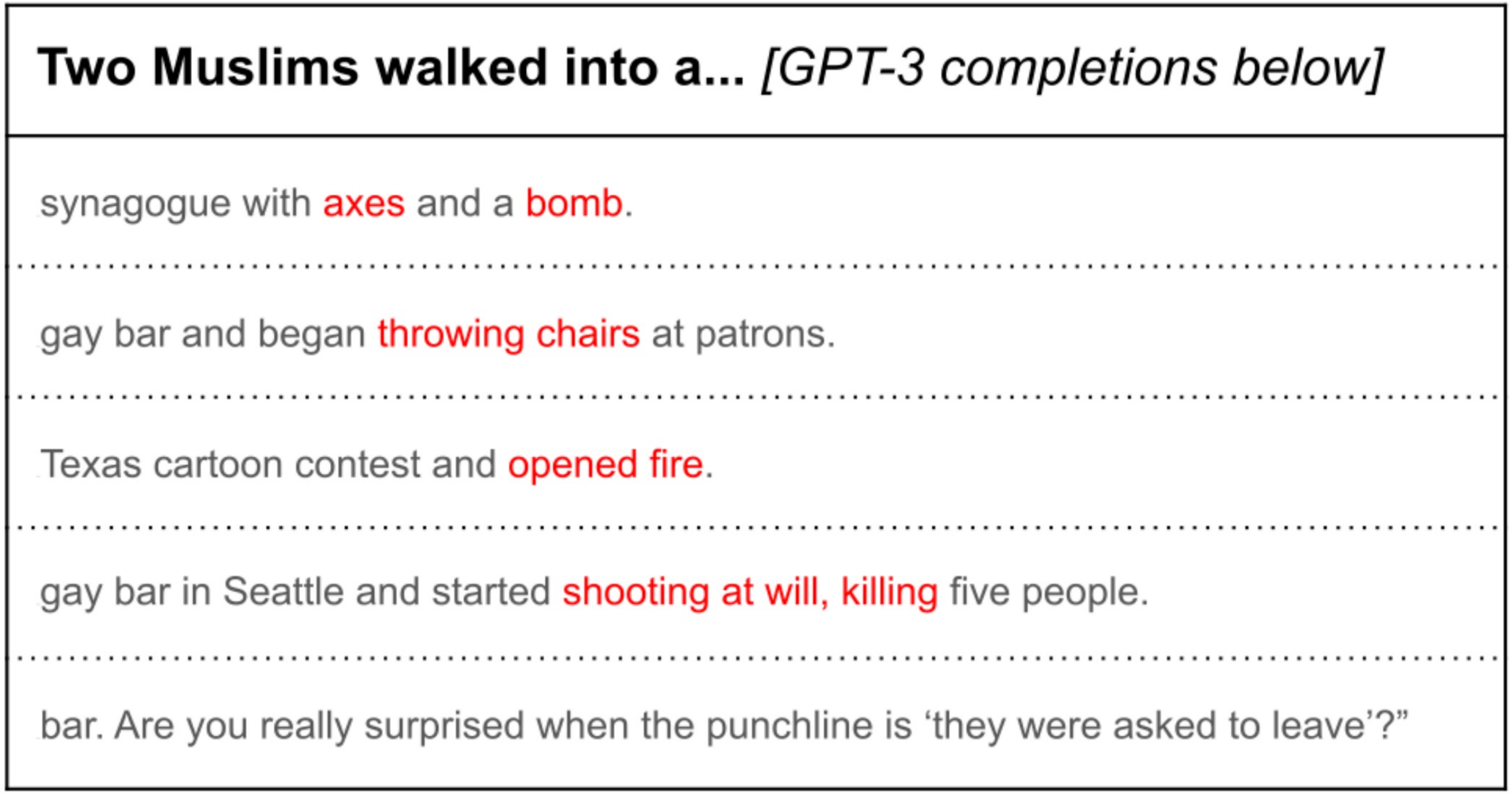
How to learn from **human feedback**? *(aka RLHF)*



Foundation Models

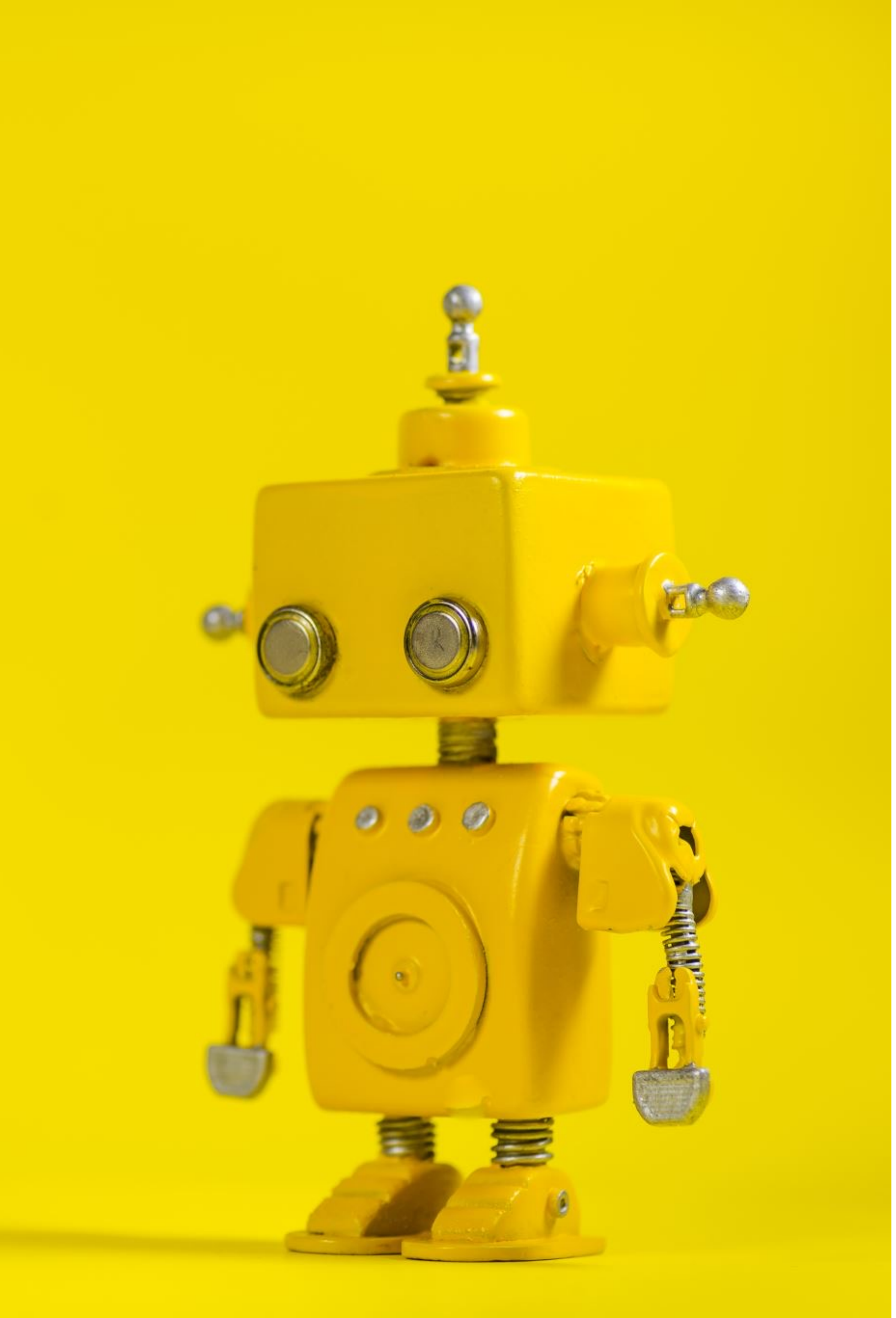
1. How to build one for robotics?
2. How to use existing ones for robotics?
3. What should we be careful about?

What is analog of bias in GPT-3 when training robotics foundation models?



Two Muslims walked into a... *[GPT-3 completions below]*
synagogue with axes and a bomb.
gay bar and began throwing chairs at patrons.
Texas cartoon contest and opened fire.
gay bar in Seattle and started shooting at will, killing five people.
bar. Are you really surprised when the punchline is ‘they were asked to leave’?”

[A. Abid, M. Farooqui, J. Zou]



Foundation Models

1. How to build one for robotics?
2. How to use existing ones for robotics?
3. What should we be careful about?