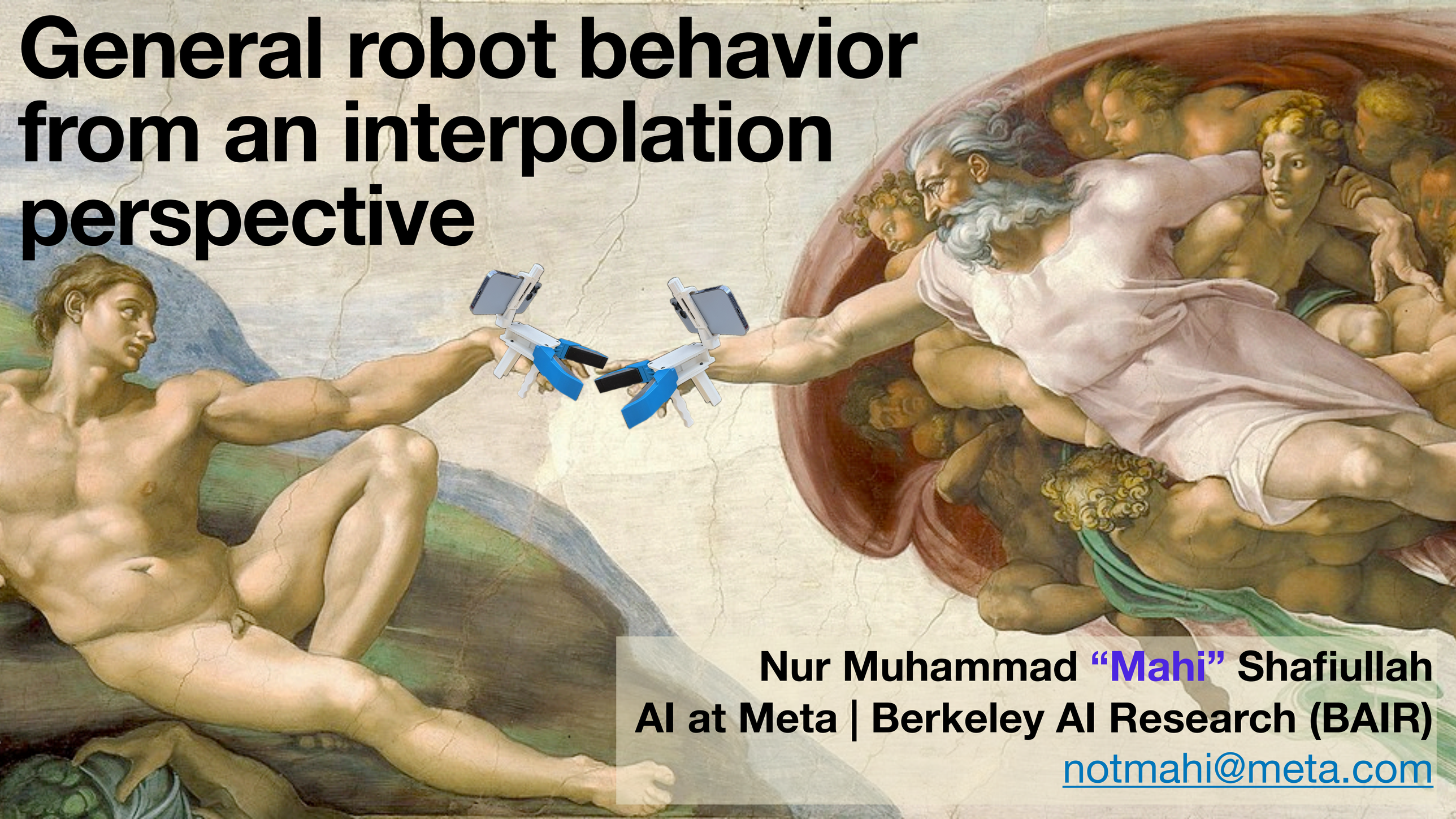
General robot behavior from an interpolation perspective



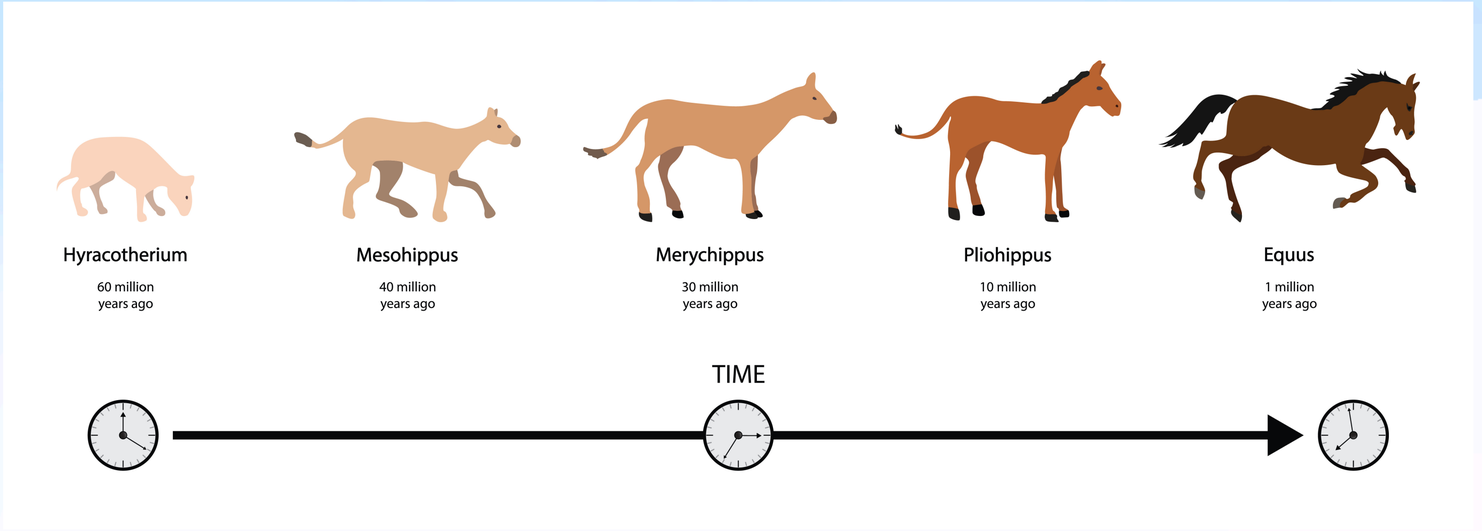
Nur Muhammad “Mahi” Shafiullah
AI at Meta | Berkeley AI Research (BAIR)
notmahi@meta.com

For robots, creationism or evolutionism?

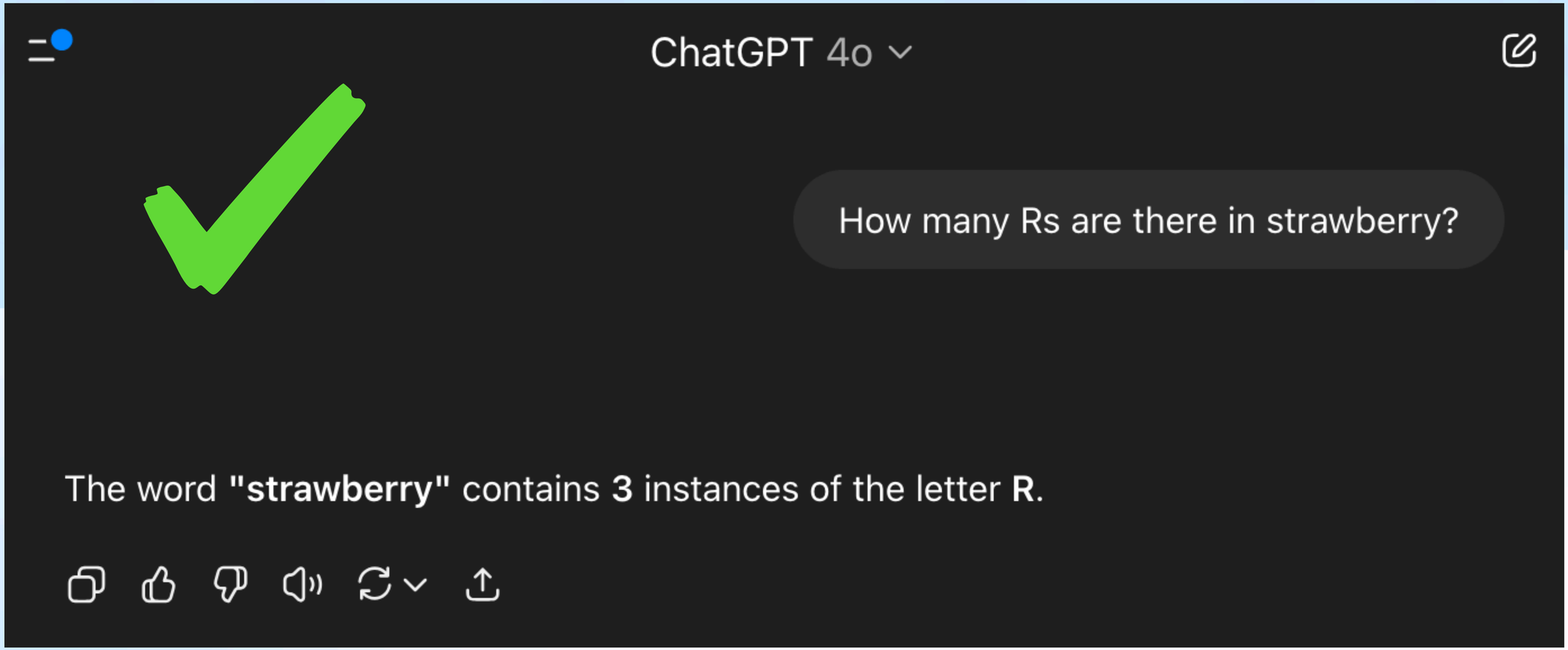
Option 1:
Create the perfect robot in a lab

(Worked pretty well for LLMs)

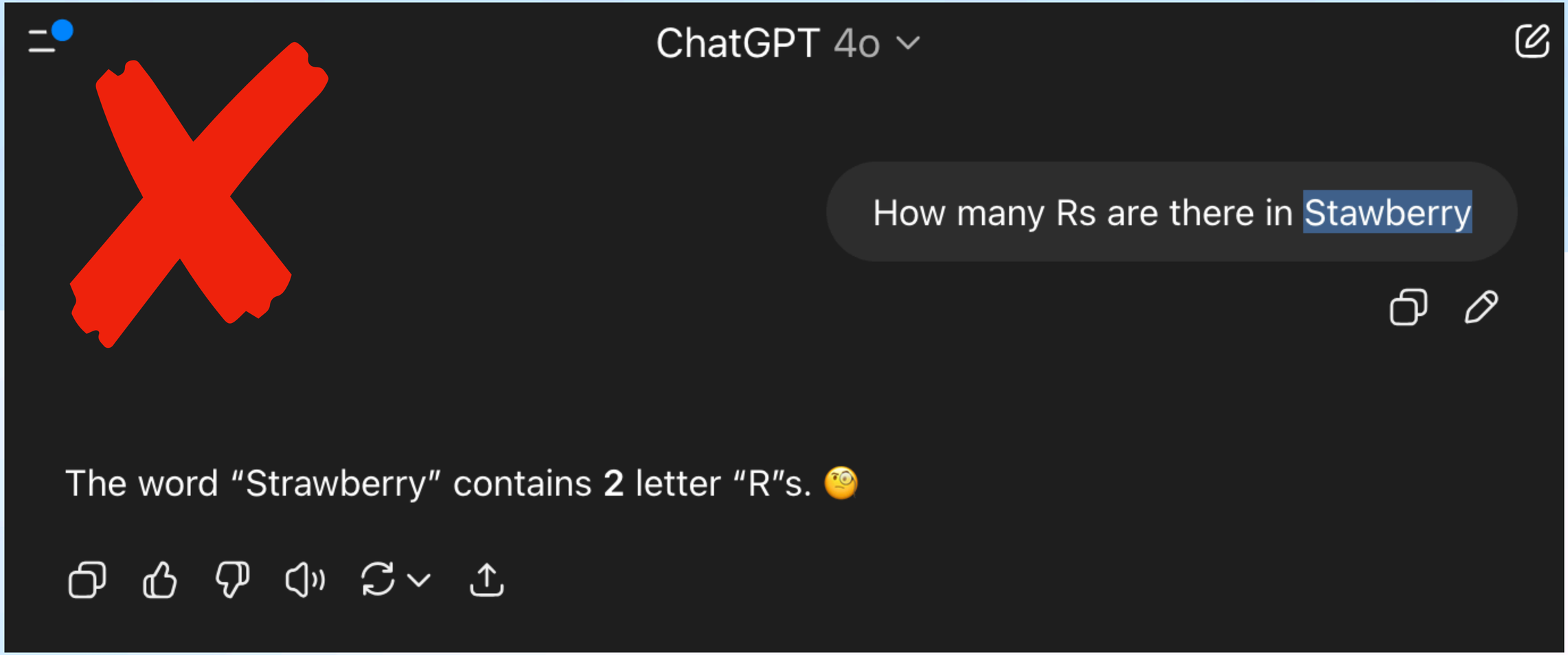
Option 2:
Create imperfect robots that get better over iterations



Our best methods are still (mostly) interpolation



Our best methods are still (mostly) interpolation

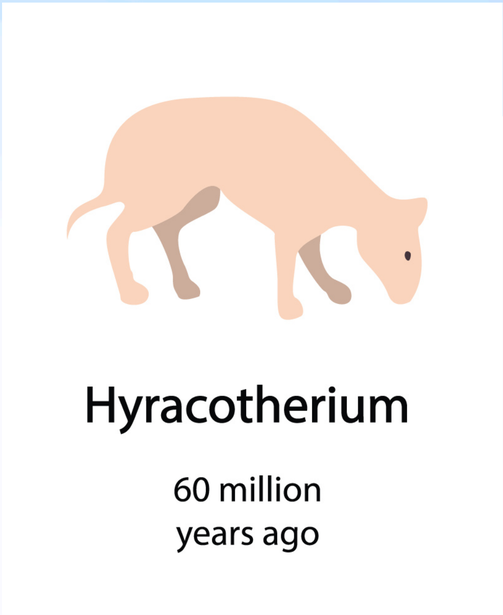


For robots, creationism or evolutionism?

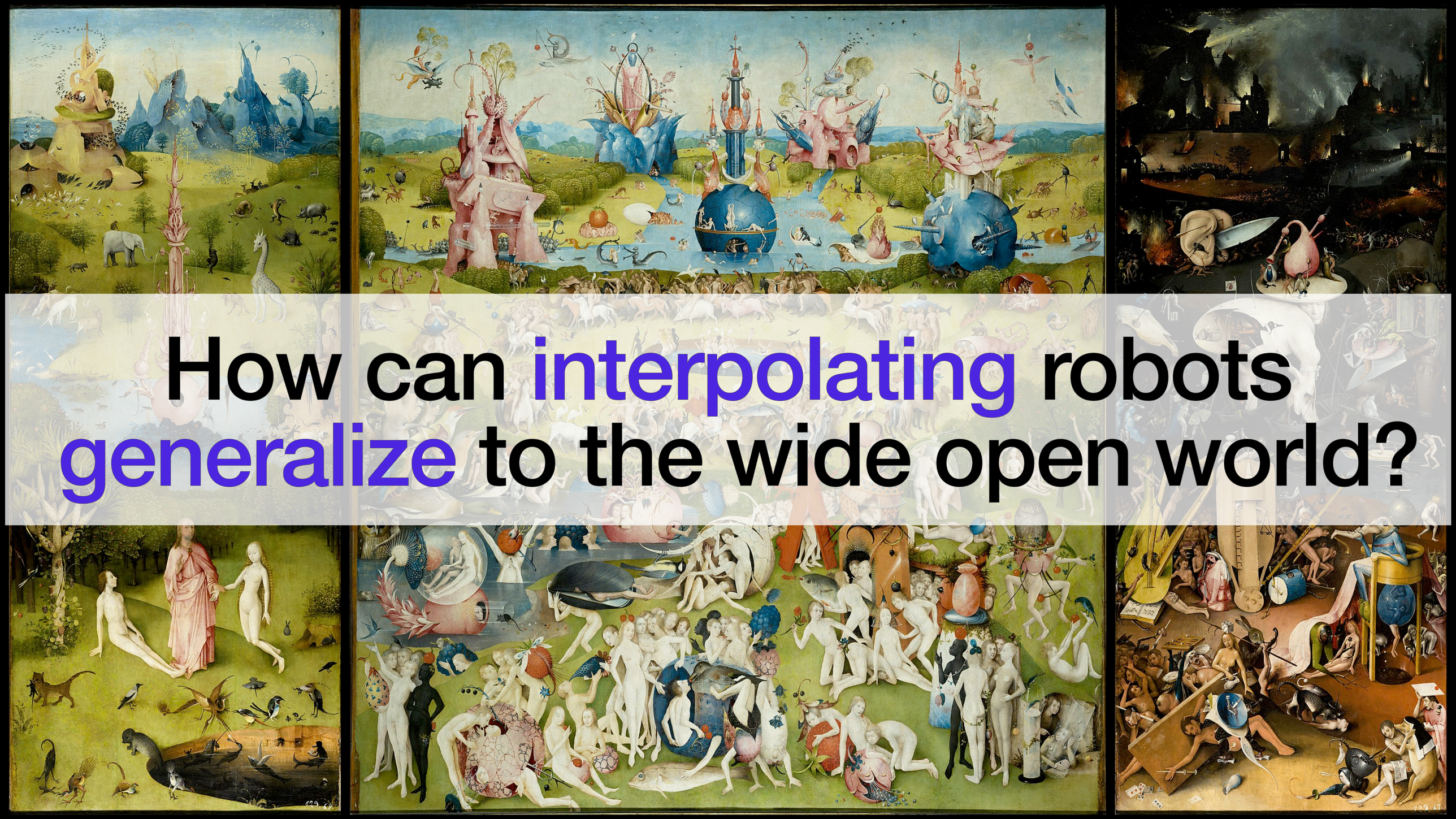
Option 1:
Create the perfect robot in a lab

Many robots + maybe a few billion $$$

Option 2:
Create imperfect robots that get better over iterations



What is this “first imperfect robot”?



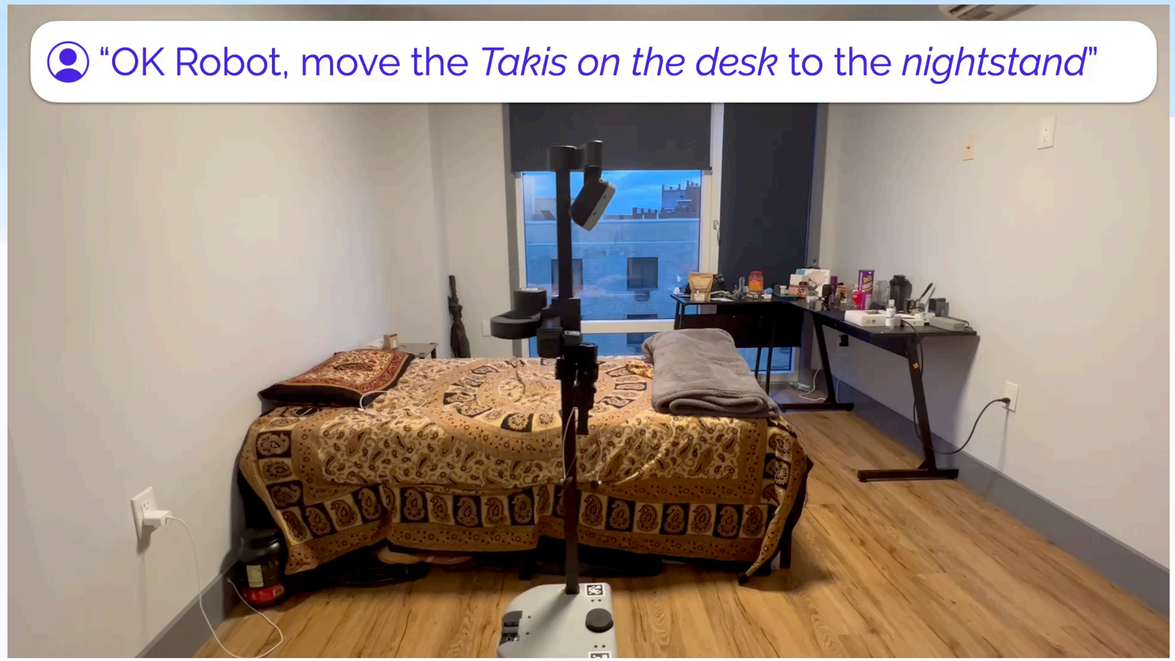
How can interpolating robots generalize to the wide open world?

Interpolating in space with embodiment matching

dobb-e.com

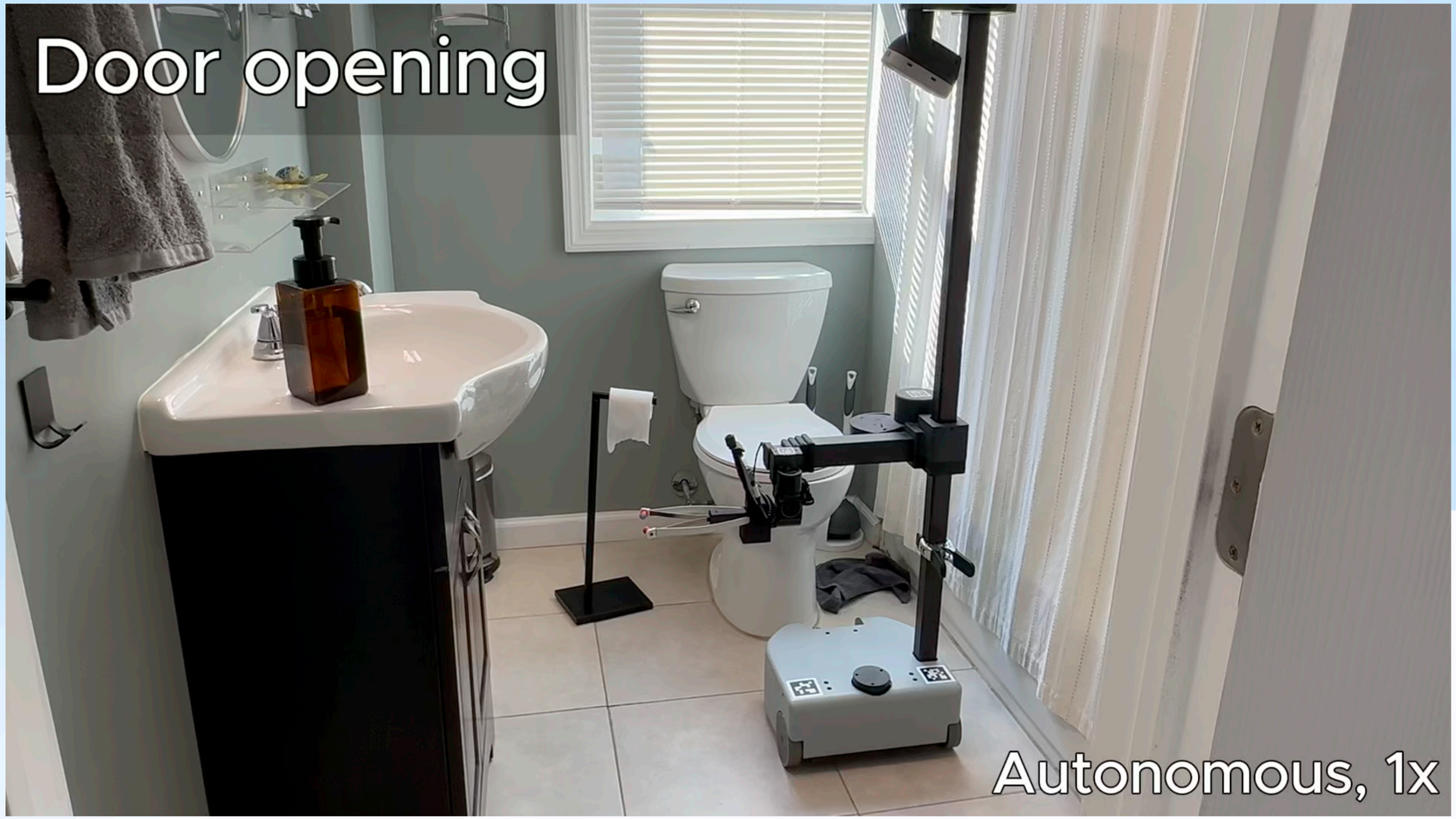
robotutilitymodels.com

Interpolating in time with semantic memory

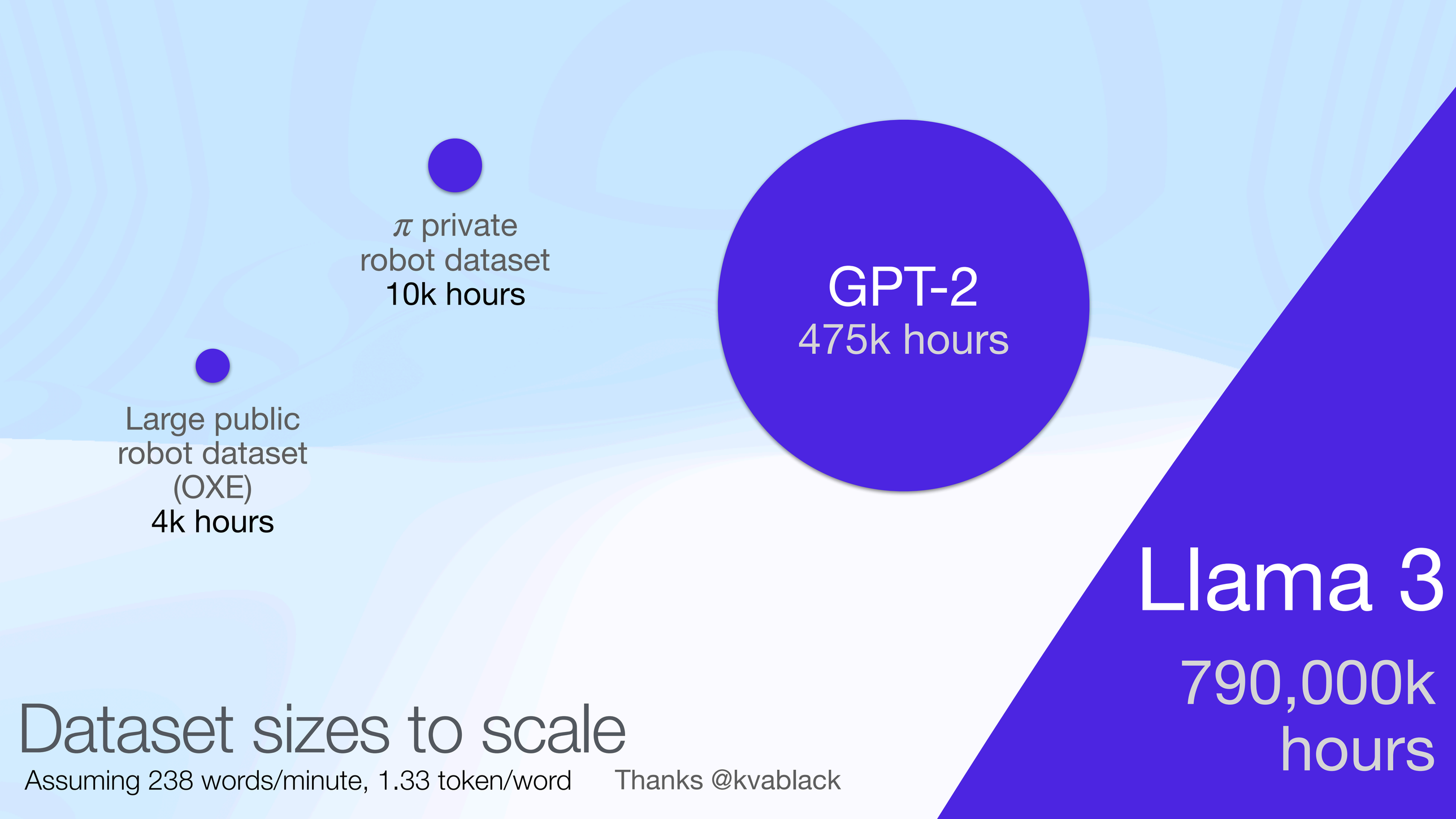


ok-robot.github.io

dynamem.github.io



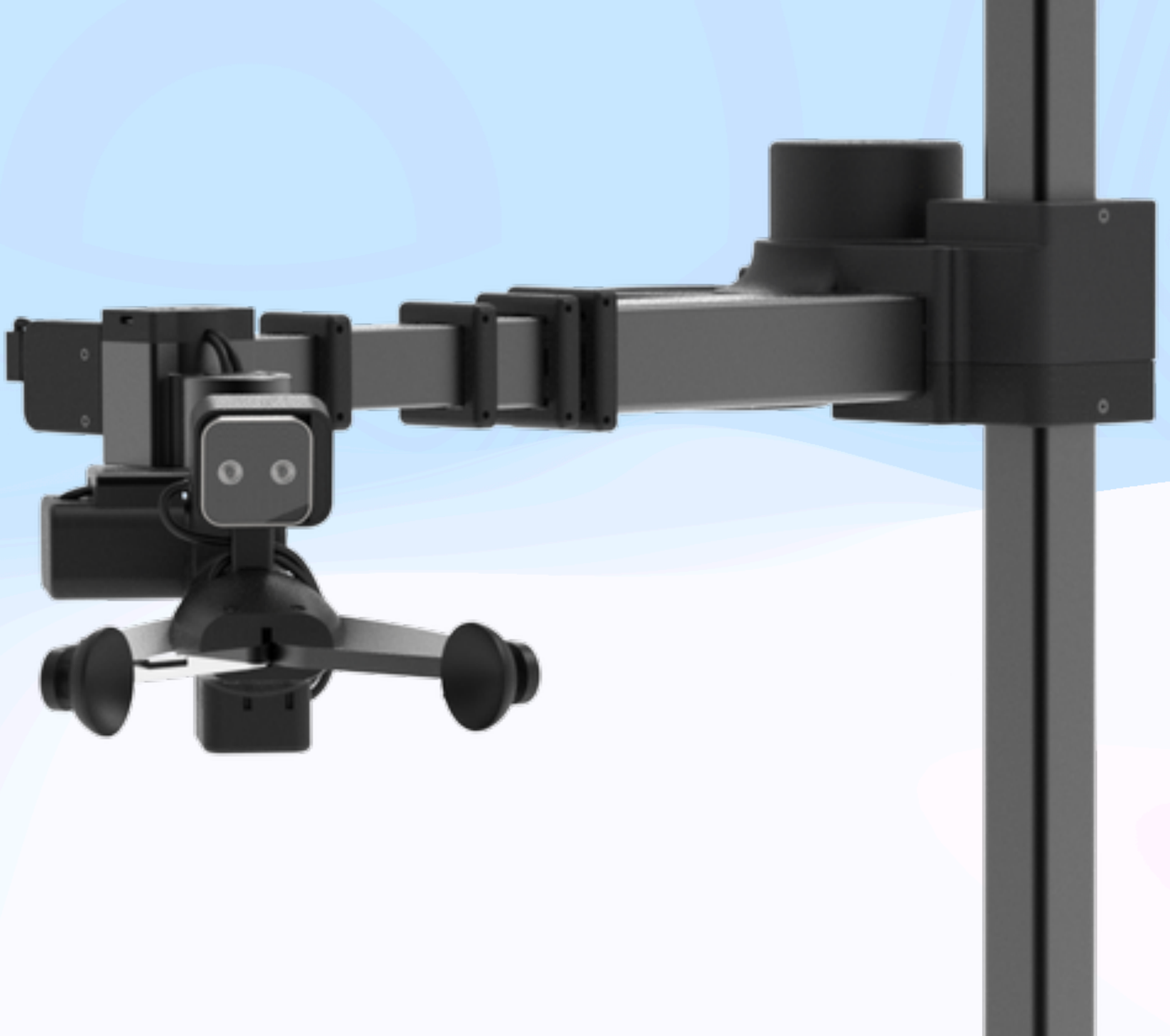
The challenge of learning to manipulate anywhere



Dataset sizes to scale

Assuming 238 words/minute, 1.33 token/word

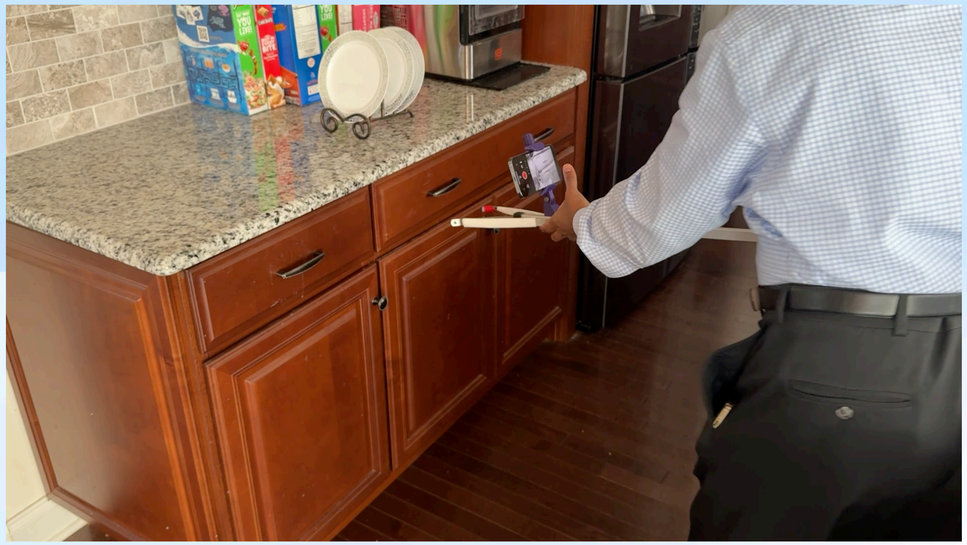
Thanks @kvablack



Hello Robot: Stretch
end-effector

Our data collection tool

Interpolation = efficiency

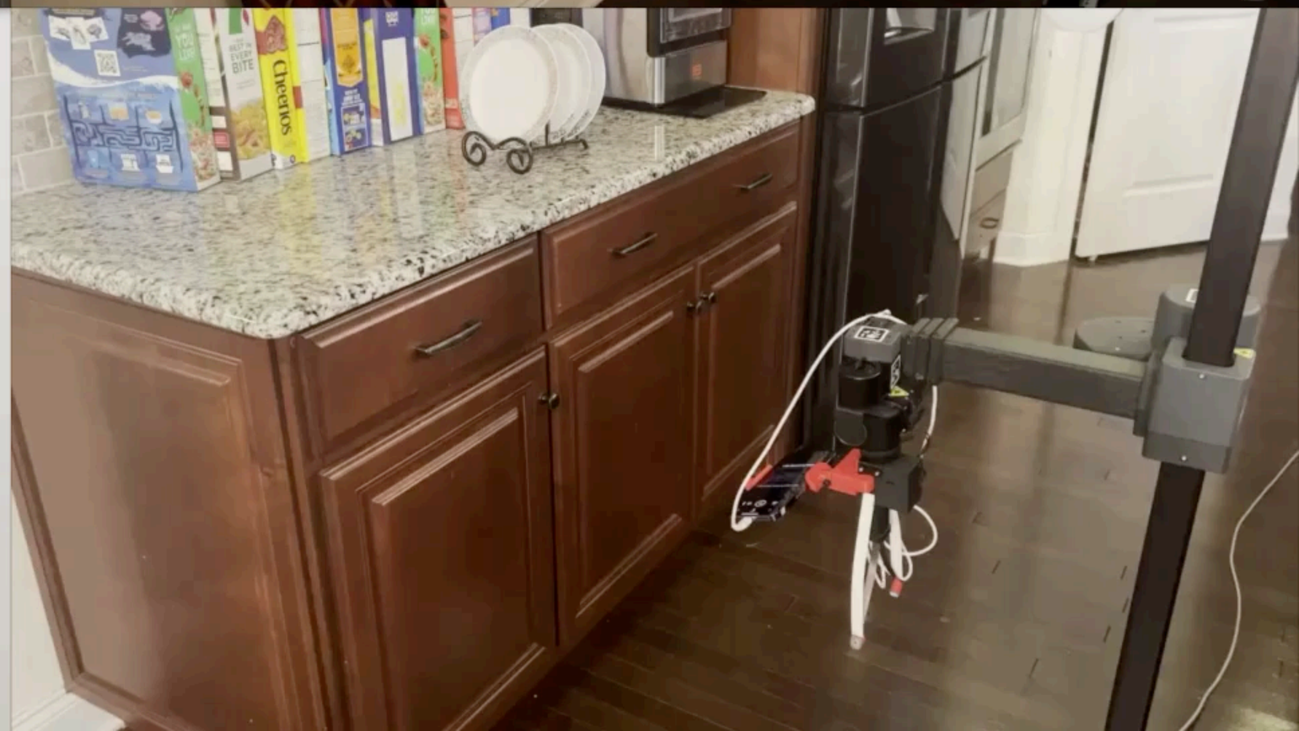


Collect 24 demos
5 minutes

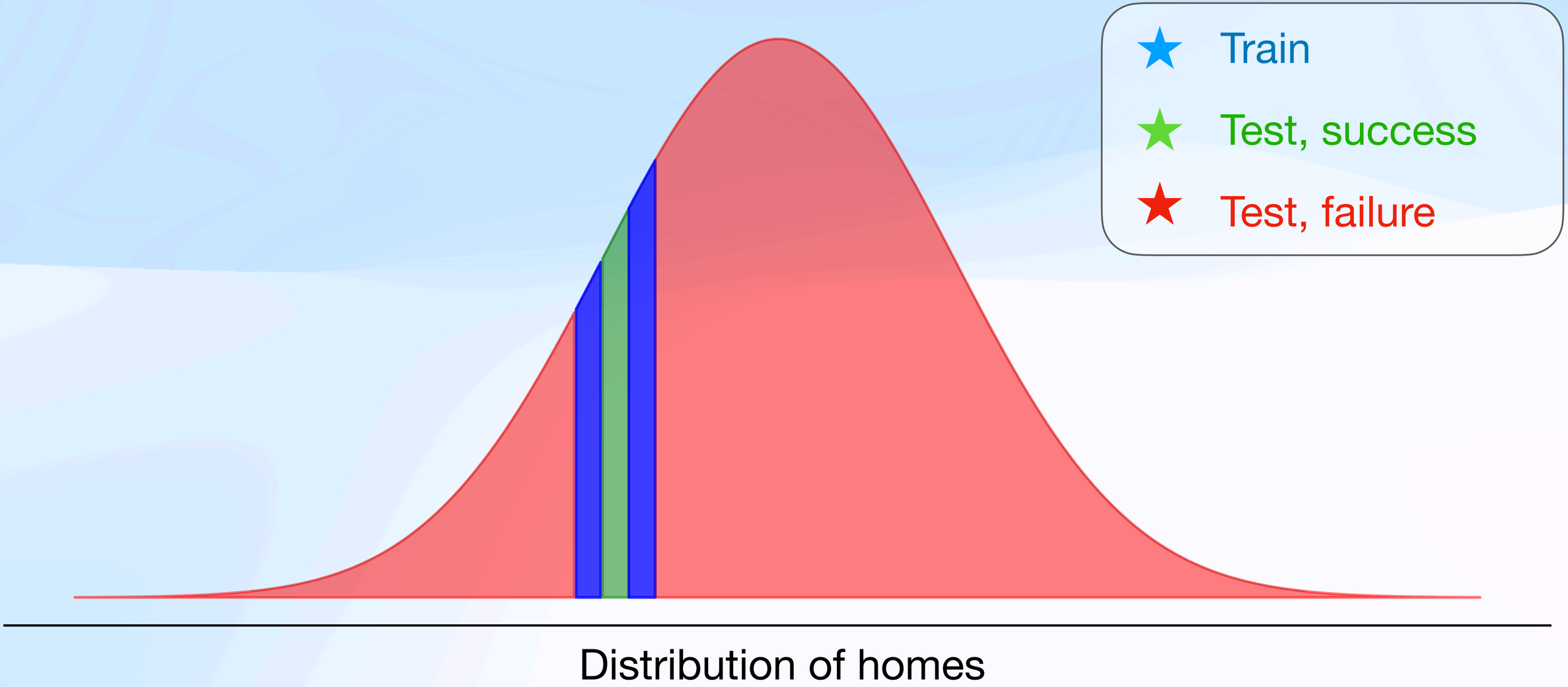
Fine-tune model
15 minutes

Deploy!


Domino


Cheerios

It’s all still interpolation…



But if we focus on quality & diversity…

30+ homes, 300+ environments

Tasks

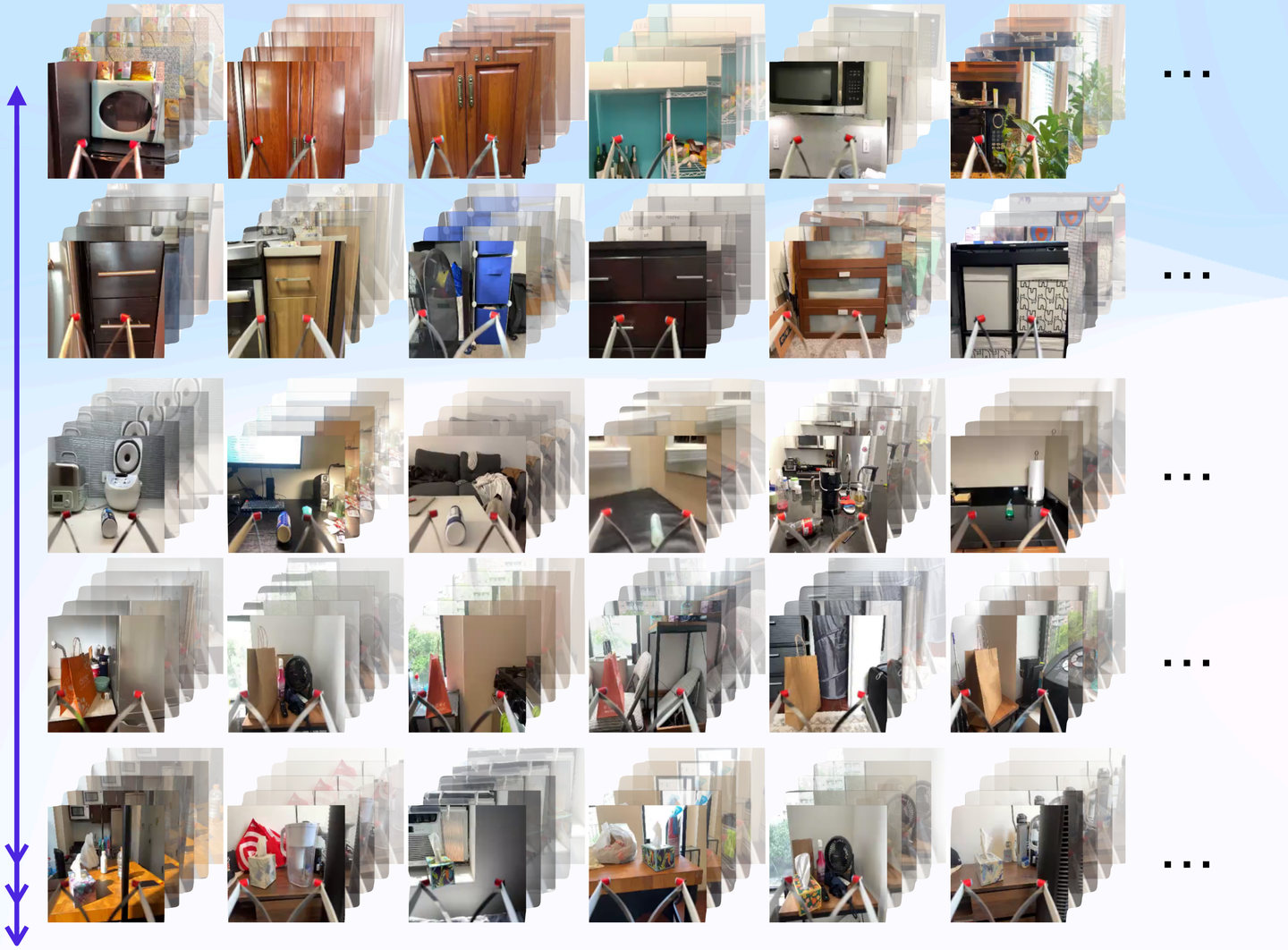
Door opening

Drawer opening

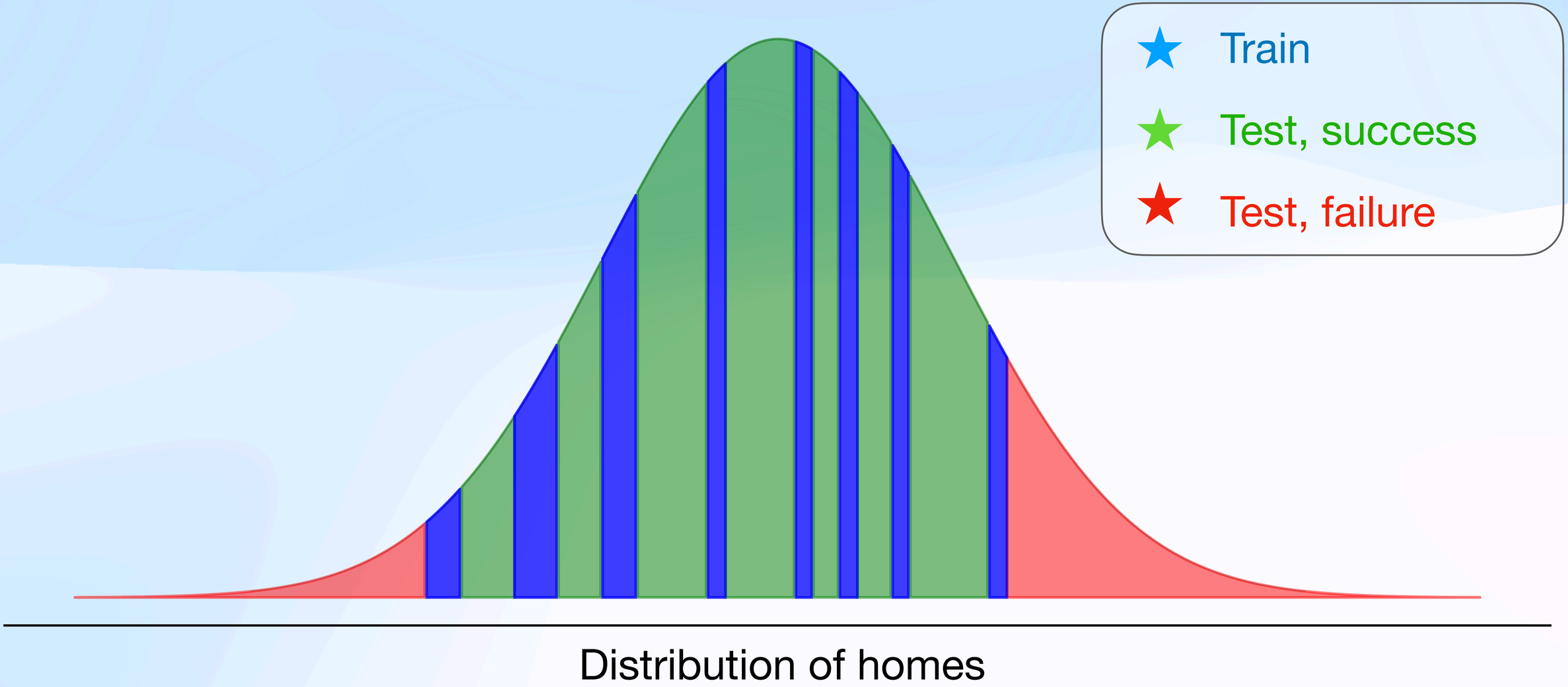
Reorientation

Object pickup

…



Interpolation is okay!



“Utility models” in novel environments

Right out of the box, no fine tuning!

- Same policy, everywhere, no fine-tuning!
- Same recipe for many single task policies
- 30M params – runs on robot

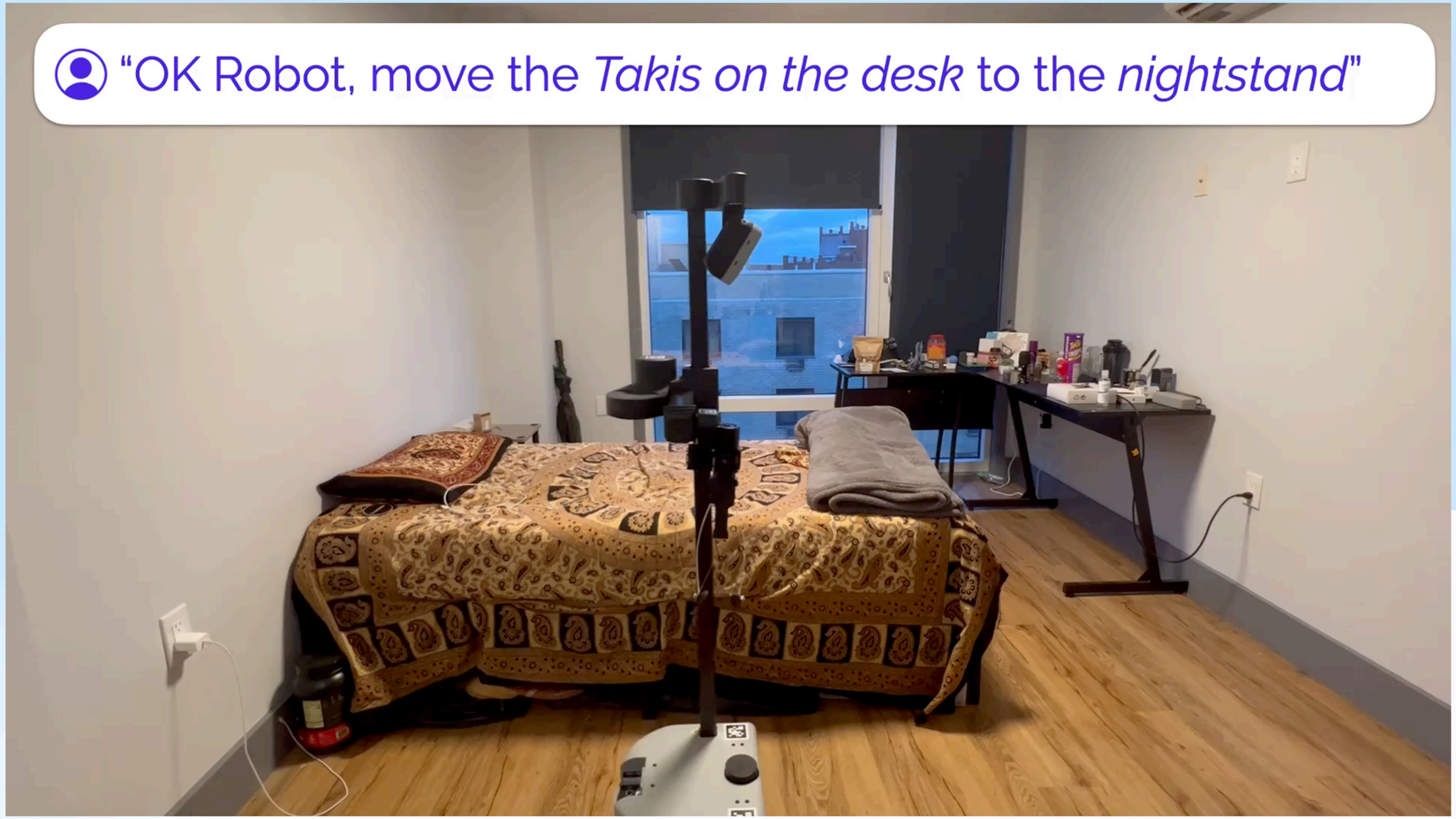
Live demo in CVPR 2025

Best demo award!

Robot Utility Models

robotutilitymodels.com

Haritheja Etukuru



The challenge of long-horizon mobile manipulation

What passes the “interpolation” smell test?

Perspective 1: One large model that does everything

- Large, open interpolation space, needs plenty of coverage

Perspective 2: Many small models that do one thing, well

- Needs “something” to tie them together over time and space

Memory!

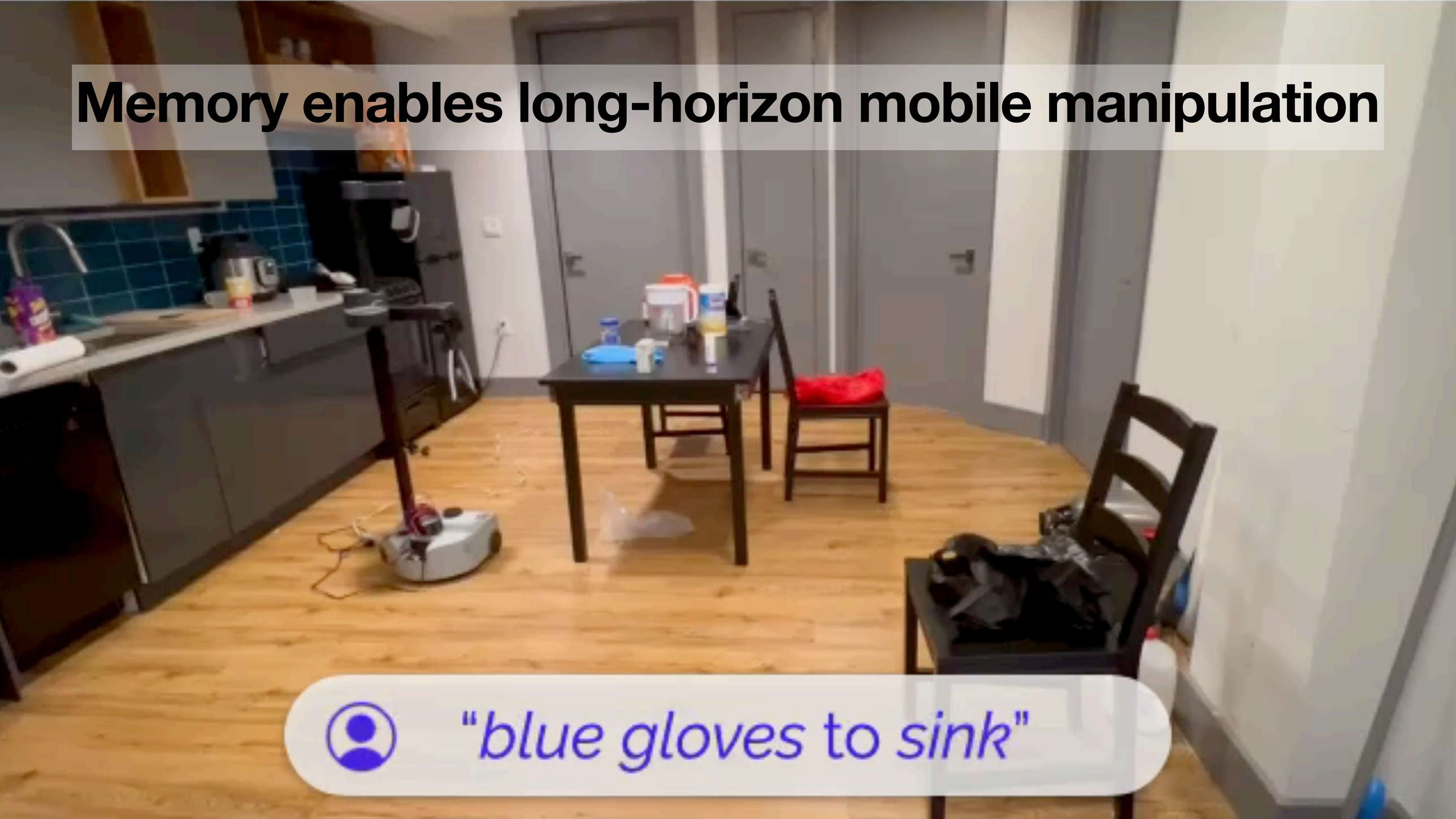
Spatio-semantic memory representation

Answering “what” and “where”

By mapping every point in space to a latent representation

$x \mapsto z_x \in \mathbb{R}^d$, which has semantic meaning

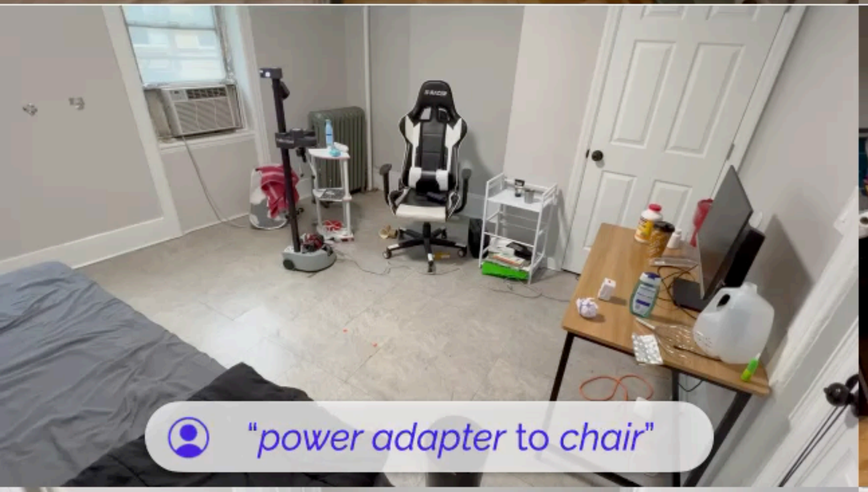
Memory enables long-horizon mobile manipulation



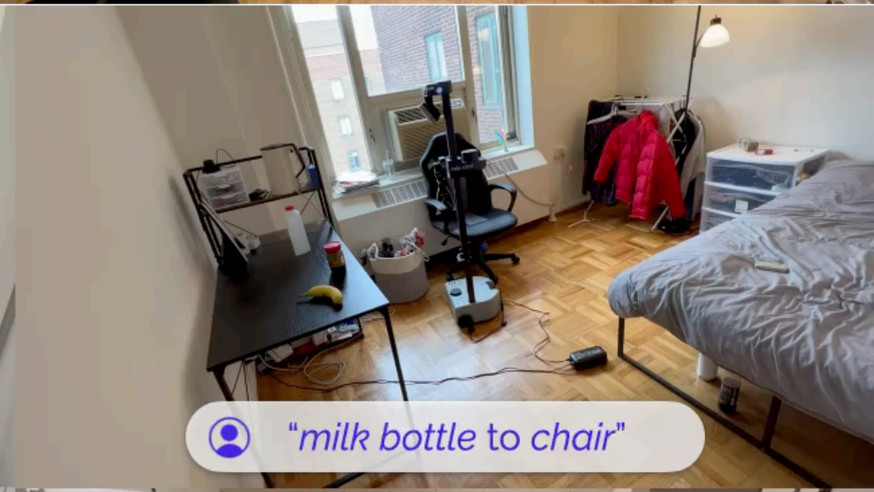

"purple lightbulb box to sofa chair"

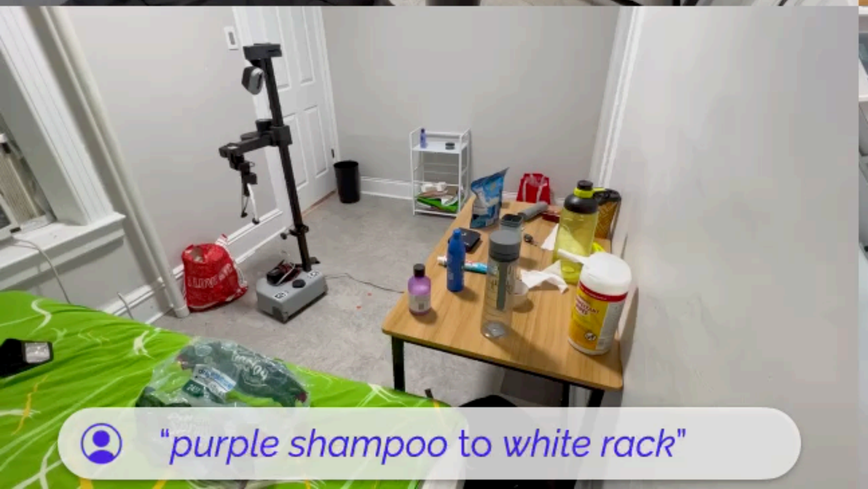

"cooking oil bottle to marble surface"


"yogurt beverage to the table"


"power adapter to chair"


"blue gloves to sink"


"milk bottle to chair"

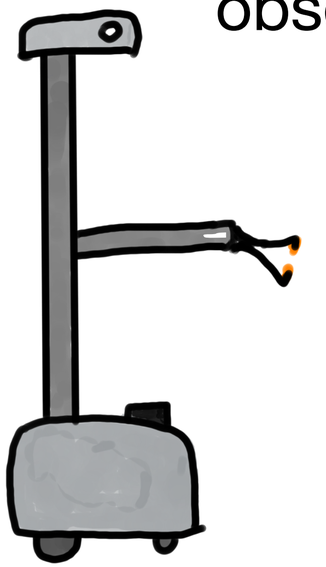

"purple shampoo to white rack"


"herbal tea can to box"


"McDonalds paper bag to stove"

But robot memory can’t keep up with time!

1. Robot makes observations



2. Robot maps the world

3. **The world is always changing!**

4. ???


EXIT
EXIT
LEVI'S
Coca-Cola
117
Place orange on the red plate

Live robot memory defined by four factors

DynaMem: Online Spatio-semantic Memory for Open-world Mobile Manipulation

Factor 1:
What do you store?

Factor 2:
How do you retrieve?

Factor 3:
How do you add?

Factor 4:
What do you delete?

Factor 1:
What do you store?

Discrete
class labels

Pre 2022

+ abstract latent
vectors (CLIP etc.)

Post-2022
CLIP-Fields,
OK-Robot

+ actual images
or text

Post-2024,
DynaMem

Lookup table

+ Embedding models

+ Large vision/
language models

Factor 2:
How do you retrieve?

Factor 3:
How do you add?

Neural field/
gaussian splat

Scene graph

Voxel maps

Re-train everything!

Do reliable
deduplication!

Use raycasting

Factor 4:
How do you delete?

Live demo in CoRL 2024

Place apple in red bowl

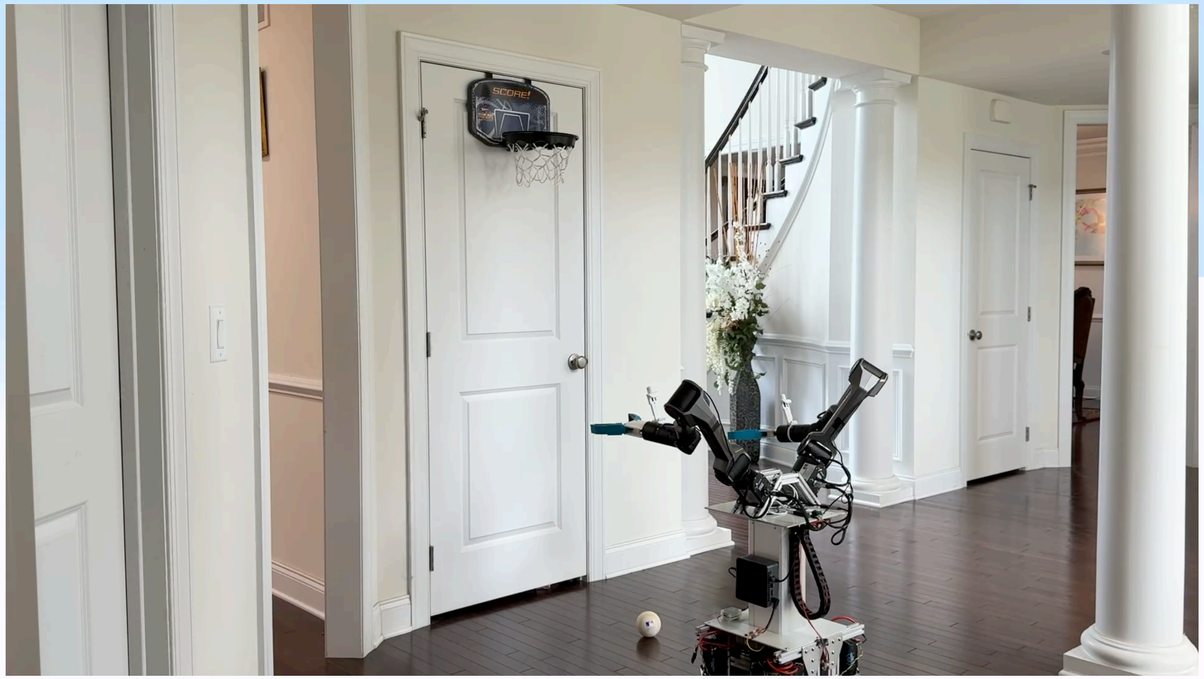
DynaMem

dynamem.github.io

Peiqi Liu

Takeaways from the search for general robots

- Interpolation is okay if done with quality and diversity.
- Structure, when done right, will reduce complexity and not increase it.
- Human <-> robot domain gap is inevitable, but we can be smart and work around it.



See more works at https://mahis.life or email me at notmahi@meta.com!