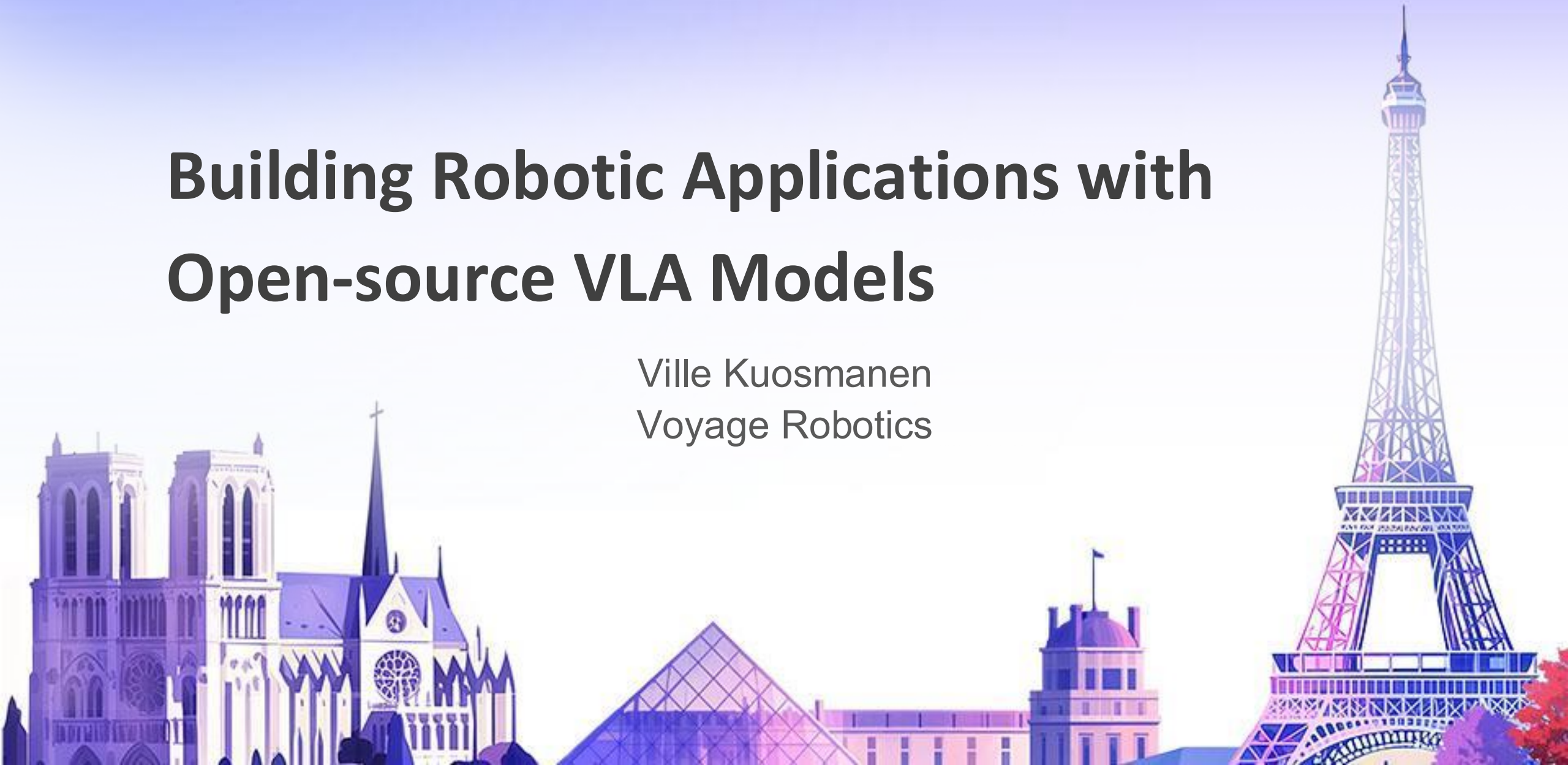
GOSIM

GOSIM AI Paris 2025

Building Robotic Applications with Open-source VLA Models

Ville Kuosmanen
Voyage Robotics

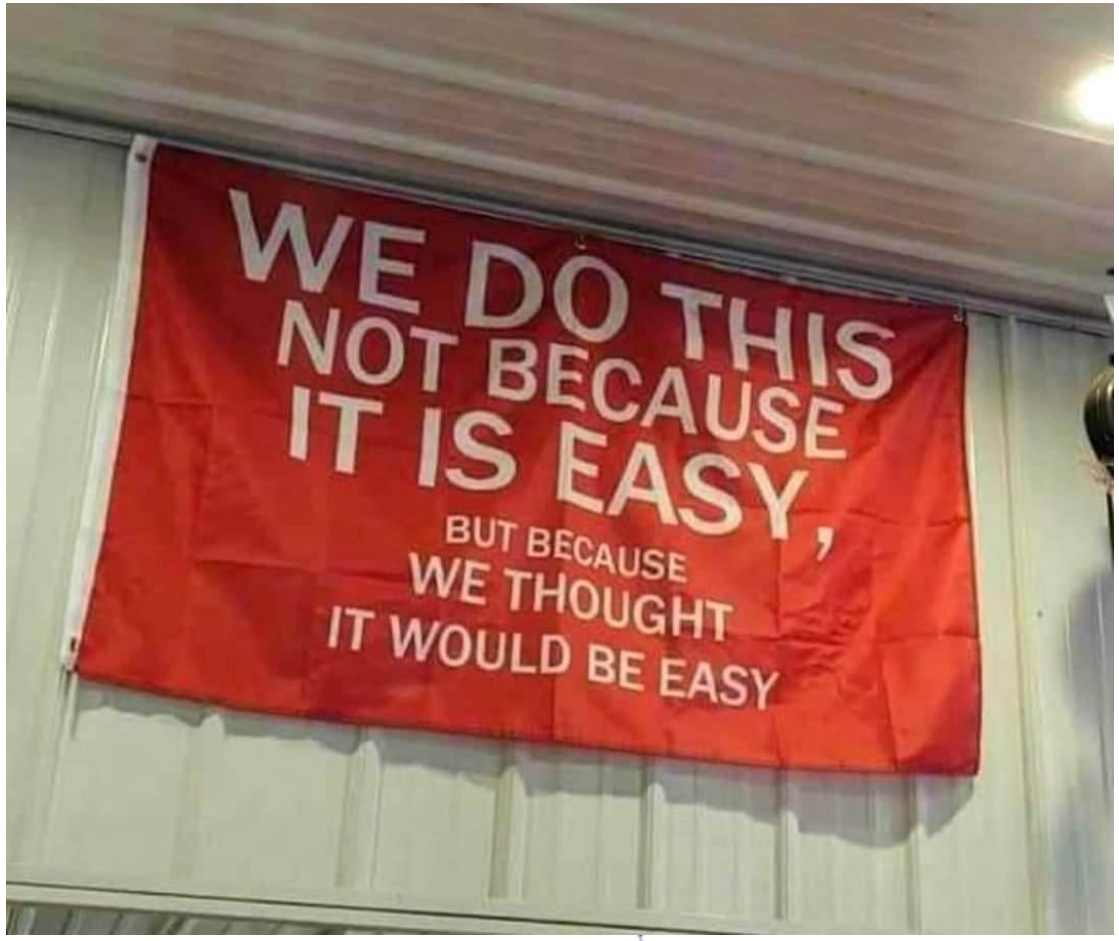



Cook Shrimp
(autonomous)
Stanford


Ville
@VilleKuosmanen · Jul 29, 2024
Promote
I bought a robot
13
12
238
16K

GOSIM


Ville @VilleKuosmanen · Jul 29, 2024
Promote
I bought a robot
13
12
238
16K

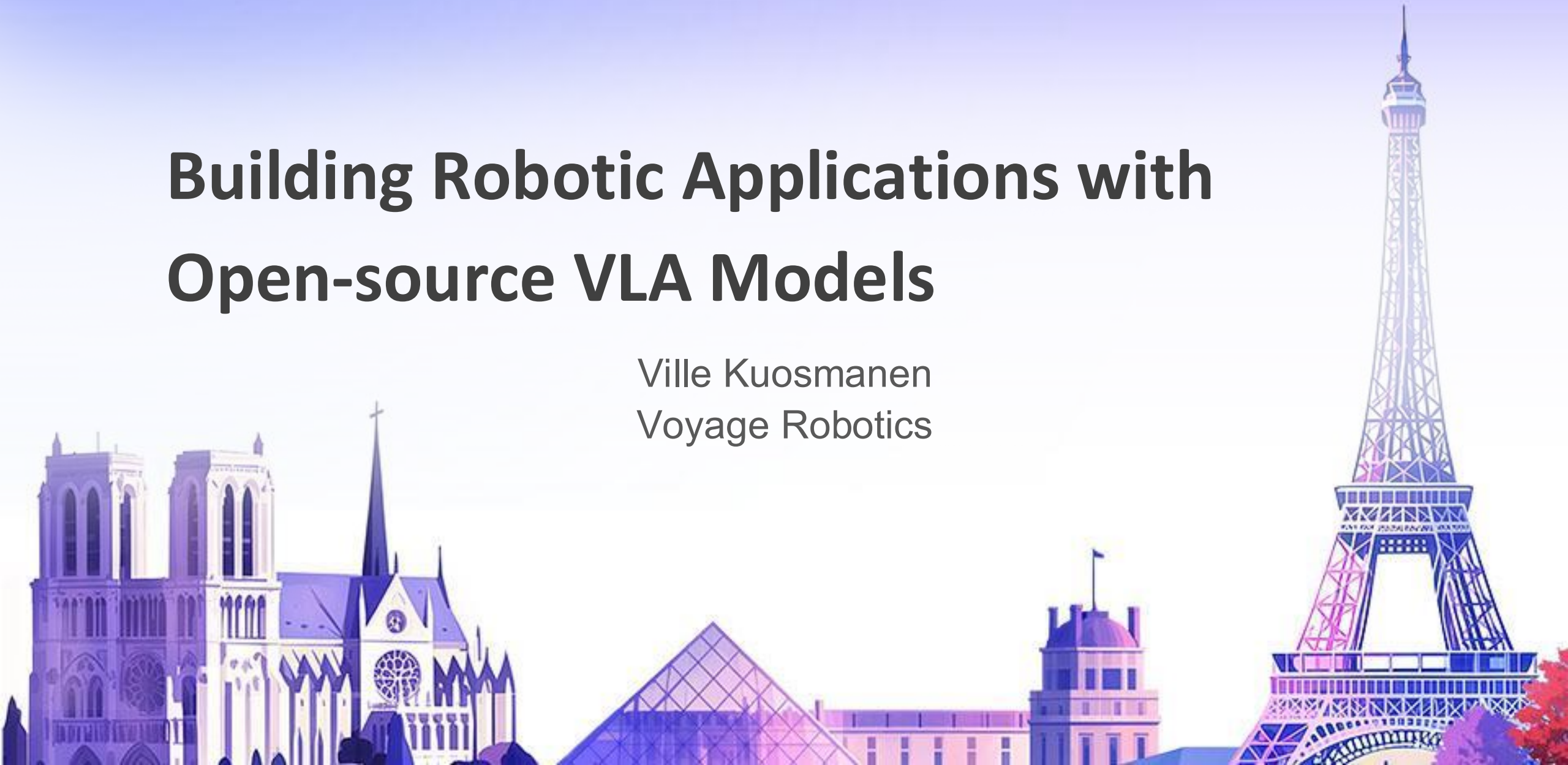

WE DO THIS
NOT BECAUSE
IT IS EASY,
BUT BECAUSE
WE THOUGHT
IT WOULD BE EASY

GOSIM

GOSIM AI Paris 2025

Building Robotic Applications with Open-source VLA Models

Ville Kuosmanen
Voyage Robotics

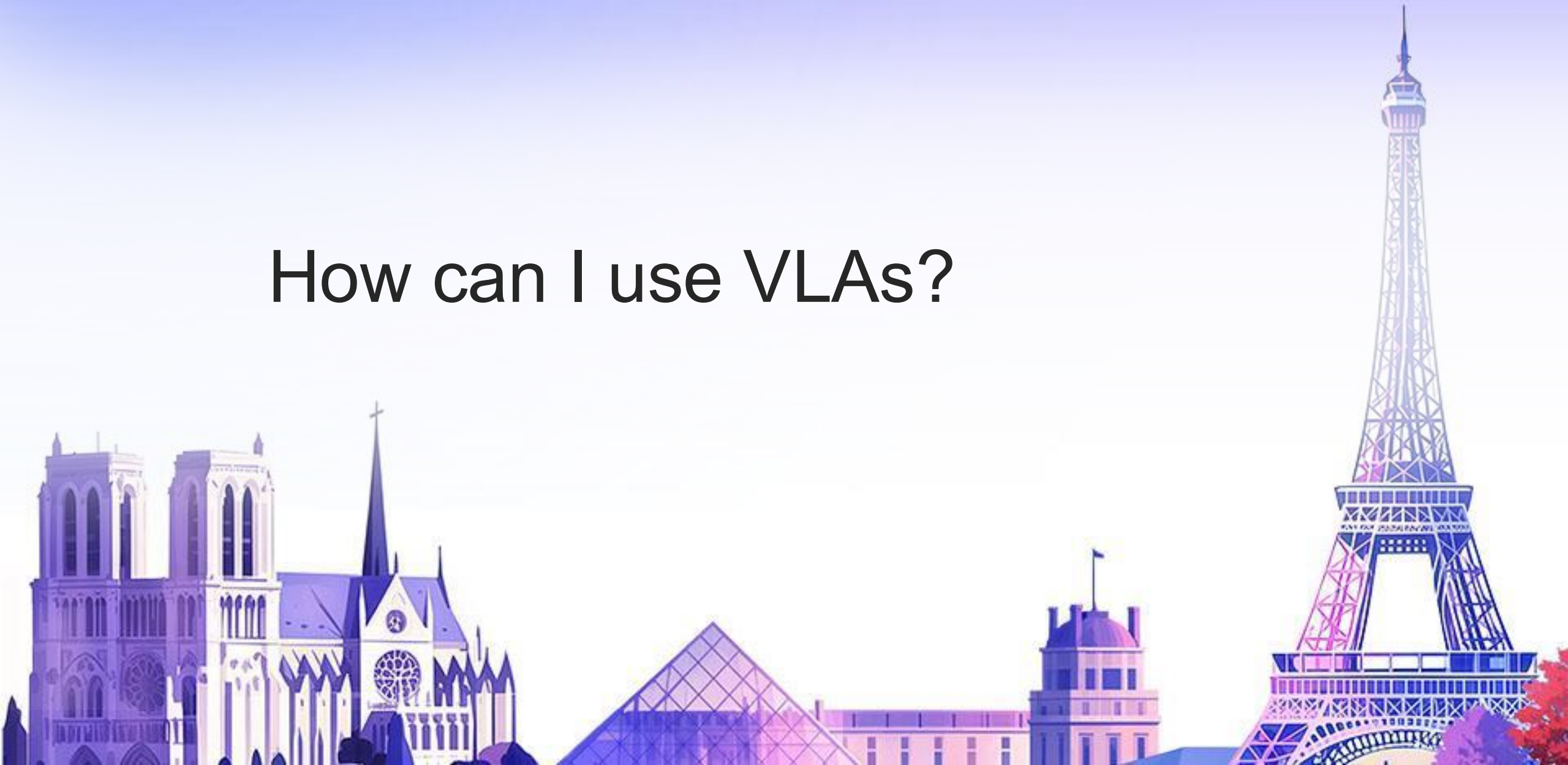


Engineered systems to end-to-end learning

GOSIM

E2e learning is eating robotics

How can I use VLAs?

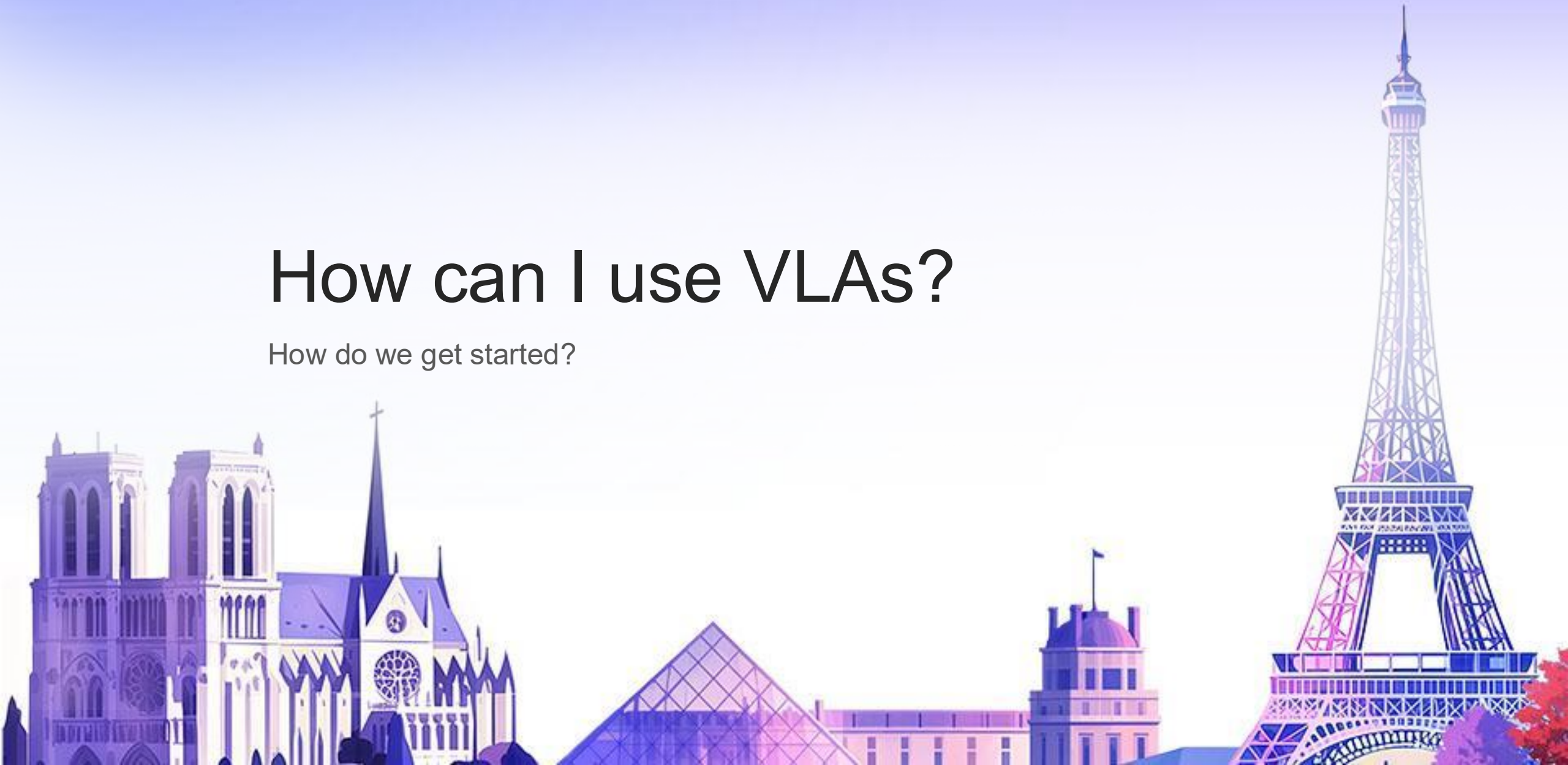


GOSIM

GOSIM AI Paris 2025

How can I use VLAs?

How do we get started?



First we need a robot 🤖

AI robotics is Python-driven

AI robotics is Python-driven

LeRobot

Fine-tuning VLAs

- Today’s VLAs don’t work zero-shot
- We need to **fine-tune** a pre-trained model with our own data

OS VLAs to choose from

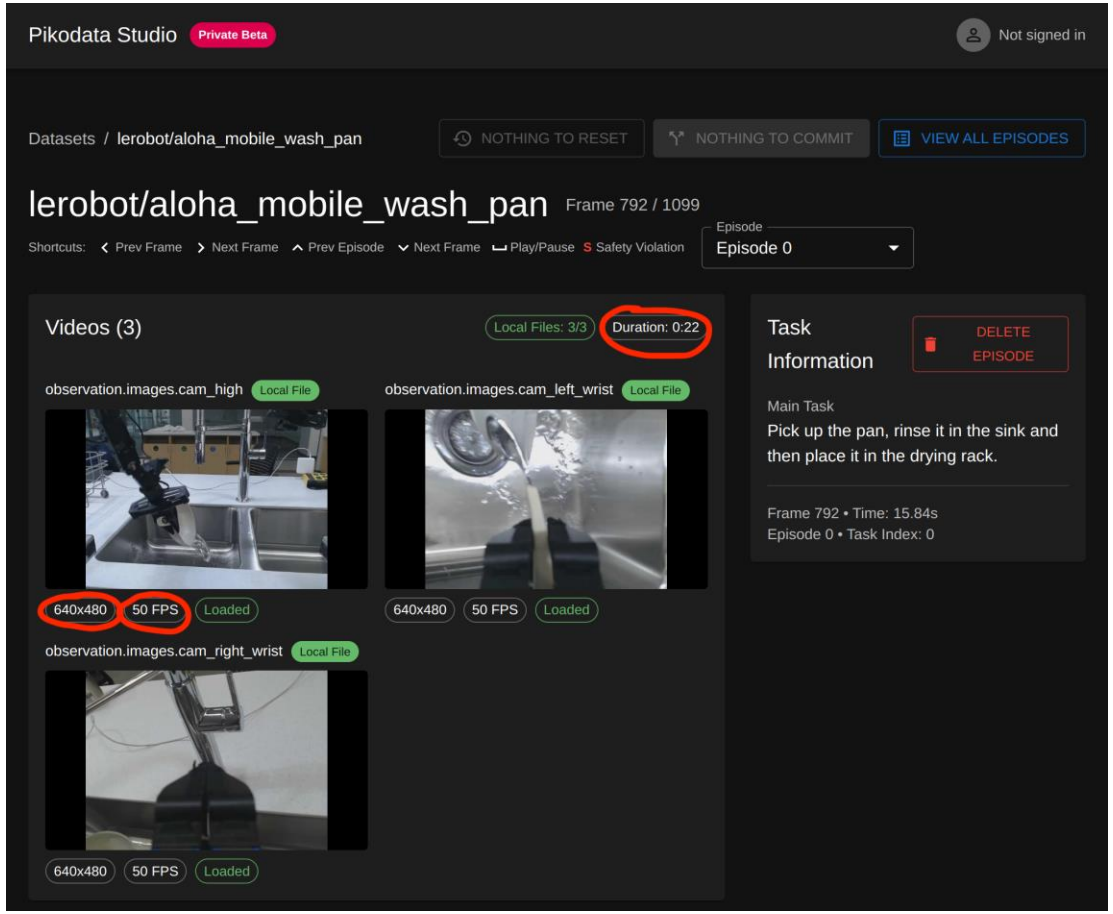
- Pi0 π
- GR00T N1
- RDT

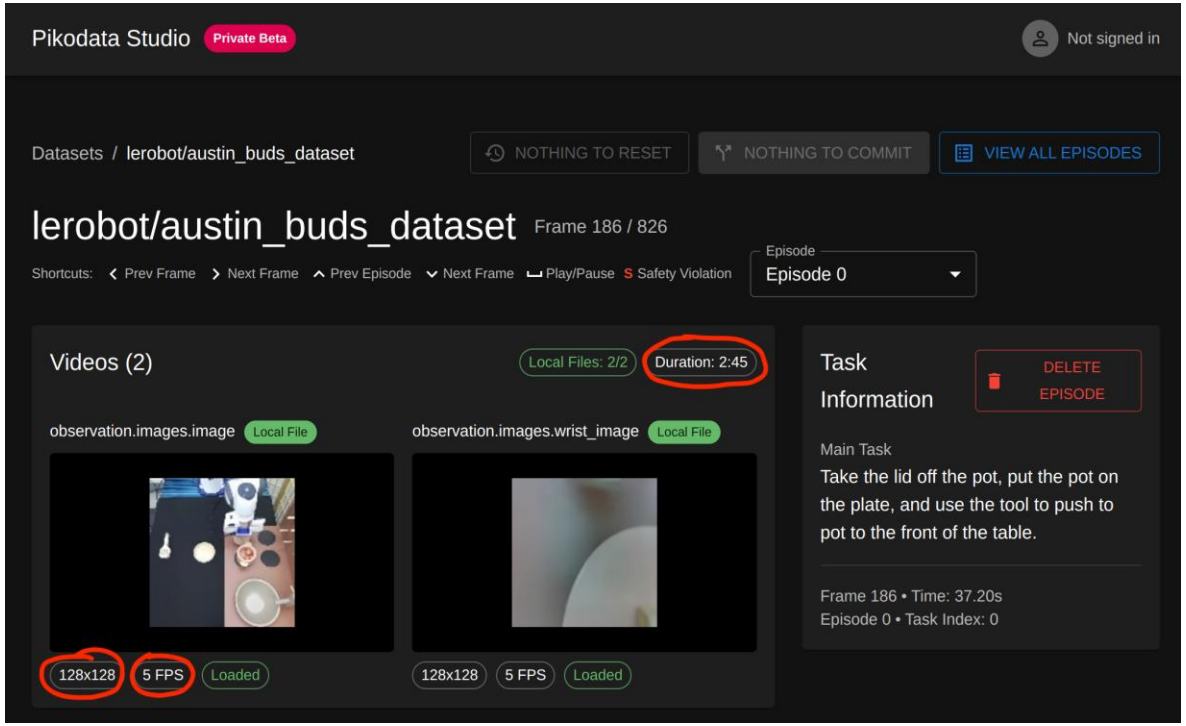
(many others…)

Robotics is a data problem

Robot data is highly valuable 🤑

Robot data is highly valuable 🤑



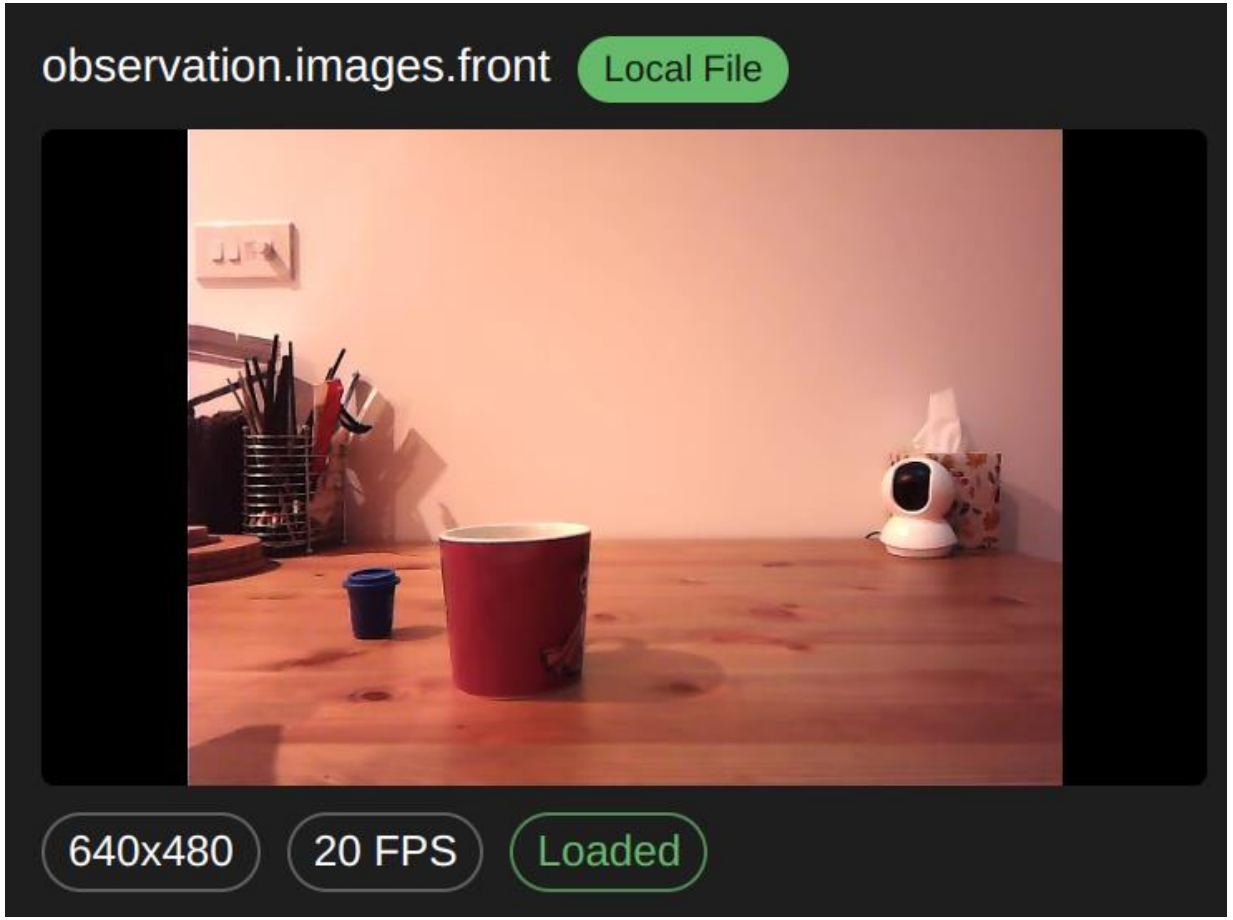


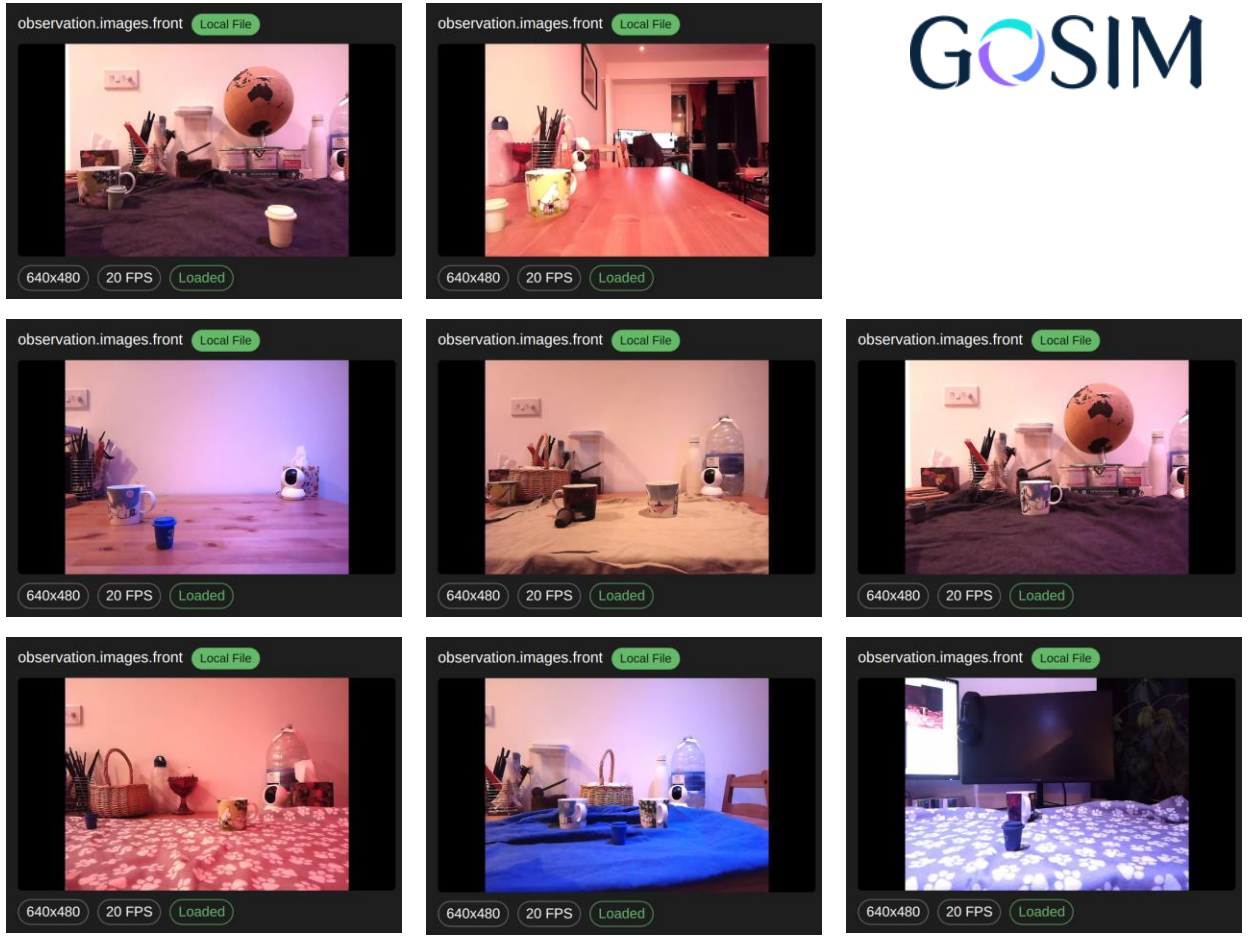
Characteristics of good robotics datasets

GOSIM

- High resolution (640 x 480 or better)
- High FPS (20 or better)
- Consistent FPS, no lag or frozen frames
- Fast movements
- Stable movements, no shaking
- Target objects visible in frames
- Human controller not visible in frames
- Wrist cameras showing manipulated object
- Not too dark

Data diversity





Metrics for dataset quality

GOSIM

- Value-order correlation

[1]Google DeepMind [2]University of Pennsylvania [3]Stanford University

- Demo-SCORE

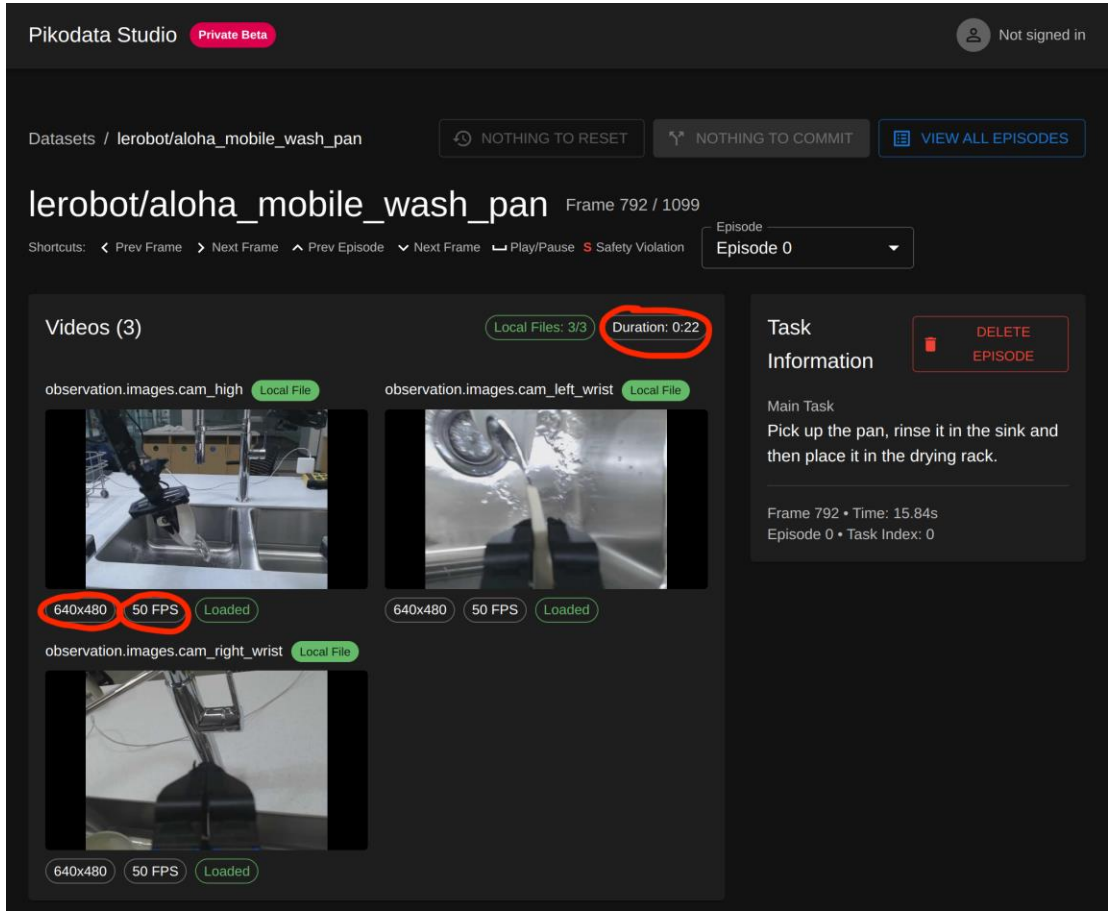
Stanford University

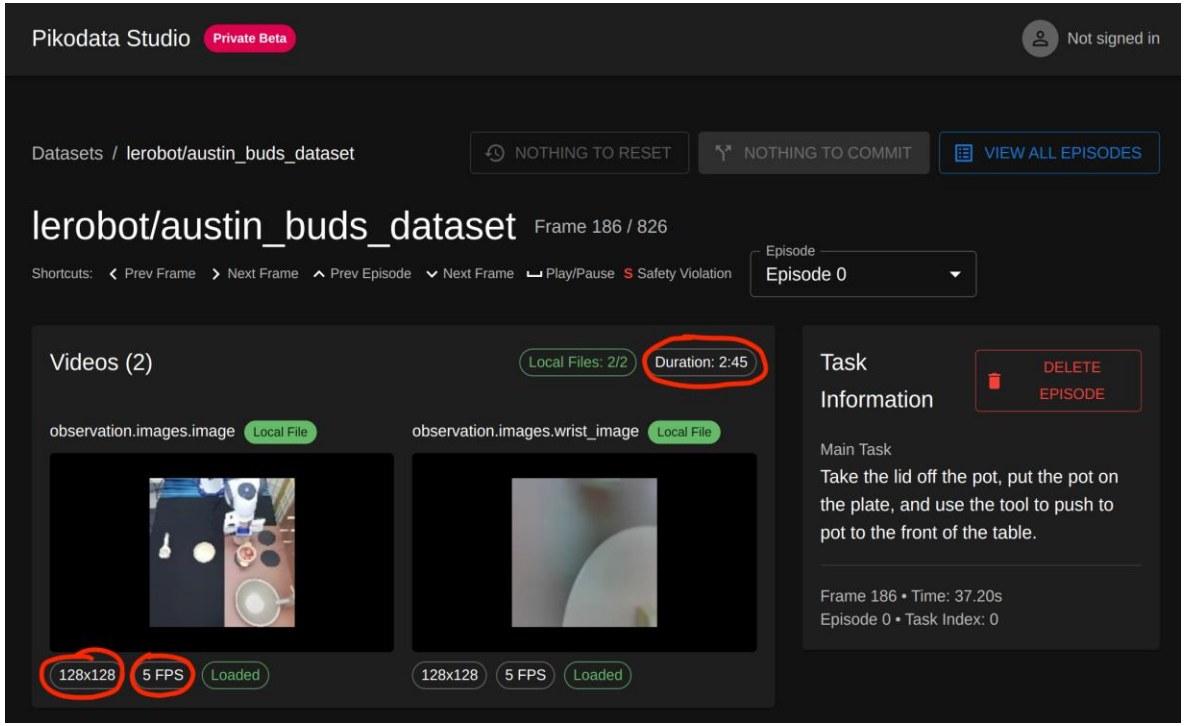
Dataset	Avg. VOC
RT-1 [5]	0.74
Dobb-E [54]	0.53
Bridge [62]	0.51
QT-OPT [28]	0.19
DROID [30]	-0.01
RoboNet [12]	-0.85

Table 1 | Average Value-Order Correlation (VOC) scores on selected OXE datasets. As shown, demonstration datasets with un-occluded camera views generally have high scores. In contrast, exploration datasets have low scores.

Robotics is a data problem

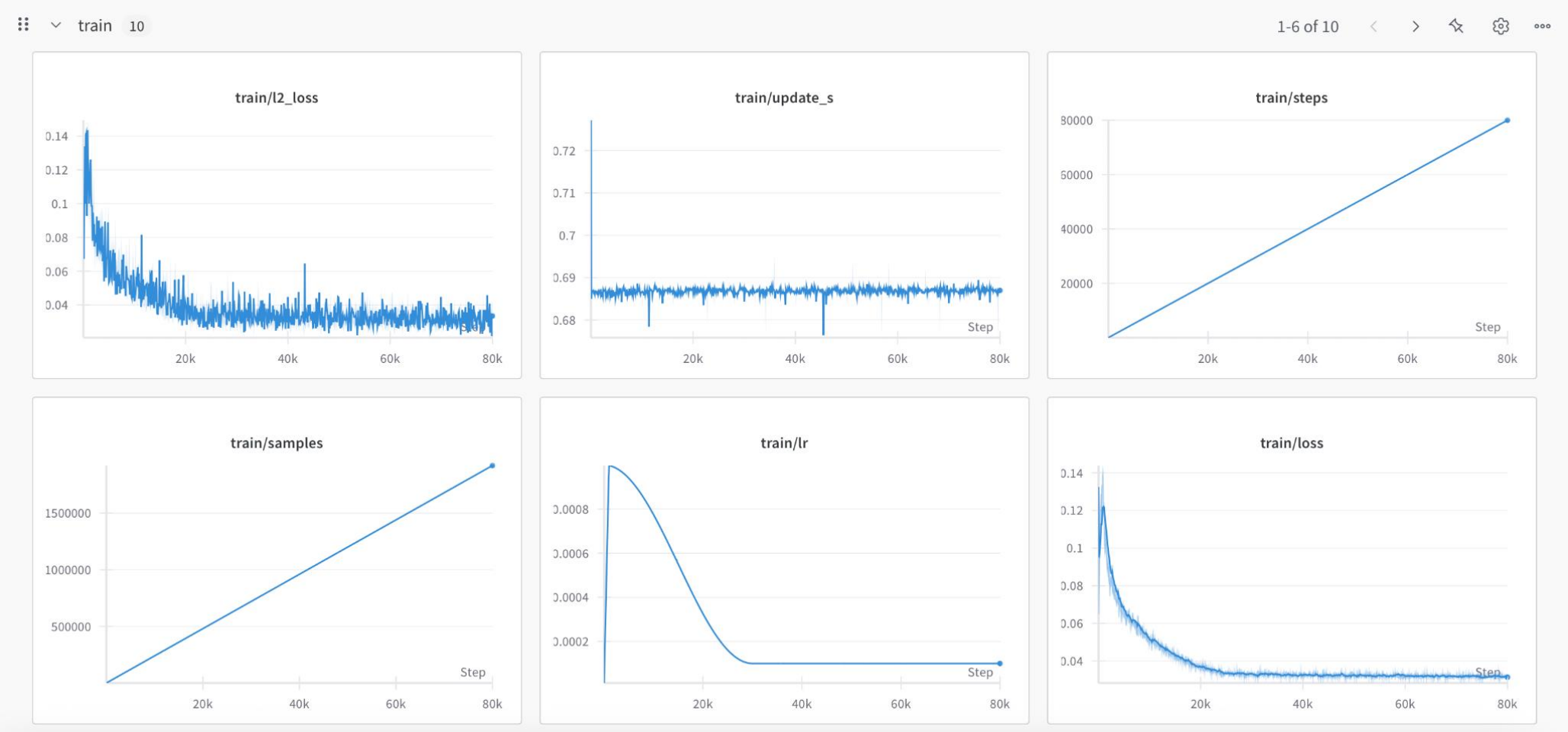
Robot data is highly valuable 🤑





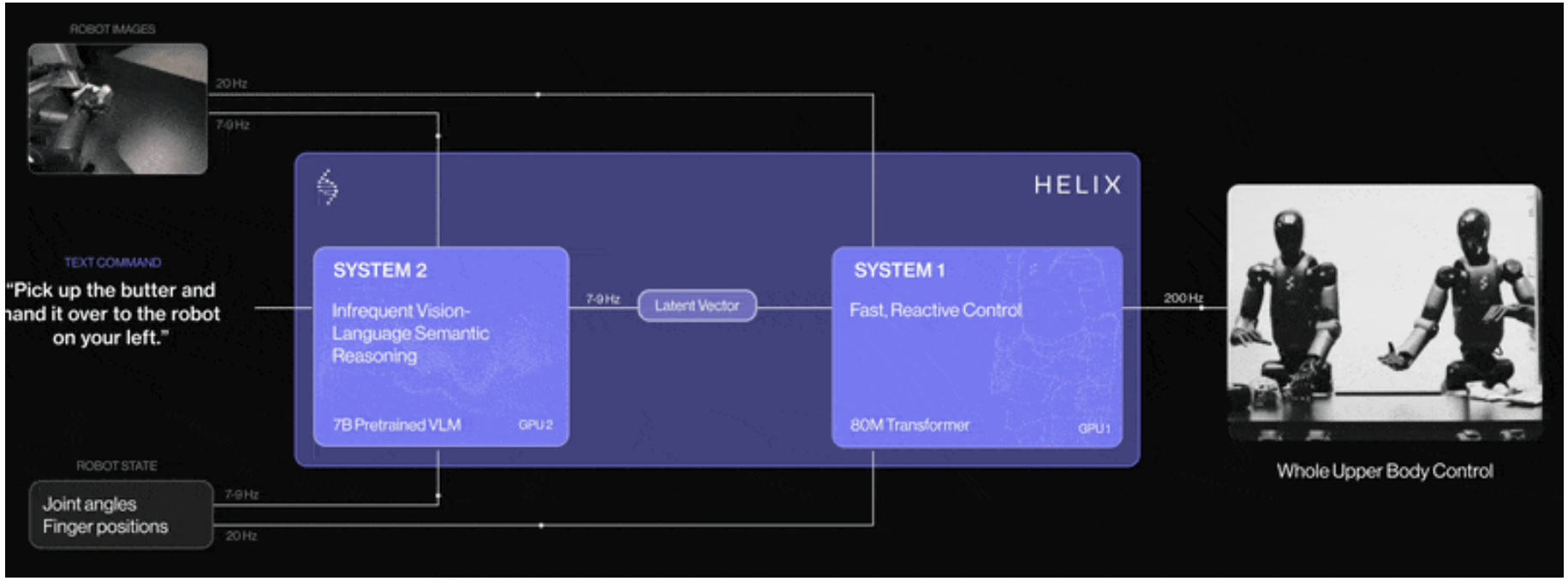
Training and inference

Training on the cloud



Inference

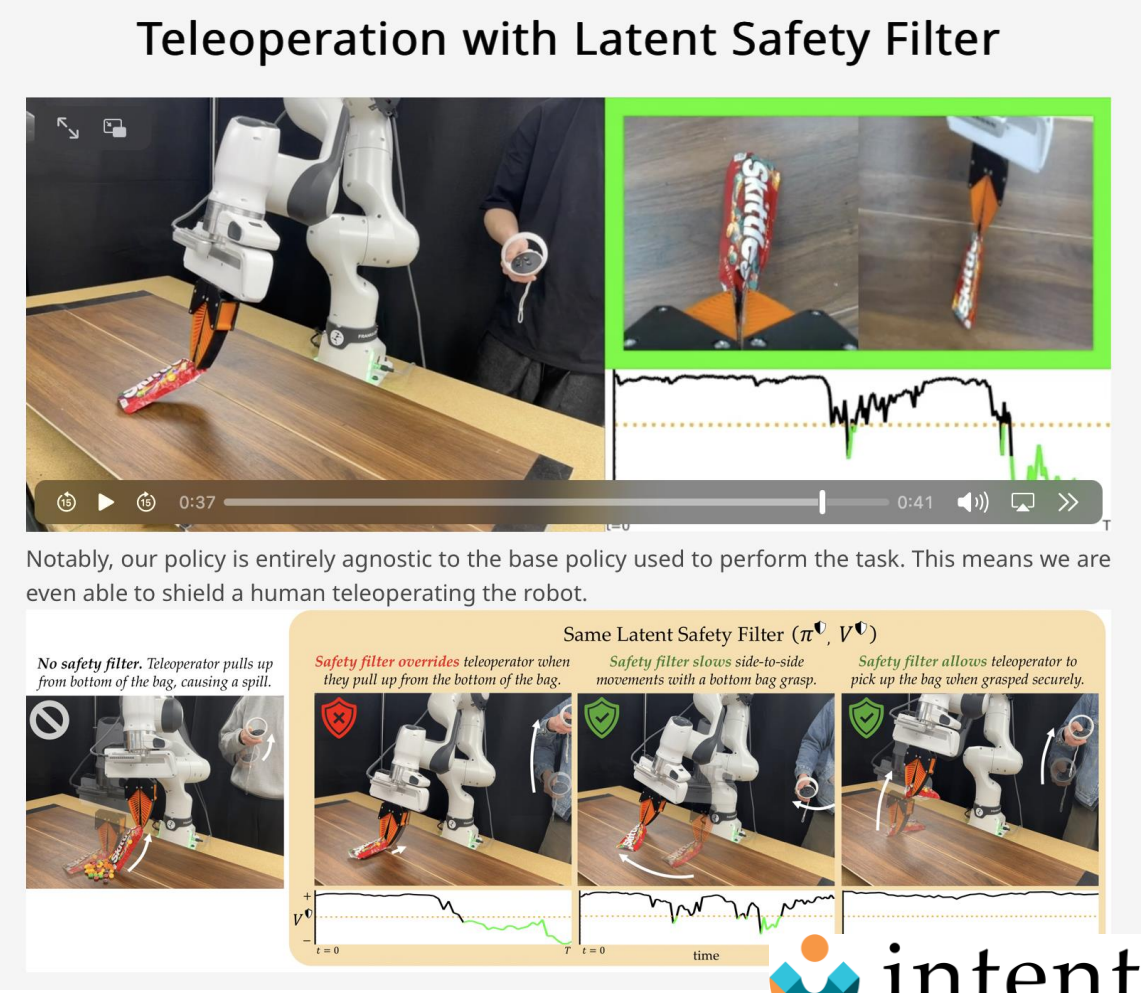
- We can ask a VLM model to prompt our VLA with a short term instruction



Pi0 demo

Supporting the VLA model

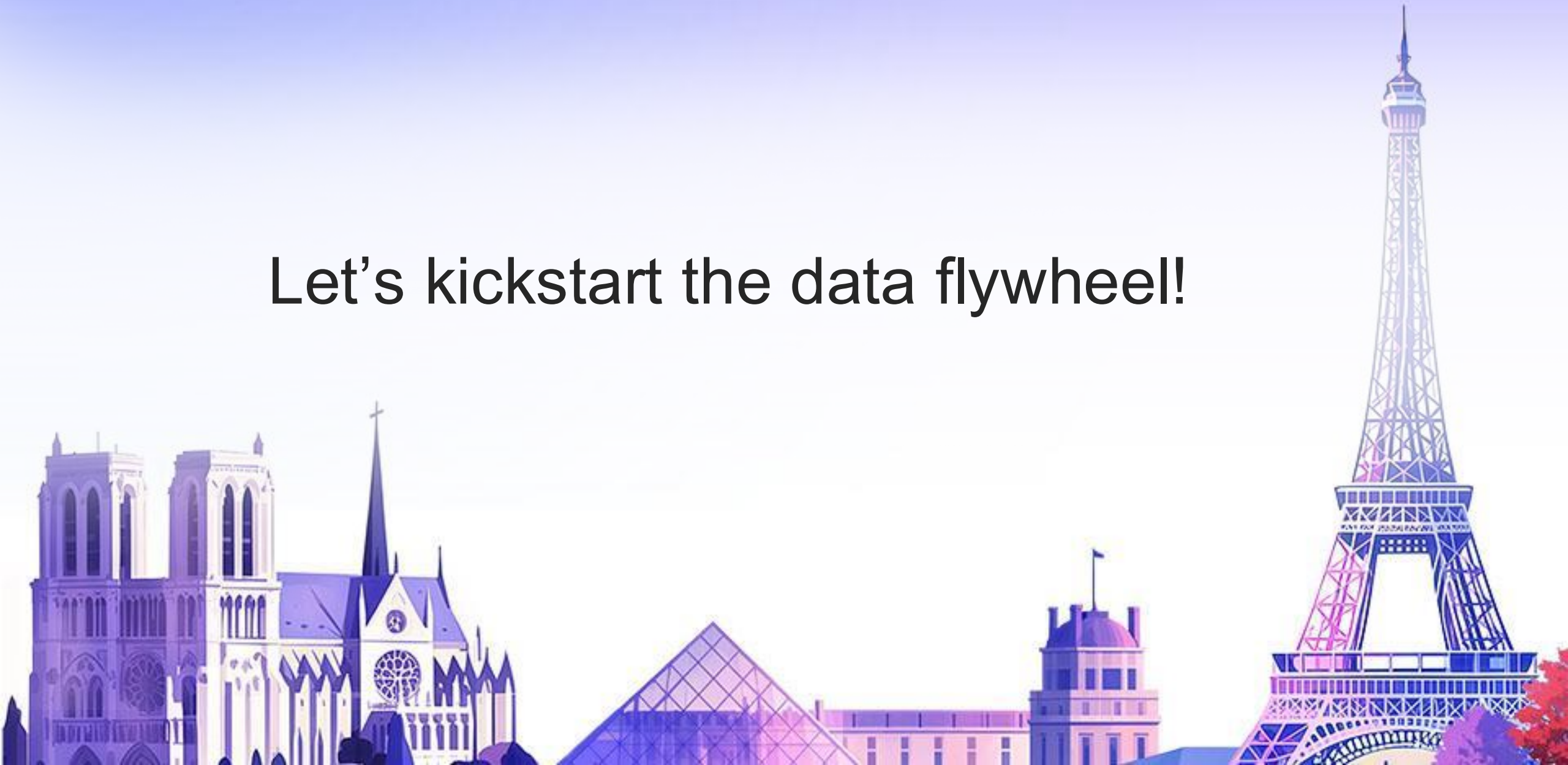
- Unable to tell when rollout is a success or failure
  - Use VLM to detect success / failure
- Safety and error detection
  - Augment imperfect model with human interventions



Verdict

- All major components of the VLA stack are open source
- Your job is to combine them
  - “Robotics is an integration problem”
- Real deployments produce real data…

Let’s kickstart the data flywheel!



GOSIM

GOSIM AI Paris 2025

THANK YOU