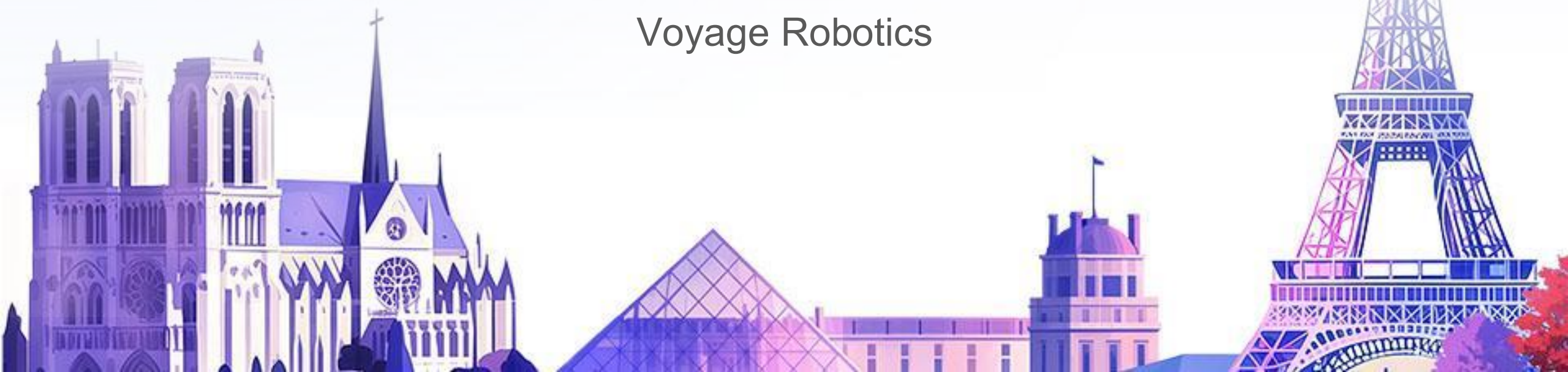
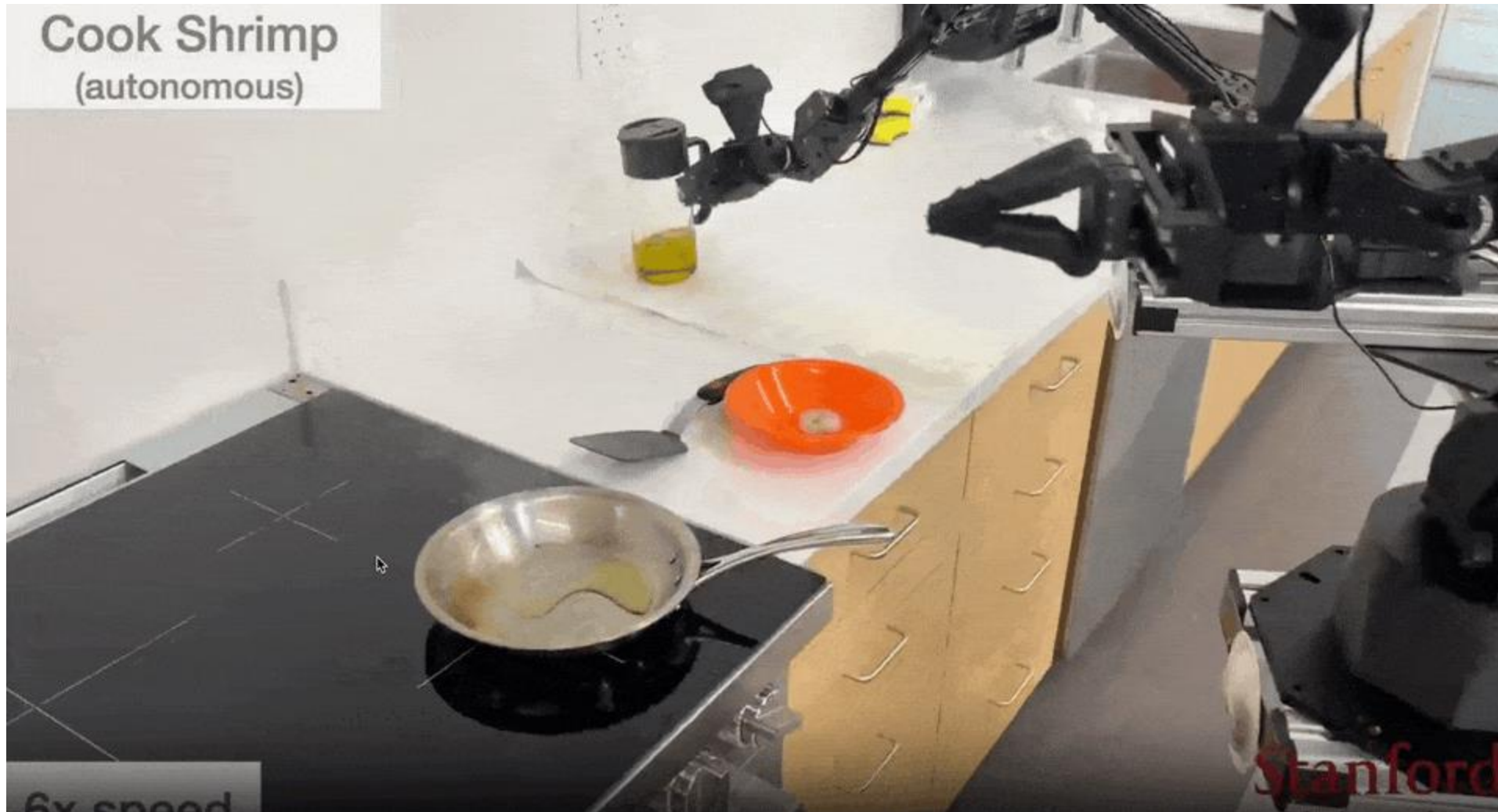


Building Robotic Applications with Open-source VLA Models

Ville Kuosmanen
Voyage Robotics







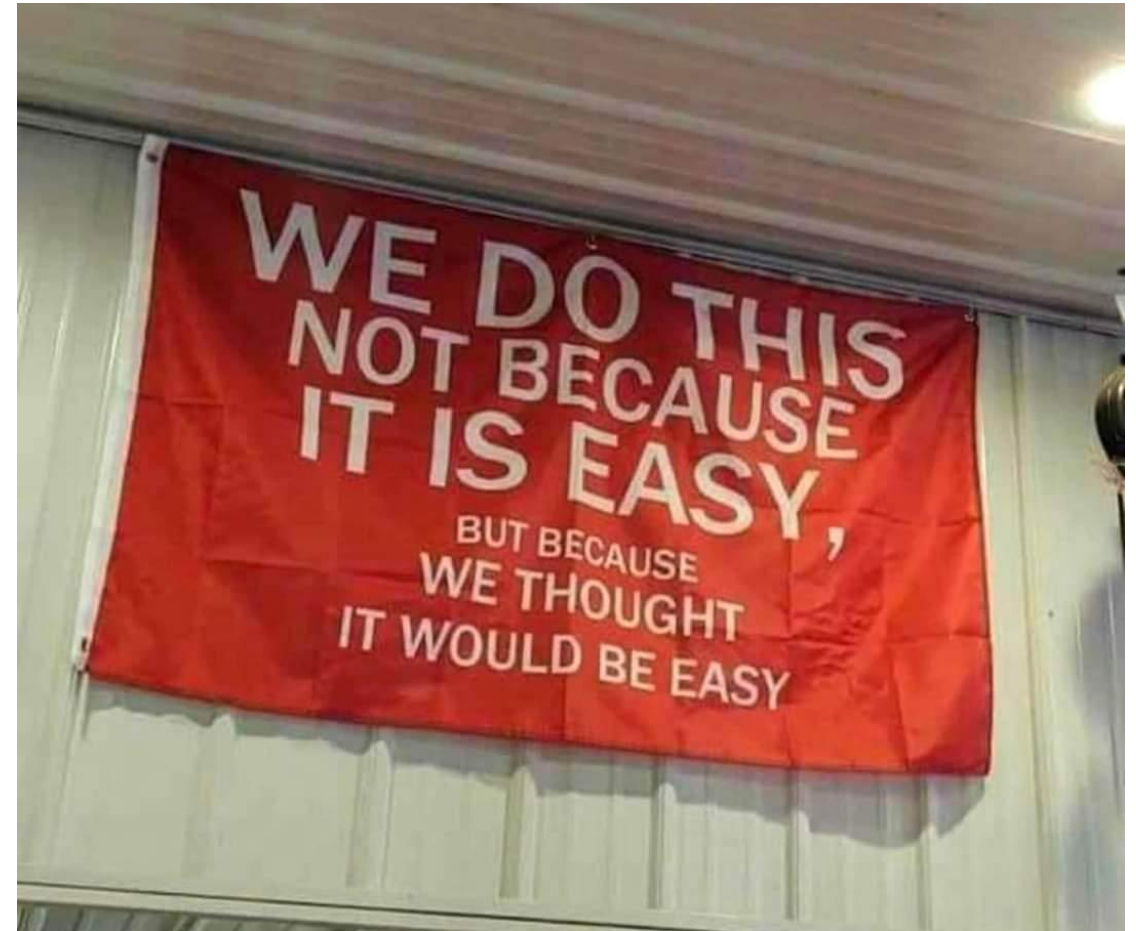
GOSIM

GOSIM AI Paris 2025





GOSIM

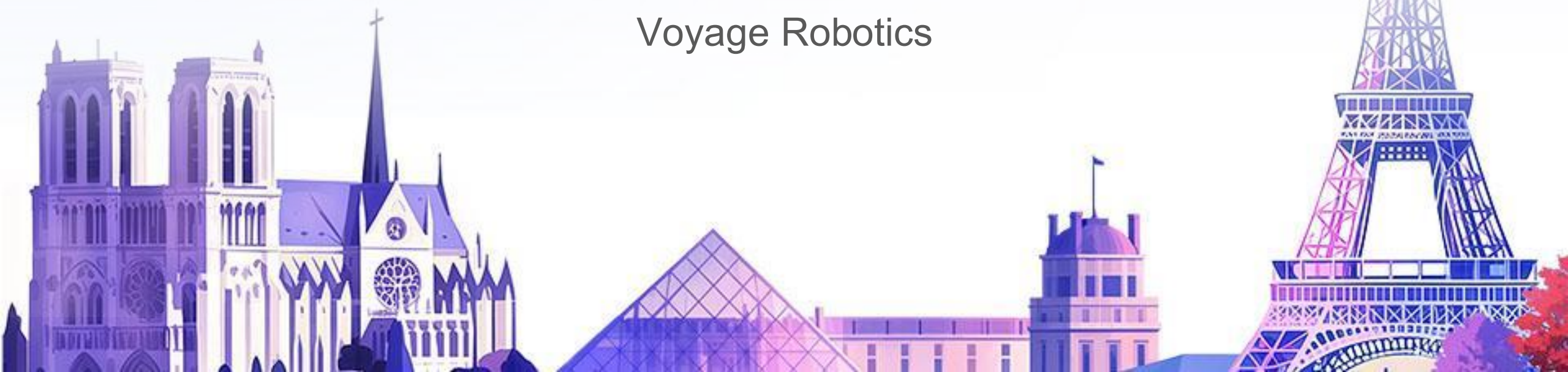


GOSIM AI Paris 2025



Building Robotic Applications with Open-source VLA Models

Ville Kuosmanen
Voyage Robotics



Engineered systems to end-to-end learning

TASKS

[<](#) [< PREV](#) [#1425 \(SEPTEMBER 24, 2014\)](#) [NEXT >](#) [>](#)



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.



E2e learning is eating robotics

GOSIM



GOSIM AI Paris 2025



How can I use VLAs?



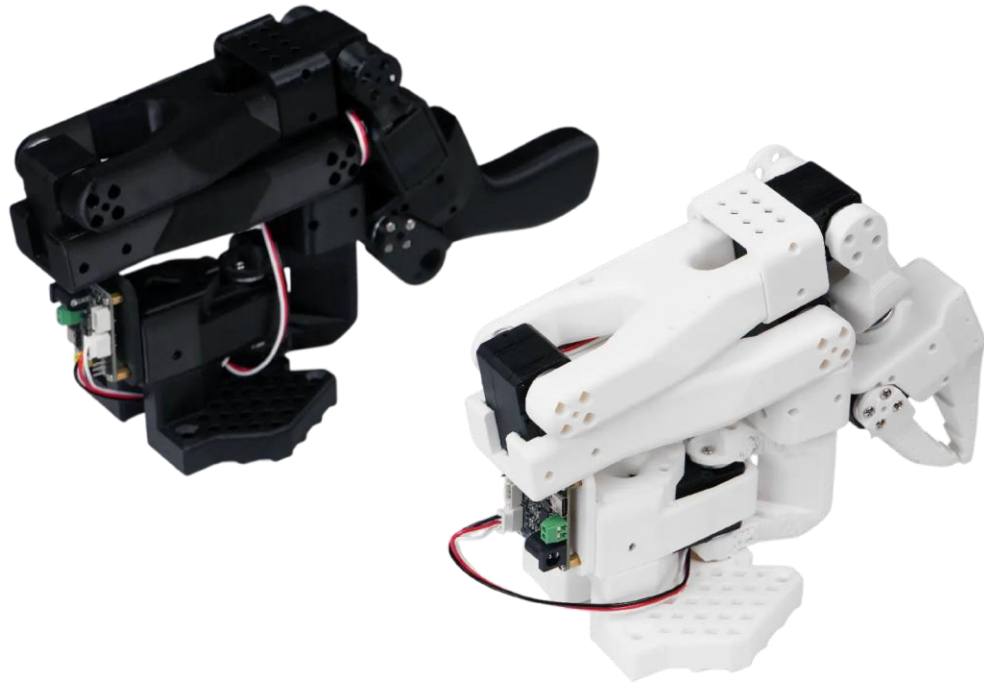
How can I use VLAs?

How do we get started?



First we need a robot 🤖

GOSIM



GOSIM AI Paris 2025



AI robotics is Python-driven



GOSIM

GOSIM AI Paris 2025



AI robotics is Python-driven



GOSIM

LeRobot

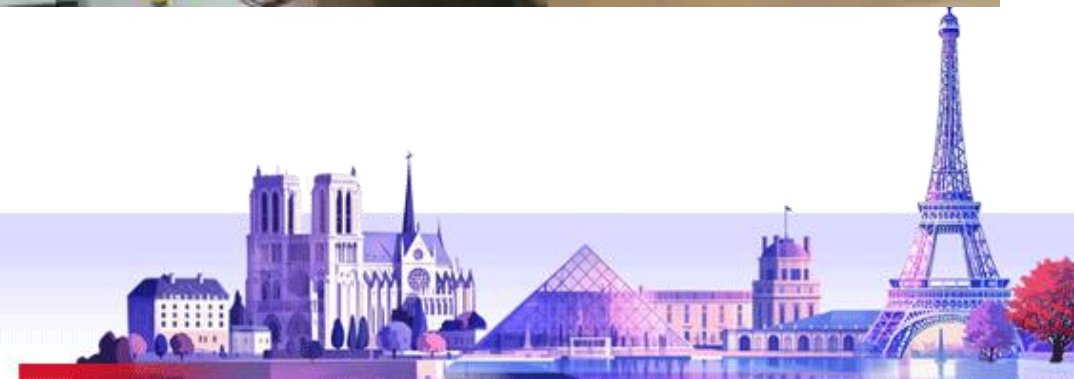


GOSIM AI Paris 2025



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 🤖💰

GOSIM

GOSIM AI Paris 2025



Robot data is highly valuable



GOSIM

Pikodata Studio Private Beta Not signed in

Datasets / lerobot/aloha_mobile_wash_pan

Frame 792 / 1099

Shortcuts: < Prev Frame > Next Frame ^ Prev Episode v Next Episode ⏮ Play/Pause ⚠ Safety Violation

Episode 0

Videos (3) Local Files: 3/3 Duration: 0:22

observation.images.cam_high Local File

640x480 50 FPS Loaded

observation.images.cam_left_wrist Local File

640x480 50 FPS Loaded

observation.images.cam_right_wrist Local File

640x480 50 FPS Loaded

Task Information

Main Task

Pick up the pan, rinse it in the sink and then place it in the drying rack.

Frame 792 • Time: 15.84s
Episode 0 • Task Index: 0

DELETE EPISODE

Pikodata Studio Private Beta Not signed in

Datasets / lerobot/austin_buds_dataset

Frame 186 / 826

Shortcuts: < Prev Frame > Next Frame ^ Prev Episode v Next Episode ⏮ Play/Pause ⚠ Safety Violation

Episode 0

Videos (2) Local Files: 2/2 Duration: 2:45

observation.images.image Local File

128x128 5 FPS Loaded

observation.images.wrist_image Local File

128x128 5 FPS Loaded

Task Information

Main Task

Take the lid off the pot, put the pot on the plate, and use the tool to push to pot to the front of the table.

Frame 186 • Time: 37.20s
Episode 0 • Task Index: 0

DELETE EPISODE

GOSIM AI Paris 2025



Characteristics of good robotics datasets

- 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

GOSIM

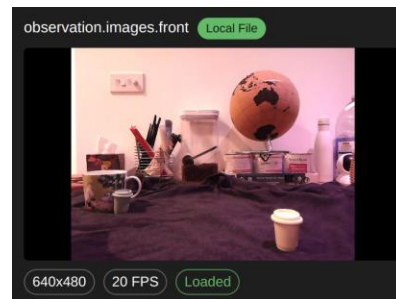
observation.images.front Local File



640x480

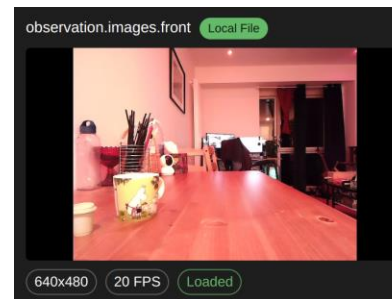
20 FPS

Loaded



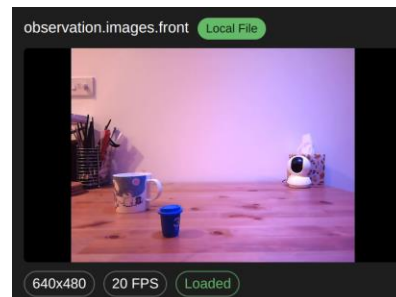
observation.images.front Local File

640x480 20 FPS Loaded



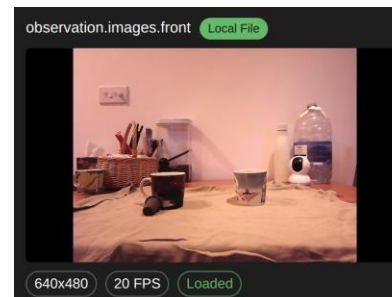
observation.images.front Local File

640x480 20 FPS Loaded



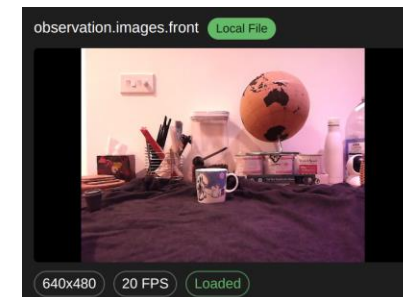
observation.images.front Local File

640x480 20 FPS Loaded



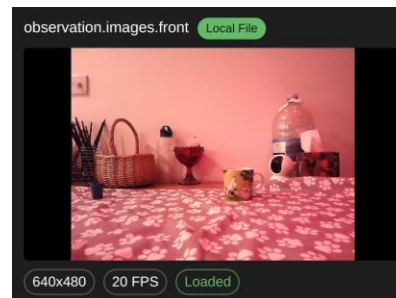
observation.images.front Local File

640x480 20 FPS Loaded



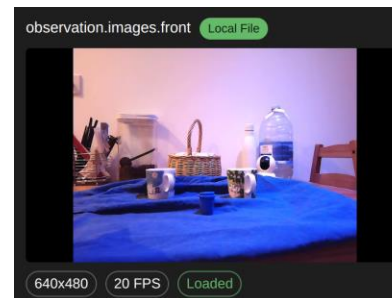
observation.images.front Local File

640x480 20 FPS Loaded



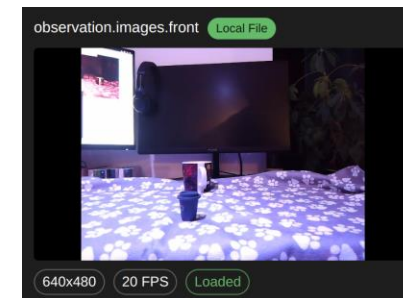
observation.images.front Local File

640x480 20 FPS Loaded



observation.images.front Local File

640x480 20 FPS Loaded



observation.images.front Local File

640x480 20 FPS Loaded

GOSIM AI Paris 2025



Metrics for dataset quality

- Value-order correlation

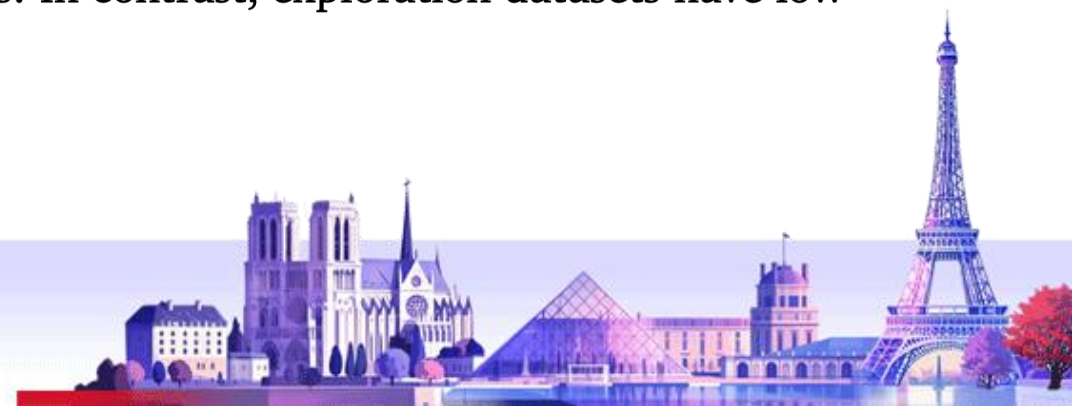
 ¹Google DeepMind  ²University of Pennsylvania  ³Stanford University

- 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



GOSIM

Pikodata Studio Private Beta Not signed in

Datasets / lerobot/aloha_mobile_wash_pan

Frame 792 / 1099

Shortcuts: < Prev Frame > Next Frame ^ Prev Episode v Next Episode ⏸ Play/Pause ⚠ Safety Violation

Episode 0

Videos (3) Local Files: 3/3 Duration: 0:22

observation.images.cam_high Local File

640x480 50 FPS Loaded

observation.images.cam_left_wrist Local File

640x480 50 FPS Loaded

observation.images.cam_right_wrist Local File

640x480 50 FPS Loaded

Task Information

Main Task

Pick up the pan, rinse it in the sink and then place it in the drying rack.

Frame 792 • Time: 15.84s
Episode 0 • Task Index: 0

DELETE EPISODE

Pikodata Studio Private Beta Not signed in

Datasets / lerobot/austin_buds_dataset

Frame 186 / 826

Shortcuts: < Prev Frame > Next Frame ^ Prev Episode v Next Episode ⏸ Play/Pause ⚠ Safety Violation

Episode 0

Videos (2) Local Files: 2/2 Duration: 2:45

observation.images.image Local File

128x128 5 FPS Loaded

observation.images.wrist_image Local File

128x128 5 FPS Loaded

Task Information

Main Task

Take the lid off the pot, put the pot on the plate, and use the tool to push to pot to the front of the table.

Frame 186 • Time: 37.20s
Episode 0 • Task Index: 0

DELETE EPISODE

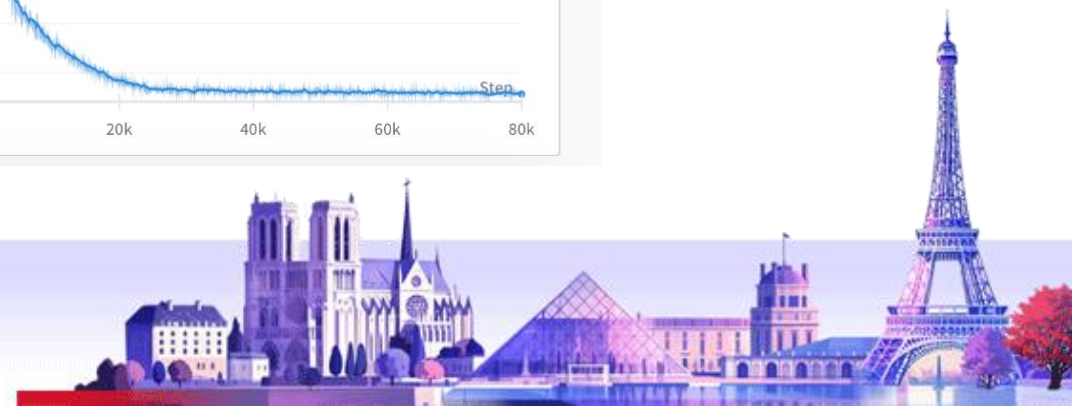
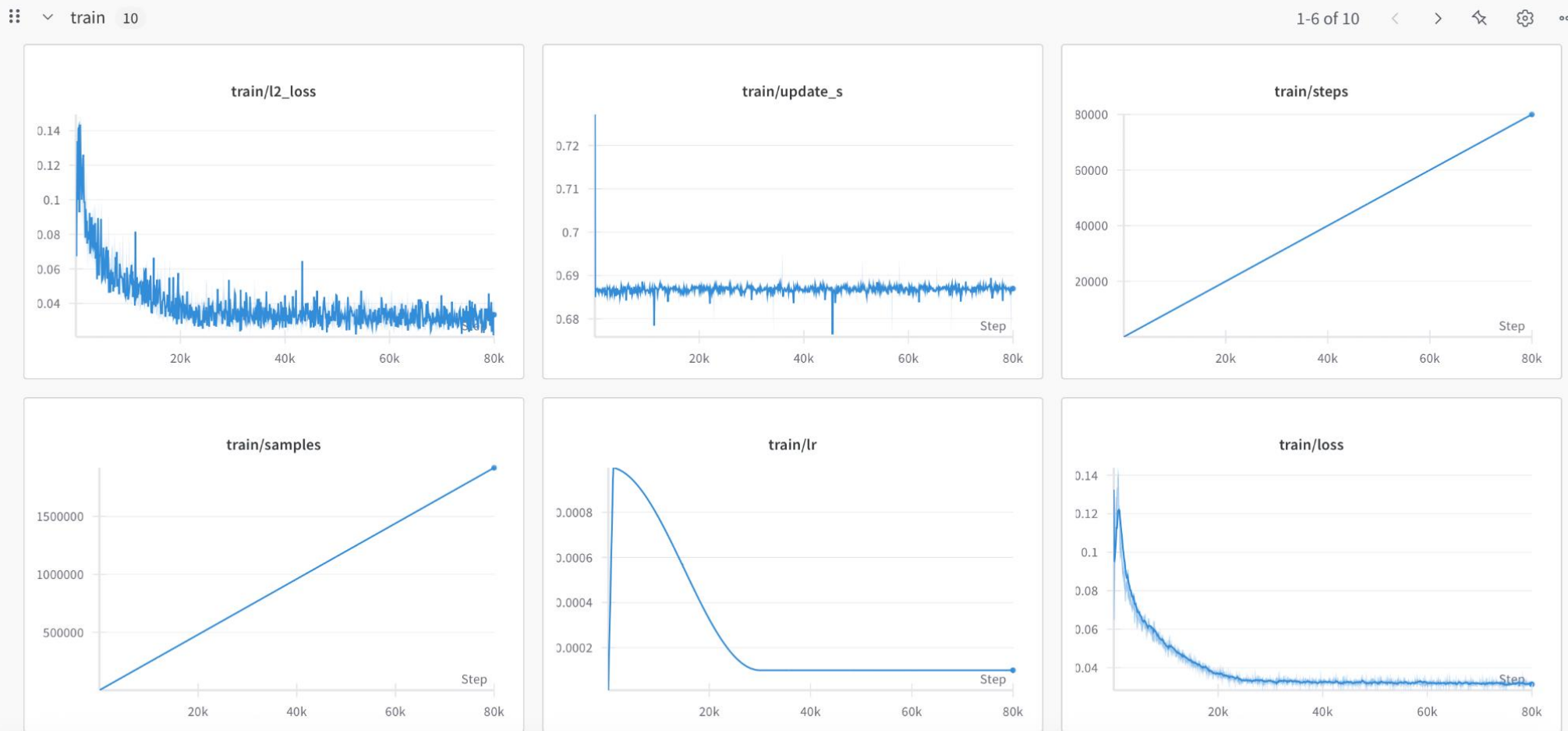
GOSIM AI Paris 2025



Training and inference

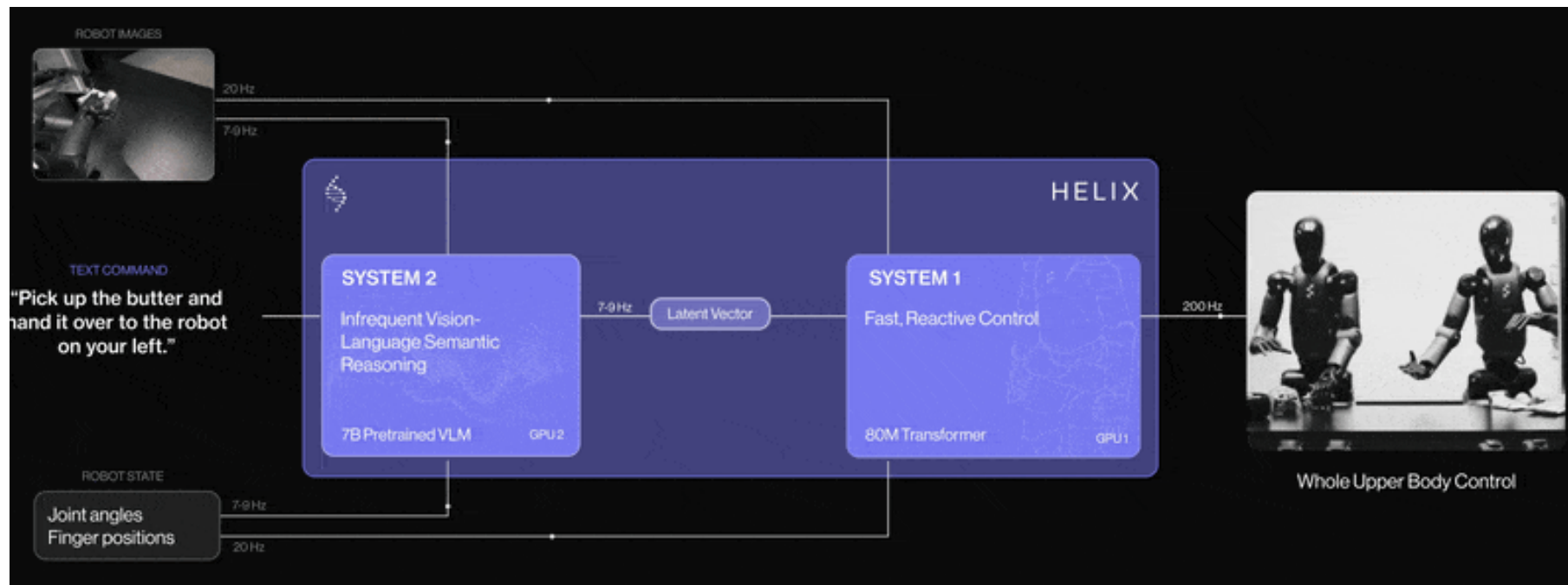


Training on the cloud



Inference

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



Pi0 demo

GOSIM



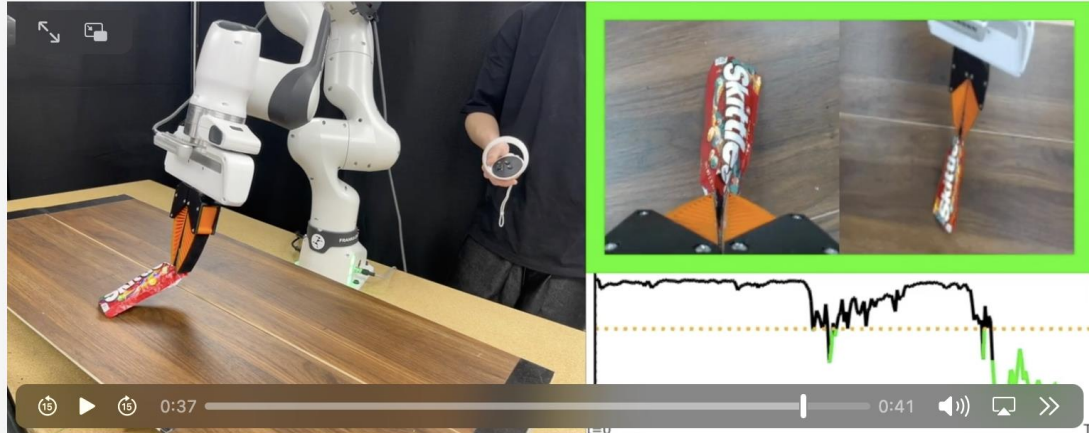
GOSIM AI Paris 2025



Supporting the VLA model



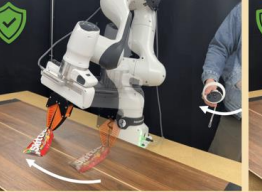
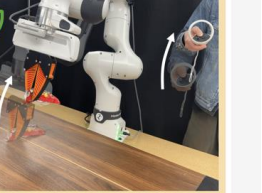


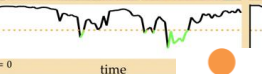

- Unable to tell when rollout is a success or failure
 - Use VLM to detect success / failure
- Safety and error detection
 - Augment imperfect model with human interventions

Teleoperation with Latent Safety Filter



Notably, our policy is entirely agnostic to the base policy used to perform the task. This means we are even able to shield a human teleoperating the robot.

Same Latent Safety Filter (π^0, V^0)

No safety filter. Teleoperator pulls up from bottom of the bag, causing a spill.	Safety filter overrides teleoperator when they pull up from the bottom of the bag.	Safety filter slows side-to-side movements with a bottom bag grasp.	Safety filter allows teleoperator to pick up the bag when grasped securely.
			
			



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!



THANK YOU

