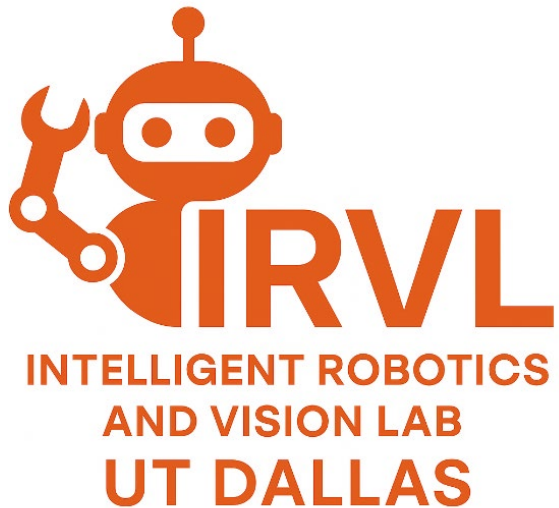


Vision-Language-Action Models: The Next Frontier for AI and High- Performance Computing



Yu Xiang

Assistant Professor

Intelligent Robotics and Vision Lab

Computer Science

9/16/2026

Comet Computing Conference 2026

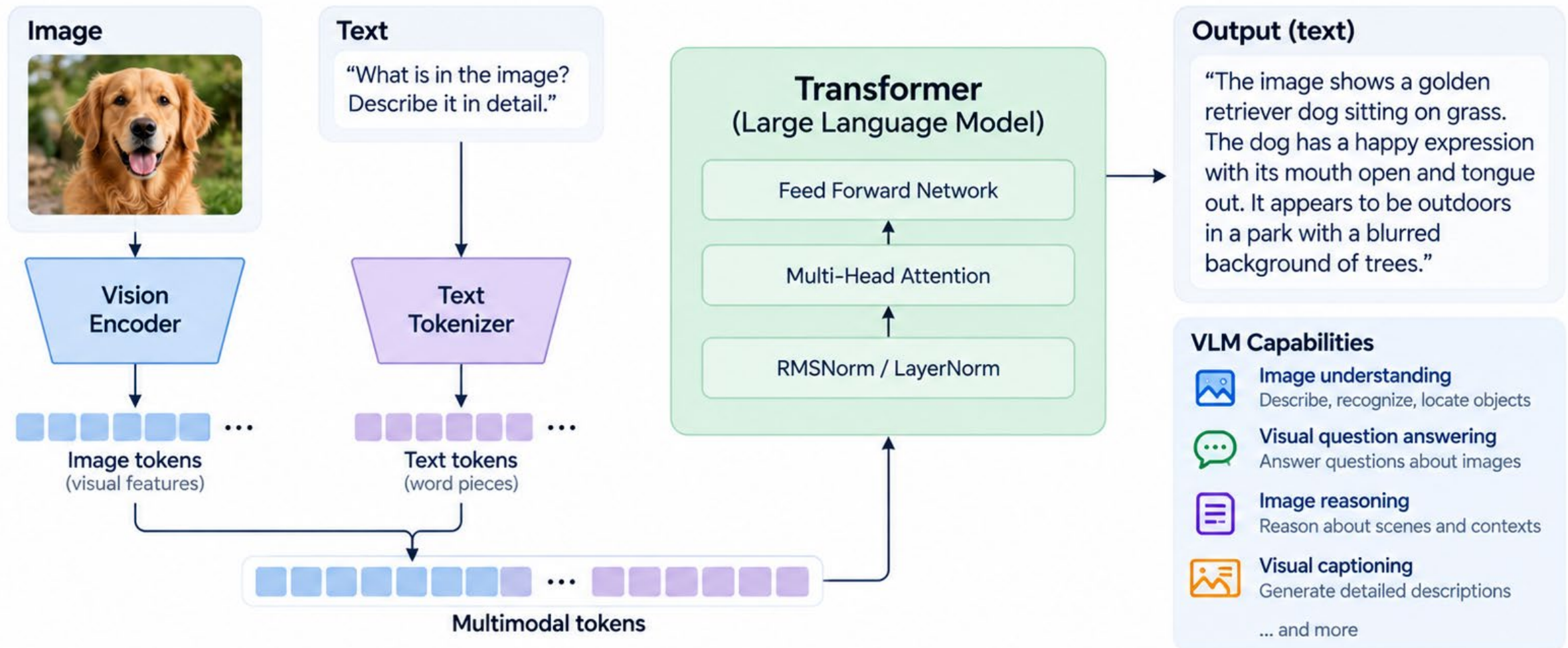
The current AI revolution started with language

- Large Language Models (LLMs)



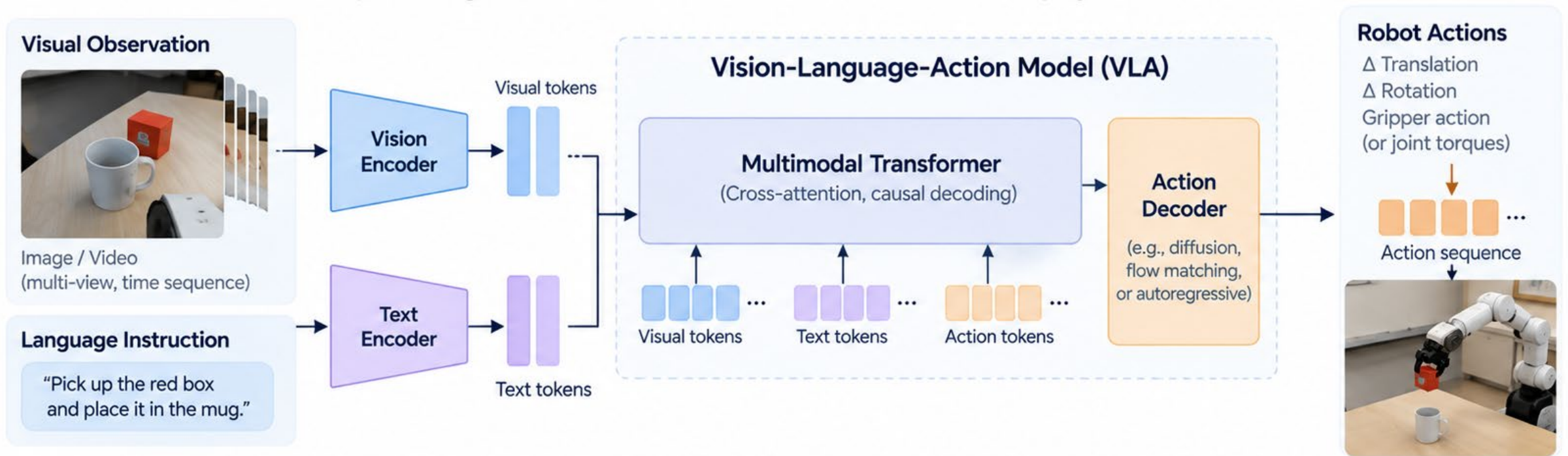
Then AI learned to see

- Vision-Language Models (VLMs)



Then AI learned to act

- Vision-Language-Action Models (VLAs)



Example Robot Manipulation Tasks



Pick and place



Open a drawer



Plug insertion



Pouring

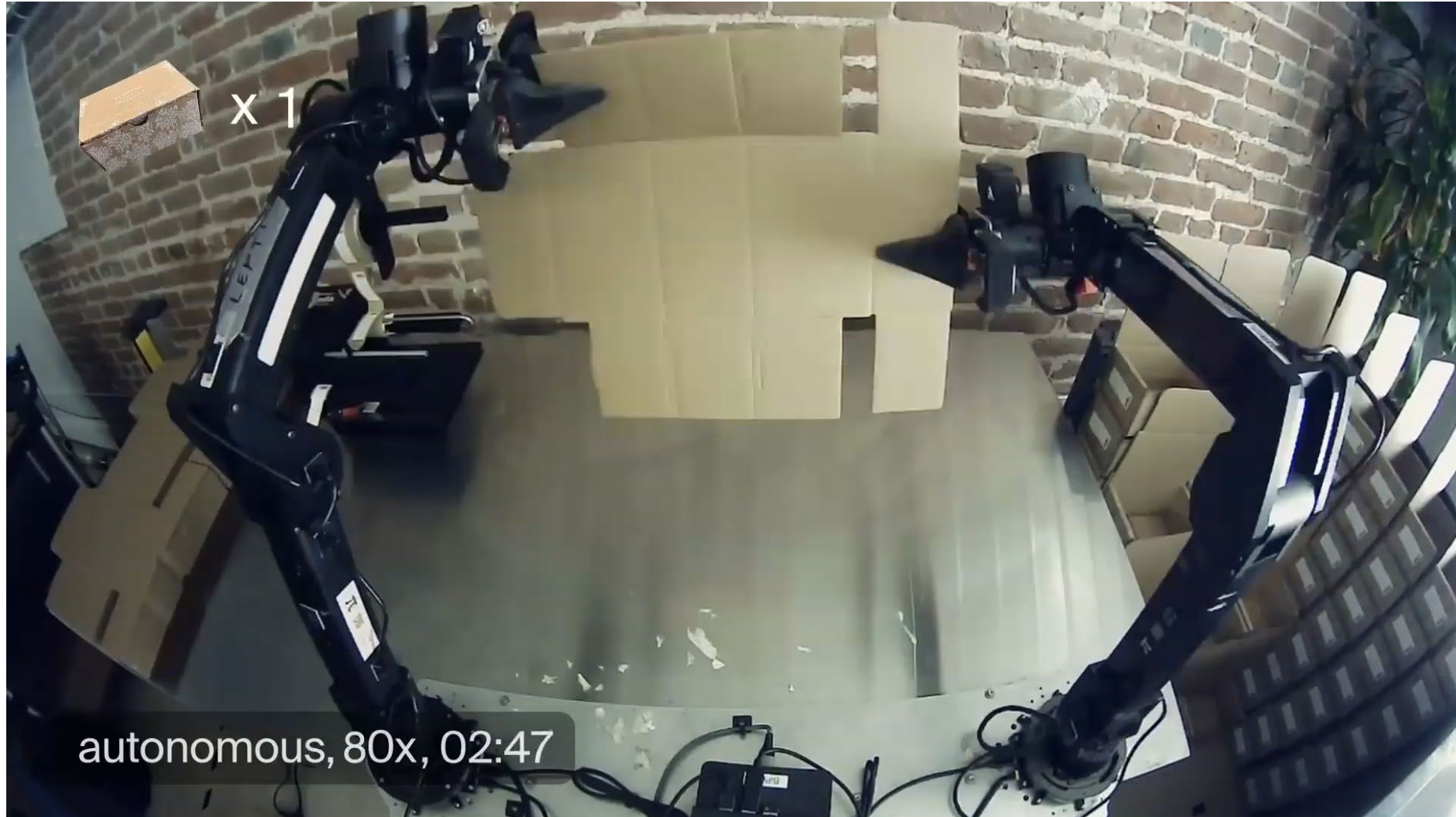


Stacking



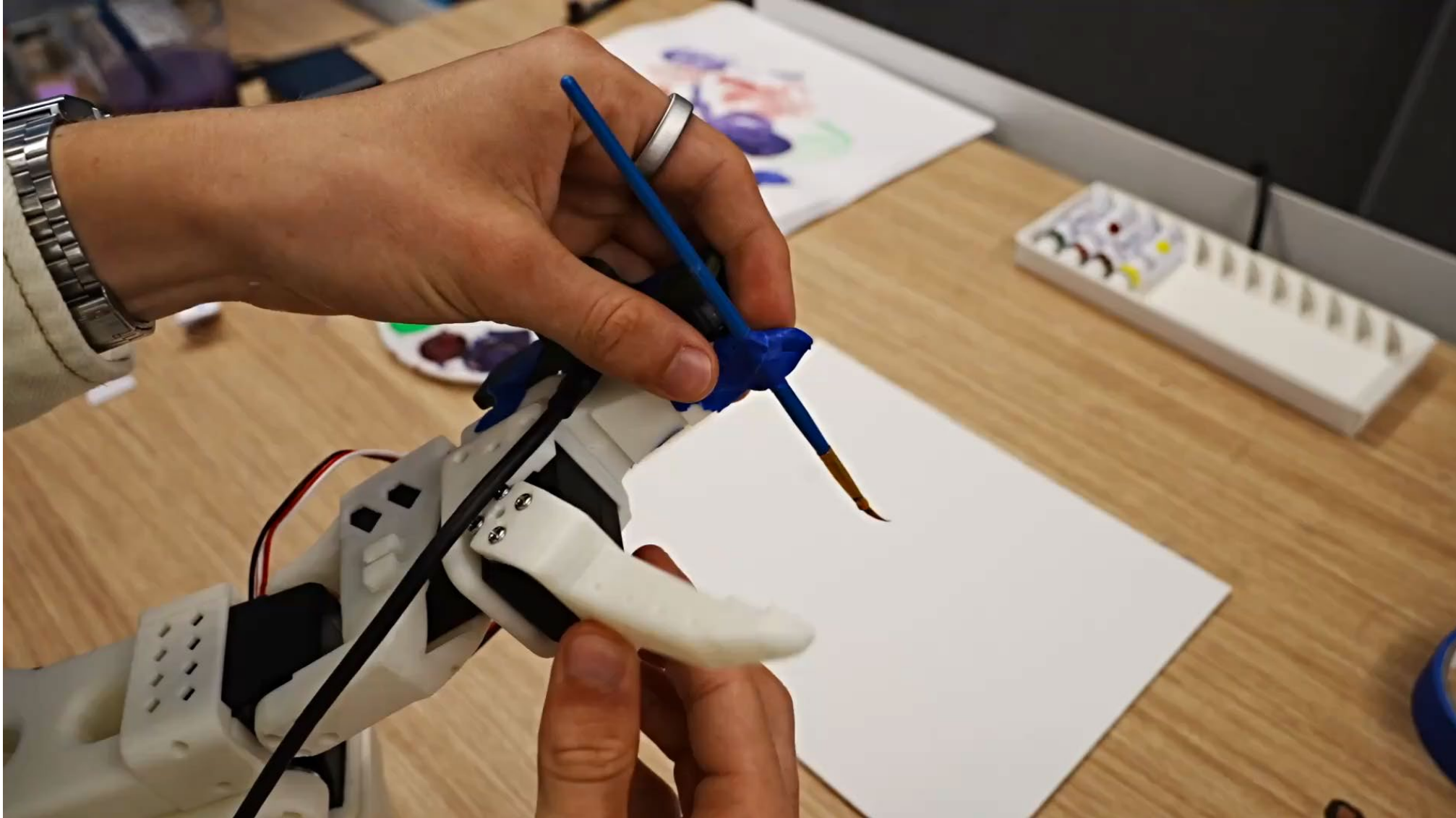
Fold cloth

VLA Example: Physical Intelligence



<https://www.pi.website/blog/pistar06>

VLA Example: GPT-6 Astra



How do we train a VLA?

Model	Model Size	Training Data	Training Approach	Compute / GPU Setup	Reference
RT-1 (Google, 2022)	3B	~130K real robot episodes	Transformer (BC)	256 TPUV4 for ~17 days	Brohan et al. 2022
RT-2 (Google, 2023)	~6B	Web data + robot data (~1M examples)	Vision-language pretraining + robot fine-tuning	Not publicly disclosed (large-scale TPU cluster)	Brohan et al. 2023
OpenVLA (Stanford, 2024)	7B	970K robot trajectories (6 datasets)	BC with action chunking	64 × A100 (80GB) for 15 days	Kim et al. 2024
π_0 (Physical Intelligence, 2024)	~3B	~970K robot episodes	Flow matching	Not publicly disclosed (estimated large GPU cluster)	Black et al. 2024
$\pi_0.5$ (Physical Intelligence, 2024)	~3B	~350K real robot episodes	Flow matching + post-training	Not publicly disclosed	Black et al. 2024

Takeaway: VLA training can require large robot datasets and dozens to hundreds of accelerators running for days or weeks.

Juno: an opportunity for UT Dallas

Advanced computing resource like Juno enable the next generation of embodied AI research, from model training to large-scale evaluation

Natural Workloads for Juno

VLA Pretraining / Fine-tuning



"Pick up the red box and place it in the mug."



Large multi-robot datasets

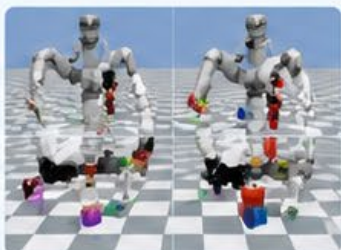


Model training and fine-tuning



VLA / foundation models

Simulation at Scale



Large-scale parallel simulation



RL and data generation



Domain randomization and environment generation

Large Evaluation Sweeps



Many tasks and scenes

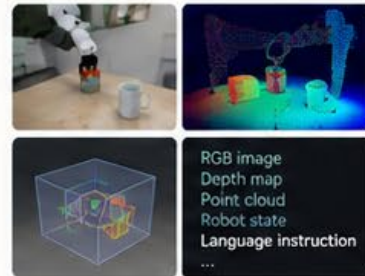


Reproducible benchmarking (e.g., VLA-Replica)



Cross-embodiment evaluation

Multimodal Data Processing



RGB, depth, point cloud, tactile, proprioception



Data storage and processing



Efficient training pipelines

UT Dallas Juno

Powering Discovery at Scale



52 NVIDIA H200 GPUs (recent upgrade)
67 total GPUs on Juno



High-performance GPU computing



Supporting research across UT Dallas



Enabling next-generation AI and robotics research

JUNO

More Compute. Greater Impact.

AI is leaving the screen.

From understanding the world to taking action in the real world.



*With advanced computing,
we can build a more capable,
embodied, and beneficial AI for real-world impact.*

PEOPLE
IDEAS
COMPUTE
REAL-WORLD IMPACT

*A Brighter Future
Together*

Thank You!

Together, we can make AI more useful in the real world.

BIGGER IDEAS.
BRIGHTER TOMORROWS