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

Yu Xiang

Assistant Professor

Intelligent Robotics and Vision Lab

Computer Science

INTELLIGENT ROBOTICS AND VISION LAB

UT DALLAS

9/16/2026

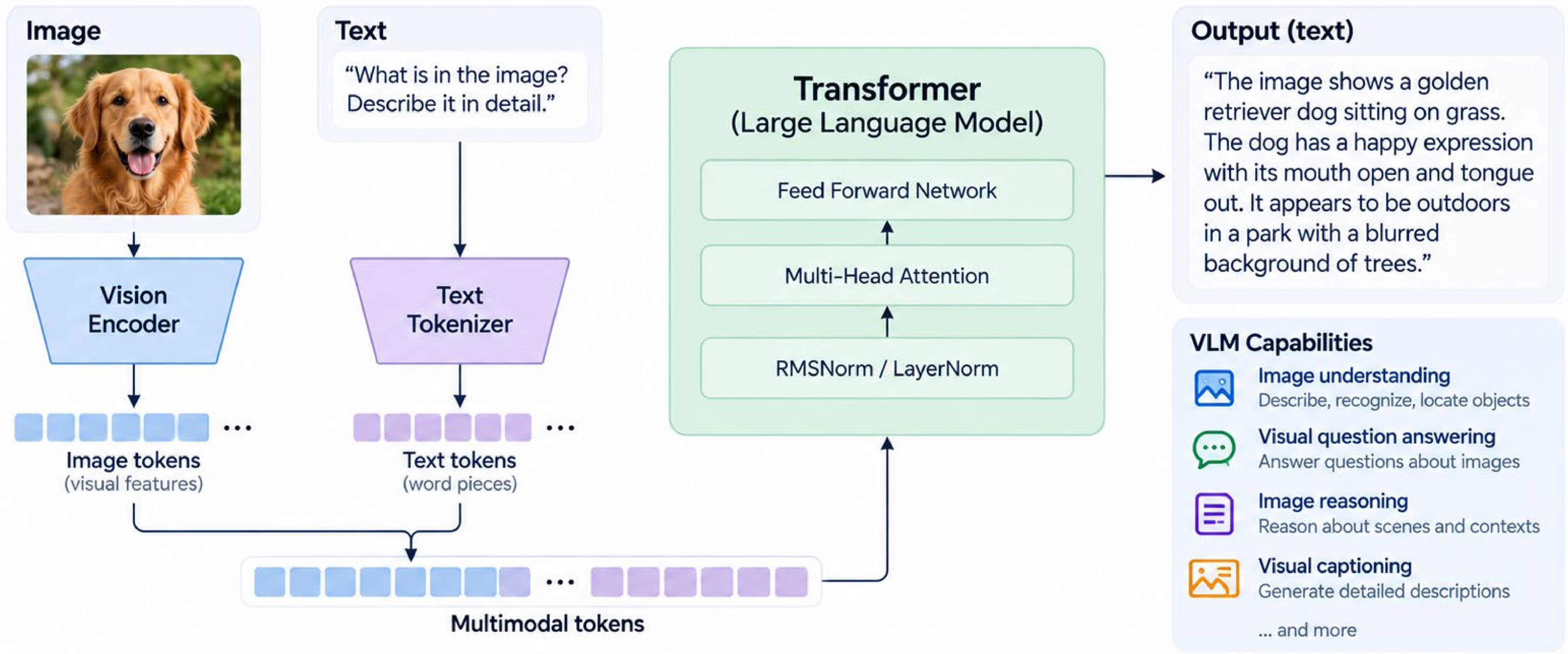
**Comet Computing Conference 2026**

# The current AI revolution started with language

- Large Language Models (LLMs)

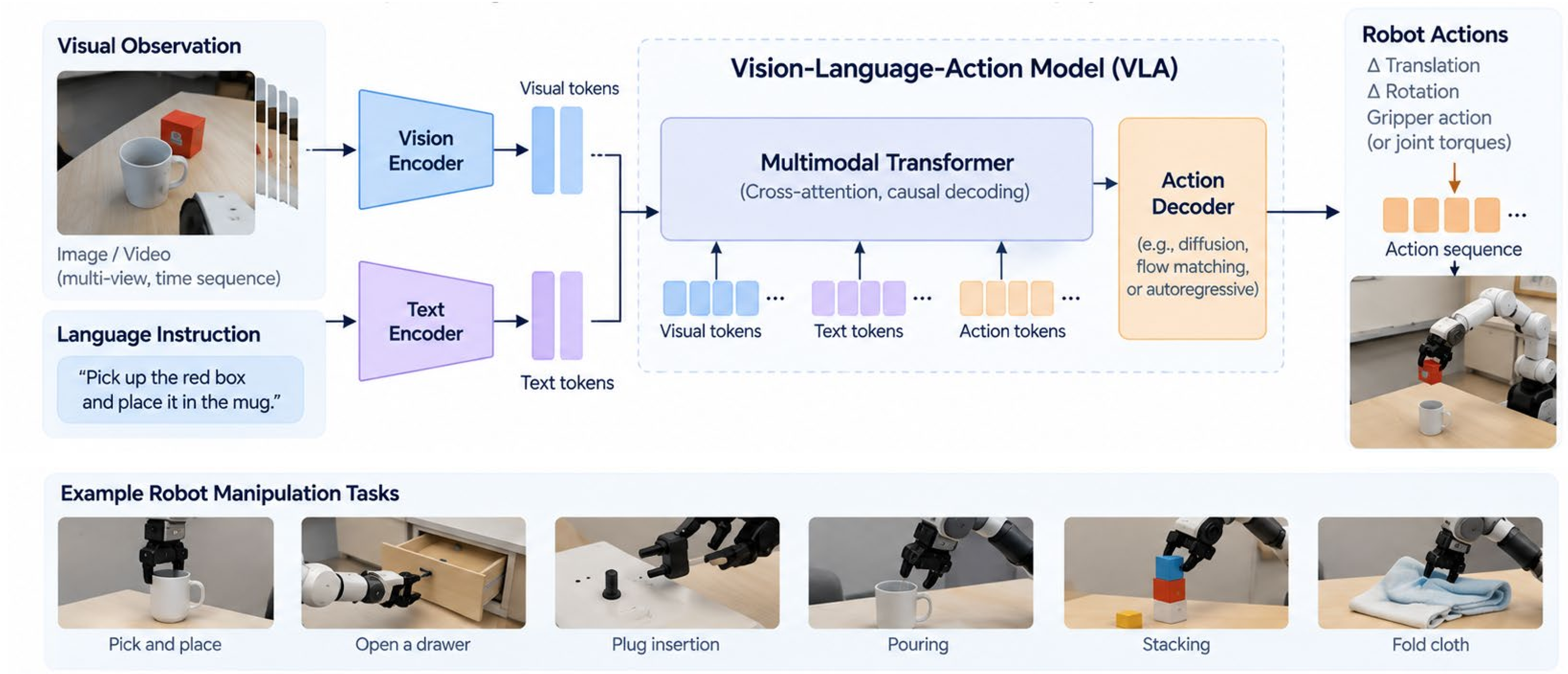
# Then AI learned to see

- Vision-Language Models (VLMs)

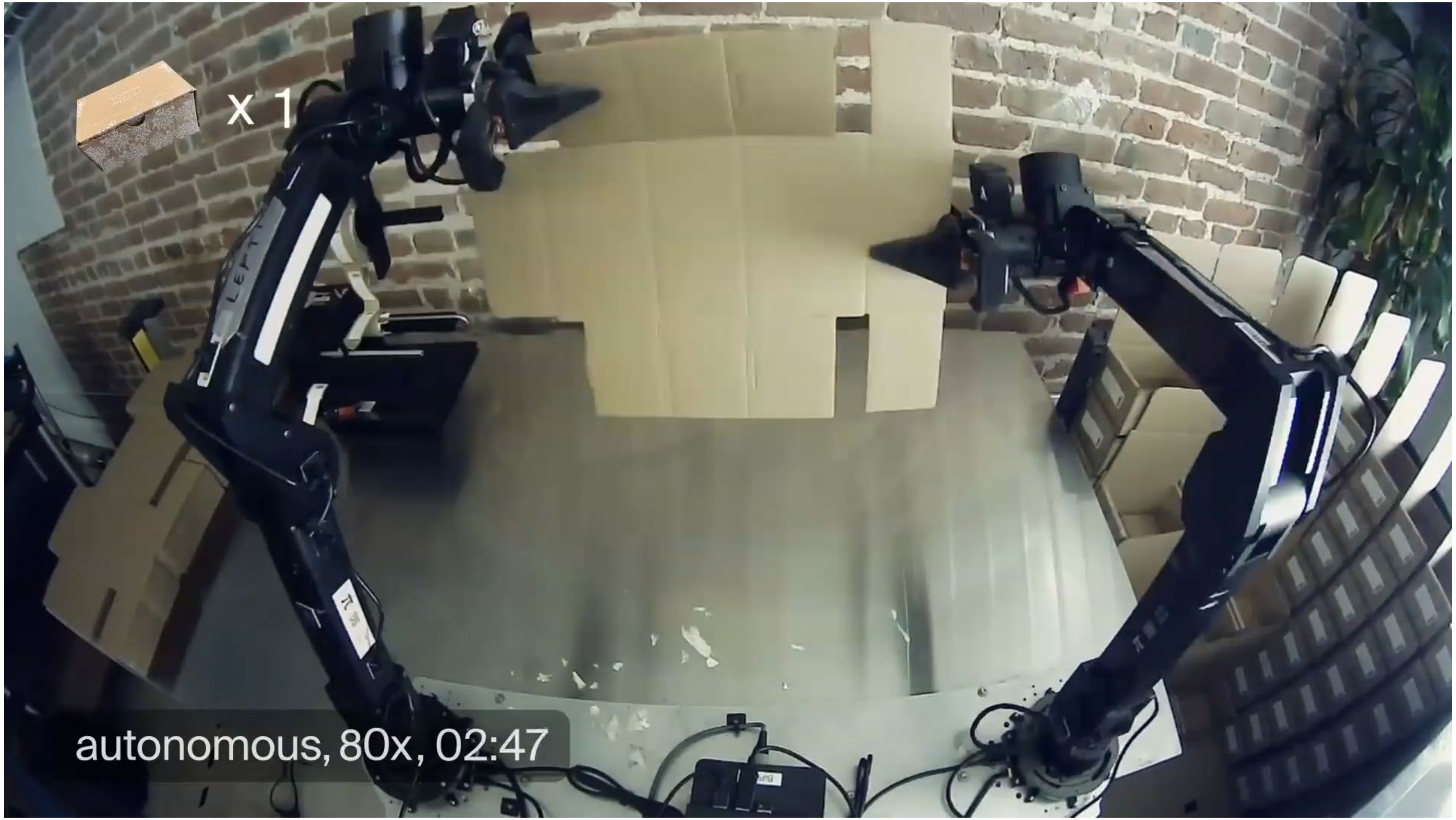


# Then AI learned to act

- Vision-Language-Action Models (VLAs)

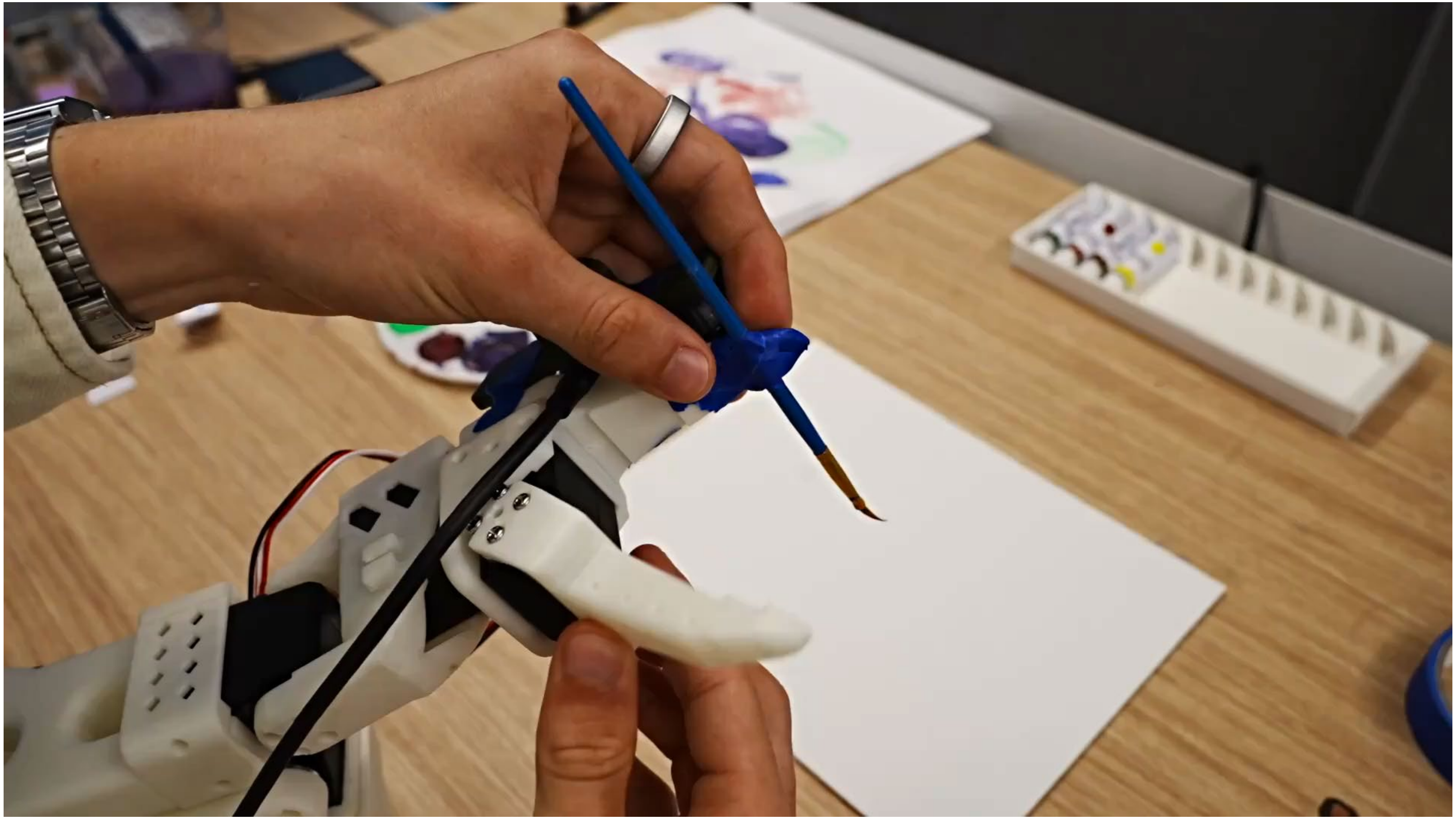


# VLA Example: Physical Intelligence




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

# VLA Example: GPT-6 Astra



https://x.com/cdngdev/status/2097339677128982873?s=20

# 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 |
| **$\pi_0$ (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**

# AI is leaving the screen.

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

THE UNIVERSITY OF TEXAS AT DALLAS

Language (LLMs) › Vision + Language (VLMs) › Vision + Language + Action (VLAs) › Real-World Impact (Embodied AI)

*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.