Robot Learning: From Imitation to Implementation

Exploring Imitation Learning, VLA Models, and Open-Source Robotics

Hamed Rahimi, PhD - Oct 2025

Agenda

Journey from Theory to Practice

- Imitation Learning: The Foundation
- Vision-Language-Action (VLA) Models: The Modern Approach
- Hugging Face: The Ecosystem
- LeRobot: The Implementation Framework
- SO-100: Real-World Application

PART 1: IMITATION LEARNING

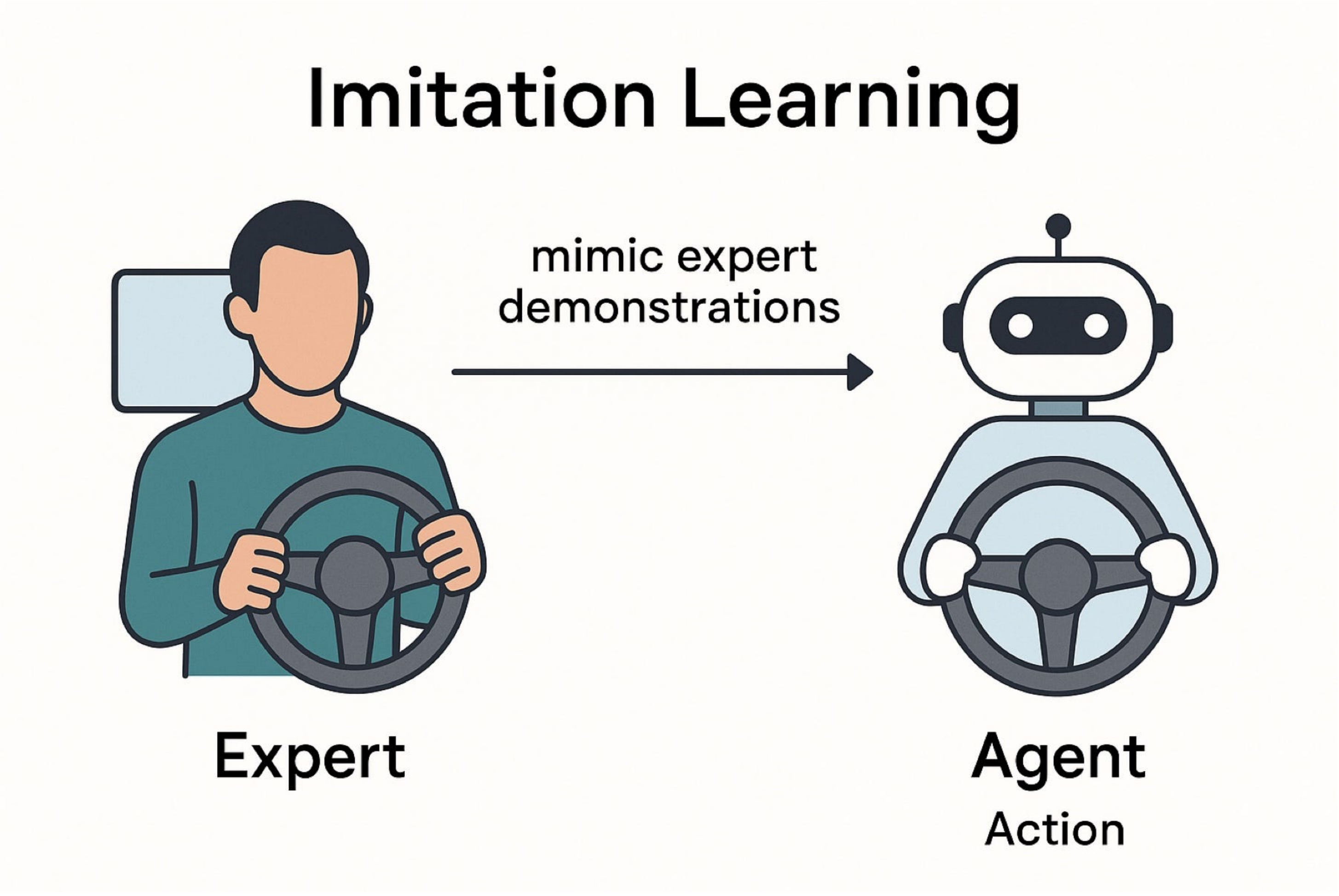
What is Imitation Learning?

Learning by Watching and Mimicking

Core Concept: Robots learn to perform tasks by observing expert demonstrations rather than explicit programming.

The Process:

- Collect demonstrations from human experts or teleoperation
- Extract patterns and behaviors from the data
- Train a policy (decision-making model)
- Deploy the learned behavior on the robot



Key Advantage: More intuitive than manually coding every behavior

Imitation Learning in Context

How Does It Compare?

Learning Type	Data Source	Best For
Supervised Learning	Labeled input-output pairs	Classification, prediction
Reinforcement Learning	Trial and error with rewards	Game playing, optimization
Imitation Learning	Expert demonstrations	Complex motor skills, manipulation

Why Imitation Learning for Robotics?

- Leverages human expertise directly
- Faster than trial-and-error RL
- More practical for complex, multi-step tasks

The Imitation Learning Pipeline
From Human to Robot

1. Data Collection

   └──> Human demonstrates task (e.g., picking up objects)

2. Feature Extraction

   └──> Record: camera images, joint angles, gripper states

3. Policy Learning

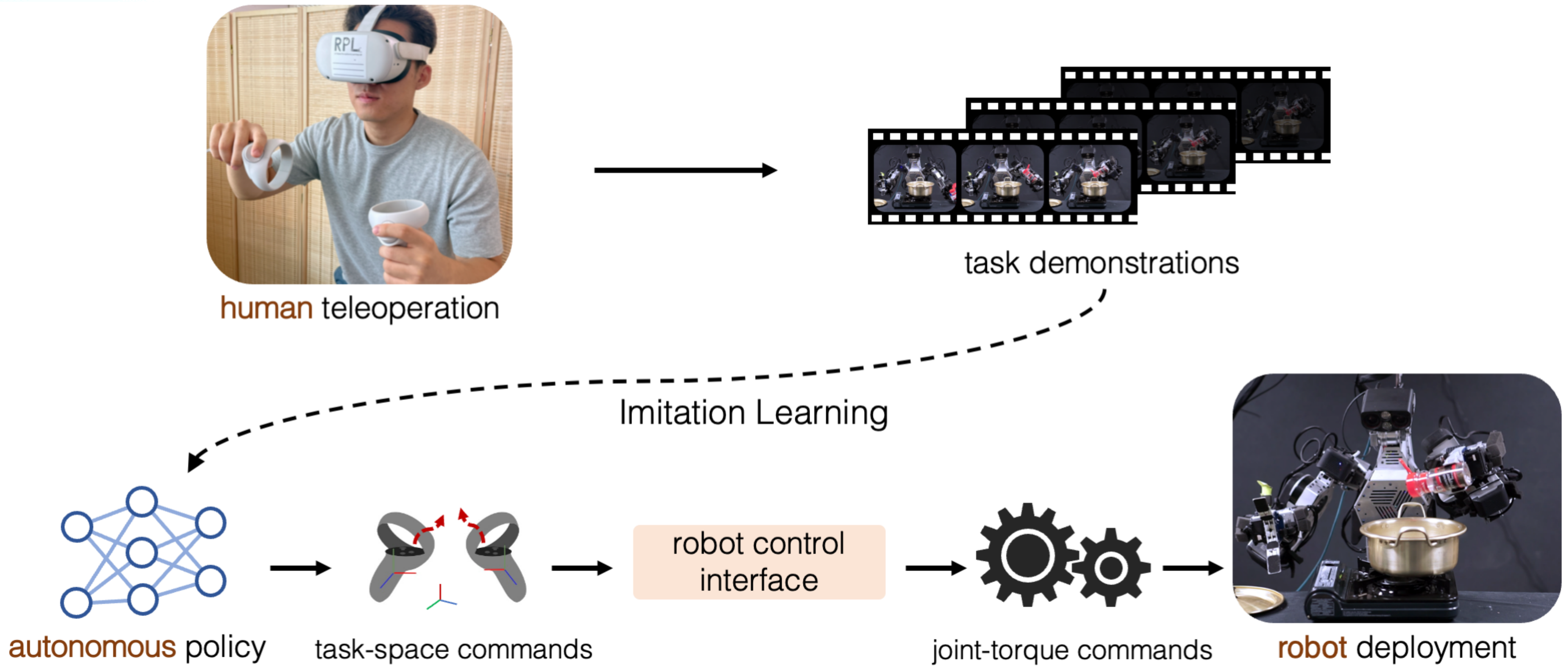
   └──> Train neural network: observations → actions

4. Deployment

   └──> Robot executes learned policy autonomously

5. Refinement

   └──> Collect edge cases, retrain, improve



Real-World Examples
Imitation Learning in Action

Robotic Manipulation:

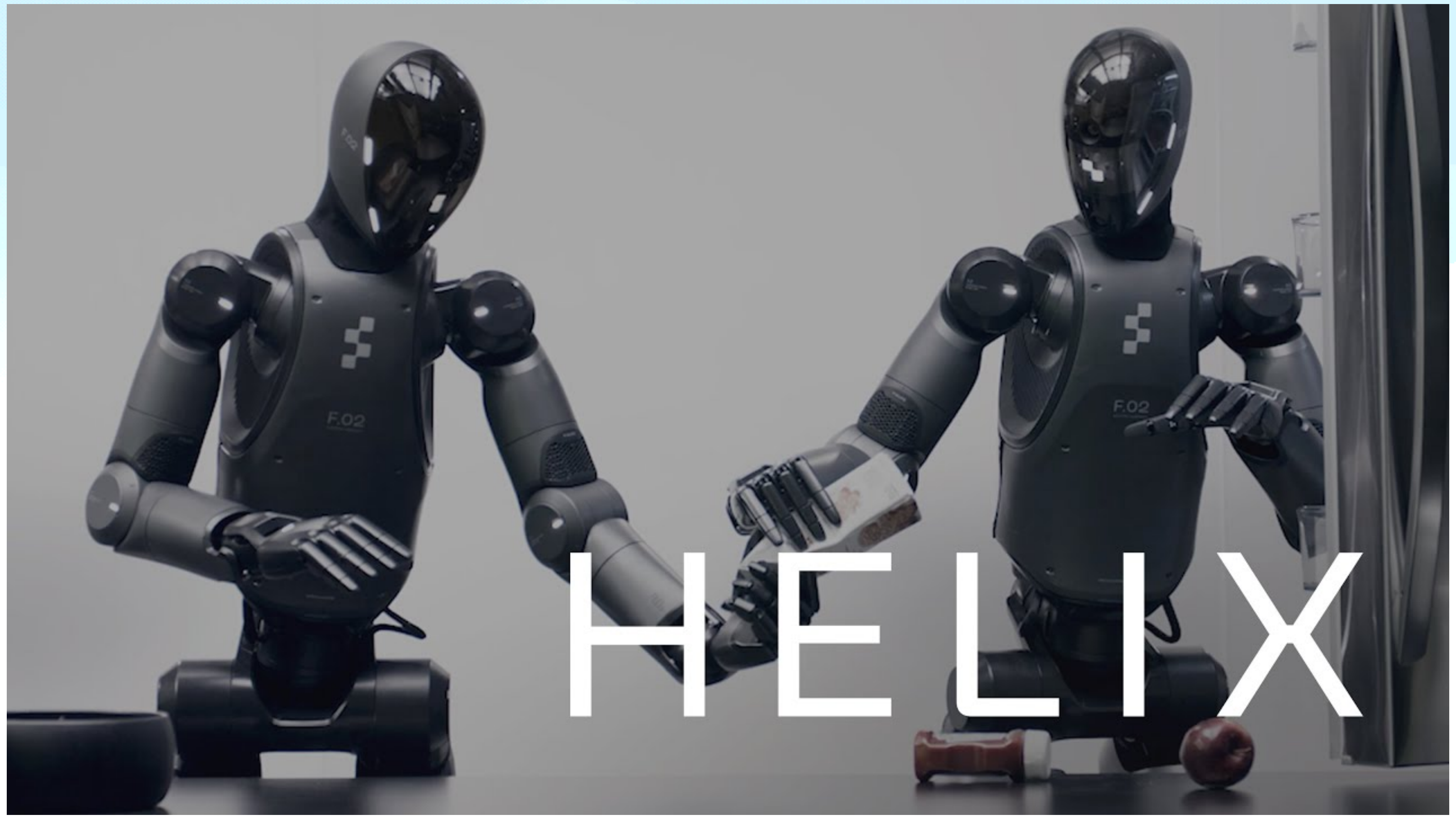
- Learning to grasp diverse objects
- Assembly tasks in manufacturing
- Food preparation and cooking

Autonomous Navigation:

- Learning driving behaviors from human drivers
- Drone flight patterns

Healthcare:

- Surgical assistance robots learning from expert surgeons



PART 2: VISION-LANGUAGE-ACTION MODELS

The Evolution to VLA Models

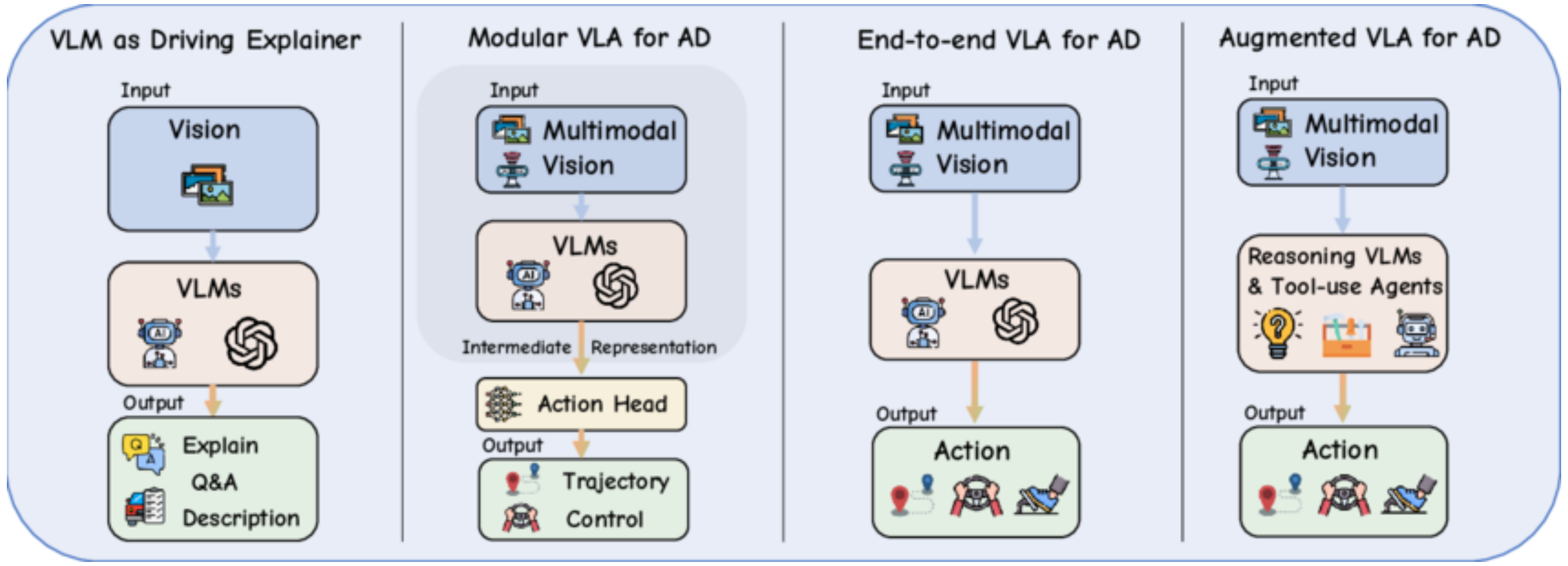
Next-Generation Robot Intelligence

The Challenge: Traditional imitation learning treats each task in isolation

The Solution: Vision-Language-Action (VLA) Models

- **Vision:** Understanding the environment through cameras
- **Language:** Following natural language instructions
- **Action:** Generating appropriate robot movements

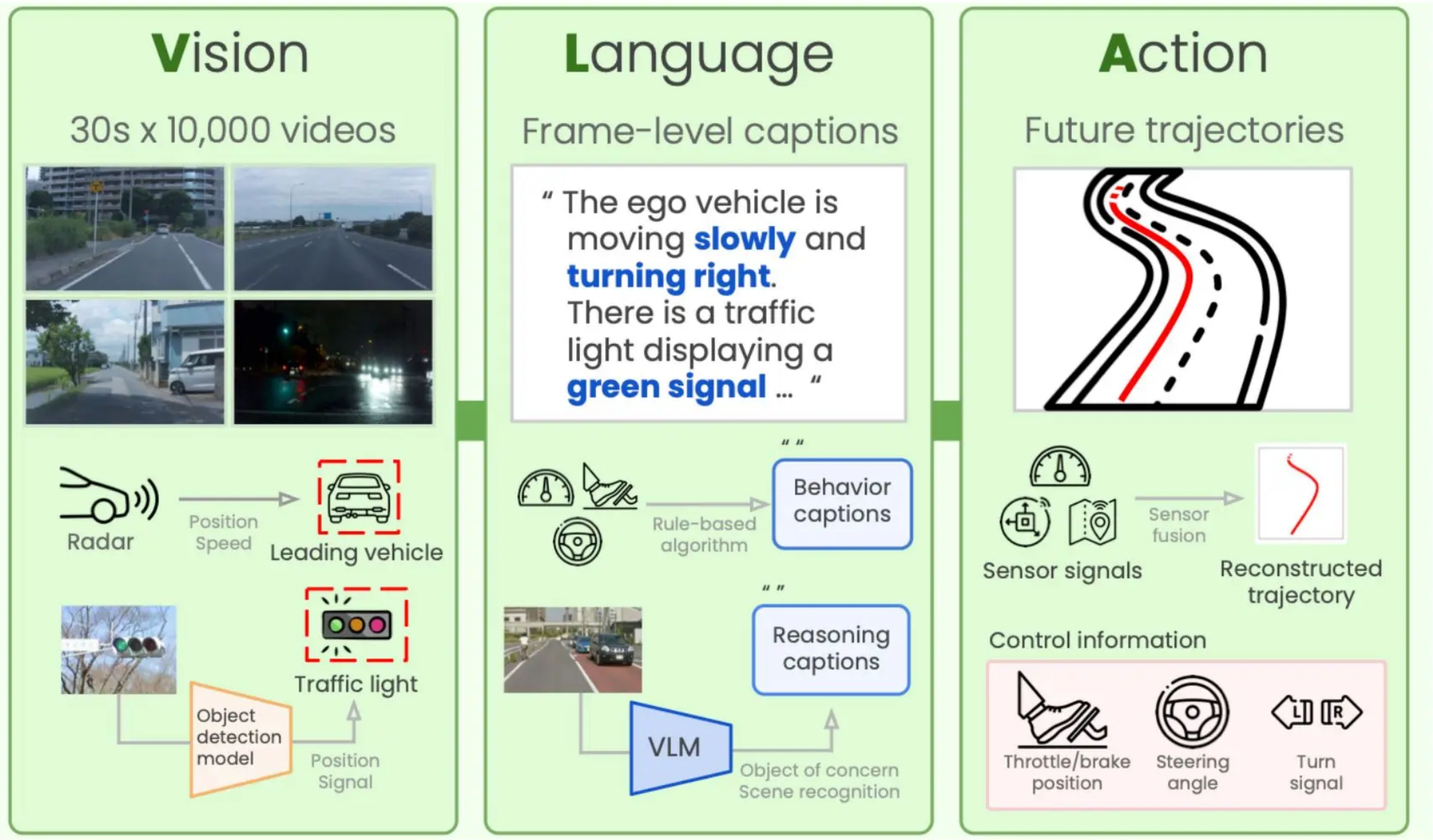
Key Innovation: One unified model that handles multiple modalities and generalizes across tasks



How VLA Models Work

Multimodal Intelligence

- **Input Layer:**
  - Camera images (what the robot sees)
  - Text instructions (what we want: "pick up the red cup")
  - Robot state (current joint positions, gripper status)
- **Processing:**
  - Foundation model backbone (Transformer architecture)
  - Cross-attention between vision, language, and joint-positions/angles
  - Learned action embeddings
- **Output:**
  - Continuous action commands (joint velocities, gripper open/close)



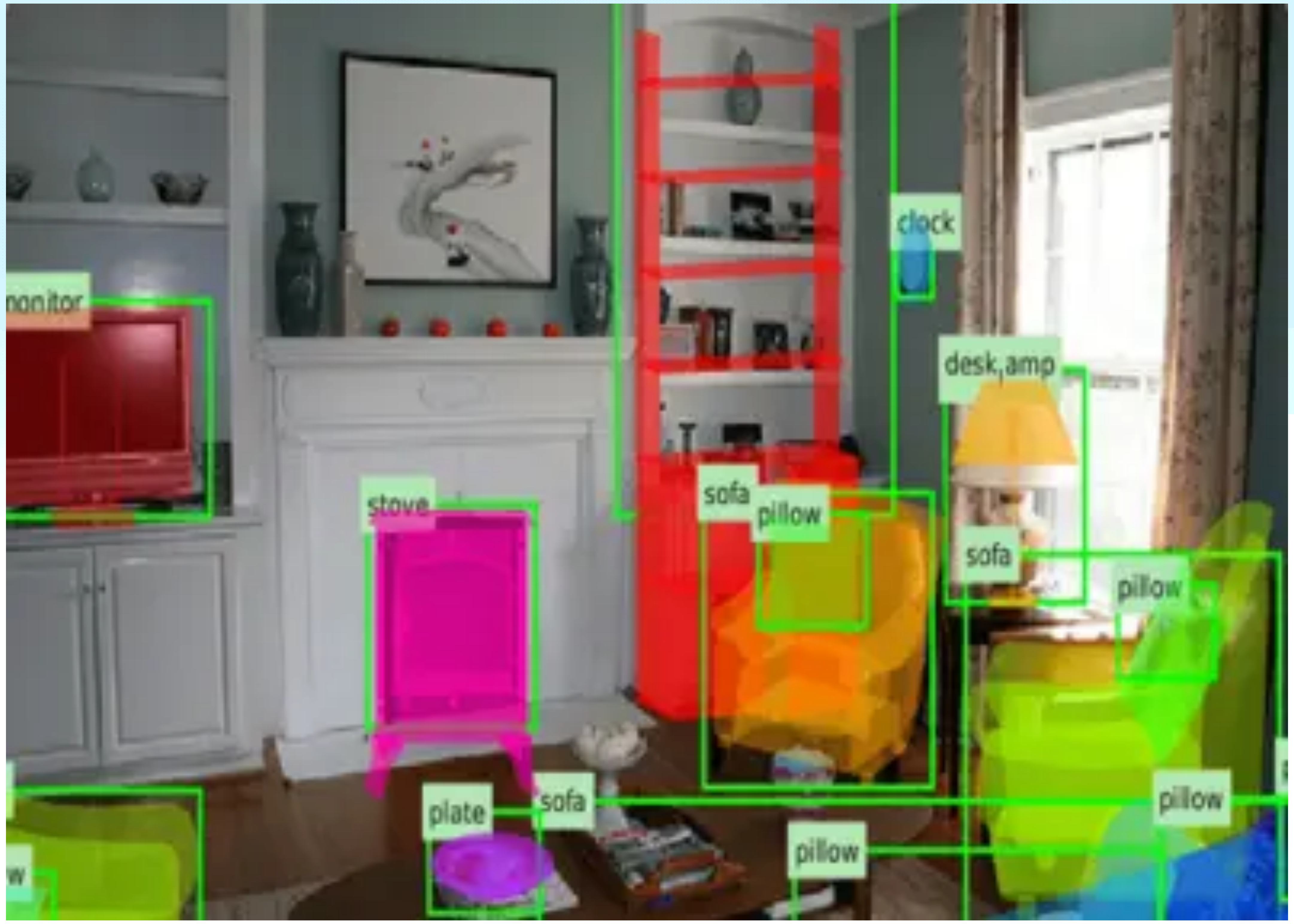
Why VLA Models Matter

The Generalization Advantage

Traditional Approach: One policy per task

- "Pick up cups" ≠ "Pick up bottles"
- Requires separate training for each object type

- **VLA Approach:** One policy for many tasks
  - "Pick up the [object]" works for cups, bottles, tools
  - Language provides context and flexibility
  - Visual grounding ensures robust perception



- **Result:** Robots that can follow instructions like "put the blue block in the red bowl" without task-specific training

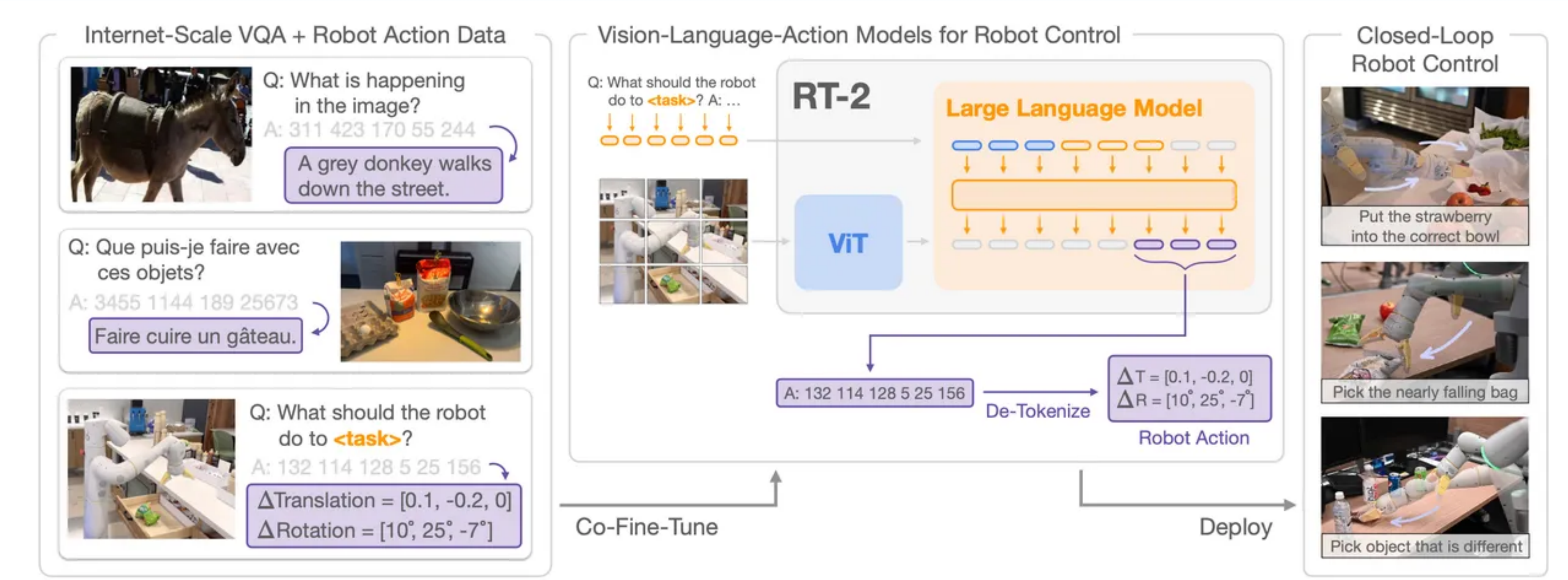
Notable VLA Models

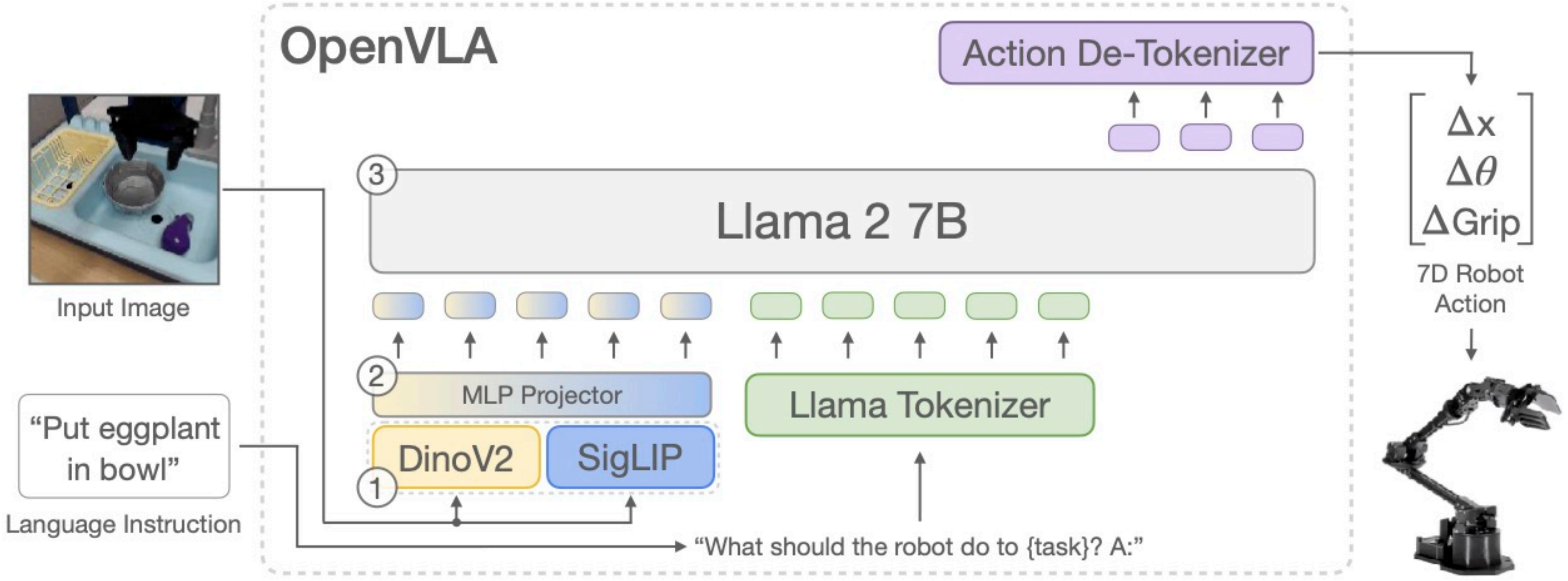
State-of-the-Art Examples

- **RT-2 (Google DeepMind)**
  - A. Combines vision-language model with robot actions
  - B. Trained on web data + robot demonstrations
  - C. Can perform 6,000+ robotic tasks
- **OpenVLA (Open Source)**
  - D. 7B parameter open-source VLA model
  - E. Built on Llama and SigLIP vision encoder
  - F. Designed for research and community use

Benefits of Open VLA Models:

- Transparency and reproducibility
- Lower barrier to entry for researchers
- Community-driven improvements

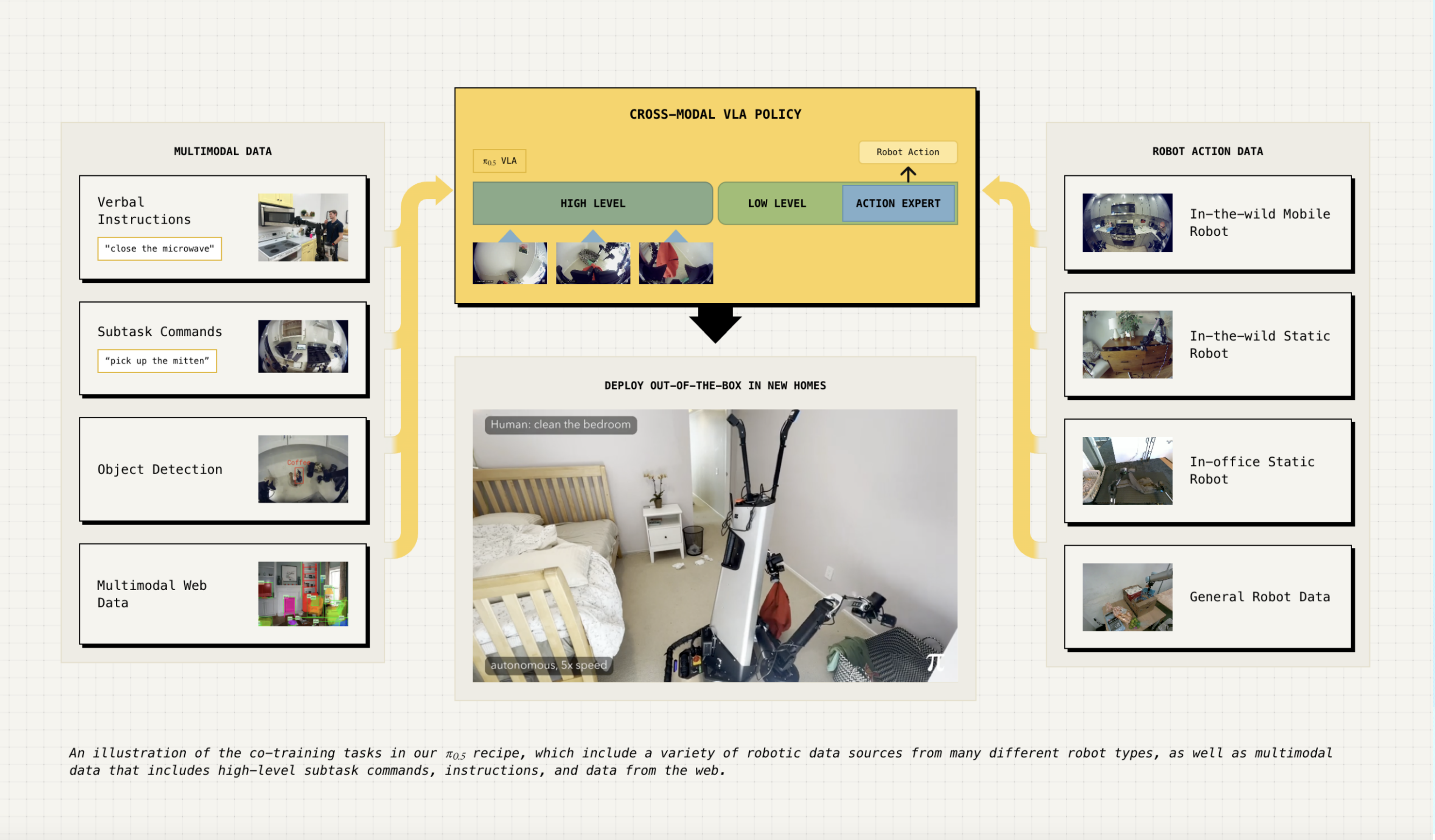


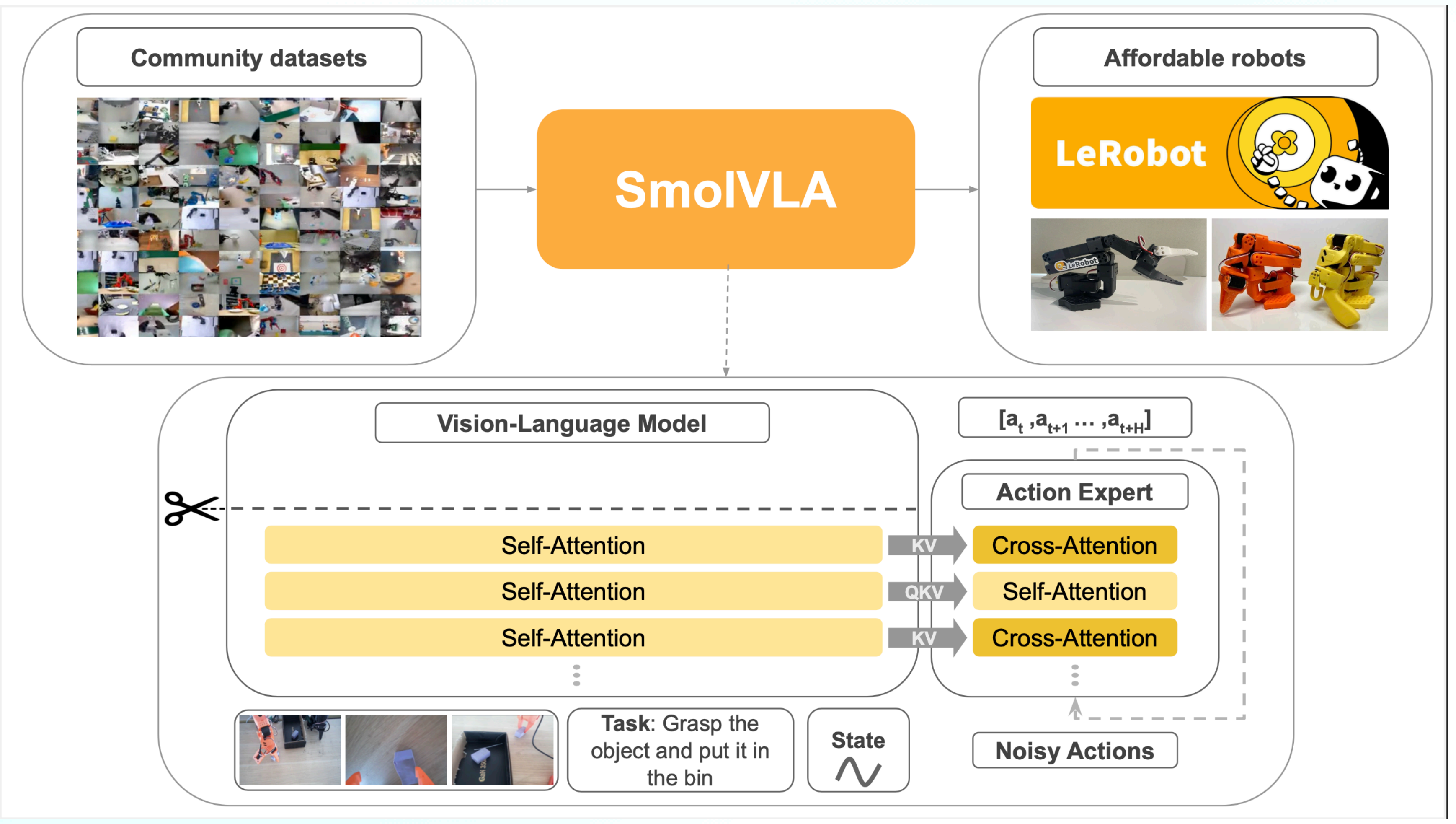


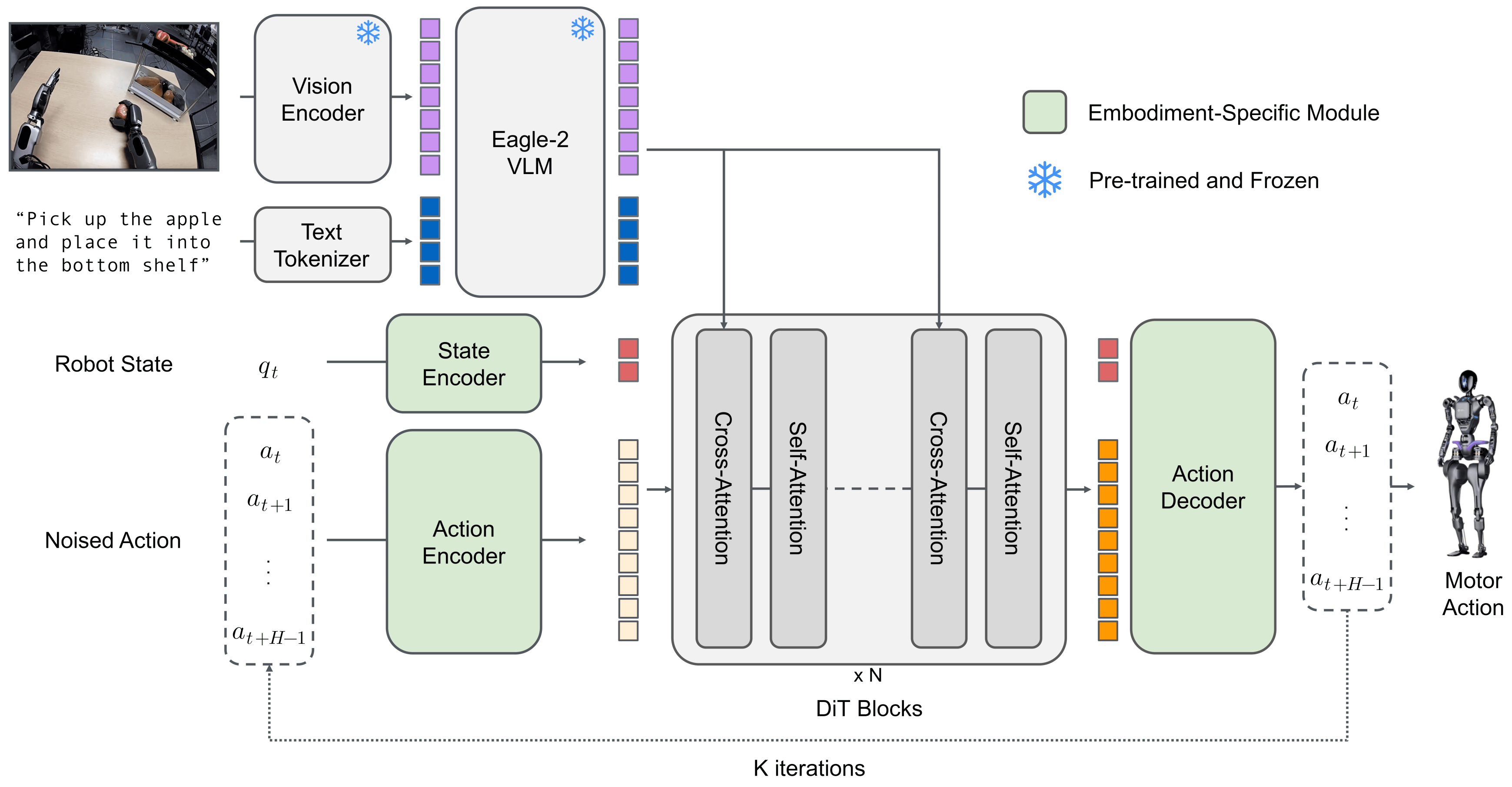
More Recent VLAs

State-of-the-Art Examples

- Pi-0 / 0.5 (Physical Intelligence)
- Gr00T - N1/N1.5 (Nvidia)
- SmolVLA (HuggingFace and Sorbonne)







PART 3: HUGGING FACE ECOSYSTEM

Introducing Hugging Face

The GitHub of Machine Learning

What is Hugging Face?

A platform for sharing, discovering, and collaborating on machine learning models and datasets

Core Components:

- 🤗 **Transformers:** Pre-trained models library (BERT, GPT, T5, etc.)
- 🤗 **Datasets:** Standardized datasets with easy loading
- 🤗 **Spaces:** Host ML demos and applications
- 🤗 **Hub:** Central repository for models, datasets, and code

Why It Matters for Robotics:

- Democratizes access to advanced AI models
- Enables reproducible research
- Facilitates collaboration across teams and institutions

Hugging Face for Robotics

Bringing Open-Source to Embodied AI

Traditional Robotics Development:

- Siloed research labs
- Proprietary datasets and models
- Difficult to reproduce results

Hugging Face Robotics Initiative:

- Open datasets of robot demonstrations
- Pre-trained robot policies ready to download
- Standardized formats and evaluation metrics
- Community contributions and improvements

Impact: Accelerating robotics research by 10-100x through shared resources

PART 4: LeRobot Framework

Introducing LeRobot

Hugging Face's Robotics Library

LeRobot: Open-Source Robot Learning Made Easy

What It Provides:

- **Pre-trained models:** State-of-the-art imitation learning policies
- **Datasets:** Thousands of robot demonstrations ready to use
- **Training tools:** Scripts and utilities for training your own policies
- **Simulation support:** Integration with physics simulators
- **Hardware integration:** Support for popular robot platforms

Philosophy: Make robot learning as accessible as NLP with Transformers

ISIR
INSTITUT
DES SYSTÈMES
INTELLIGENTS
ET DE ROBOTIQUE

LeRobot Features

Complete Robot Learning Toolkit

1. Dataset Collection & Management

- Record robot demonstrations with standard format
- Load pre-existing datasets from Hugging Face Hub
- Visualize and inspect data quality

2. Policy Training

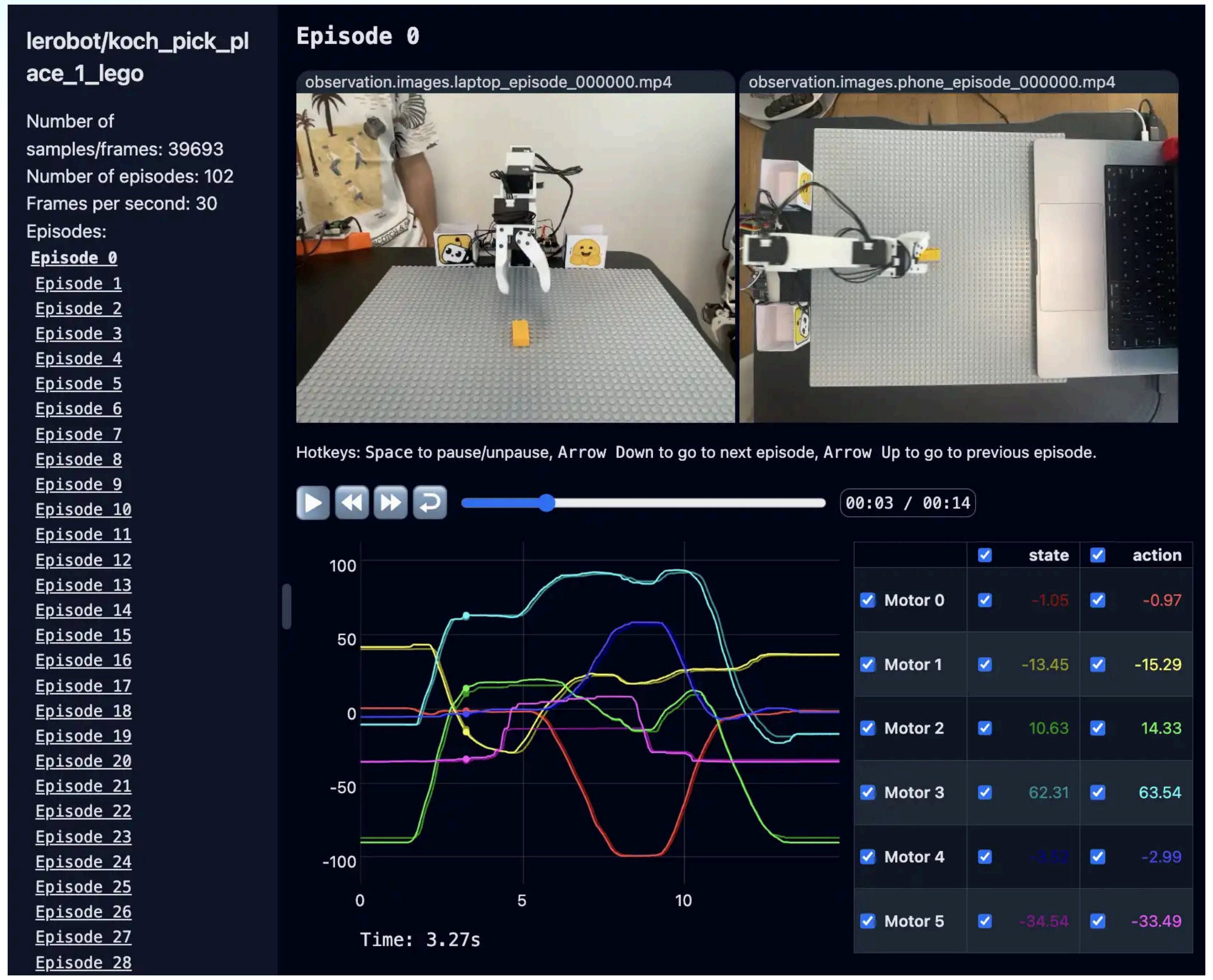
- Supported algorithms: ACT, Diffusion Policy, VQ-BeT
- Distributed training support
- Hyperparameter optimization tools

3. Evaluation & Deployment

- Simulation testing before real robot deployment
- Real-time policy inference
- Performance metrics and logging

4. Community Contributions

- Share your datasets and trained models
- Reproduce published research
- Collaborate on improving policies



LeRobot Supported Hardware

Works with Popular Robot Platforms

[Low-Cost] Robotic Arms:

- *SO-100/101 (our focus!)*
- Koch v1.1
- Aloha platforms
- Franka Emika Panda

[Low-Cost] Mobile Manipulators:

- LeKiwi

PART 5: SO-100 ROBOT

Meet the SO-100/101

Open-Source, Low-Cost ARM Robot

What is SO-100? An affordable, open-source arm robot designed for research and education

Key Specifications:

- **Cost**: ~$200 (vs. $50k+ for commercial arm manipulators)
- **Design**: Open-source CAD files and assembly instructions
- **Actuation**: 6+ degrees of freedom
- **Computation**: Runs on standard laptop or embedded PC

Philosophy: Democratizing manipulation robotics research

SO-100 + LeRobot Integration

Bringing It All Together

Data Collection

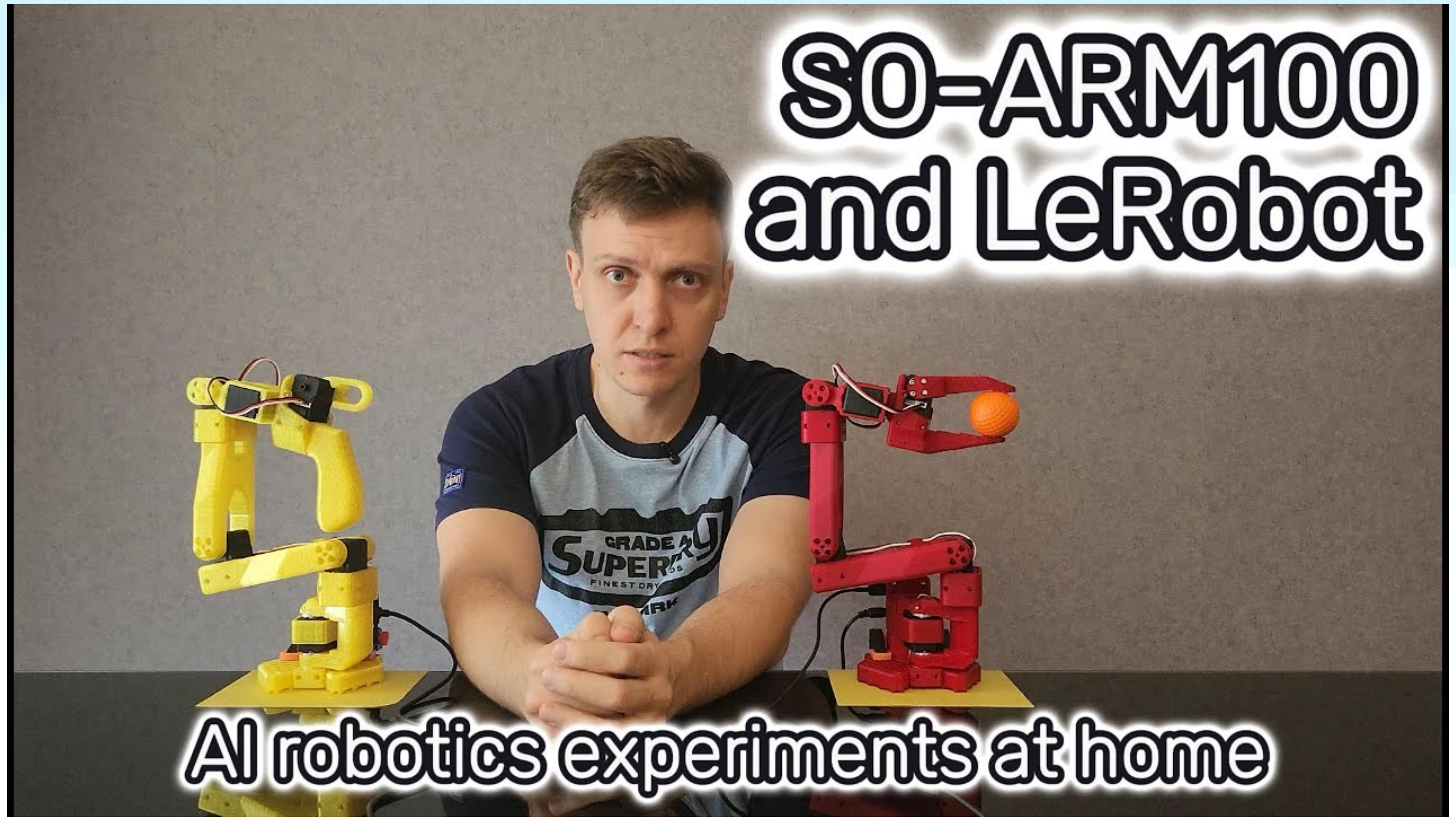
- Teleoperate SO-100 to demonstrate tasks
- Record with LeRobot data collection tools
- Store in standard format on Hugging Face

Training

- Load SO-100 demonstration dataset
- Train VLA policy using LeRobot
- Leverage pre-trained vision-language models

Deployment

- Download trained policy to SO-100
- Robot executes tasks autonomously
- Follows natural language instructions



Result: An arm robot that learns from demonstrations and understands commands

Thanks, questions?

hamed.rahimi@sorbonne-universite.fr