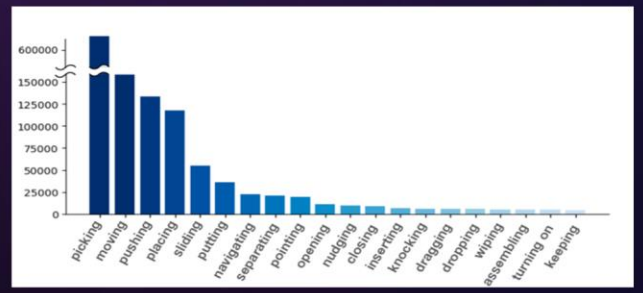
How to Transform Data into Knowledge?

Serving the Future with AGI

Zerith Robotics

Embodied AI Datasets

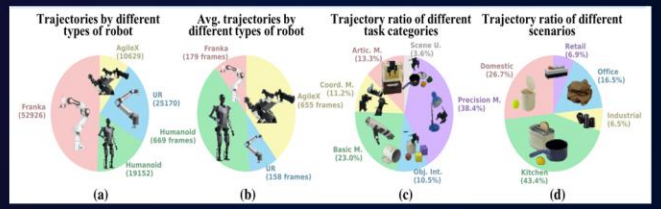
Open X-Embodiment



- **> 1 million trajectories**

  The distribution is uneven, with a majority of simple skills.

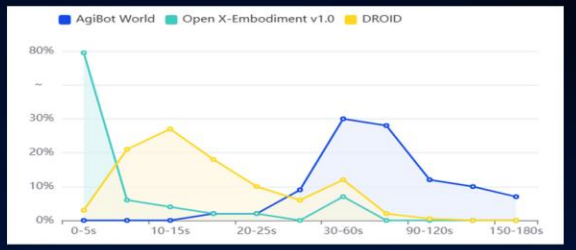
RoboMIND



- **Improve data consistency**

  Enhance the scientific nature of task classification

AgiBot World



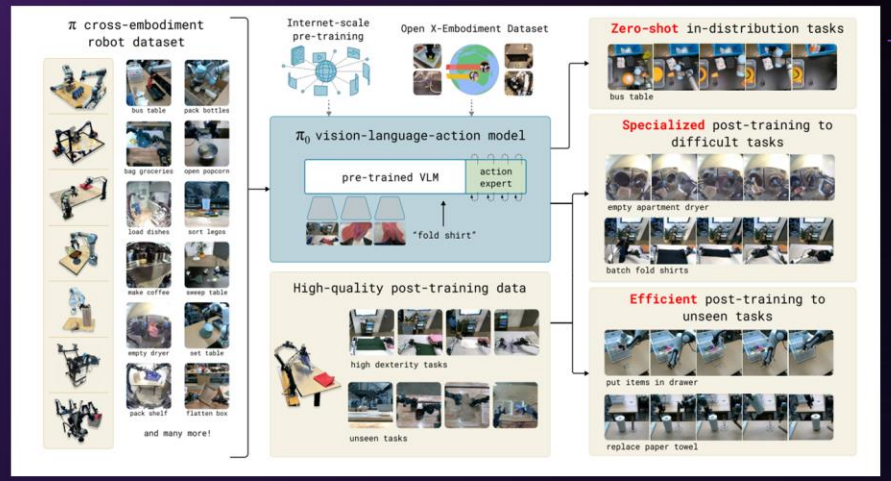
- **Increased long-term tasks and object diversity**

However, all of them are devoid of knowledge labels.

Co-evolution of Models and Data

ZERITH 智次方

π_0



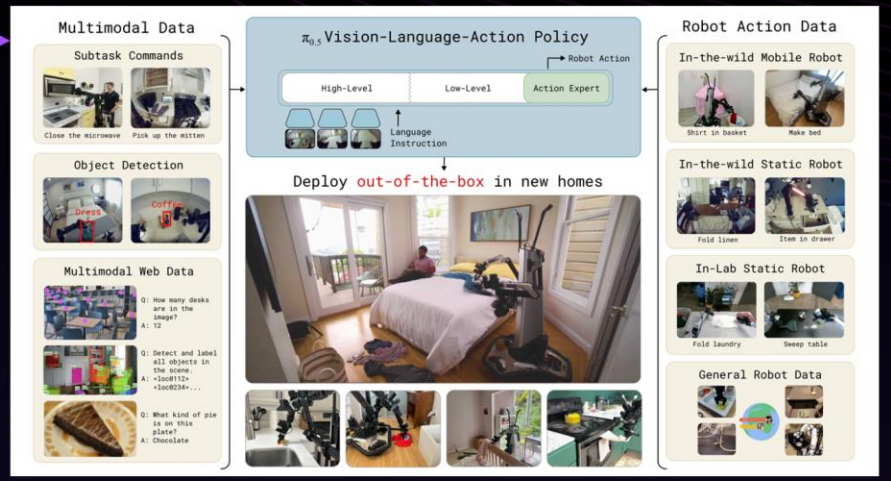
+ Environment Variation

+ Visual Grounding

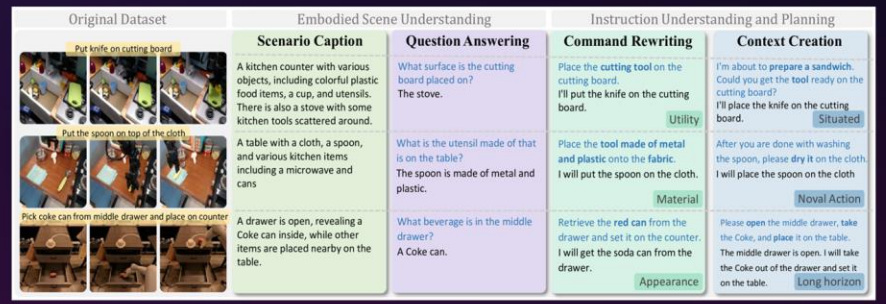
+ Subtask Predition

+ VQA

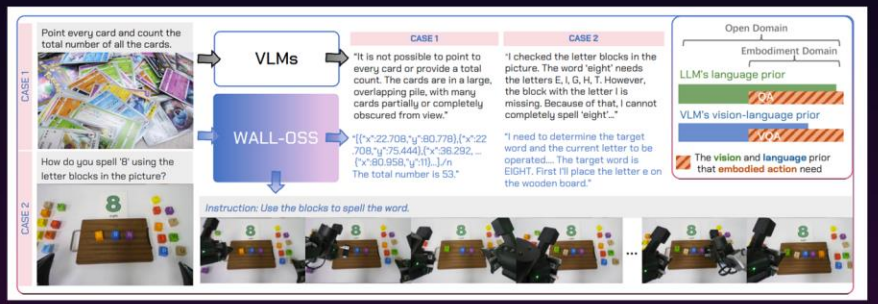
$\pi_{0.5}$



InstructVLA



WALL-OSS



Pure robotics data -> understanding, grounding, reasoning

Zerith Robotic Data Standard

Task Description: Fill the vase with water.

Scene Description: This image shows a modern bathroom sink area. The sink is white, rectangular, and built into a glossy countertop. A sleek, metallic faucet is positioned at the top left of the sink, and the drain hole is located in the center. On the right side of the counter, there is a small vase containing white flowers with yellow centers. The background wall appears to be light gray tile, and part of a mirror is visible at the top of the image.

Planning:

1. Pick up the vase.
2. Move the vase under the faucet.
3. Turn on the faucet.
4. Turn off the faucet.
5. Put the vase back in its original place.

Visual Grounding:

Q In the [SUBTASK(pick up the vase)] task, what objects are being focused on and where are they located?

A The vase is located at [0.704, 0.517, 0.839, 0.698].

Temporal Order:

Q After the robot wipes the sink area, what should it do next?

A It should return the cleaning cloth or tool to its storage spot.

Trajectory Flow:

[[(0.801, 0.552), (0.812, 0.632), ...]]

Spatial Understanding:

Q Where is the vase of flowers relative to the sink?

A It is on the right side of the sink, near the edge of the counter.

Physical Common Sense:

Q What will happen if the faucet is turned on and the drain is blocked?

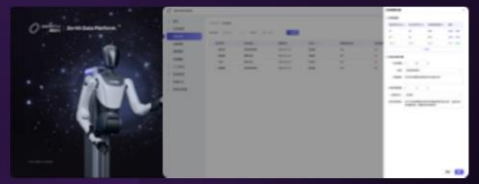
A Water will fill the sink and eventually overflow onto the counter.

Zerith Chain

ZERITH 零次方

Task Allocation

Task Creation, Collection Example Uploading, and Task Delegation

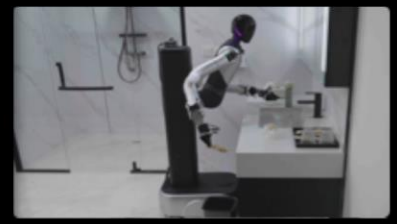


Data Collection

Self-developed VR Data Collection Application — Become a Data Collector in Just 10 Minutes

Deploy

Direct Deployment of Cloud Model Checkpoints to Robots for Real-world Validation



Evaluation

Integrated Simulation Benchmark for Efficient, Low-cost Model Evaluation and Optimization Feedback

Data Annotation

Human-in-the-loop Semi-automatic Knowledge Annotation
Supporting Multiple Out-of-the-box Tools such as GPT, GroundingSAM...

Zerith V1

Eye Cameras

Latent Encoder

is keyframe?

Memory Queue

Task Description

Clean the washbasin

Cognitive System

- Understanding
- Reasoning
- Planning
- Coarse Visual Grounding

subtasks

- Sparse
- Latent
- Planner

future key frames latent

3–4hz

sparse connect

5–7hz

Behavior System

- Real Time Inference
- Fine-grained Spatial Perception

30Hz

Proprioception

real time action


ZERITH
零次方
机器人
来干活了！
ZERITH H1

Benchmark

Instruction Following

The agent accurately executes tasks based on natural language instructions. This capability thoroughly evaluates semantic disambiguation, interactive communication, and reasoning skills.

Spatial Understanding

The agent perceives and comprehends spatial relationships between objects, such as relative position and orientation.
This is crucial for adapting to 3D real world.

Long-horizon Planning

The agent plans and makes decisions over extended sequences of actions in complex environments.
This is essential for tasks like cleaning and cooking, where long-term coordination is needed.

Force Control

The agent precisely controls the forces applied during object manipulation to prevent damage to itself, the object, or the environment. This is vital for tasks involving dense contact, such as wiping or door opening.

Generalization

The agent transfers knowledge to new, unseen environments or tasks, demonstrating adaptability and robustness.

Hours	Tasks	Scene variants
2000	100	50

- Bedroom 20%
- Bathroom 15%
- Balcony 15%
- Laundry Room 10%
- Living Room 20%
- Bar Counter 10%
- Kitchen 10%

observations

- Five camera streams
- Joint States (Position, Velocity, Torque)
- Chassis States
- IMU
- Finger Tactile Force

Zerith Product Matrix

Upcoming product

6D Tactile Force Gripper

Desktop-grade 6-DOF Robotic Arm

ZERITH H1

5× Cameras (3× RGBD, 1× Stereo); 1× LiDAR; 1× IMU

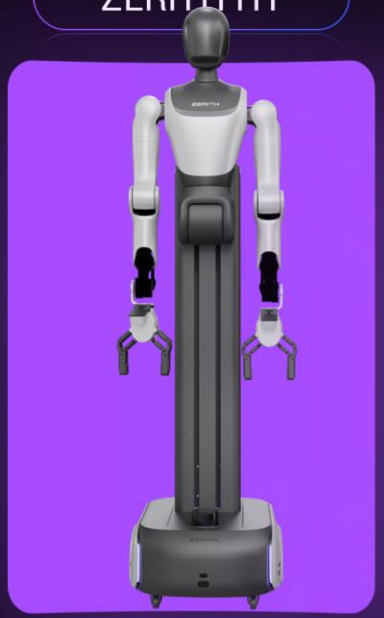
6D Force/Torque Sensors (optional)

23 DOF for Whole-Body Control (WBC)

Height up to 2.3m

Arm span up to 2.1m

ZERITH H1



ZERITH Z1

ZERITH Z1

Capable of traversing 10+ complex terrains

Over 75% of the robot's exterior surface is covered with flexible materials

Ultra-low latency voice interaction capability

Zerith chain

Efficient End-to-End Data-to-Knowledge Conversion

H1	Height: 0.9–1.74 m	Arm reach: 2.15 m	Vision sensors: 3×RGB-D, 2×wide-angle RGB	Per-arm DoF: 7	Per-arm payload: 1.5 kg	End-effector: Two-finger gripper
Z1	Dimensions (standing): 1450 mm × 500 × 220 mm	Endurance: 2 h	Head DoF: 2	Total DoF: 31	Waist DoF: 3	Terrain adaptability: 10+