AgiBot World Colosseo: Large-scale Manipulation Platform for Scalable and Intelligent Embodied Systems

Team AgiBot-World

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

AGIBOT

上海创智学院 Shanghai Innovation Institute

Highlights

- **AgiBot World dataset**, a multifarious robot learning dataset accompanied by opensource tools to advance research on policy learning at scale.
- **GO-1**, a scalable robot foundation policy using latent action representations to unlock web-scale pre-training on heterogeneous data.
- **ADC**, an adversarial data collection pipeline that redefines robotic data acquisition through real-time, bidirectional human-environment interactions

AgiBot World

- 1M+ Trajectories
- 3,000+ Objects
- 217 Tasks

Robot Platform: AgiBot G1

- All-Purpose Sensor Setup
  - RGBD Cameras
  - Visuo-tactile Sensor
- 6-Dof Dextrous Hand
- Dual-arm Humanoid

Based on AgiBot G1, we construct AgiBot World—a high-quality robot manipulation dataset embracing a human-in-the-loop framework.

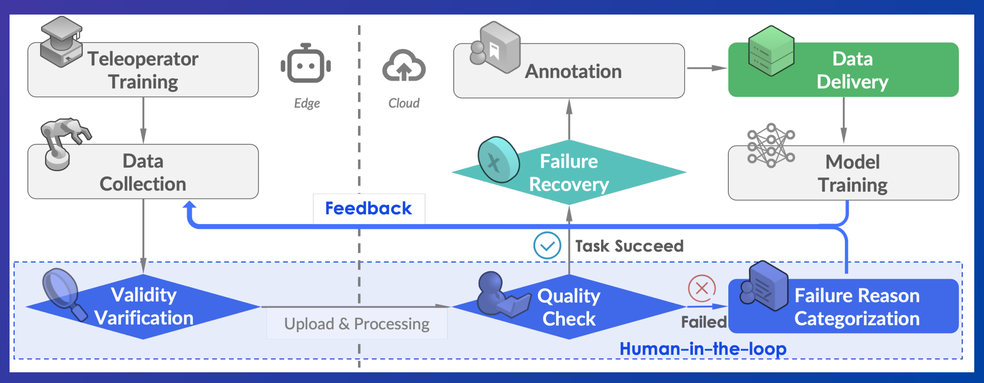


Figure 1. Data collection with human-in-the-loop

Dataset Statistics

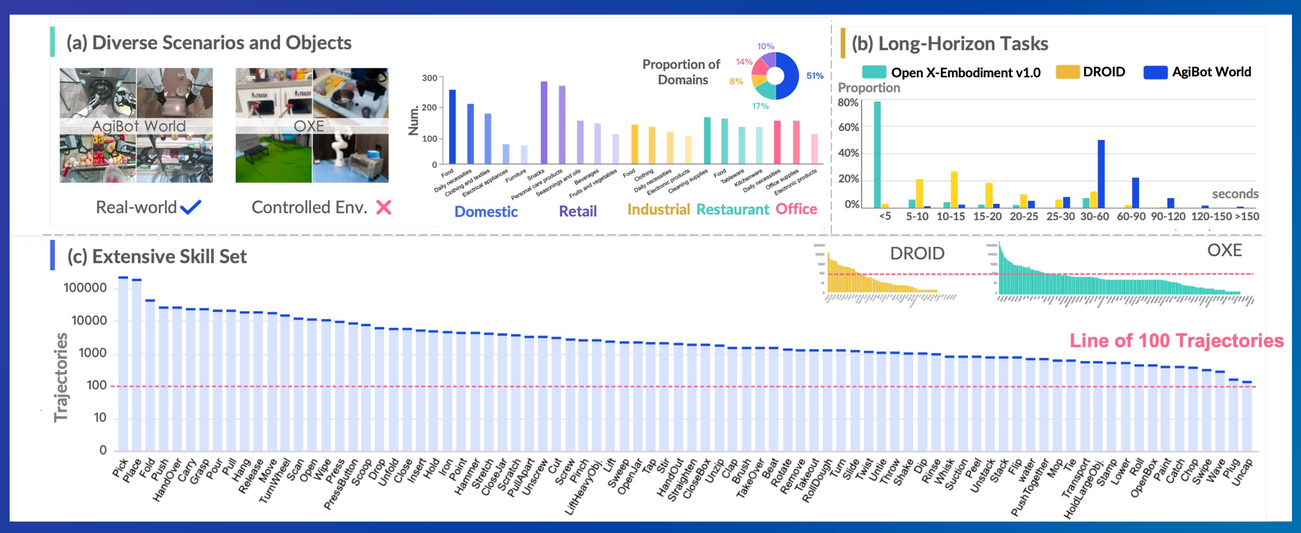


Figure 2. Dataset Statistics. AgiBot World dataset covers the vast majority of robotic application scenarios and interactive objects. Our dataset features long-horizon tasks and focuses on valuable atomic skills, each supported by a minimum of 100 trajectories.

GO-1

We propose GO-1, a hierarchical Vision-Language-Latent-Action (ViLLA) framework pretrained on web-scale heterogeneous data.
It predicts latent action tokens to enable embodiment-agnostic long-horizon planning and efficiently bridge general-purpose vision-language models (VLMs) with robotic decision-making.

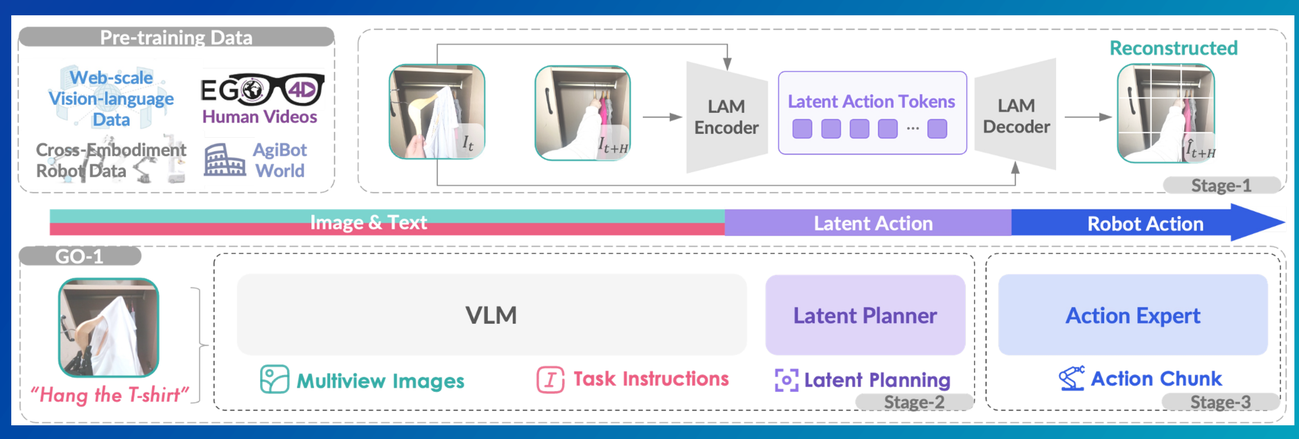


Figure 3. GO-1, a scalable robot foundation policy.

Experiments

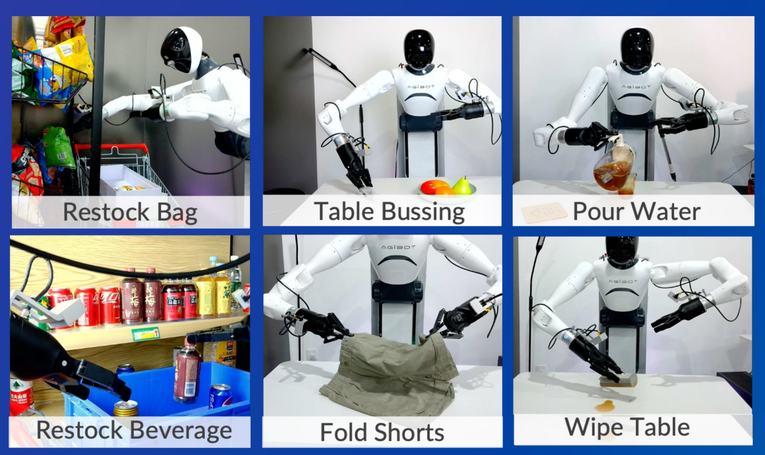


Figure 4. Experiment Setups for policy evaluation

Figure 6. Policies pre-trained on our dataset (236h for alpha) outperform those trained on OXE(~2000h).

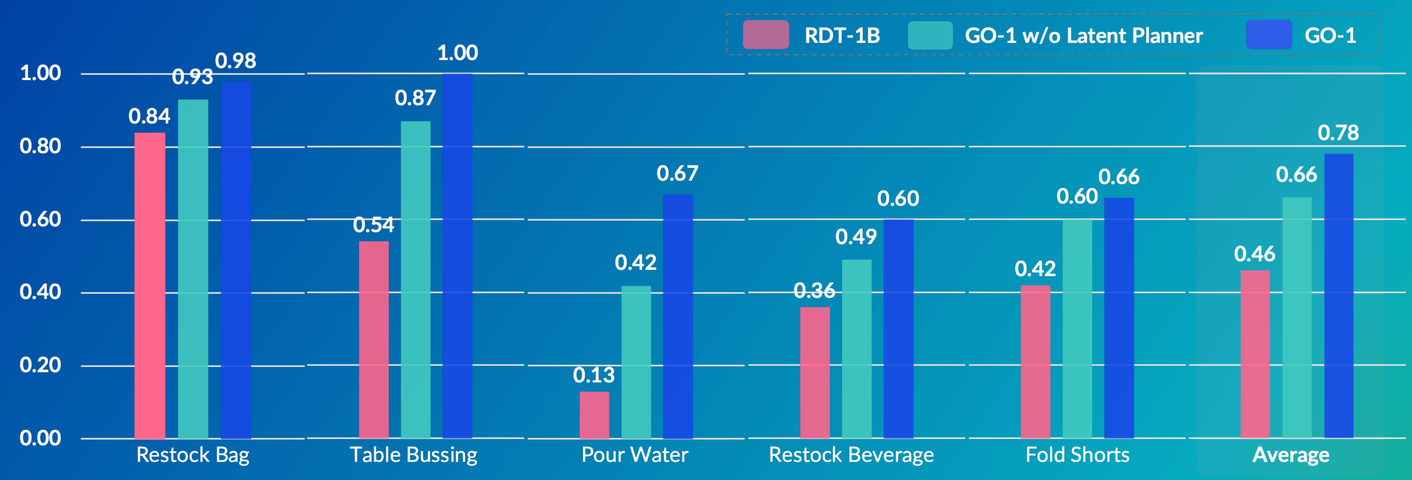
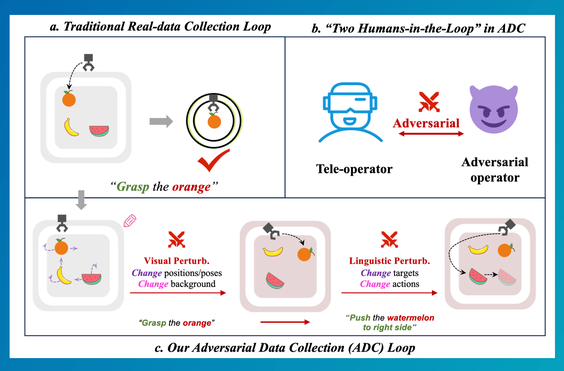


Figure 5. We evaluate different models pre-trained on AgiBot World dataset, where GO-1 significantly outperforms previous SOTA policy RDT, and the inclusion of the latent planner in the ViLLA model further improves performance.



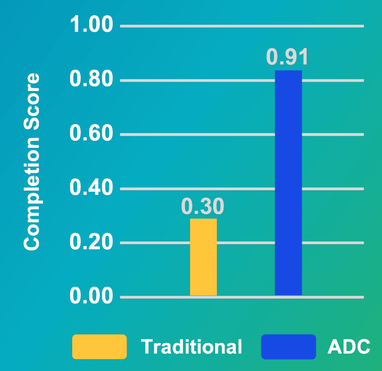


Figure 7. We present Adversarial Data Collection (ADC), a framework that transforms robotic data acquisition through real-time human-environment interactions, effectively bridging data-driven learning and practical robot deployment.

RSS AC Meeting

2025.04