

Di Wu

24 years old | RL, Embodied agent
13966610828 | diwu7012@gmail.com



Education

Sep. 2020 - Sep. 2024	Software Engineering (Bachelor)
Tongji University	
Sep. 2024 - Present	Computer Science (Master)
Tongji University	

Honors & Awards

- CET-4 626 point, IELTS 7.5 , National Scholarship for Graduate Students (1/65).

Research

Generative Multi-Agent Collaboration in Embodied AI: A Systematic Review, 2025

Accepted by IJCAI 2025, Sole first author. This work presents the first systematic survey on collaborative behaviors among multiple agents in embodied scenarios from the perspective of the generative-model era. We propose that agents may exist as either embodied entities or internal components within foundation models. This leads to a distinction between external and internal collaboration, which we analyze through key system designs and future directions.

On-Policy Reinforcement Fine-Tuning with Offline Reward for Multi-Step Embodied Planning, 2025

Accepted by ACL 2026 Main Conference, Sole first author. This is the first work to apply RLVR-based reinforcement fine-tuning to embodied planning tasks. Unlike conventional reasoning tasks, embodied planning involves a multi-turn, interactive environment. We therefore introduce a simple yet effective offline reward method to address sparse rewards in multi-turn tasks and the high cost of interaction. We provide a theoretical proof of the algorithm’s upper bound and achieve SOTA performance on EmbodiedBench. This work has been cited by notable studies such as ERA and Pelican-VL.

REvolve-VLA: Reflection-Evolving Reward Functions for Efficient VLA Post-Training, 2026

Ongoing work. Develops reflection-guided reward shaping and agent-driven evolution of reward functions and RL pipelines for efficient, self-evolving VLA post-training.

Benchmarking Offline Reinforcement Learning for Vision-Language-Action Models, 2026

Ongoing work. Offline RL can leverage suboptimal data more effectively than behavior cloning, yet its application to VLA models remains underexplored and lacks a unified evaluation platform comparable to D4RL. This project aims to build a standardized benchmark to systematically evaluate how offline RL improves VLA policies across different data-quality levels, while investigating the training and utilization of value models.

TextAtari: 100K Frames Game Playing with Language Agents, 2024

arXiv preprint. Constructs text-based environments for 23 Atari games using AtariARI to evaluate long-horizon decision-making and planning over up to 100K steps.

Startup Experience

Wange Intelligence (万格智元)

Participated in the Tsinghua University-affiliated startup project “Wange Zhiyuan,” which has secured tens of millions of RMB in seed funding. The team was awarded the Gold Prize in the 2025 China International College Students’ Innovation Competition, First Prize in the 2025 HICOOL Global Entrepreneurship Competition, and 1st Place Nationwide in the 2024 Amazon Generative AI Innovation Challenge. Website: [万格智元 | 端侧 AI 推理引擎与企业私有化部署](#)

Project experience

Apr 2024 - Sep. 2024	Tsinghua-Alibaba Research on Recommendation System
Participated in the Tsinghua × Alibaba Recommendation System Project, where we optimized the recommendation model. Offline metrics improved significantly, including +10pt AUC, +11pt Top 0.3% Hit Rate, and +60pt DiffRate. These enhancements led to substantial real-world business gains, with +6.04% conversion rate, +12.97% GMV per 1,000 impressions, and +8.25% exposure value in online production. The project also resulted in two research publications.	

Jul 2025 - Mar 2026

Tongji University & GALAXIS Embodied AI Collaboration Project

Led an collaboration project between Tongji University and Galaxis, targeting intelligent automation in warehouse handling and sorting scenarios. The goal is to build a generalizable VLA model and real-robot system tailored to real-world warehouse task. Over the course of 5 months, led a team of undergraduate students to complete the full-stack development from scratch, covering the entire R&D pipeline:

- **Real-World Hardware Integration:** Built on the Realman dual-arm wheeled robot platform. Completed key hardware adaptations including gripper and camera integration, hands-on debugging of hardware issues (e.g., USB disconnection), and custom gripper retrofitting for warehouse-specific manipulation.
- **Simulation System Development:** Developed a warehouse logistics simulation environment using Isaac Lab, aiming to establish a task benchmark specifically tailored to warehouse manipulation and sorting tasks.
- **Data Collection & Replay System:** Designed and implemented a high-efficiency data collection and replay framework supporting real-time visualization, trajectory synchronization, and post-processing for large-scale training.
- **Model Training and Deployment:** Built the end-to-end VLA training and deployment pipeline based on pi05, implemented on-device inference code with latency optimization to ensure responsive performance in real-world.
- **Long-Horizon Sorting Demo:** Constructed a multi-stage warehouse sorting demo involving perception, grasping, transporting, and placing, achieving stable real-robot execution of long-horizon tasks.

May 2026 - Sep. 2026

Agentic Intelligence Lab, The University of Hong Kong

Research Intern

Worked as a Research Intern at the HKU Agentic Intelligence Lab and contributed to the following projects:

- (1) **Embodied Foundation Model Pre-training:** Pre-trained an embodied foundation model using thousands of hours of real-world robotic data and egocentric data, with the goal of investigating the scaling effects of egocentric data on pre-training. Primarily responsible for data collection and processing, data-quality filtering, and model pre-training.
- (2) **Reinforcement Learning Post-training for Embodied Models:** To address the low sample efficiency and slow convergence of existing VLA RL post-training methods, explored a reflection-based reward-shaping mechanism that incorporates reflection signals generated during policy execution into VLA online RL post-training, thereby improving the overall VLA-RL post-training pipeline.

Social Practice & Service

- Short exchange program at Imperial College London
- Participant: CEAI 2025;CMAS 2025;CMAS 2026; IJCAI 2025 Volunteer ; IROS 2026; ICML 2026
- Reviewer: ICLR 2026, CVPR 2026, NeurIPS 2025 MIT-LLM Workshop, ICLR 2026 LLA Workshop, NeurIPS 2026 Meta-Agents Workshop