

吴笛

24岁 | 男 | RL, Embodied agent
diwu7012@gmail.com



教育背景

2020-09 ~ 2024-09	同济大学	软件工程（本科）
2024-09 ~ 至今	同济大学（推免）	计算机科学与技术（硕士）

荣誉证书

- 英语四级626分，雅思7.5，本科优秀学生奖学金若干，研究生国家奖学金（1/65）

科研成果

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

IJCAI 2025中稿，独立一作，首篇在大模型时代视角下，对具身情景下的多智能体协作进行的系统性综述。我们指出，如今的Agent可以外化在一个具身实体上，也可以指代于一个内部的Foundation model，因此，我们将协作分为了外部协作与内部协作，并系统阐述了这样的系统的关键模块设计，应用与未来展望。

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

ACL 2026 main中稿，独立一作。首篇将RLVR强化微调应用在具身规划任务中。与传统的math推理任务不同，具身规划是一个多轮的，交互式的环境，因此，我们应用了一种简单但有效的离线奖励方式，解决多轮奖励稀疏 & 具身交互昂贵的问题。我们给出了此算法理论上的upper bound证明，并在EmbodiedBench上取得了SOTA的效果。此工作被ERA, Pelican-VL等知名工作引用。

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

Ongoing Paper. 探索 (1) 利用Reward Shaping将反思信号融入到VLA RL后训练中，提高RL效率 (2) 利用 agent 自进化reward function和整体RL后训练系统。从而实现针对VLA-RL后训练模块的Agentic Self-evolving

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

Ongoing Paper. Offline RL相较于BC能更有效地利用次优数据，但其在VLA中的研究仍处于早期阶段，缺乏类似D4RL的统一评测平台。因此，我们拟构建标准化Benchmark，系统评估不同数据质量下Offline RL对VLA的策略提升，并探究Value Model的训练与使用。

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

Arxiv预印本。基于AtariARI构建23款游戏文本化Env，评测语言智能体在最长可达 10 万步环境下的超长时序决策/规划能力。

实习经验

2024-04 ~ 2024-09	清华&阿里推荐系统项目	实习参与
参与清华&阿里的推荐系统项目，通过优化推荐模型与排序策略，离线指标显著提升（AUC +10pt, Top0.3% Hit rate +11pt, Diffrate +60pt），并在真实线上业务中带来转化率 +6.04%、千展 GMV +12.97%、曝光价格 +8.25% 的收益。		

2025-07 ~ 2026-03	同济&凯乐士具身合作项目	group leader
参与同济大学 × 凯乐士具身智能合作项目，面向仓储搬运/分拣场景推进智能化落地，目标产出对标仓储场景，跨出简单桌面任务的通用 VLA模型与真机系统。5 个月内带领本科生完成实验室 从 0 到 1 的真机系统构建与 VLA 训练部署，覆盖完整研发闭环如下：		
- 真机硬件适配：基于 Realman 双臂轮式机器人，完成夹爪和相机等关键硬件适配，例如拆机排障、更换改造夹爪适配物流场景等 - 数采系统构建：自主搭建高效数据采集与replay框架，支持可视化采集、同步与后处理功能 - 模型训练与部署：基于pi05完成从训练到部署的 VLA 流程搭建，构建本地部署代码并优化推理速度，提升真实场景下的执行效率。 - 长程物流分拣 Demo：构建多阶段物流分拣 Demo，涵盖识别、抓取、搬运、投放等完整流程，产出5min级别物流分拣demo。		

2026-05 ~ 2026-09	香港大学 Agentic Intelligence-Lab	Research Intern
在HKU Agentic Intelligence Lab 作为 Research Intern，参与以下项目工作：(1) 具身基模预训练：拟使用数千小时的真机机器人数据与Ego数据，进行具身基座模型的预训练，探索Scaling Ego数据对于预训练的效果。其中主要负责数据采集处理与质量筛选，以及模型预训练的相关工作。(2) 具身强化学习后训练：针对现在的VLA RL后训练样本效率低，收敛速度慢的问题，拟提出一种基于反思的Reward shaping机制，将训练执行过程中的反思信号加入到VLA Online RL的后训练中，从而优化VLA-RL的后训练流程。		

创业经验

万格智元	核心初创团队
参与清华大学万格智元创业项目，已获数千万规模种子轮融资，获得中国国际大学生创新大赛（2025）金奖、HICOOL全球创业大赛（2025）一等奖亚马逊生成式ai创新挑战赛(2024) 全国第一。团队基于自研的Multi-Agent Task Distribution和端侧非GPU推理优化两大核心技术，成功将强大的AI能力直接部署于个人设备。公司官网： 万格智元 端侧 AI 推理引擎与企业私有化部署	