冯雪扬

xueyangfeng@ruc.edu.cn · (+86) 15698396778 · Google Scholar · GitHub

教育经历

中国人民大学高瓴人工智能学院，北京 2023 – 至今

人工智能专业博士研究生；导师：陈旭（准聘副教授）；预计 2028 年 6 月毕业；GPA：3.93/4.00

西北工业大学软件学院，西安 2019 – 2023

软件工程专业工学学士；GPA：3.77/4.00

研究概述

研究聚焦 **LLM Agent Post-training 与 Long-Horizon Agents**，关注智能体如何从交互经验与反馈中持续改进策略，并在复杂真实环境中可靠完成任务。主要研究方向包括：

- **长程任务执行：** Computer-Use Agent（浏览器场景）、技能发现与组合、由短到长的探索与策略训练。
- **动态编排与执行纠偏：** Subagent-as-Tool 多智能体协作、执行中的目标与 Harness 调整、基于事后反馈的编排策略优化。
- **主动交互与智能体后训练：** 主动澄清、搜索工具调用、用户模拟；Agentic RL、DPO、On-Policy Distillation，以及多轮信用分配与过程监督。

以**第一作者发表 6 篇**、**共同第一作者发表 1 篇** LLM Agent 相关论文，其中 **CCF-A 5 篇**（NeurIPS 3 篇、ACL 1 篇、WWW 1 篇），另有 EMNLP、NAACL 各 1 篇。

实习经历

华为诺亚方舟实验室，深圳 研究实习生 2024.06 – 2026.03

导师：戴全宇。围绕浏览器长程任务与面向用户的智能体开展研究，涉及技能探索、工具使用、主动交互、长期记忆与策略训练。

- **Computer-Use Agent：浏览器长程探索与训练**。提出 SCOPE，从历史执行轨迹中提取可复用技能与失败经验，通过技能发现、编排与扩展逐步形成复杂任务的执行能力，并结合 agentic on-policy distillation 优化策略。在 WebArena、WebVoyager 和 WebShop 上，相较各设置下的最强基线取得 3.9%–18.6% 的相对提升（论文发表于 NeurIPS 2026）。
- **主动用户交互与工具使用（小艺探店）：** 构建用户模拟器，利用反馈信号辅助多轮信用分配，通过 DPO 优化对话式推荐效果与用户侧可感知的体验指标（论文发表于 ACL Findings 2025）；引入生成式内在动机，联合优化搜索工具调用、主动澄清与回复生成能力（论文发表于 WWW 2026）。
- **长期记忆与个性化（小艺对话）：** 构建长期记忆与偏好利用评测框架，研究个性化记忆对助手决策行为的影响，识别不合理的偏好利用行为（论文发表于 NeurIPS 2026 E&D Track）。
- **多目标强化学习与用户反馈：** 构建情感陪伴用户模拟器，跟踪多轮情绪与认知需求变化；通过需求感知的多目标强化学习，将延迟反馈转化为逐步优化信号，改善长期用户状态与回复支持质量（论文发表于 NeurIPS 2026）。

北京智源人工智能研究院，北京 研究实习生 2026.05 – 至今

导师：李朝卓。围绕长程任务的动态编排、执行纠偏与编排策略训练开展研究。

- **动态编排与策略优化：** 提出 POrchestra，以 Subagent-as-Tool 将任务委派给具有独立上下文的子智能体，支持执行过程中主动请求调整目标与 Harness，并保持局部执行连续性；基于这一机制进一步提出 On-Policy Orchestration Distillation（OPOD），将调整请求归因到先前的编排决策，结合事后反馈、过程监督与最终奖励优化编排策略。在 GAIA、GAIA2 和 SWE-bench 上验证（已投稿 ICLR 2027，审稿中）。

论文成果

[1] **From Local Skills to Long-Horizon Web Tasks: Progressive Skill Exploration for LLM Web Agents**.
Xueyang Feng, Xiaohe Bo, Quanyu Dai, et al.
NeurIPS 2026（已录用，CCF-A）.
面向 Computer-Use Agent 的浏览器长程任务，提出 Skill-Composed Progressive Exploration（SCOPE），通过技能发现、技能编排与技能扩展，将长程任务探索分解为从局部技能到完整任务能力的渐进学习过程；利用失败

经验引导新技能探索，并通过 agentic on-policy distillation 将技能链内化到策略中。在 WebArena、WebVoyager 和 WebShop 上取得 3.9%–18.6% 的相对提升。

[2] **How Does Personalized Memory Shape LLM Behavior? Benchmarking Rational Preference Utilization in Personalized Assistants**.
Xueyang Feng, Weinan Gan, Xu Chen, et al.
NeurIPS 2026 E&D Track（已录用，CCF-A）.
构建面向个性化智能体的偏好利用评测框架，系统研究长期记忆对 LLM Agent 决策行为的影响，并揭示个性化过程中潜在的非理性偏好利用现象。

[3] **Optimizing Multi-Turn Interactive Recommendation Agents via Generative Intrinsic Motivation**.
Xueyang Feng, Jiakai Tang, Xu Chen, et al.
WWW 2026.（CCF-A）
提出生成式内在动机机制，在稀疏反馈下联合优化对话 Agent 的搜索工具调用、主动澄清与回复生成能力，提升主动探索和多轮交互效果。

[4] **Expectation Confirmation Preference Optimization for Multi-Turn Conversational Recommendation Agent**.
Xueyang Feng, Jingsen Zhang, Jiakai Tang, et al.
ACL Findings 2025.（CCF-A）
构建用户模拟器，基于期望确认机制生成反馈信号，辅助多轮交互中的信用分配，并通过 DPO 优化对话式推荐效果与用户侧可感知的体验指标。

[5] **Need-Aware Multi-Objective Reinforcement Learning for Emotionally Intelligent LLM Agents**.
Xiaohe Bo*, **Xueyang Feng***, Rui Li, Zihang Tian, Xu Chen, Quanyu Dai.
NeurIPS 2026（CCF-A）.
提出心理学驱动的用户模拟器 MindFlow 与需求感知多目标强化学习框架 NeMo，动态权衡情绪安抚与认知引导目标；通过前向模拟将延迟反馈转化为逐步优化信号，改善长期用户状态与回复支持质量。

[6] **Proactive Agentic Orchestration**.
Xueyang Feng, Chaozhuo Li, Xiaohe Bo, Jiakai Tang, Chenxi Li, Qiwei Ye, Xu Chen.
ICLR 2027（已投稿，审稿中）.
智能体编排通过将任务委派给具有独立上下文的子智能体，扩展系统的有效执行时程。针对现有方法在创建子智能体时固定目标与 harness、难以适应动态环境的问题，提出主动编排框架 POrchestra，以 Subagent-as-Tool 支持子智能体在不中断局部执行的情况下主动请求目标与 Harness 调整；提出 OPOD，将调整请求归因到先前的编排决策，并将事后反馈转化为过程监督。在 GAIA、GAIA2 和 SWE-bench 上，相较最强基线，任务成功率平均相对提升 9.1%；结合 OPOD 后，小规模编排器相较现有训练基线的成功率平均相对提升 10.0%，并取得更优的成本与性能权衡。

[7] **Improving Retrospective Language Agents via Joint Policy Gradient Optimization**.
Xueyang Feng, Bo Lan, Quanyu Dai, et al.
NAACL 2025.（CCF-B）
提出联合策略梯度优化框架，对智能体的任务推理与回顾反思过程进行协同训练，提升其基于环境反馈进行策略级自我改进的能力。

[8] **Large Language Model-based Human-Agent Collaboration for Complex Task Solving**.
Xueyang Feng, Zhi-Yuan Chen, et al.
EMNLP Findings 2024.（CCF-B）
研究复杂任务中的人机协作决策，并利用强化学习优化智能体请求人类干预的策略。

[9] **A Survey on Large Language Model Based Autonomous Agents**.
Lei Wang, Chen Ma*, **Xueyang Feng***, et al.
Frontiers of Computer Science, 18(6), 2024.
较早系统梳理 LLM-based Agent 研究框架与关键技术的综述工作，Google Scholar 引用 5000+ 次。

* 表示共同第一作者。

荣誉与奖励

国家奖学金（2020–2021）；西北工业大学优秀毕业生、陕西省优秀毕业生（2023）；中国人民大学二等学业奖学金（2023–2025）。