

🎓 教育背景

- ▶ 硕士, 苏州大学, 自然语言处理实验室, 计算机科学与技术学院 2024.09 - 至今
- ▶ 学士, 齐鲁工业大学, 智能科学与技术专业, 数学与统计学院 2020.09 - 2024.07

💼 实习经历

🐼 百度 / 应用模型技术部 (AMU) / 多模态感知与理解组 2025.09 - 2026.03

- ▶ **文心-闪电 VLM 基座多阶段渐进式预训练与后训练** | 面向 AI 搜索的专精多模态 MoE 大模型基座预训练与后训练能力
 - **项目背景:** 建设搜索专精多模态 MoE 大模型, 优化 AI 搜索场景下大模型应用的效果, 提升用户满意度。
 - **我的职责:** 基于 **Megatron** 参与数据建设、架构适配与训练迭代; 使用 **MapReduce** 完成大规模预训练数据清洗与质量评估; 负责适配 InternViT/RiceViT 与 Lightning LLM/Qwen3 LLM 用于消融实验, 并实现 Megatron-HF 权重双向转换; 参与搭建 **H200 千卡训练通路** 与 EvalScope 评测流水线, 依据评测迭代 SFT 训练数据配比和策略。构建 MPO 偏好数据, 并基于 DeepSpeed 框架开展 **MPO 算法** 后训练, 专项提升数学推理能力。
 - **成果:** 模型在 VLM 领域通用 Benchmark 上, 对比 **Qwen3-VL-30B-A3B** 模型, 综合性能达到 **105.19%**; 其中 STEM 领域性能达到 **124.11%**, 平均思维链输出 token 长度下降 **18.7%**; 模型推理速度提升 **23.1%**。
- ▶ **文心-闪电 VLM 手机百度文心助手拍照解题业务后训练** | 面向 K12 拍照解题场景的 VLM 业务后训练与在线优化能力
 - **项目背景:** 建设面向手机百度文心助手的拍照解题场景 (**K12 题目**), 效果与性能综合最优的 VLM 模型。
 - **我的职责:** 围绕真实用户拍照解题数据构建数据飞轮与 **MPO 偏好样本**; 在 **VeRL** 中适配文心-闪电 30B-A3B MoE 模型, 并搭建百卡 **GSPPO** 训练通路; 同时搭建离线测评平台以迭代奖励与约束, 从而缓解 **Reward Hacking**; 优化 **LLM-as-a-Judge** 调度以减少训练气泡, 并修改 **vLLM** 图像处理逻辑以压缩视觉 Token、降低推理成本。
 - **成果:** 经过 MPO 训练的模型在拍照解题测评集上的准确率达到 **81%**, 超过线上模型的 79%; 后续经过 GSPPO 训练的模型准确率提升至 **85%**, 达到 **Seed1.6 水平**。上线后 PV 显著正向提升 **1.37%**、搜次显著正向提升 **1.12%**。

📁 项目经历

- ▶ **具备前瞻规划与自进化能力的 LLM Agent 强化学习框架** | Agentic RL 开源项目 2025.06 - 2025.09
 - **项目背景:** 在开展长程 Agent 研究过程中, 定位并修复 VeRL 主库 **Agentic RL 训练链路 bug** (PR 合入主干), 随后基于 **verl-agent** 框架扩展训练框架, 解决 ALFWorld 等动态环境中 LLM Agent 只依赖当前观测、缺少环境预演能力, 导致探索效率低和无效动作比例高的问题。引入 World Model 的思想解决上述问题。
 - **核心工作:** 设计并实现 **World Model** 驱动的前瞻规划模块, 在多层 rollout 中对每个状态采样多个候选动作, 调用 World Model 批量预测候选动作后的 **next observation**, 再结合 **规则奖励 / 奖励模型** 对预测状态打分, 选择期望收益最高的动作执行。从而模拟人在决策前在大脑中的前瞻规划, 选择预期结果最好的动作。
 - **实验效果:** 在 Qwen2.5-1.5B Base 上, 协同训练 40 steps 后成功率达到 50.78%, 收敛后最高达到 **93.84%**; 训练日志显示 World Model loss 从 **0.824 降至 0.155**。执行 ALFWorld 任务的平均 step 数在训练收敛后下降了 **70%**。

🔬 科研经历

- ▶ **EffectMem: Effectiveness-Driven Memory Reuse for Long-Horizon LLM Agents**
一作, **EMNLP 2026 已录用, CCF-B**
 - **研究背景:** 长程 LLM 智能体需要复用历史交互轨迹辅助规划与执行, 但现有记忆增强方法多依赖语义相关性检索, 缺少对当前状态适用性和执行结果的检查。
 - **主要问题:** 语义相关的历史轨迹未必在当前任务中可执行, 容易因对象、状态或动作约束不匹配导致错误复用。
 - **解决方案:** 提出 EffectMem, 将成功轨迹构造包含 **全局适用条件**、**子目标动作块** 和 **预期后置条件** 的可执行合约; 推理时通过有效性门控路由筛选适用记忆, 在合约感知执行中结合 **约束注入**、**后置条件验证** 和 **差异引导纠错**。
 - **效果:** 在 ALFWorld、ScienceWorld 和 WebShop 上超过 ReAct、ExpeL、AutoGuide、CFG 等基线; GPT-4-Turbo 设置下成功率分别达到 **93.8%**、**79.0%**、**73.1%**, WebShop 平均奖励达到 **85.8**。
- ▶ **ANTI-EXCEPTION ACTION GENERATION FOR AUTOMATIC PLANNING**
一作, **ICASSP 2026 已录用, CCF-B**
- ▶ **RubricFlow-Mem: Rubric-Guided Self-Evolving Curriculum for Long-Term Memory Agents**
一作, **ICLR 2027 在投**