

刘茂琦

Post-Training · Continual Learning

+86 18537205012 | liumaoqi@bupt.edu.cn | estrellajer.github.io | Estrellajer | OpenReview



教育背景

北京邮电大学 · 人工智能学院

智能科学与技术 · 硕士研究生 · 导师：方全教授

2025.09 – 至今

硕士新生学业奖学金

北京邮电大学 · 人工智能学院

人工智能 · 工学学士

2021.09 – 2025.06

校级优秀毕业生

实习经历

字节跳动 · 抖音电商客服

2026.06 – 至今

LLM 算法实习生 · Mentor：何俊伟

北京

客服切换新架构，以单一 Agent 模型替换各场景独立优化的模型。独立负责完整场景的 RL 后训练：数据清洗筛选 → 奖励体系构建 → 模型训练与上线。

- **数据构建**：从线上会话中按“真实执行动作 + 事后结局”倒推构建训练与奖励标注数据，清洗筛选可用会话，好结局标注一致率 14% → 90%+。
- **解决性奖励**：对会话“是否真正解决”建模，蒸馏小模型作解决性 RM (F1 0.86)，并用 TabPFN 挖掘真实会话数据与解决性指标的关系；将轨迹级结局奖励分配到 step 级动作，实现细粒度信用分配。
- **体验性奖励**：将 Pointwise 打分改为换位 Pairwise 对比，叠加业务协议与动作白名单硬约束，显著提升 Reward 的组内排序能力，为 GRPO 提供有效优势信号。
- **训练与上线**：以双维奖励进行 GRPO 训练，售前场景推全上线，会话解决性 +3.2pp、体验性 +2pp、交互阻塞率 -7pp。
- **后训练算法探索**：系统探索 Rubric-Based RL、OPD 等算法，提出协议级评分细则 ProRubric 抑制 Reward Hacking（一作，论文 [1]）；与合作者完成 Agentic RL 工作，用 TabPFN 实现轨迹级到 step 级的 Policy Optimization；沉淀多篇技术博客（见个人主页）。

论文发表（均为第一作者）

[1] Scoring Higher, Answering Worse: Mitigating Reward Hacking in Rubric-Based RL via Protocol-Level Rubrics

Maoqi Liu, Junwei He, Bowen Zhang, et al.

Under Review

发现 Rubric-RL 的问题不在判据本身而在加权求和：判据可相互补偿，策略漏掉关键决策也能靠无关内容补分，覆盖率上升而恰当性下降（“分数更高、回答更差”）。提出离线将清单重组为协议级维度，全满足且未触发否决才计分，不改优化器；恰当性较 Rubric-RL +10.8，4B / 8B 七基准平均均最优。

[2] CoDe-LoRA: Mitigating the Orthogonality Dilemma in Continual Learning of LLMs via Knowledge Consolidation and Decoupling

Maoqi Liu, Quan Fang, Yufei He

EMNLP 2026 · CCF-B

发现严格正交隔离在抑制遗忘的同时剥离了任务间的共享成分，且任务越相似损害越大（“正交困境”）。提出免回放的双分支 LoRA：共享分支以自适应零空间投影整合共性，专家池以原型语义路由解耦特性、无需任务 ID；标准 CL 基准平均精度 80.6%，超过多任务联合训练，长任务序列上较 O-LoRA +7.0。

[3] NOTAM-Evolve: A Knowledge-Guided Self-Evolving Optimization Framework with LLMs for NOTAM Interpretation

Maoqi Liu, Quan Fang, Yuhao Wu, et al.

AAAI 2026 · CCF-A

指出 NOTAM 深度解析不止文本抽取，还需动态知识对齐（如查询机场跑道构型）与基于规章的推理。提出知识图谱增强检索 + 以模型自身错误构造偏好对的 SFT / DPO 迭代自进化，无需人工推理标注；7B 模型四任务平均准确率 0.458 → 0.762，接近 GPT-4o。

[4] Knots: A Large-Scale Multi-Agent Enhanced Expert-Annotated Dataset and LLM Prompt Optimization for NOTAM Semantic Parsing

Maoqi Liu, Quan Fang, Yang Yang, et al.

ADVEI 2025 · SCI Q1 · CCF-B

将 NOTAM 从表层抽取推进到需领域推理的语义解析，构建覆盖全球 194 个飞行情报区、12,347 条专家标注的数据集；以多智能体 + 辩论框架自动发现字段 (F1 92%，传统方法 57%)，并发现 few-shot 带来最大提升 (F1 55.0 → 81.7)，而 CoT 收益不稳定。

荣誉与个人陈述

- **北京市自然科学基金**：“启研”计划资助（本科生科研项目，5 万元）2024.09
- 熟悉 verl / VeOmni 等 RL 框架，可快速学习复现前沿方法；常写技术博客，ByteTech Lv5 作者、AI 知识达人（Top 3%）。