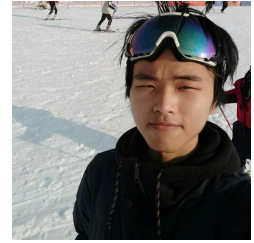


钟龙广

手机：16675151404 · 邮箱：2813409972@qq.com

微信：GGLS-258-369

个人主页: <https://mrggls.github.io>



教育背景

中山大学，计算机技术，硕士

2023.09 - 2026.06

- 导师：权小军
- 研究方向：轻量化大模型、异构模型融合、人类偏好对齐、大模型强化学习

西安电子科技大学，软件工程，本科

2019.09 - 2023.06

- GPA：3.9/4.0
- 荣誉/奖项：西安电子科技大学数学建模校赛一等奖，校一等、二等奖学金

实习经历

月之暗面 RL Team

2025.6 - 至今

主导 Agent-RL 场景下的过程价值模型 (Process Value Model) 研发

- 动机：Agent-RL 任务具有长链条多轮决策、环境回报稀疏两大约束，传统 token-level value 机制把每个词元都当作独立决策单元，导致对于长达数千词元的单轮响应，逐词元的价值信号不仅稀疏且充满噪声且难以对齐宏观意图与最终回报。直觉上，将价值近似的最小粒度从 token 提升到完整的 Agent 响应更符合现代 Agent-RL 的需求
- 做法：基于 57k+ 问题、236 万条内部游戏轨迹与 21 万条 DeepResearch 轨迹，离线预训练 14B 规模 PVM。随后在线阶段采用 Actor-Critic 框架 (PPO + GAE) 交替更新 Policy 与 PVM
- 离线评估效果：对于双人对抗游戏，PVM 能够实现 90%+ 的对局胜利预测的准确率；在 DeepResearch 场景下 PVM 的离线评估效果 (82.4) 远胜 pass@1 (65.3)，且已经很接近 pass@8 (88.4)
- 在线更新收益：在游戏上将收敛轮次从 60 \rightarrow 30，且显著提升峰值奖励；在 DeepResearch 场景下峰值 accuracy 超越基线 5 个点

参与内部 Agent RL 训练框架的搭建和算法讨论

- 在 Kimi-Researcher/K2 训练框架完成 PVM 的接入和在线更新，并验证了 PVM 的有效性
- 完成 TextArena 环境的接入，包含 90+ 游戏场景的实现
- 实现 GSPO、PVM 的训练支持

科研经历

Model Fusion / 模型融合方向

- FuseRL: Dense Preference Optimization for Heterogeneous Model Fusion [Paper]
- 提出异构模型融合框架，利用奖励分数动态调整不同数据源在 SFT 与偏好学习阶段的学习强度，最大化源模型利用率
- AAAI 2026, Under Review, 第一作者
- FUSECHAT: Knowledge Fusion of Chat Models [Paper] [Code] [HF Daily Papers] [Blog]
- 设计两阶段异构对话大模型知识融合方法 (成对知识融合 + 模型合并)，将 6 个开源异构对话大模型能力汇聚至单一模型
- EMNLP 2025 Main, 第二作者
- Weighted-Reward Preference Optimization for Implicit Model Fusion [Paper] [Code] [HF Daily Papers]
- 通过隐式加权奖励实现多源模型融合，无需显式参数平均即可提升目标模型性能
- ICLR 2025 Poster, 第三作者
- FuseChat-3.0: Preference Optimization Meets Heterogeneous Model Fusion [Paper] [Code]
- 交替优化不同源模型的内部响应偏好，构建覆盖数学、代码、指令遵循的后训练数据与评测框架
- SCI-FM @ ICLR 2025, 第三作者

Model Pruning / 剪枝方向

- **BlockPruner: Fine-grained Pruning for Large Language Models** [\[Paper\]](#) [\[Code\]](#)
- 发现 MHA 与 MLP 模块重要性差异，提出按模块重要性进行细粒度结构搜索的剪枝方法，显著优于以往结构化剪枝
- *ACL 2025 Findings*, 第一作者

Efficient Reasoning / 推理效率方向

- **ThinkSwitcher: When to Think Hard, When to Think Fast** [\[Paper\]](#)
- 构建轻量级自适应框架，让大推理模型根据任务复杂度在“快/慢”思考模式间动态切换，提升简单任务推理效率
- *EMNLP 2025 Findings*, 第二作者