

梅灵睿

联系方式: (+86)18512377787

代码仓库: github.com/Meirtz

邮箱地址: meilingrui22@mails.ucas.ac.cn

学术主页: [Google Scholar \(1500+ 引用\)](https://scholar.google.com/citations?user=Meirtz)

个人网站: me.meirtz.com/about

常驻位置: 北京市



教育经历

中国科学院计算技术研究所 | 硕士、博士研究生 / 研究方向: 大语言模型

2022.09—2027.07

智能算法安全全国重点实验室 | 计算机科学与技术

重庆交通大学 | 工学学士

2015.09—2020.06

土木工程学院 | 土木工程 · 计算机科学与技术

研究兴趣

- **Agentic RL 与可验证后训练**: 研究面向 Computer Using Agent、Code Agent 与 Deep Research Agent 的 RLVR/RFT 数据构造、可执行环境、程序化 reward 与训练稳定性。
- **大语言模型长上下文与记忆机制**: 研究可微工作记忆、条件计算、上下文工程, 提升模型在长程推理和复杂任务中的可靠性。
- **大模型安全与可信对齐**: 研究 LLM 安全生成、幻觉/越狱判别、上下文忠实性与多模态 RAG 偏差。

部分科研经历

大语言模型, 可信人工智能, 语言模型智能体 (2023.11—至今)

一作论文

- [1] Gated Differentiable Working Memory for Long-Context Language Modeling. [ACL 2026, Oral]
- [2] a1: Steep Test-time Scaling via Environment Augmented Generation. [ACL 2026]
- [3] HiddenGuard: Fine-Grained Safe Generation with Specialized Representation Router. [ACL 2026]
- [4] A Survey of Context Engineering for Large Language Models. [HuggingFace 月榜第二, MIT 科技评论等多家媒体报道]
- [5] “Not Aligned” is Not “Malicious”: Being Careful about Hallucinations of Large Language Models’ Jailbreak. [COLING 2025]
- [6] SLANG: New Concept Comprehension of Large Language Models. [EMNLP 2024]
- [7] Separations and Allocation for Conditional Memory and Computation. [NeurIPS 2026, Pending]
- [8] Can We Trust Learned Speculative-Decoding Controls? [EMNLP 2026, Pending]

合著论文

- [1] Reward and Guidance through Rubrics: Promoting Exploration to Improve Multi-Domain Reasoning. [ICML 2026, Spotlight]
- [2] YuE: Scaling Open Foundation Models for Long-Form Music Generation. [ICLR 2026]
- [3] Parameters vs. Context: Fine-Grained Control of Knowledge Reliance in Language Models. [ICLR 2026]
- [4] AdaptFlow: Adaptive Workflow Optimization via Meta-Learning. [EMNLP 2025]
- [5] Who is in the Spotlight: The Hidden Bias Undermining Multimodal Retrieval-Augmented Generation. [EMNLP 2025, Oral]
- [6] Decoding by Contrasting Knowledge: Enhancing LLMs’ Confidence on Edited Facts. [ACL 2025]
- [7] Context-DPO: Aligning Language Models for Context-Faithfulness. [ACL 2025]
- [8] EvoStealer: Differential Evolution for Prompt Template Stealing Against Text-to-Image Synthesis. [ACL 2025]
- [9] Is Factuality Enhancement a Free Lunch For LLMs? Better Factuality Can Lead to Worse Context-Faithfulness. [ICLR 2025]
- [10] Adaptive Token Biased: Knowledge Editing via Biasing Key Entities. [EMNLP 2024]
- [11] LPNL: Scalable Link Prediction with Large Language Models. [ACL 2024]
- [12] Detecting Out-of-distribution Samples via Variational Auto-encoder with Reliable Uncertainty Estimation. [Neural Networks 145, 199-208]

工作经历

腾讯 | 大语言模型部 / T3 / 青云计划研究实习生

2025.11—至今

- 参与 Computer Using Agent 的可验证 RLVR 数据飞轮与沙盒环境建设, 实现 agentic 数据进化方法构造可扩展的 agent 任务与匹配资产集, 打通可扩展的环境生成和维护、动作执行、轨迹回放、自动验收与失败过滤闭环, 形成可用于 RFT/RL 的高质量训练闭环并参与训练。
- 参与腾讯混元 Hy 2.0 与 3.0 及后续快思考模型 OPD 合版生产训练, 负责数据配比、训练配置、评测回归与异常样本归因, 基于 loss/reward、能力评测和 case replay 迭代训练方案; 试验性 model merge 方案探索。
- 参与文科能力 RM 训练优化实验, 构建偏好数据、user-feedback reward model 训练和校准评测链路, 跟踪人文社科、长文本理解与复杂问答场景的元宝石线上反馈和自建主观评测效果并反哺数据迭代, 持续优化模型在人文社科领域的回答质量。

- 参与 Skywork Deep Research Agent v2 数据飞轮与 SFT/RFT/RL 训练，DRS 任务数据飞轮建设；按多样性、唯一性、可验证性与挑战性构造高难任务，结合 GRPO、动态课程学习和线索式 Dense Reward 提升深度信息检索与复杂任务执行能力。
- 参与模型 RFT/SFT 训练及推理范式设计实现，设计数据配方、工具调用轨迹与可验证答案抽取流程，优化从 teacher rollout、训练样本筛选到离线评测的闭环。
- 参与并行思考与生成式结果验证机制设计，通过多候选路径、Tournament Ranking 与熵自适应剪枝扩展 test-time scaling；BrowseComp 38.7%、GAIA 79.1，刷新多项 Agent 任务 SOTA。

- 设计和训练领域特定 agent 基座模型，围绕领域任务构造训练数据与离线评测集，优化指令格式、工具调用策略和推理链路，支持模型在垂直场景中的能力适配。

- 参与提出 AdaptFlow/NL-MAML 工作流元优化框架，将 agentic workflow 表示为可学习的自然语言程序，基于任务聚类构造 subtask-level inner-loop adaptation，并通过 outer-loop 反思与聚合学习可泛化 workflow 初始化；在 QA、代码与数学推理 8 个基准上平均得分 68.5，优于 AFlow/ADAS 等 workflow search baselines。
- 参与提出 VEM 环境无关 GUI Agent 强化学习框架，将 GUI 决策建模为离线 state-action value estimation，先用语义进展信号训练 Value Environment Model，再冻结 VEM 以 PPO 引导 action policy 探索，避开在线交互、next-state simulation 与手写 reward；在 AITW 上以更交互成本达到 SOTA/近在线 RL 性能。

- 结合大语言模型表征优化小红书主页推荐算法，包括内流 + 外流生成式推荐算法模型。
- 业务数据 CTR 提升 1.5‰，显著改善用户体验和平台效果。

- 设计开发驾驶舱信号感知和道路信息感知算法，包括传感器融合、车载测试软件开发、模型部署和硬件性能优化。
- 参与制定 C-IAISI（中国保险汽车安全指数自动驾驶评估标准），相关研究赋能国家智能汽车集成系统试验区。

- 优化卷积脉冲神经网络，重写原有算子并整合入 Kaldi 框架，初步取代原有算子（+26.4%），并能在部分检测任务中取得比 LSTM 更佳的效果（+4.5%）。
- 设计开发 EEG 处理 pipeline，用于脑电信号解码和多模态生成。

能力栈

- 大模型训练与后训练: SFT, RFT, RLVR, RM, OPD 合版训练；熟悉 Megatron-LM, MBridge, veRL, PyTorch, Transformers。
- Agent 数据飞轮: 任务生成、环境状态构造、程序化 reward/verifier、teacher rollout、数据筛选、benchmark 评测与训练反馈闭环。
- 研究方向: Agentic RL、Computer Using Agent、Code Agent、Deep Research Agent、长上下文记忆机制、LLM 安全与可信对齐。

获奖实践

- 学术服务: 担任 NeurIPS 2025, ACL 2025, NAACL 2025, ICML 2024 等顶级会议审稿人；ICLR 2025 Notable Reviewer。
- 竞赛获奖: 中国科学院大学 2024-2025 年度三好学生；中国科学院大学学业奖学金；2020 SEC 中国场地拉力赛（亚军），2018 国际大学生类脑计算大赛（三等奖，唯一本科生），HACKxTHU 2018 一等奖，HACKxFDU 2017（2/32），HACKxSJTU 2017（第二名），RoboCup 2013（个人和团队世界冠军）等。
- 其他头衔: 中国汽车摩托车运动联合会 A 级场地赛车手 / 北京中汽联赛车俱乐部耐力车手 / 中国科学院大学特约摄影师。