

温学儒

wenxueru2022@iscas.ac.cn · Google Scholar · GitHub

教育背景

中国科学院大学, 北京	2023 – 至今
计算机应用技术博士在读 (预计 2028 年 6 月毕业); 导师: 何笨教授 中文信息处理实验室, 中国科学院软件研究所	
吉林大学, 长春	2019 – 2023
理学学士 唐敖庆理科试验班 (计算机方向)	

研究工作

CoVRL：无验证器强化学习 [1]	第一作者, ICML 2026
针对探索低效与思维链-答案不一致, 提出耦合先验与后验的变分强化学习框架。Qwen2.5-7B 在数学与通用推理基准上的平均准确率由 37.8% 提升至 50.2% (+12.4 个百分点), 较最优比较基线提高 2.3 个百分点。	
奖励模型评估 [4]	第一作者, ICLR 2025 Spotlight
系统分析评估准确率与下游 RL 效果的关系, 发现相近准确率可对应显著不同的策略表现; 从评估设计与 Goodhart 效应解释奖励过优化风险, 揭示仅凭准确率选择奖励模型的局限。	
Cheems：中文奖励模型体系 [5]	第一作者, ACL 2025
构建中文奖励模型训练与评估体系, 发布人机协作偏好数据集 CheemsPreference 和人工评估基准 CheemsBench; 评估开源模型的中文能力, 并训练中文奖励模型。	
RLFH：自我对齐与幻觉缓解 [6]	第一作者, ACL 2025
将生成内容分解为原子事实, 结合外部知识构造 token 级稠密奖励, 支持无人工标注的在线策略自我对齐; 在 HotpotQA 等基准上降低幻觉并提升回答准确性。	

实习经历

小红书 (2025 年 7 月起 Ace 顶尖实习生)	2023 年 9 月 – 至今
奖励模型迭代 : 结合偏好反馈、幻觉检测与过程奖励迭代 RLHF 奖励模型, 评估下游效果, 并探索可扩展监督机制。	
Math Verification Agent : 基于 Lean、Python 与自然语言验证器构建 agent, 进一步通过 SFT 与 agentic RL 增强验证能力, 提升 ProofBench 表现。	
SWE Coding Agent : 搭建 Docker/Kubernetes 采样链路, 实现大规模轨迹获取与验证, 并通过 SFT 评估数据效果; 实现 TITO、R3 与 partial rollout, 支持 agentic RL, 提升 SWE-bench Verified、Multilingual 与 Pro 表现。	
Automated Posttrain : 在云端 GPU 集群上构建大规模数据合成链路, 利用 world model 进行定向合成, 并在训练与推理侧实现分段 tokenization, 提升 PostTrainBench 表现。	

中文信息处理实验室	2023 年 3 月 – 2023 年 8 月
大模型预训练	
- 基于 Megatron-DeepSpeed 实现 LLaMA 训练支持、Checkpoint 转换与数值精度对齐 - 集成 Kernel Fusion 算子及 Flash Attention 等训练优化技术, 提升训练吞吐量	

荣誉奖励

中国科协青年科技人才培育工程博士生专项计划入选者	2025
中国科学院大学院长奖学金优秀奖 (中国科学院系统)	2026 年 8 月

论文成果

-
- [1] **Coupled Variational Reinforcement Learning for Language Model General Reasoning**
Xueru Wen, Jie Lou, Yanjiang Liu, Hongyu Lin, Ben He, Xianpei Han, Le Sun, Yaojie Lu, Debing Zhang. ICML 2026
- [2] **Decoupling Reasoning and Confidence: Resurrecting Calibration in Reinforcement Learning from Verifiable Rewards**
Zhengzhao Ma, Xueru Wen, Boxi Cao, Yaojie Lu, Hongyu Lin, Jinglin Yang, Min He, Xianpei Han, Le Sun. ICML 2026
- [3] **Learning from Failures: Correction-Oriented Policy Optimization with Verifiable Rewards**
Mengjie Ren, Jie Lou, Boxi Cao, Xueru Wen, Hongyu Lin, Xianpei Han, Le Sun, Xing Yu, Yaojie Lu. COLM 2026
- [4] **Rethinking Reward Model Evaluation: Are We Barking up the Wrong Tree?**
Xueru Wen, Jie Lou, Yaojie Lu, Hongyu Lin, Xing Yu, Xinyu Lu, Ben He, Xianpei Han, Debing Zhang, Le Sun. ICLR 2025 Spotlight
- [5] **Cheems: A Practical Guidance for Building and Evaluating Chinese Reward Models from Scratch**
Xueru Wen, Jie Lou, Zichao Li, Yaojie Lu, Xing Yu, Yuqiu Ji, Guohai Xu, Hongyu Lin, Ben He, Xianpei Han, Le Sun, Debing Zhang. ACL 2025
- [6] **On-Policy Self-Alignment with Fine-grained Knowledge Feedback for Hallucination Mitigation**
Xueru Wen, Jie Lou, Xinyu Lu, Yuqiu Ji, Xinyan Guan, Yaojie Lu, Hongyu Lin, Ben He, Xianpei Han, Debing Zhang, Le Sun. ACL 2025
- [7] **Transferable Post-training via Inverse Value Learning**
Xinyu Lu, Xueru Wen, Yaojie Lu, Bowen Yu, Hongyu Lin, Haiyang Yu, Le Sun, Xianpei Han, Yongbin Li. NAACL 2025
- [8] **AutoAlign: Get Your LLM Aligned with Minimal Annotations**
Xinyu Lu, Dong Xu, Chunkang Zhang, Xinyan Guan, Junxiang Wang, Qingyu Zhang, Pengbo Wang, Yingzhi Mao, Hao Xiang, Xueru Wen, Zichao Li, Yaojie Lu, Hongyu Lin, Le Sun, Xianpei Han. ACL 2025
- [9] **Critic-CoT: Boosting the Reasoning Abilities of Large Language Model via Chain-of-Thought Critic**
Xin Zheng, Jie Lou, Boxi Cao, Xueru Wen, Yuqiu Ji, Hongyu Lin, Yaojie Lu, Xianpei Han, Debing Zhang, Le Sun. ACL 2025
- [10] **The Devil Is in the Details: Tackling Unimodal Spurious Correlations for Generalizable Multimodal Reward Models**
Zichao Li, Xueru Wen, Jie Lou, Yuqiu Ji, Yaojie Lu, Xianpei Han, Debing Zhang, Le Sun. ICML 2025
- [11] **Offline Pseudo Relevance Feedback for Efficient and Effective Single-Pass Dense Retrieval**
Xueru Wen, Xiaoyang Chen, Xuanang Chen, Ben He, Le Sun. SIGIR 2023
- [12] **End-to-End Entity Detection with Proposer and Regressor**
Xueru Wen, Changjiang Zhou, Haotian Tang, Luguang Liang, Hong Qi, Yu Jiang. Neural Processing Letters 2023
- [13] **Type-supervised Sequence Labeling Based on the Heterogeneous Star Graph for Named Entity Recognition**
Xueru Wen, Changjiang Zhou, Haotian Tang, Luguang Liang, Yu Jiang, Hong Qi. arXiv 2022

开源工具

- **Skynet**: 原生 macOS agent 工作台，整合本地与 SSH 会话、后续任务队列及代码审查。
- **Python Hot Patch Skill**: 保留内存状态的 CPython 热补丁，支持安全点执行与独立验证。
- **Wait Skill**: 事件驱动的 agent 等待与任务协调，支持被动监控、定时循环及依赖工作流。
- **JSONLine Viewer**: VS Code 扩展，支持 JSONL/NDJSON 格式化预览与大文件索引导航。