



# 韩槿一



☎ 15224975562    ✉ jinyihan099@gmail.com    🏠 <https://jinyihan99.github.io>

## 🎓 教育经历

华东师范大学 (复旦大学联合培养) 学术博士 计算机科学与技术学院	2023.09 – 至今
东华大学 学术硕士 计算机科学与技术学院	2020.09 – 2023.03
河南财经政法大学 本科 计算机科学与技术学院	2016.09 – 2020.06

## 📖 研究方向

**大模型深度推理与效率优化:** 研究通过训练范式与推理策略优化提升大模型复杂推理能力, 包括 Deep Reasoning、Test-time Scaling、推理路径优化与推理效率优化 (Reasoning Efficiency)

**Agentic RL:** 研究基于 Agentic Reinforcement Learning 的智能体能力训练, 包括 Tool Use、Planing与持续自我进化, 探索具备自主决策与长周期任务执行能力的智能体系统。

## 📁 实习经历

字节跳动 中国交易与广告 筋斗云人才计划 (多模态大模型算法工程师) 2025.07 – 至今

### 01 - Reasoning-based SafetyGPT-GRM 奖励模型训练

- **背景:** 面向电商治理与安全场景, 研发具备推理能力的通用奖励模型 (GRM), 用于对前置模型生成的 CoT 推理结果进行反思校验、质量过滤与可解释分析, 提升领域数据质量与模型推理可靠性。
- **内容:** (1) 提出推理判定范式  $PRA \rightarrow A'R'$ , 使 GRM 能够基于问题  $P$ 、推理过程  $R$  与答案  $A$  生成反思推理  $R'$  并输出判定结果  $A'$ , 实现对模型推理质量的自动评估, 并应用于二次反思校验、模型校准等业务领域; (2) 针对传统 critic 模型在业务落地中存在的推理路径依赖与误差累积问题, 提出判定结构  $PRA \rightarrow A'R'A''$ , 使首 token  $A'$  直接承担高置信度筛选任务, 从而支持高精度阈值过滤与高质量样本挖掘; (3) 为了进一步优化 GRM 的首 token 卡阈值优化的能力, 设计 SFT + RL Merge Loss 联合训练框架。对首 token 使用 SFT Loss 保证判定稳定性, 对剩余推理过程使用 GRPO loss
- **效果:** (1) 在 Ecom-GRM-Reasoning-Bench 上, 对多模型输出进行反思校验, 相比直接生成答案平均提升准确率 +6.3 ~ +6.75 PP; (2) 在领域 Benchmark 中, 首 token  $A'$  的高精度筛选 Recall 达 26.86%, 相比  $A''$  提升 +26.77 PP; (3) 在 90% Precision 条件下, Merge Loss 相比纯 GRPO 训练 Recall 提升 +2.9 PP

### 02 - 仲裁智能化算法 (客服专项)

- **背景:** 面向电商仲裁场景, 建设智能化仲裁能力, 在申诉率不提升、用户满意度不下降的前提下替代部分人工客服, 并提升仲裁处理时效与判定准确率。
- **内容:** (1) 参与并落地 SOP 规则驱动的仲裁决策系统, 将原端到端大模型升级为可强规则 follow 的模块化多模型架构, 通过原子决策节点组合完成仲裁方案生成; (2) 训练并上线 policy 检索模型、时效判断模型、判证模型与方案生成模型, 实现 标签识别  $\rightarrow$  SOP 规则检索  $\rightarrow$  时效判断  $\rightarrow$  举证有效性判定  $\rightarrow$  售后方案生成的自动化仲裁流程; (3) 构建辅助识别模型, 包括首次反馈时间识别模型以及预测仲裁单是否能协商一致, 补充仲裁关键决策信息
- **效果:** 核心模型已上线应用于仲裁系统, 支撑客服仲裁自动化决策流程, 关键模块中 policy 检索模型 ACC 88.78%; 细分场景+商品类型识别 91.08%; 时效判断模型 ACC 98.95%, 判证模型 ACC 90.78%, 方案模型 ACC 89.49%

## 🔧 项目经历

01 - 大模型专业思维提升技术研究 华为 技术核心人员 2024.12 – 2025.12

- **背景:** 以 o1 为代表的深度推理模型在复杂任务中展现出显著优势, 但开源模型在复杂推理能力上仍存在明显差距, 因此探索提升开源模型深度推理能力的方法, 以增强其在复杂任务中的表现。
- **内容:** 模拟人类解决复杂问题时的思维过程, 设计并定义一系列思维算子 (如 规划、问题分解、信息提取、假设与启发、反思、小结与转折等); 结合 Inference Scaling Law, 在模型生成过程的关键位置插入思维算子, 引导模型按照指定思维模式进行多步推理, 从而提升复杂问题求解能力。
- **效果:** 在 AIME24 数学竞赛数据集上, Qwen2.5-32B 准确率达到 33%, 相较原始模型提升 +10 PP。

02 - 基于大模型的代码智能化技术 华为 子模块技术负责人 2023.11 – 2024.11

- **背景:** 构建具备“检索—复用—改写”能力的代码大模型, 使其生成流程更接近人类程序员的编程思维模式。

- **内容:** 通过 **合成低资源代码数据**提升模型在低资源场景下的代码生成能力; 结合 **智能代码检索**获取与需求相关的高质量代码片段, 增强模型代码复用能力; 最后通过 **风格一致性微调**, 引导模型生成与上下文规范一致的代码。
- **效果:** 成功构建面向代码生成场景的 **Copilot 系统**, 集成 **代码生成、代码补全与代码风格优化**等核心能力。

### 03 - R1-Zero 深度思考能力高效复现 (GRPO) [项目介绍](#) **主要贡献人** **1.1K Star** 2025.03 – 2025.04

- 开源社区 **最早复现 R1-Zero 自发反思能力**的项目之一, 基于 **GRPO 强化学习方法**实现深度思考能力训练, 核心实现仅约 **200 行代码**, 仅依赖基础深度学习库, 具备良好的 **轻量性与可复现性**。
- 显著降低 R1 复现训练成本: 仅需 **5 步强化训练**即可使 **Qwen2.5-3B 准确率**稳定在 **60%+**, 峰值约 **70%**; 训练至 **30 步**时模型格式遵循能力接近 **100%**, 并在 **200 步内**模型表现出 **“Aha Moment”**。

### 04 - 基于深度思考的大模型自主工具学习方法 [开源项目](#) **主要贡献人** 2025.04 – 2025.09

- 利用大模型的 **深度思考能力**探索复杂任务中的 **主动式工具使用策略**, 使模型能够自主判断是否调用工具、选择合适工具, 并将工具结果整合到推理过程中, 从而扩展模型能力边界。
- 设计两种工具使用机制: **“边想边干” (Think-Act)** 与 **“专业分工” (Planner-Executor)**; 通过 **强化学习训练**赋予模型自主工具调用能力, 使其在生成过程中能够动态选择工具辅助问题求解。
- 模型在复杂任务中能够 **主动调用工具**, 在简单任务中保持 **无工具推理**, 并在未见任务上展现良好的 **泛化能力**; 在 **9 位数以上复杂数学问题**上, 相较基线模型 **Qwen2.5-7B-Instruct**, 准确率提升 **+16.7 PP**。

## 论文列表

[\*: 导师第一作者, 学生为第二作者; †: 通讯作者; 横线加粗: 本人; 曾用名: Haixia Han]

#### [1] [A Stitch in Time Saves Nine: Proactive Self-Refinement for Language Models](#)

Jinyi Han, Xinyi Wang, Haiquan Zhao, tingyun li, Zishang Jiang, Sihang Jiang, Jiaqing Liang, Xin Alex Lin, Weikang Zhou, Zeye Sun, Fei Yu, Yanghua Xiao  
2026, ICLR.

#### [2] [Your Models Have Thought Enough: Training Large Reasoning Models to Stop Overthinking](#)

Jinyi Han, Ying Huang, Ying Liao, Haiquan Zhao, Zishang Jiang, Xinyi Wang, Xikun Lu, Guanghao Zhou, Sihang Jiang, Jiaqing Liang, Weikang Zhou, Zeye Sun, Fei Yu, Yanghua Xiao  
2026, ICLR.

#### [3] [Selective Expert Guidance for Effective and Diverse Exploration in Reinforcement Learning of LLMs](#)

Zishang Jiang, Jinyi Han, tingyun li, Xinyi Wang, Sihang Jiang, Zhaoqian Dai, Ma Shuguang, Fei Yu, Jiaqing Liang, Yanghua Xiao  
2026, ICLR.

#### [4] [Difficulty Is Not Enough: Curriculum Learning for LLMs Fine-tuning Must Consider Utility](#)

Zishang Jiang, Jinyi Han, tingyun li, Xinyi Wang, Sihang Jiang, Jiaqing Liang, Xiaojun Meng, Jiansheng Wei, Yanghua Xiao  
2025, AAAI.

#### [5] [Structured Reasoning for Large Language Models](#)

Jinyi Han, Zixiang Di, Zishang Jiang, Ying Liao, Jiaqing Liang, Jie Wang, Zheming Yang, Zhi Li, Yongqi Wang, Xiaofeng Ji, Yanghua Xiao, Miao Liu  
2026, ACL. (Under Review, Work Done in ByteDance Internship)

#### [6] [Knowing How Certain It Is: Confidence Estimation Throughout LLM Generation](#)

Jinyi Han, Tingyun Li, Shisong Chen, Xinyi Wang, Jiaqing Liang, Yanghua Xiao  
2026, Artificial Intelligence. (Under Review)

#### [7] [Small Language Model Can Self-correct](#)

Haixia Han, Jiaqing Liang, Jie Shi, Qianyu He, Yanghua Xiao  
2024, AAAI.

#### [8] [Think Thrice Before You Act: Progressive Thought Refinement in Large Language Models](#)

Chengyu Du, Jinyi Han, Yizhou Ying, Aili Chen, Qianyu He, Haokun Zhao, Haoran Guo, Sirui Xia, Jiaqing Liang, Zulong Chen, Liangyue Li, Yanghua Xiao  
2025, ICLR.

#### [9] [CDS: Data Synthesis Method Guided by Cognitive Diagnosis Theory](#)

Haokun Zhao, Jinyi Han, Jiaqing Liang, Xiaojun Meng, Jiansheng Wei, Yanghua Xiao  
2025, ACL.

- [10] [Not All Negative Samples Are Equal: LLMs Learn Better from Plausible Reasoning](#)  
Zixiang Di, Jinyi Han, Shuo Zhang, Ying Liao, Zhi Li, Miao Liu, Xiaofeng Ji, Yongqi Wang, Zheming Yang, Ming Gao, Bingdong Li, Jie Wang  
2026, ICML. (Under review, Equal Contribution, Work Done in ByteDance Internship)
- [11] [From Outcomes to Actions: Leveraging Hindsight for Long-Horizon Language Agent Training](#)  
Zishang Jiang, Tingyun Li, Jinyi Han, Xinyi Wang, Sihang Jiang, Yizhou Ying, Xiaojun Meng, Jiansheng Wei, Jiaqing Liang, Yanghua Xiao  
2026, ICML. (Under Review)
- [12] [CEM: A Data-Efficient Method for Large Language Models to Continue Evolving From Mistakes](#)  
Haokun Zhao, Jinyi Han, Jie Shi, Chengyu Du, Jiaqing Liang and Yanghua Xiao  
2025, CIKM.
- [13] [Don't Tell the Answer, Truly Guide the Reasoning During RL Rollouts](#)  
Xinyi Wang, Jinyi Han, Zishang Jiang, tingyun li, Jiaqing Liang, Sihang Jiang, Zhaoqian Dai, Ma Shuguang, Fei Yu, Yanghua Xiao  
2026, ACL. (Under Review)
- [14] [ADaPT: Token-Level Decoupling for Efficient Large Reasoning Models](#)  
Tingyun Li, Zishang Jiang, Jinyi Han, Xinyi Wang, Sihang Jiang, Han Xia, Zhaoqian Dai, Ma Shuguang, Fei Yu, Jiaqing Liang, Yanghua Xiao  
2026, ACL. (Under Review)
- [15] [SED: Self-evaluation decoding enhances large language models for better generation](#)  
Ziqin Luo, Jinyi Han, Haokun Zhao, Guochao Jiang, Chengyu Du, Tingyun Li, Jiaqing Liang, Yanghua Xiao  
2024, Preprint.
- [16] [Dynamic Clustering Based Contextual Combinatorial Multi-armed Bandit for Online Recommendation](#)  
Cairong Yan, Haixia Han\*, Yanting Zhang, Dandan Zhu, Yongquan Wan  
2022, Knowledge-Based Systems.
- [17] [CoCoB: Adaptive Collaborative Combinatorial Bandits for Online Recommendation](#)  
Cairong Yan, Jinyi Han†, Jin Ju, Yanting Zhang, Zijian Wang, Xuan Shao  
2025, Dasfaa.
- [18] [Thompson Sampling with Time-Varying Reward for Contextual Bandits](#)  
Cairong Yan, Hualu Xu, Haixia Han, Yanting Zhang, Zijian Wang  
2023, Dasfaa.

## 🏆 竞赛经历

- 
- 01 - 快慢思考混合模型训练方法 华为火花奖 技术负责人 2025.09 – 2026.01
- 背景: 针对 Test-time Scaling 慢思考范式中的“过度思考 (Overthinking)”问题, 提出快慢思考混合训练方法, 使模型能够根据问题难度自适应控制推理长度, 在保持推理能力的同时减少冗余生成。
  - 内容: 设计基于 GRPO 的两阶段 Rollout 训练策略, 构建包含长/短推理路径以及正误反馈的对比学习信号, 引导 Policy 模型学习生成正确且更简洁的推理过程, 从而提升推理效率。
  - 效果: 在 AIME24、MATH500、GPQA-D 等推理基准上, 在保持官方准确率不下降的前提下, 实现输出 Token 数量减少 40%+, 显著提升模型推理效率与实际部署可行性。
- 02 - 基于认知诊断的大模型持续进化与知识更新研究 华为火花奖 技术负责人 2024.09 – 2025.02
- 背景: 现有大模型持续训练通常采用“评估—合成—训练”的迭代范式, 但评估多停留在答案级别, 难以细粒度刻画模型能力, 导致模型真实能力与错误根源难以被有效诊断。
  - 内容: 引入认知诊断 (Cognitive Diagnosis) 思路, 以知识点组件的形式细粒度刻画模型能力; 构建任务知识点库并建立问题—知识点映射关系, 在评估阶段定位模型错误的知识点来源, 从而指导针对性数据合成与持续训练。
  - 效果: 在数学推理、代码补全及综合学科测试等多个基准任务上, 相比现有合成数据训练方法取得更优性能表现。

## 基金情况

---

面向大语言模型的持续学习方法研究 项目负责人	优秀博士生学术创新能力提升计划	2024.09 – 至今
基于大模型的智能教材问答关键技术研究 项目负责人 已结项	人工智能赋能心理/教育学科交叉研究项目	2025.05 – 2025.11
面向大语言模型的持续学习方法研究 项目负责人 已结项	博士研究生科研创新基金	2023.12 – 2024.12

## 获奖情况

---

上海市优秀毕业生	2023.03
东华大学优秀硕士毕业论文	2023.03
学业奖学金一等奖学金	2020.10-2022.10
河南省优秀学位论文	2021.07