张皓宇

hzhang2838-c@my.cityu.edu.hk
howeyz.github.io | GitHub / howeyz | Google Scholar

教育背景

香港城市大学 计算机科学系 · 计算机科学（直博） 2025.09 – 至今
中南财经政法大学 统计与数学学院 · 金融数学（本科） 2021.09 – 2025.06

实习经历

Tencent* 腾讯科技（深圳） | *CDG · AMS 2026.07 – 至今

基于 LLM 与在线强化学习的查询改写检索

问题背景：面向微信朋友圈、视频号等场景的纯查询（QO）与看后搜（SAF）需求，针对改写与下游检索目标脱节、长尾泛化不足的问题，构建**查询 → 改写 → 文档（q2q2d）**链路，以实际检索反馈优化改写列表。

方法：❶ SFT 冷启动：依据用户消费行为筛选真实改写意图，构建 QO / SAF 多指令数据集，在 **Qwen2.5-0.5B / 3B** 上完成 SFT。**❷ Listwise 在线强化学习：**一次生成完整改写列表，按各改写的 span 将元素级奖励分配至对应 token；以**边际检索收益**抑制重复改写导致的 **reward hacking**，结合格式、去重、命中与文档覆盖奖励优化列表质量，以真值锚定缓解熵坍缩。**❸ 近线部署：**离线预生成高频查询的改写并存入 KV 缓存，**10% 查询覆盖约 75% 检索流量**；采用 **SAF → QO** 级联查询，并通过流式任务补充缓存未命中项，降低在线推理开销。

结果：仅用不足 **10k** 条 RL 训练样本，Qwen2.5-0.5B 在 QO 离线测试集上相较 SFT，**GTHR@30 从 34.04% 提升至 39.75%，独占命中率从 5.12% 提升至 6.18%**。

科研经历

大语言模型推理与强化学习 2025.06 – 至今

Fishing Out Free Riders: Shapley-Based Reward Attribution for Parallel Reasoning [共一 · 预印本] 论文↗

问题背景：并行推理中各路径共享结果奖励，冗余与误导路径难以区分，易产生 **reward hacking**。

方法：提出 **Parallel Shapley**，以生成式奖励模型评估路径组合，采用 **Monte Carlo Shapley** 估计边际贡献；联合路径级与结果级奖励进行 **GRPO** 训练，引导模型生成有效、互补的推理路径。

结果：Qwen3-4B-Base 的 AMC23 Mean@16 达 **72.8%**，较 Parallel-R1-Unseen 提升 **3.1%**。

Medical Reasoning Enhancement via Self-Corrected Fine-grained Reflection [共一 · 预印本] 论文↗

问题背景：医学推理过程标注昂贵、错误步骤难以定位，模型缺乏学习有效反思所需的细粒度反馈。

方法：提出 **Med-REFL**，利用 **ToT** 叶节点反馈估计节点价值并辅助定位错误，保留共同前缀探索替代分支，按步骤与轨迹的价值差筛选反思，改写为第一人称的中途与事后纠错轨迹；将有效 / 无效反思及正确 / 错误推理轨迹构成偏好对，采用 **LoRA-DPO** 训练内化自纠错能力，无需额外训练 PRM。

结果：Huatuo-o1-8B 在 MedQA（5 选项）上由 **69.59% 提升至 73.74%**，7 项医疗基准平均提升 **3.59%**。

TC-RAG:Turing-Complete RAG's Case study on Medical LLM Systems [二作 · ACL 2025 Oral · CCF-A] 论文↗

问题背景：医学多步问答缺乏明确的检索停止条件，错误证据与冗余信息持续累积，影响推理可靠性。

方法：提出 **TC-RAG**，以显式状态监控和栈式记忆统一检索、推理与规划；通过 **Push / Pop** 管理记忆，结合 **Backtrack / Summary** 清除错误、压缩冗余，并以不确定性阈值控制检索终止。

结果：医学预训练 Qwen-32B 在 MMCU-Medical 上 EM 达 **93.15%**，较同底座 ReAct 提升 **4.96%**。

时空图建模、图检索增强学习 2023.03 – 至今

围绕动态图与时空图，研究检索增强、分布外泛化、图域适应、图对比学习等。

一作 | ***IEEE TPAMI*** [*CCF-A · 中科院一区 Top*] Pattern in Motion: Retrieval-Augmented Learning for Dynamic Spatio-Temporal Graphs.

一作 | ***NeurIPS 2025*** [*CCF-A*] STRAP: Spatio-Temporal Pattern Retrieval for OOD Generalization.

一作 | ***ECML-PKDD 2025 Oral*** [*CCF-B*] Rethinking Graph Domain Adaptation: A Spectral Contrastive Perspective.

共一 | ***NeurIPS 2026*** [*CCF-A*] Decoupled Spatio-Temporal Gaussian Scene Graph for 4D Cooperative Driving Reconstruction.

一作 | ***IEEE TITS 2024*** [*CCF-B · 中科院一区 Top*] Contrastive Learning-Based Deep Embedded Clustering.

一作 | ***Knowledge-Based Systems 2024*** [*中科院一区 Top*] ConUMIP: Continuous-time Dynamic Graph Learning.

一作 | ***Pattern Recognition 2024*** [*CCF-B · 中科院一区 Top*] IMCSN: An Improved Neighborhood Aggregation Interaction Strategy.