

鲁泽含

学校：香港科技大学（广州） 专业：数据科学与分析（DSA） 状态：本科四年级在读
电话：19252073106 | 邮箱：zlu561@connect.hkust-gz.edu.cn | GitHub：github.com/Kumoris

求职意向

人工智能（MLLM、NLP） / Agent 应用 / 隐私安全方向

教育背景

数据科学与分析（本科） 香港科技大学（广州）
预计 2027 年 6 月毕业 本科四年级在读

实习经历

华为技术有限公司实习生 2026.07.08-2026.08.31
具体岗位：AI 算法工程师

科研与项目经历

Code Review New：面向 CodeHub 的多 Agent 协作代码检视系统 2026.07-2026.08

华为实习期间项目成果 / Agentic Workflow / Harness Engineering

- 面向华为 CodeHub 代码检视场景，设计并实现 Reviewer-Adversary-Synthesizer 多 Agent 协作系统，完成大规模部署验证与实际交付。
- 按 8 类检视维度动态路由最多 5 个 Reviewer 并行执行，由 Adversary 完成四级可信度裁定、Synthesizer 完成事实验证、去重与严重度校准。
- 本地在 19 个真实 MR 上开展评测，输出 8 条候选 finding（含 meta-rule），其中 5 条通过缺陷证据验证；Reviewer 层有效缺陷精度为 25%，通过门控过滤 43.8% 无效输出。
- 本地评测端到端耗时：小型 MR 为 6-12 分钟、中型 MR 为 19-31 分钟、大型 MR 多 Agent 为 70-90 分钟；单条 finding 平均耗时为 1-4 分钟。

链路推荐与社会影响最大化研究项目 2025-2026

图算法研究 / ACM WSDM 主会待评审

- 面向大规模社交网络，研究在独立级联模型下通过概率链路插入提升距离衰减社会影响的 DIMA 优化问题。
- 参与 ADIL 核心算法设计，基于距离感知反向可达集构建无偏影响估计器，并引入边际增益维护（MG）与低概率边剪枝（LPE）；算法以 $1 - \delta$ 概率获得 $(1 - 1/e - \varepsilon)$ 近似保证。
- 参与算法剪枝策略的消融分析：在 GRQC 数据集上，MG 与 LPE 将贪心选择耗时由 4,198.433 秒降至 0.266 秒，实现超过 10,000 倍加速。

端侧大语言模型 KV Cache 语义压缩与动态调度系统 2025-2026

推理优化 / 算法设计

- 相关发明专利申请：CN202610081890.6。
- 面向端侧长上下文推理中的显存占用与生成延迟问题，参与设计结合 RAG、语义重要性评估与 KV Cache 动态调度的推理优化方案。
- 参与 Chunk KV 算法设计，将连续 Token 按 10 / 16 Token 划分为语义 Chunk，通过 Attention 重要性评分、Top-K 片段选择与跨层索引复用，动态保留高价值上下文并淘汰冗余缓存。
- 在 10%-30% KV Cache 保留率下，方案指标显示可实现显存占用最高降低 70%、推理吞吐量提升 26.5%、推理延迟降低 20.7%，同时保持长文本语义完整性与生成稳定性。

其他科研与项目经历

AscendFlow Law 全模态法律智能系统	2024-2025
RAG / Multi-Agent	
- 面向律师及企业法务的合同处理场景，构建融合视觉文档解析、法律知识检索与多智能体协作的法律 AI 系统，支持复杂合同解析、法律依据召回及任务流程编排。 - 负责视觉文档 RAG 技术方案设计，集成 PDF-Extract-Kit 的版面检测、公式识别与中文 OCR 能力，将复杂 PDF 转换为结构化 Markdown，并完成文档上传、解析、预览与检索链路。	

多语种多语音跨模态大模型训练部署系统	2024-2025
ASR / TTS / LID	
- 相关发明专利申请：CN202511072126.4。 - 参与基于 wav2vec 2.0 与 1B 参数 Transformer 的统一语音大模型训练方案设计，采用 ASR / TTS / LID 多任务微调，覆盖 1,107 种语言的 ASR / TTS 与 4,017 种语言的 LID。 - 参与 GPU-CTC 对齐与端侧部署优化，采用 16 kHz 标准化、15 秒切分及 CER 过滤，使错误率降低 32%、速度较 CPU 提升约 8 倍；结合 2M 参数 Adapter、FP16→INT8 量化、分区加载与流式解码降低部署成本。	

待评审论文

Maximizing Distance-decaying Social Influence via Link Recommendation	
ACM WSDM / Full Paper (Main Track)	待评审
X. Chen, Z. Lu, W. Yu, and J. Tang.	

竞赛获奖

- ASC 世界大学生超级计算机竞赛（ASC24）：二等奖
- “挑战杯”全国大学生课外学术科技作品竞赛：省级银奖

知识产权

- 外观设计专利 4 项（已授权）
- 发明专利 2 项（其中 1 项已授权）

专业技能

Agent 工程	具备 Multi-Agent 角色编排、动态路由、Tool / Skill 调用、上下文工程、评测集构建、Fail-Closed 门控及 Harness Engineering 实践；完成 RAG、向量检索与复杂文档 Agent 工作流搭建。
AI / ML 与多模态	熟悉 PyTorch、CNN、Transformer，具备跨模态对比学习与表征对齐、Hungarian 匹配、跨模态检索评测、多模态视频理解及 KV Cache 优化实践；使用 Hugging Face / Kaggle 完成模型训练、微调与评测。
编程、工程与科研	熟练使用 Python、NumPy、Pandas、MySQL / SQL 开展数据清洗、特征统计、算法原型及训练 / 评测，具备 C++ 实现与性能测试经验；熟悉 Linux、Git / GitHub、Docker、Ollama 及 AutoDL / 阿里云 GPU 部署，具备论文复现与消融分析能力，IELTS 6.5。