

毕启航

15972505305 | 1430097658@qq.com



教育经历

武汉大学（985） 电子信息（专硕） 免试研究生	2026.9 — 2028.6
武汉大学（985） 计算机类 网络空间安全	2022.9 — 2026.6
GPA: 3.85 / 4.0 (18/143, Top 12%), 校级奖学金 * 2	

竞赛获奖

- 花旗杯金融创新大赛（全国二等奖，第七名）
- 全国大学生数学竞赛（省一等奖）

实习经历

上海蔚来汽车有限公司 | 爬虫开发实习生

项目背景：蔚来竞品分析团队此前依赖人工 Excel 核对外部车型与内部主数据，数据更新周期长达数周。独立开发了一套自循环竞品情报采集与车型对齐系统，为内部竞品看板提供准实时数据支撑。

- 反爬与数据采集**：针对目标站返回 HTTP 200 但实际被拦截的“假 200”问题，逆向分析 SSR 页面结构，通过检查 `__NEXT_DATA__` 标签是否有效来判断是否真正拿到数据；配合 5 个代理 IP 轮换与重试，20 线程并发稳定抓取 4,700+ 款车型的配置与销量
- 车型模糊对齐**：先通过品牌、年款、指导价等硬指标收敛候选集，再用 Token-Dice 与 SequenceMatcher 加权计算车型名相似度；每次对齐后重点抽查相似度最低的 100 条数据人工复核，根据 Bad Case 持续迭代匹配规则，实现外部车型与内部主数据的 1:1 自动对齐
- 定时调度与容错**：基于 Docker 编写定时脚本，设定每周一凌晨自动运行；加入步骤失败自动中止与核心表 7 日备份轮转，避免脏数据覆盖线上看板

项目经历

me-Agent 一个人职业 AI 数字分身

在线体验: hireme.ccwu.cc

- Agent 问答状态机**：基于 FastAPI + LangGraph 实现多轮问答状态图，支持简历、项目卡片与历史对话的多工具并行检索；设置工具参数校验、重复调用拦截与最大 4 轮调用限制，防止 Agent 陷入死循环
- 证据约束与溯源展示**：实现 EvidenceRegistry 证据注册模块，每轮问答按 ID 收集并去重检索结果，格式化后作为白名单传给模型；前端生成对应的证据引用卡片，支持答案溯源与无依据问题友好拒答
- 混合检索与代码定位**：采用 ChromaDB 向量粗召回 + 本地 Reranker 精排的两阶段检索，针对项目代码符号和专有名词增加正则精确匹配兜底，兼顾自然语言语义与关键词准确性
- LLM 路由与稳定性**：封装统一的 LLM 客户端，支持多 Endpoint 故障切换与熔断器（连续 3 次失败冷却 60 秒）；通过 SSE 实现 Token 流式输出与 Agent 思考步骤的可视化展示

反电信诈骗大数据平台 | 打击新型网络犯罪实验室

项目背景：为公安反诈中心开发的预警分析平台，对接真实业务的通信行为数据，用于实时提取高危线索。

- 特征筛选与轻量模型**：从千万级原始通信流水（约 2GB）中，按地域关键词与交互规则过滤出 8 万条高价值数据，人工标注 2.2 万条核心样本；训练 54MB 的轻量 SVM 分类器部署在 CPU 上做实时推理，在极低算力下实现高危线索的高召回
- 慢 SQL 优化**：排查 13 秒慢查询发现原有的（省，市，区）联合索引在按区县单独查询时脱离了最左前缀原则导致全表扫描；拆分联合索引并按高频查询字段重构索引结构，使慢查询平均时延降低 80%

其他

- 开源贡献**：为 1.5k Star 开源项目 `blockads-android` 贡献代码。排查并修复 Root 模式下虚拟网卡与物理网卡的路由冲突；通过 Linux 内核将 53 端口的 DNS 流量重定向至代理服务，重构服务分发器实现广告拦截与第三方 VPN 共存，PR 已合入主分支
- 英语六级 566 分，能流畅阅读英文技术文档