

蔡梓钊



求职方向：AI 应用开发实习 / 大模型应用开发实习

教育背景：广州应用科技学院 · 人工智能本科 · 2027 届

到岗时间：可立即到岗，每周 5 天，实习 3 个月以上

意向地区：上海 / 深圳 / 广州

电话：13715861123 邮箱：2752768312@qq.com

GitHub: github.com/Yaoniguan-Money 个人主页: field-signals.pages.dev

核心技能

LLM / Agent : LangGraph、Tool Calling、Structured Output、Prompt、Guardrail
RAG / 检索 : Embedding、pgvector、BM25、RRF、Reranker、Evidence Trace
后端 / 工程化 : Python、FastAPI、PostgreSQL、WebSocket/SSE、Docker、pytest/CI

项目经历

顺德 12345 多频工单智能研判平台 | 顺德黑客松 2026.08

企业解决方案赛道 | 真实政务场景 | 已获政府侧认可，正推进业务对接

- 基于 128,278 条真实 12345 工单，针对同一现实问题跨表述、主体/地点别名及跨时间重复投诉导致文本相似度误聚类的问题，构建事件级研判系统；支持 Excel/CSV 自动映射、分块断点续传与幂等导入，全量入库 128,278/128,278，0 失败。
- 设计规则分段 + LLM Structured Output + Evidence Trace + Gazetteer 实体归一，将当前投诉、历史背景、部门回复和当前诉求拆解为独立现实事件，并对模型证据回查原文，降低处置/回复文本被误建为新事件。
- 构建 pgvector 候选召回 -> LLM SameEvent -> consistency-aware Event Graph；排除同工单自比较，以主体、地点、问题、时间与硬冲突判定事件同一性，并用 complete-link guard 阻止传递式误聚类；完成 React 研判工作台及人工纠错、AuditLog、CSV 导出闭环。

Verifiable Company Research Agent | LangGraph + RAG 2026.03 - 04

证据型企业研究 Agent

GitHub: github.com/Yaoniguan-Money/Verifiable-Company-Research-Agent

- 针对企业投研资料分散、结论难核验及来源不足时模型易幻觉的问题，构建覆盖检索、事实抽取、交叉核验与带引用报告生成的研究 Agent；基于 LangGraph 设计 17 节点工作流与 4 类条件路由，证据不足时执行补充检索、重写或拦截。
- 搭建 Dense Embedding + BM25 双路召回、RRF 融合与 Reranker 精排的 Hybrid RAG；57 条冻结检索集 Recall@5 从 90.4% 提升至 95.2%，表格事实抽取 micro F1 达 96.9%，建立 450 项后端测试。

AI 同声传译助手 | Electron + Streaming ASR / LLM 2026.06

实时 ASR / LLM 桌面字幕应用

GitHub: github.com/Yaoniguan-Money/AI-Simultaneous-Interpretation-Assistant

- 针对实时跨语言沟通中的字幕等待、工具切换与上下文丢失，开发 Electron 透明字幕桌面端，打通音频采集 -> RingBuffer -> WebSocket 流式 ASR -> 语义分句 -> LLM SSE -> IPC，并设计快/慢双通道兼顾低延迟翻译与术语/摘要上下文回注。
- 以 8 槽 RingBuffer 替代忙时直接丢帧策略，在 30 组固定负载离线对照中将音频块平均交付率由 59.25% 提升至 81.53%；基于 10 条固定英文输入 × 3 轮真实 DeepSeek 调用，流式首个可用中文 P50 为 604ms，较同响应缓冲至完成的配对中位等待缩短 203ms (30/30 成功)。

竞赛与荣誉

2026中国国际大数据产业博览会官方配套赛事「贵客松」AI黑客松 三等奖

2026.08

中国计算机设计大赛 AI 智能体赛道省三等奖

2026.03 - 07