

# 蔡杰超

计算机科学与技术本科 | 香港大学计算机硕士 | Agent/大模型应用开发

✉ 2651159710@qq.com ☎ (+86) 136-0270-5148

🌐 www.caijiechao.com 📄 github.com/computersniper in jiechao-cai

求职意向: AI Agent 研发、大模型应用工程化、LLM 平台开发



## 🎓 教育背景

北师香港浸会大学 (BNBU) & 马来亚大学 (交换) 2022.9–2026.6

计算机科学与技术本科 (全英教学, 专业前 15%, 荣获二等奖学金)

- 核心课程: 数据结构与算法、操作系统、数据通信与网络、机器学习、自然语言处理、深度学习与神经网络等

香港大学 2026.9 – 2027.8 (预计)

硕士 计算机科学 (MSc in Computer Science)

- 计划专注于大模型与自然语言处理、AI/Agent 应用、强化学习与对齐、模型微调方向

## 🔧 专业技能与 AI 特质

- AI 编程顶级玩家:** Cursor、Claude Code、Trae 重度资深用户 (日耗 Token 数百万, 多平台付费会员); 极强的 Prompt 编写与调优能力, 熟练掌握 LLM 生成生产级代码的工程范式。
- Agent 与大模型工程:** 深入理解 LLM 能力边界; 熟练掌握 Context Engineering、RAG 知识库构建; 拥有结合 MCP (Model Context Protocol) 与 Skill 开发的 Agent 落地经验。
- 核心技术栈:** Python(精通)、Java、TypeScript、Kotlin; 熟悉 FastAPI、SQLAlchemy、Docker、asyncio 异步编程; 具备全栈开发能力及扎实的数据结构与网络基础。

## 📁 实习 & 项目经历

犀牛国际教育 | AI 平台研发实习生 2026.01–2026.03

- 公司为教育内容服务商, 长期为腾讯、阿里、小红书等 AI 厂商提供奥赛级高质量数学题数据。
- 全栈开发, 负责前后端架构与核心功能开发、测试。构建数据清洗与智能标注管道, 集成大模型实现题目质量检测 (原创性/难度/严谨性等); 搭建 LiteLLM 网关, 优化调用性能并实现用量统计与自动化日报。

**MathTasks 数学题库智能生产平台** 2026.01–2026.02

Python, FastAPI, Next.js, PostgreSQL, Vector Search, Playwright, Agent 开发, 大语言 & 多模态大模型

链接: stem-align.com

- 核心架构与工作流:** 基于有限状态机 (FSM) 与 RBAC 实现 5 角色全链路协作, 解决并发流转冲突; 构建含 30+ 字段的题库模型及 5 维自动评分体系。
- 异步队列与实时推送:** 搭建难度评估、向量查重等异步任务队列, 支持万级题库高效去重; 结合 SSE 流与 Redis 状态缓存, 实现前端进度实时推送与断线恢复。
- AI 质量评估:** 设计基于大模型的“对抗验证”机制 (多次采样校验阈值), 将 LLM 的概率性输出转化为确定性的难度评估指标。
- 自动化测试:** 基于 Playwright 实现多角色端到端测试; 集成多模态大模型解决验证码识别, 并设计 Tesseract/手动干预的降级策略 (Fallback), 保障主流程鲁棒性。

**企业级 LLM 统一网关与 API 管理平台** 2026.02–2026.04

Python, LiteLLM, Docker, PostgreSQL, Redis, Asyncio, Langfuse

- 网关路由与高可用:** 统一接入 OpenAI/Anthropic/Doubao 等 8+ 厂商 20+ 模型; 配置多 Key 负载均衡与自动故障转移 (Failover), 平滑应对上游限流与网络异常。
- 性能压测优化:** 编写并发压测脚本评估并调优网关在 50+ 并发下的首字延迟与吞吐量; 优化长时推理连接配置, 稳定支撑每日数千次复杂 Agent 任务。
- 成本治理与可观测:** 集成 Langfuse 实现 Token 级追踪; 开发自动化定时脚本, 深度解析聚合网关日志, 通过飞书 Webhook 自动推送多团队/多项目精细化成本日报。
- 规模与流量:** 累计处理 1 亿 + Tokens 请求, 稳定支撑每日数千次复杂推理调用。

北师香港浸会大学 | 教学助理 2025.09–2025.12

- 担任《数据结构与算法》课程助教, 负责辅导课讲解、作业批改与实验课指导, 帮助学生掌握核心算法。
- 主导开发 Multi-Agent AI 助教系统, 从课程资料中提取结构化知识构建 RAG 知识库, 实现 50%+ 常见问题自动化解答。