

孔维轩 (Weixuan Kong)

✉ Jackkong29@Gmail.com ☎ 18796611232 🌐 github.com/Jackela

教育经历 (Education)

悉尼大学 (The University of Sydney)

计算机科学硕士 (Master of Computer Science)

2024 年 2 月 - 2025 年 12 月 (预计)

WAM: 77.4 / 100

阿尔戈马大学 (Algoma University)

计算机科学学士 (Bachelor of Computer Science)

2020 年 9 月 - 2023 年 6 月

WAM: 88.8 / 100, 荣获荣誉学位 (Cum Laude)

核心技能 (Skills)

编程语言

Java, Python, JavaScript/TypeScript, C#, SQL

框架与库

FastAPI, Spring Boot, React, Node.js, Next.js, PyTorch, Langchain4j, LlamaIndex, Haystack

数据库

MySQL, PostgreSQL, MongoDB, SQLite, DynamoDB, Milvus, ChromaDB

云原生与 DevOps

AWS (Lambda, S3, API Gateway), Docker, Kubernetes, Terraform, Git, CI/CD (GitHub Actions)

架构与方法论

工程实践: 领域驱动设计 (DDD), 微服务架构, SOLID, 分布式系统, RAG

项目管理: 敏捷开发 (Agile/Jira), MoSCoW

产品战略与创新理论: 第一性原理思维 (First-Principles), 创造性破坏 (Schumpeter), 破坏性创新 (Christensen), 跨越鸿沟 (Moore)

项目经历 (Projects)

CAPSTONE 毕业设计: 临时学术人员工时管理系统

2025 年 8 月 - 至今

产品负责人与架构师

- 战略思考:** 正独立承担产品负责人与架构师双重角色, 旨在通过数字化解决企业内部工时管理及财务审计痛点。前期规划已完成核心方案设计, 采用 MoSCoW 定义和管理 MVP 功能优先级。架构上, 规划兼具扩展性与稳定性的“模块化单体”架构, 利用 DDD 和清晰的限界上下文, 并借鉴“绞杀榕模式”为未来向微服务平滑演进奠定坚实基础, 旨在平衡短期交付与长期战略扩展性。通过撰写超过 20 页的技术与战略备忘录, 系统化论证了技术选型, 并制定了分阶段交付与风险管理策略。

个人独立项目: AI 招聘文员 (AI Recruitment Clerk)

2025 年 6 月 - 至今

产品设计者 & 全栈开发者

- 产品哲学:** 这是我对“低端市场破坏性创新”的刻意实践。面对功能臃肿、价格昂贵的传统 ATS 系统, 我反其道而行之, 刻意只聚焦于“简历-JD 匹配度”这一个核心性能维度, 为那些被大型系统忽略的小微企业和独立猎头, 提供了一个轻量、高效、‘刚刚好’的解决方案。
- 产品管理:** 独立撰写了包括商业需求文档 (BRD)、产品需求文档 (PRD) 和软件需求规格说明书 (SRS) 在内

的全套产品文档，完整实践了从商业目标到产品功能的转化过程。

- **核心 AI 功能:** 实现了从简历中智能提取关键信息，并与岗位描述 (JD) 进行深度语义匹配，最终生成包含匹配度评分和优劣势分析的综合报告。
- **技术架构:** 采用微服务架构，前端使用 Angular，后端通过 NestJS 网关连接多个独立的 AI 服务 (简历解析、JD 提取、报告生成)，并通过 Docker 进行容器化部署。([GitHub Link](#))

个人独立项目: AI 增强 PDF 学术工具 (AI Enhanced PDF Scholar)

独立开发者与产品设计师

- **战略定位:** 本项目的核心是对“新市场破坏性创新”的一次探索。我观察到传统 PDF 工具的核心价值是‘阅读’，因此我通过 RAG 技术，开创了一个全新的性能维度——‘对话式分析’，从而为学术研究创造了一个此前不存在的、非消费性的新市场。
- **核心功能:** 集成 LlamaIndex/Haystack 和 Google Gemini API，实现对 PDF (包括扫描版) 内容的智能问答；并设计了独特的智能引用系统，自动解析 APA, MLA 等格式并构建文档间的引用网络。
- **技术架构:** 采用前后端分离模式 (React/TypeScript + FastAPI/Python)，通过 WebSocket 进行实时通信。后端采用高度模块化的异步处理流水线，依次调用 OCR、文档结构化、向量化等可插拔处理器。
- **工程实践:** 项目完全容器化 (Docker)，支持高性能生产环境部署 (Gunicorn + Uvicorn)，拥有全面的测试体系 (单元、集成、E2E)，并通过 GitHub Actions 实现 CI/CD。([GitHub Link](#))

个人独立项目: MCP 学术 RAG 服务器 (MCP Academic RAG Server)

核心后端开发者

- **项目介绍:** 作为“AI 增强 PDF 学术工具”的后端核心，此项目专注于构建一个可被其他 AI 助手调用的标准化 RAG 服务。它通过统一的 MCP 协议，提供高效、准确的学术文献检索与内容生成能力，是实现多 AI 助手协作的关键组件。
- **核心功能:** 实现了基于 Haystack 的检索增强生成 (RAG) 管道，能够从大量学术文献中检索相关信息并生成高质量的回答。提供了标准化的 API 接口，支持外部 AI 助手通过 MCP 协议进行调用，实现知识共享与复用。
- **技术架构:** 后端基于 FastAPI/Python 构建，采用模块化设计，集成了文档解析、向量化存储和检索等核心组件。利用异步处理机制提升并发性能，确保在高负载下的响应速度。
- **工程实践:** 项目完全容器化 (Docker)，便于部署和扩展。具备完善的测试覆盖 (单元测试、集成测试)，并通过 CI/CD 流程确保代码质量和版本迭代的稳定性。([GitHub Link](#))

个人独立项目: Novel-Engine (多智能体互动小说生成器)

2025 年 7 月 - 至今 (进行中)

独立开发者与系统架构师

- 一个基于 Context engineering 的多智能体 AI 叙事模拟系统，旨在生成动态的战役叙事和故事内容。
- **系统设计:** 构建了包含 DirectorAgent (游戏主控)、PersonaAgent (角色 AI) 和 ChroniclerAgent (叙事 AI) 的多智能体协作系统，通过 Google Gemini API 驱动角色基于其性格特征进行智能决策。
- **性能优化:** 设计并实现了“缓存协议”，通过 LRU 文件缓存和基于 SHA256 哈希的 LLM 响应缓存，避免了超过 85% 的重复计算和 API 调用；并采用 GZip 压缩优化 API 带宽。
- **生产级特性:** 实现了线程安全的并发处理 (ConfigLoader 采用单例模式)、优雅降级 (AI 失效时自动切换至备用算法) 和全面的分级日志记录系统。([GitHub Link](#))

LLM A/B Testing Platform - 大语言模型对比评估平台

2024 年 8 月 - 至今 (进行中)

全栈开发工程师 / AI 产品技术负责人

- 设计并开发了标准化的大语言模型对比评估平台，实现了多模型、多维度的自动化 A/B 测试。该平台采用微服务架构，集成了 OpenAI、Anthropic、百度文心等主流 LLM API。
- 核心功能:** 异步任务处理: 基于 FastAPI 和 Celery 实现高性能异步处理，支持并发模型调用，提高测试效率。多维度评估框架: 构建了基于多 Judge 交叉验证的评估体系，涵盖准确性、相关性、创造性等指标。可视化 Dashboard: 使用 Streamlit 构建交互式 Dashboard，提供直观的可视化报告和统计分析。
- 技术架构:** 后端基于 FastAPI 和 Celery 实现高性能异步处理。通过 LLM-as-Judge 机制替代传统人工评估，有效解决了 LLM 模型选择中的客观性评估难题。完整的实验设计包含了标准数据集处理、成本效益分析和可视化报告生成。(GitHub Link)

学术复现项目: 深度学习文本分类框架

2025 年 7 月 - 2025 年 8 月

独立研究员与算法工程师

- 精确复现 VDCNN 和 DenseNet 两篇顶级深度学习论文的文本分类成果，并将其重构为一个生产级的实验框架。
- 工程化设计:** 采用“配置即代码”的核心思想，通过 YAML 文件驱动整个实验流程（模型选择、数据集、超参数等），实现了代码与配置的完全分离，具备极强的可复现性和扩展性。
- 架构模式:** 严格遵循 SOLID 设计原则和工厂模式 (Factory Pattern)，将数据处理、模型定义、训练评估等模块高度解耦，使得添加新模型或数据集无需修改核心代码。(GitHub Link)

学校课程项目: 云原生 AI 影像标注系统

2025 年 2 月 - 2025 年 6 月

独立开发者与架构师

- 为大规模 AI 影像标注任务，独立设计并部署了一个完整的事件驱动、无服务器 (Serverless) 系统。
- 性能优化:** 利用 AWS S3 Events, EventBridge 和并行 Lambda 函数，成功将单批次处理时延从 45 秒降至 8 秒 (-82%)。
- 部署自动化:** 通过 IaC (基础设施即代码)，使用模块化的 CloudFormation 模板编排了整个云基础设施，将复杂的云环境配置时间从数小时缩短至 15 分钟。(GitHub Link)

行业战略分析报告: 零售业中的数字人

2025 年 2 月 - 2025 年 6 月

核心分析师 & 报告整合负责人

- 作为核心成员及整合负责人，主导完成了一份关于“数字人”在零售行业应用的深度战略分析报告。
- 分析框架:** 运用“分布式创新”和“API 平台战略”等高阶理论框架，深度剖析了德国电信与欧莱雅等顶级企业的不同创新模式，对比了其在平台范围、API 策略、治理模型上的根本差异。
- 个人贡献:** 负责整体对比分析的结构搭建和主要撰写，并承担了整合不同理论部分、确保报告系统性与流畅性的最终职责，展现了优秀的战略分析与系统整合能力。

分布式医疗微服务系统

2024 年 8 月 - 2024 年 11 月

Agile 项目经理 & 核心后端开发者

- 在三人团队中，基于 AWS Lambda 和 Langchain4j，共同开发了一个无服务器化的 AI 医疗微服务系统。
- 项目管理:** 作为团队的 Agile 项目经理，使用 Jira 进行需求排期、任务分解和迭代跟踪，将团队协作效率提升了约 15%。
- AI 集成:** 系统集成了 Langchain4j 与大语言模型，为医生端提供智能健康咨询功能，为病人端提供智能注册

服务。

- 架构设计:** 系统被明确拆分为 Doctor-Service 和 Patient-Service 两个核心服务，遵循“单一职责”和“数据库-每-服务”的设计模式。(GitHub Link)

分布式商店平台 (Saga 模式实现)

2024 年 2 月 - 2024 年 6 月

架构师 & 核心后端开发者

- 架构权衡:** 为解决分布式系统中跨服务的数据一致性难题，我深入研究并手动设计、实现了 Saga 模式。这个决策深受 Geoffrey Moore《跨越鸿沟》理论的启发，坚信一个产品的技术架构在服务“早期采纳者”阶段，就需为未来进入“主流市场”的高并发挑战预作准备，这是一种对产品生命周期的长期负责，以确保数据最终一致性和长期稳定性。
- 技术栈:** 后端采用 Java Spring Boot 微服务集群，服务间通过 gRPC 进行高性能通信；前端采用 Next.js 和 TypeScript。(GitHub Link)

数学建模项目：自适应混合随机算法

2025 年 7 月 - 2025 年 8 月

独立研究员与算法工程师

- 针对在线优化中的两级滑雪租赁问题，设计并实现了一种创新的自适应混合随机算法。
- 核心创新:** 提出了基于成本结构动态调整的自适应参数，使得算法在多场景下均能收敛至近乎理论最优的竞争比率。
- 研究方法:** 结合了严格的理论推导（势函数分析）与大规模的实验验证（100,000 次/实验的蒙特卡洛模拟）。(GitHub Link)

在线购物平台 (MERN Stack)

2025 年 2 月 - 2025 年 6 月

核心开发者

- 运用 MERN 技术栈，构建并交付了一个功能完善、代码结构清晰的全栈电子商务平台原型。
- 核心功能:** 实现了用户认证、商品浏览、购物车、订单处理等完整流程，并为管理员提供了后台管理面板。
- 架构模式:** 在前端采用了 MVVM 设计模式，通过自定义 ViewModel Hooks 将 UI 逻辑与视图层有效分离，提升了代码的可维护性。(GitHub Link)

实习经历 (Experience)

北京愿景明控集团

2023 年 6 月 - 2023 年 8 月

人工智能团队实习生 (AI Team Intern)

- 跨语言服务集成:** 独立设计并工程化实现 Java 工具类，成功桥接 Java 主体业务与 Python AI 服务，将模块集成时间从数天缩短至数小时。
- 技术知识分享:** 主导关于大语言模型 (LLM) 的内部技术培训，有效提升团队整体技术认知。

罗姆尼光电系统技术 (广东) 有限公司

2022 年 5 月 - 2022 年 7 月

工程助理实习生 (Engineering Assistant Intern)

- 协助工程项目任务的执行与跟进，积累了团队协作与沟通经验。

个人简介 (Profile)

我是一名坚定的“第一性原理”思考者，拥有扎实的计算机科学背景与前瞻的产品架构思维。我的思维模型受益于长期跨学科的系统性阅读，自 2019 年以来，我投入超过 3300 小时，构建了以计算机科学为深度，人文社科为广度的“T 型知识结构”。

这种结构并非是为了追求广度，而是让我能够像一名“思想侦探”，去探寻那些被忽略的、未被解决的根本性困境。我坚信，我们这一代工程师的真正机遇，在于理解技术作为“创造性破坏”的引擎，为被成熟市场忽略的“沉默的大多数”，构建全新的价值网络。

我的所有独立项目，都是我对这一信念的笨拙但真诚的实践，也是我在“后台”对未来世界的系统性构想。我渴望加入一个同样拥有“颠覆者”基因的团队，将我对技术的热情与这一产品哲学相结合，共同打造能定义下一个十年的产品。