

王增焕

深圳 ✉ w905840774@gmail.com ☎ 17665252185
📧 wzhnbsixsixsixsix



教育背景

- 2022.09-2024.07 广东工业大学 | 学士 | 计算机科学与技术 | 主修课程: 数据结构、计算机网络、操作系统
2024.09-2026.06 埃克塞特大学 (QS 前 150) | 学士 | 计算机科学 | 主修课程: AI (Deep Learning)、Hadoop
2026.09-2028.07 香港城市大学 (QS 前 100) | 硕士 | 人工智能 | 主修课程: Machine Learning、NLP

实习经历

- 深圳华大基因科技有限公司 | 前沿技术研发部 (X) | Agent 后端开发实习生 2025.7-2025.11
- Agent workflow与检索性能优化: 基于 Dify 构建问诊助手与血型文献 Agent, 重构 PubMed 检索链路, 将关键流程耗时从 10 分钟降至 1 分钟, 并完成 Milvus 向量库部署。
 - 评估与数据构建: 搭建 LLM 评估与并发测试脚本, 定位并缓解知识图谱并发超时问题; 复现并部署 SciAgents, 开发轮询接口并迁移至 Dify; 同时按 Pasa 论文标准构建 10w+ RL 高质量训练数据。
- 杭州胜事影视文化有限公司 | AIGC 内容部 | AI Agent 开发 2025.5-2025.7
- AI 内容生产 workflow: 基于 Coze 搭建百万字小说改编 Agent 流程 (导入、篇幅规划、剧本生成、原文匹配评估), 显著提升动态漫/短剧脚本生产效率。
 - 陪聊系统全栈与 RAG 记忆: 主导 LumiLove 前后端分离架构, 前端 Next.js + React + Tailwind, 后端 Spring Boot + MySQL, 独立 FastAPI 负责 RAG 对话; 基于 LangChain + ChromaDB 实现对话记忆检索并部署至 AWS (EC2/Amplify, JWT 鉴权, RDS 托管)。

项目经历

- CompanyCopilot — 企业知识库与多智能体研究助手 2026.01 - 2026.03 [CompanyCopilot](#)
- 项目简介: 面向企业场景的 AI Copilot 全栈平台, 前端为开源项目 Chatbot UI, 后端为基于 Agentscope 的 Agent 系统。集成知识库问答 (RAG)、企业背景调研能力; 支持多格式文档入库、语义检索
- 核心技术栈: Python, FastAPI, AgentScope, Qdrant, MinerU, SKILLS, MCP, Function Calling, Supabase, Docker
- 多模态 RAG 检索链路: 利用 MinerU 构建文档解析链路, 支持 PDF / Markdown / CSV / 图片等多模态数据解析并进行向量化入库, Chunk 采用 paragraph+512-token 切分, 并且为每个 Chunk 生成文档级上下文描述并拼接后进行向量化 (Anthropic 提出); 检索层结合 Query Rewriting, 采用 Hybrid Search (Dense(本地部署 Bge-m3)+BM25), Top-10 召回后接入轻量多模态 Reranker(qwen3-vl-rerank) 精排。
 - 多智能体协作与任务路由: 基于 AgentScope 构建 Researcher/Analyst/Compiler/Guidance 四角色闭环, 覆盖证据采集、分析提炼、报告生成与意图路由, 稳定支持企业调研与尽调流程。
 - 工具化与外部增强: 通过 ReActAgent + Function Calling 接入检索工具与 Firecrawl MCP 外部搜索, 并以 SKILL.md 与 slash commands (/research、/due-diligence) 固化 workflow, 提升调用效率与结果一致性。
- Trump Persona 微调 (QLoRA + SFT) 2026.2-2026.3
- 项目简介: 基于 Kaggle 开源 Trump 推文数据集构建 persona 指令微调流程, 围绕数据清洗、蒸馏扩增与 QLoRA 训练, 验证小参数模型在风格化文本生成中的可控迁移能力。
- 核心技术栈: Python, LangChain, LLaMA-Factory, LoRA, Instruction SFT, Alpaca 格式, 数据蒸馏/增强
- 数据构造与蒸馏扩增: 基于 Kaggle 开源 Trump 推文数据集, 使用 Python + LangChain 搭建清洗、噪声过滤、长度截断与 Alpaca 转换流程; 并基于 qwen3.5-397b-a17b 进行数据蒸馏与增强, 将高质量 instruction-output SFT 数据从原始 10K 扩增至 50K。
 - SFT-QLoRA 训练: 使用 LLaMA-Factory + 4-bit QLoRA 对 Qwen2.5-1.5B 进行有监督微调, 训练过程中显存占用降低约 4x, 训练过程中 Loss 平稳下降并完成收敛。
 - 效果: 模型可根据不同目标生成符合 Trump 语言风格的辱骂文本, 风格一致性达 92%, 语气夸张性达 88%, 验证了 persona SFT 在可控风格迁移中的有效性。

个人技能

编程语言: Python, Java, JavaScript/TypeScript, SQL

LLM / Agent / RAG: 大模型微调 (SFT/LoRA) , 深度学习基础,Transformer 架构, SKILLS, MCP, Dify, Coze, LangChain, LangGraph, AgentScope, PyTorch

后端开发: FastAPI, Spring Boot, Django, RESTful API, JWT, MyBatis-Plus, Redis Stream, Lua 脚本, 分布式锁 (Redisson) ,TCP/IP 网络协议 ,Linux 系统

数据库 / 向量库 / 中间件: MySQL, PostgreSQL, Redis, Qdrant, Milvus, ChromaDB

工程化 / 云: Docker, CI/CD, AWS EC2, AWS Amplify, Nginx, Linux

AI 辅助开发: 熟练使用 Claude Code (Open Code CLI)、Antigravity(Cursor) 等 AI 编程工具, 具备 Prompt 工程能力, 建立 AI 辅助开发规范体系, 开发自定义 Skill 指令和 MCP, 实现 AI 深度参与开发全流程。

语言能力: 英语 (IELTS 6.5), 具备良好的英文沟通与阅读能力。