单玉昆

男 | 22岁 19192422920 yshan@bu.edu

教育经历

波士顿大学 本科 计算机科学 2021-2025

QS世界排名88名

工作经历

意仕腾人工智能科技 AI算法开发工程师 2025.05-至今

- 作为后端核心开发，主导AI数字人教学项目的需求分析与架构设计，使用FastAPI构建高性能RESTful API，集成GPT、RAG、ASR/TTS，ollama联网搜索（Searxng）等AI能力，并通过负载均衡与全链路异步化改造，支撑高并发实时语音交互。
- 对Ollama模型服务实施压测与调优，设计部署负载均衡代理，结合全链路异步化（ainvoke），降低平均响应时间70%，提升系统吞吐量近10倍。
- 独立负责英语发音测评算法迭代，重构音素解析与相似度计算逻辑，增强对连读、省略等自然语音现象的处理，评分准确率提升约40%，并封装为标准接口集成至产品线。
- 开发多种K12口语题型（如图片描述、情景对话）后端逻辑与动态题目生成算法，实现智能离题判断与交互流程。

实习经历

BU Spark! 软件工程师 2023.09-2023.12

波士顿, 马萨诸塞州

协调一个由 1 名用户体验设计师和 3 名开发人员组成的跨职能团队，成功启动并推动一家初创企业发展，致力于为宠物主人提供关于宠物友好餐厅和周边目的地的精准信息。

使用 Ionic React 开发了一款功能全面的移动端全栈应用程序，并使用 PostgreSQL 作为后端数据库，服务于波士顿地区超过 100 名宠物主人。

个人优势

人工智能与机器学习:

OpenAI GPT API, RAG, Agents, Function Calling, 大语言模型应用开发, 语音识别, 语音合成, Ollama, 发音测评算法, NLP工程化, 模型部署与压测

编程语言:

Python, Java, JavaScript, TypeScript, C, Assembly, SML, Bash, SQL

后端开发:

FastAPI, Node.js, Express.js, TRPC, Prisma, RESTful API开发, 高并发与异步编程, 负载均衡， Jmeter

前端与全栈,数据库:

React.js, Bootstrap, Passport.js, PostgreSQL, MongoDB, SQLite

云与DevOps:

Docker, Nginx, Git, Jira, Postman, GNU/Linux, 环境配置, CI/CD

开发工具与其他:

Winsurf, Gemini-CLI, Xcode, LaTeX, JSON/XML, Microsoft Office, Figma