

王帅



## 基本信息

姓名: 王帅 电话: 15928682059 邮箱: jcanha1@foxmail.com

## 教育背景

2020.10-2024.07 西南民族大学 人工智能专业  
英语四级

## 技能能力

- 编程语言: Python 熟练掌握 Numpy、Pandas、Pytorch、Opencv、NLTK、spaCy 等常用 Python 框架和库。
- 开发环境: 熟练掌握 Linux 下 Python、AI 项目的开发与环境配置部署。
- AI 开发:
  - 熟悉 ResNet 等常见深度学习模型及 Transformer 架构。
  - 熟练掌握 LlamaIndex、Langchain 框架, 精通 RAG、Agent 工程实践, 熟悉 function calling、MCP 等技术。
  - 精通 Prompt Engineering, 熟悉 CoT 等技术以提升大模型性能与稳定性。
  - 具备 LLM Fine-tuning、性能测试与评估经验。
  - 熟悉 AWS Bedrock、Azure AI、Vertex AI 等云平台, 熟练调用 Anthropic、Gemini、OpenAI 等模型 API。

## 开发工具

- 熟练使用 VScode、Cursor、Poetry、Github 等代码开发管理工具。

## 工作经历

2024.7-2024.9 元数智慧科技有限公司 Python 后端开发

工作内容: 根据公司业务部门的需求, 搭建知识图谱、构建 Agent

- 根据业务需求构建知识图谱及 Agent。
- 基于 LlamaIndex 与 RAG 技术搭建智能客服系统, 实现了长记忆、多用户并行及持续学习功能。
- 运用 Prompt Engineering (含 CoT) 深度挖掘大模型潜力, 优化模型输出与外部 API 调用稳定性。

2024.9-至今 思维巡航(深圳)网络技术有限公司 AI 应用开发工程师

- 负责根据企业用户需求定制与优化 AI 模板。
- 主导通过 Fine-tuning 低性能 LLM 及应用结构化输出技术, 在保证业务效果的同时, 成功降低了 95% 的模型成本。
- 持续跟踪前沿 LLM 技术并推动业务迭代, 开发自动化测试工具, 从多维度评估 LLM 在具体业务场景下的性能。
- 通过多源异构数据融合与多路召回策略协同优化, 提升 AI Chat 的上下文理解能力和生成内容的准确性、丰富度。

## 项目经历

- 项目 1: 智能客服助理, 基于 LlamaIndex 与 Ollama, 使用 RAG 和 FunctionCalling 技术搭建智能客服助理, 通过提示工程赋予角色身份实现情感对话, 并拥有长记忆功能可实现多轮对话, 通过混合向量检索以及关键词检索, 使效率更高。使用 CoT 技术, 使模型可以自主思考、调用外部 API 更稳定。
- 项目 2: 实现了基于 LLM 的语音响应, 并利用实时 API 显著降低了端到端语音交互延迟。
- 项目 3: 利用多模态模型进行视频理解, 实现通过上传视频或录屏自动生成文档和分析关键帧。
- 项目 4: 搭建智能化 LLM 测评框架, 构建多维度评估基准体系, 实现模型性能的快速量化测评与动态追新。

## 自我评价

- 具备强烈的责任心、敬业精神、优秀的团队协作与沟通能力。
- 拥有扎实的数据分析与处理能力, 以及出色的持续学习和快速适应能力, 能高效投入新领域并执行任务。
- 熟悉 Transformer 架构及 PyTorch 框架, 对 AI 技术抱有浓厚兴趣, 并积极进行前沿模型(如 Anthropic、Gemini、OpenAI 等)的追踪、评测与学习。
- 具备 AI 应用开发及 LLM 优化、评估的实践经验, 熟悉主流云 AI 平台(AWS Bedrock、Azure AI、Vertex AI)的 API 调用, 致力于通过技术解决实际问题并创造价值。

