

赖为乾

男 | 年龄: 28岁 | 13430147593 | arkin-dev@qq.com

4年工作经验 | 求职意向: AI/大模型应用工程师 | 期望薪资: 20-22K | 期望城市: 东莞



自我评价

• 职业背景

4年计算机视觉及LLM大模型开发工作经验，具备计算机软件行业职业成长能力，拥有小技术团队管理经验，熟悉项目开发流程，掌握Python、Golang、TypeScript、C/C++ 软件开发语言。

• 专业技能

1. 具备计算机视觉和LLM大模型研发案例成功的经验，拥有项目独立开发实战经验。
2. 理解朴素贝叶斯、SVM等机器学习算法，理解注意力机制、Embedding、Transformer、Bert、GPT等模型基本原理和架构。
3. 熟悉YOLO系列、Mobilenet系列、SSD等检测算法。
4. 熟悉使用LangChain、Eino、VLLM等大模型框架，卷积神经网络使用及PyTorch深度学习框架。
5. 掌握模型的压缩量化方法，Lora、QLora等微调技术，掌握长文本训练，以及定制化输出等。

OpenAI开发

1. 熟练使用 OpenAI 接口进行调用和定制化开发，包括模型选择、调参以及语言和多模态的结合
2. 基于 OpenAI 接口开发过一个支持多模态输入的 AI Chatbot App。

LangChain开发

1. 熟悉 Prompt Engineering 技术：Zero-shot、Few-shot、COT、ReAct、Reflexion、Prompt Chaining等。
2. 掌握 LLM、Chat、Models、PromptTemplates、OutputParse、Chains 等组件的使用。
3. 熟悉 Verbose 日志，Debug 模式和 LangSmith 平台等调试监控技术。
4. 能够基于 LangServe 搭建服务，扩展 API 供客户端调用。
5. 能够使用 LangChain Chat History 组件实现聊天历史持久化。
6. 能够自定义 FunctionTool，实现工具的同异步调用。实现聊天、天气查询和网页搜索的 Agent 调用功能。
7. 熟练使用 LangChain 构建本地向量数据库，实现存储与查询。熟练使用 Chroma 等向量数据库实现 Similarity Search、RAG 增强检索。
8. 能够利用 Streamlit、Loader、Splitter、Embedding、Vector Store、Retrievers 等组件，开发具备记忆能力的 RAG 问答机器人。实现历史记录感知功能和对话历史状态管理。

Hugging Face 模型训练与微调

1. 熟悉使用 Hugging Face API 调用模型，实现文本生成、文本分类和文本问答等 NLP 任务。

2. 熟练使用本地模型对下游任务做针对性的模型设计，并训练评估和测试。
3. 掌握模型微调的三种基本模式（全量/增量/局部），使用 BERT 微调过中文评价情感分析数据集。
4. 掌握通过更改模型配置信息并初始化模型，来实现超长文本训练方法，实现文本多分类任务。
5. 熟练根据下游任务通过新增全连接层至顶输出类别数来进行相应的模型设计。掌握冻结编码层，只对嵌入层和全连接层进行训练。
6. 熟练使用远程服务器对 DeepSeek 进行 Lora 微调测试评估、模型合并与量化、部署量化后的模型。
7. 熟练使用 ModelScope 在线训练平台进行数据下载和模型调用，在线训练过 GPT2 模型。
8. 理解 LoRA 微调的基本原理，能够使用 LLaMA-Factory 基于自定义数据集进行微调训练。

• 职业优势

崇尚终身学习及自我赋能，具备计算机视觉模型和现在流行大模型应用开发能力，积极将成熟的GIS系统和LLM大模型结合，实现更加智能的地理信息系统，让非专业人士也能熟练使用。希望在更多行业可以有落地项目。

• 综合素质

系统全面的逻辑思维和分析能力，高效闭环，快速的适应能力，极强的承压抗压及应变能力，具备团队协作意识，拥有良好的职业操守及企业忠诚度。

项目经历

• GeoAI智能体内部研发项目 (2024.11-2025.07)

◦ 项目简介：

针对传统GIS系统操作复杂、需要专业技术人员操作等问题，基于大模型技术研发时空智能体系统。结合北斗网格编码引擎与MCP协议，实现自然语言指令驱动的GIS系统时空任务自动化，包括无人机调度、空间查询等核心功能，推动产品智能化转型，让非专业人士也能通过自然语言与GIS系统交互。

◦ 技术方案：

基于DeepSeek大模型进行指令微调与私有化部署，学习时空网格任务概念和GIS领域知识。设计并实现MCP服务及前端展示接口，构建智能体与业务系统的无缝对接。开发GIS数据解析与上下文接口，将复杂时空数据（遥感影像、矢量数据等）转化为模型可处理的结构化信息，实现智能体对空间数据的深度理解。

◦ 项目成果：

成功实现自然语言下的时空网格任务推理执行，模型任务理解精度达60%+，首次将大模型技术应用于GIS时空任务自动化领域。开发的MCP服务支持无人机调度、空间查询等典型任务的智能化，系统可同时响应数百个任务请求，实现高并发智能服务。GIS数据解析与上下文接口显著提升任务执行准确率约40%，大幅降低人工干预需求，为GIS产品智能化转型提供了完整的技术解决方案。

• 松山湖无人机低空警务智能应用系统 (2024.10-2025.07)

◦ 项目简介：

针对松山湖密集路段无人机频繁巡航场景下，传统人工监控交通事故效率低下、人力不足等问题。构建百架次无人机并发AI检测系统，基于计算机视觉和多模态大模型技术，实现交通事故和异常人员的

智能识别与自动取证，为低空经济警务应用提供智能化解决方案。

- **技术方案：**

采用YOLOv8小模型结合TensorRT推理加速构建实时检测引擎，实现<280ms超低延迟交通事故检测。基于北斗网格空间编码实现多无人机识别去重，确保事件唯一性记录。搭建基于Qwen2.5-VL多模态大模型的智能审查系统，通过指令工程和视觉-文本融合技术，实现检测结果的智能复核与优化。

- **项目成果：**

成功构建支持百架次无人机并发的AI检测系统，单帧交通事故检测延迟降低至<280ms，满足连续巡查的实时性要求。多模态大模型智能审查系统识别事故的平均精度达60%，处理效率提升至每小时复核上百起事件，大幅减轻人工误检处理工作量。基于北斗网格编码的事件去重机制确保每起事故唯一记录，提高事件准确性，为低空警务智能化应用提供了完整的技术解决方案。

• 省水利厅"广东智慧河长"AI智能检测系统 (2022.03-2023.12)

- **项目简介：**

针对广东省水利厅河道"四乱"治理的智能巡查需求，传统人工检测效率低下、检测标准不一致、缺乏自动化分析能力等问题。基于计算机视觉和深度学习技术开发河道"四乱"智能检测系统，实现遥感影像自动化分析，检测准确率达人工的80%，处理速度提升4倍，为智慧水利建设提供AI技术支撑。

- **技术方案：**

采用YOLO系列目标检测算法和语义分割技术构建智能检测模型，从零构建万张级"四乱"遥感检测数据集，建立标准化数据标注、清洗、增强流程。通过模型轻量化与推理加速优化（量化部署 + ONNX/TensorRT），结合Docker微服务化和分布式部署架构，实现大规模遥感影像的高效处理和高并发在线检测服务。

- **项目成果：**

成功构建行业内可复用的万张级"四乱"数据集，形成标准化的数据资产。AI检测模型预测准确率达人工检测的75-80%，单幅影像检测耗时从12秒缩短至3秒，整体效率提升4倍+。通过分布式部署支持每小时处理数千张瓦片影像，实现高并发在线巡查，为河道治理提供可靠的AI技术支撑，显著提升监管效率和智能化水平。

教育背景

理学学士 (2017-2021)

地理信息科学 (GIS) 专业是一门融合地理学、计算机科学与大数据分析的交叉学科。

嘉应学院 - 梅州, 广东

- 毕业论文《[基于 Faster R-CNN 和无人机多光谱影像的梅州柚树单株识别方法研究](#)》
- IEEE 期刊《[Detecting and Mapping Individual Fruit Trees in Complex Natural Environments via UAV Remote Sensing and Optimized YOLOv5](#)》
- 普通期刊《[城市景观格局变化对生态系统服务价值的影响](#)》
- 发明专利《[基于无人机图像深度学习的植株识别制图方法和系统](#)》- 实质审查

个人博客: <https://arkin-developer.github.io/blog>