

个人简历



基本信息

姓 名：杨斌
电 话：18375289639
邮 箱：18375289639@163.com
就职状态：在职

工作年限：五年
毕业院校：贵州商学院
专 业：物联网工程
学 历：统招本科

求职意向

工作性质：全职
目标职能：大模型开发工程师

地点：深圳
目标薪资：面议

专业技能

- 熟练使用 Java、Python、Scala、Shell 编程语言。
- 熟练使用 Hadoop、Hive、HBase、ClickHouse、ElasticSearch 进行大数据平台搭建和数据分析。
- 熟练使用 SparkCore、SparkSQL、SparkStreaming 进行实时/离线数据分析，进行 Spark 调优。
- 熟练使用 FlinkDataStream、FlinkSQL、FlinkCEP 进行实时/离线数据分析及复杂事件处理。
- 熟悉阿里云 DataWorks、MaxCompute 进行实时数据同步，离线数据分析及相关作业运维与管理。
- 熟悉常见数据库 Oracle、MySQL、PostgreSQL、MongoDB、Redis。
- 理解 CNN、RNN、注意力机制、Embedding、Transformer、BERT、GPT 等模型原理和架构。
- 熟练掌握 LangServe 搭建服务，LangChain 模型包装器、提示词语模版开发，扩展 API 接口供客户端调用和 LangChain ChatHistory 组件实现聊天历史持久化。
- 熟练使用 LangChain 数据加载器和输出解析器，如 JSON、XML、YAML 等格式输入输出和自定义输入输出。
- 熟练使用 FAISS、Chroma、Pinecone、Zilliz 等向量数据库的 Similarity Search、RAG 增强检索和 Embedding 优化。
- 熟练使用 LangChain 自定义工具，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。
- 熟练使用 Agent 和 LangGraph 核心组件，理解 Human-in-the-loop 和 React Agent 实现思路，实现过高阶 Agent 用法，包括 Multi-Agent Systems 和 Plan-and-Execute 策略。
- 熟练使用 HuggingFace 和 ModelScope 的 API 调用模型和数据集处理，实现文本生成、文本分类和文本问答等 NLP 任务。
- 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。
- 掌握 Llama3、Ollama 进行模型部署、功能使用、调用模型。
- 理解 LORA 微调的基本原理，能够使用 Unsloth、LLaMA-Factory 基于自定义数据集进行微调训练。
- 熟练掌握 Dify 部署和配置，实现基于 LLM 的 RAG 增强对话交互、文本生成、多 Agent 任务分解、工具调用、多节点 workflow 编排。

工作经历

2020.03-2023.04 贵阳天依惠禾大数据有限公司。
2023.07-2025.03 润建股份有限公司。

职位：大数据开发工程师
职位：大数据开发工程师

项目名：人像大数据情报系统

项目简介：通过前端埋点在贵阳市各个街道，各个路口的摄像头，电子围栏，以及贵阳市大多数小区每户人家 WIFI 设备和门禁设备，经过抓拍，感应，聚类引擎聚类、提纯得到的 VID, IMSI, WIFI, CAR, GRD 数据构建人像大数据系统，实现高效的人脸检测定位识别比对及快速识别辨别特定人员真实身份系统。

职责简介：

职责一：负责前端埋点产生的 VID, IMSI, WIFI, CAR, GRD 数据使用 Flume 采集文件数据抛到 Kafka 消息中间件及落地到 ODS。

职责二：负责使用 Flink 对接 Kafka 读取相应 Topic 主题数据清洗存储至 ClickHouse 的 DWS。

职责三：负责使用 Spark 对 ODS 每天的 VID, IMSI, CAR 数据清洗存储 DWD。

职责四：负责使用 Spark 对 DWD 每天的 VID, IMSI, CAR 同行数据挖掘 DWS。

职责五：负责使用 Spark 统计 DWS 同行人员指标(日同行，周同行，月同行，7 天内连续 3 天同行)数据存储 ADS 并使用 Sqoop 同步到 MySQL。

贡献和成果：

- 1、使用 Hadoop、Spark、Flink 构建了数据处理和分析的框架，实现了高效的数据处理和存储。
- 2、利用 Spark 数据分析和数据挖掘进行同行人员数据开发，提供了准确的用户行为预测。
- 3、成功开发并部署了人像大数据情报系统同行人员侦查模块，帮助人像大数据更精准预测同行用户。
- 4、该系统为公安提供了有价值的洞察和人员行为分析，更精准定位同行者和同行人员地址。

项目名：深圳能源集团大数据平台

项目简介：因深圳能源集团长期发展要求，企业数据日益增多，在互联网、大数据、AI 大模型大时代下，集团及分公司以传统方式进行数据汇总上报，存在易出错且效率低下，存档麻烦，数据汇总困难，无法实现自动化管理等问题和现状，占据极大资源，导致集团处理数据成本日益递增，为解决此数据问题，并提高数据处理效率，根据数据分析企业困难点来量化企业目标，更好的了解企业收入和支出，实现对产品销量、渠道价值和未来销售预测。

职责简介：

职责一：负责前期系统调研和数据摸底。

职责二：负责指标清单数据溯源、并确定取数逻辑和取数范围。

职责三：负责大数据平台模型设计和数据开发，包括 ODS、DWD、DWS、ADS 层，并进行模型文档化。

职责四：负责大数据平台数据服务提供和数据服务文档化。

职责五：负责与后端定开页面对接和需求功能讲解，QUICK BI 可视化页面开发和调整，并生成数据分析报告。

职责六：负责任务托管配置、任务预警、数据质量(DQC)检测和配置。

贡献和成果：

- 1、使用阿里云 DataWorks+MaxCompute 构建数据仓库，日均数据处理提升 80%，ETL 任务调度成功率提升至 99.8%。
- 2、使用 MaxCompute SQL 进行数据清洗、聚合逻辑，报表产出效率提高 80%。
- 3、通过 Quick BI 设计数据可视化报表 20+，支持运营部门实时分析用户行为与业务指标。

项目名称：深圳能源智慧能源 AI 中枢系统

项目简介：随着智能化的出现，依托深圳数字经济与能源双转型战略，深度融合人工智能技术与公司管理全流程，构建覆盖 清洁电力、综合能源、碳资产管理的智能决策系统。

职责简介：

职责一：负责深圳能源财务域数据，工程域数据，战略与运控域数据，懂办域数据的清洗、转化，维度补充和归档整理。

职责二：负责深圳能源大模型微调，训练，解决大模型幻觉，可解释性，缺乏常识推理等问题。

职责三：负责深圳能源大模型 RAG 知识库建设，解决大模型知识匮乏，数据时效性问题。

职责四：负责 dify 平台的工作流编排、Agent 智能代理的任务分解、推理、工具调用。

贡献和成果：

- 1、部署 Ollama 模型管理，减少较小的资源占用。
- 2、部署 LLaMA-Factory 模型高效微调框架，减少服务器资源使用。
- 3、构建深圳能源集团 RAG 知识系统，召回准确率提升至 91%。
- 4、深圳能源综合能效提升 35%。

项目名称：财务公司财智通 AI 决策支持系统

项目简介：随着金融数字化转型加速，传统财务公司面临效率瓶颈、风险控制难度提升、数据分析滞后等挑战。为提升核心竞争力，我司启动 AI 赋能战略，深度融合人工智能技术与财务管理全流程，构建智能化、自动化、可预测的新型财务服务体系。

职责简介：

职责一：负责财务公司财报数据，资料数据，发票数据清洗，转化，维度补充和归档整理。

职责二：负责财务公司大模型微调，训练，解决大模型幻觉问题。

职责三：负责财务公司大模型 RAG 知识库建设，解决大模型知识匮乏问题。

职责四：负责 LangChain 进行 ChatTemplate 开发工作，提高 AI 响应的相关性和准确性。

职责五：负责 LangGroup 的 Agent 智能体构建，进行拆解任务，分步解决，信息存储与回忆。

贡献和成果：

- 1、部署 Ollama 模型管理，减少较小的资源占用。
- 2、自动化处理发票核验、报销审核、账务录入等重复性工作
- 3、效率提升：单笔业务处理时间缩短 70%，错误率下降 95%，降低合规风险罚款 30%。

技能证书

- 全国计算机二级（C 语言）。
- 程序员（软考初级 Java 认证）。
- Oracle Java 认证（OCJP）。
- 戴尔 JAVA 认证。

自我评价

- 1、熟悉常见业务处理，能快速融入团队。
- 2、性格开朗，头脑灵活，逻辑思维能力强，思路清楚。
- 3、学习能力强，喜欢看一些技术书籍及专研新技术。
- 4、喜欢逛一些技术论坛，如 CSDN。
- 5、对工作尽职尽责，乐于从事有挑战性的工作。