个人简历

基本信息

姓　名：杜文辉　　年　龄：26
电　话：15958306649　　年　限：3
邮　箱：d1686385160@163.com　　学　历：本科

专业技能

- **Agent 系统开发：**具备多智能体系统设计与开发经验，熟悉基于 A2A 及 MCP 等协议的智能体间协作与任务编排机制。
- **RAG 系统开发：**拥有从零到一搭建检索增强生成系统的全流程能力，熟悉文档解析、向量化、语义检索与答案生成各环节，熟练运用 Milvus 等向量数据库构建高性能检索服务。
- **大模型微调：**掌握大模型的微调技术(LoRA)，能够根据下游任务进行模型适配与性能优化。
- **NLP 自然语言处理：**掌握自然语言处理核心技术，熟悉文本分类、语义理解、序列标注等任务，具备将大语言模型应用于实际业务场景的能力。掌握模型压缩技术,比如模型蒸馏和量化
- **深度学习：**熟练掌握 RNN、CNN、LSTM、GRU 等经典深度学习网络，深入理解 Transformer 架构，并能够运用 PyTorch 框架完成模型设计、训练与部署。
- **机器学习：**熟悉逻辑回归、线性回归、随机森林、XGBoost 等经典机器学习算法，具备扎实的建模思维与解决实际问题的能力。
- **编程与数据处理：**熟练掌握 Python 编程语言，具备良好的工程规范和开发能力。熟练运用 Pandas、NumPy 等工具进行高效的数据清洗、分析与特征工程。

工作经历

2024.09-2026.01　　浙江和达科技股份有限公司　　大模型应用开发工程师

1. 负责开发基于 A2A 与 MCP 协议的多智能体协作系统，实现复杂任务的自动化分解、调度与执行。
2. 主导从 0 到 1 搭建企业级 RAG 知识库系统，完成文档解析、向量化嵌入与语义检索全流程开发，并基于 Milvus 实现高性能向量检索服务。
3. 负责将多智能体决策层、RAG 知识服务层与 SCADA/GIS 业务层进行解耦设计与松耦合集成，构建稳定可靠的一体化智能平台。

1. 参与了基于大模型与知识图谱的智能维修系统开发。主要负责模型本地化部署、RAG 检索链优化及数据处理管道构建，助力系统实现故障修复时间降低 30%，问答准确率达 92%的目标。

2. 通过完成数据标注、模型量化与蒸馏等基础工作，直接支撑了问答系统准确率的提升与服务延迟的降低。

项目经验

项目一:智慧水务多 Agent 协作系统

- **项目背景**:

当前公司各业务系统独立部署，形成了突出的“数据孤岛”问题。这一现状导致城市水网调度难以协同，应急响应迟缓，制约了整体运营效率与服务水平的提升。为从根本上打破数据壁垒、实现智能协同，我们构建了一套基于 A2A 与 MCP 协议的多智能体协作系统。

- **涉及技术**: LangChain，A2A 协议，MCP 协议，RAG，Milvus，Mysql 等
- **项目职责**:

1. 实现基于 A2A 协议的多智能体协作架构，定义中枢协调智能体与分区控漏官、供水调度官、管网运维官、大屏指挥官等专业 Agent 的通信规范，支撑水务应急场景下的高效协同。

2. 负责 MCP 工具生态的规划、封装与标准化，规划并统一 MCP 协议标准，推动各业务子系统（SCADA、工单、GIS、可视化大屏、水力模型、知识库）完成适配器开发，将其能力封装为统一的 MCP 工具。

3. 负责智能语音交互指挥模块的集成与场景化应用，主导阿里 FUN-ASR 语音识别模型的集成。

4. 主导应急预案知识库与多智能体系统的集成，确保“应急指挥官 Agent”能通过 MCP 协议精准检索、匹配并推荐处置预案

5. 保障系统整合、性能优化与持续演进，统筹大模型、语音引擎、MCP 服务与现有水务业务系统的整体集成，确保系统稳定。

- **项目成果**:

1. 它实现了从“被动响应到“主动智能”的运营模式变革，从“经验驱动”到“数据与模型双驱动”的决策方式升级。

2. 应急响应效率提升,将爆管、水质等事件的平均协同决策与初报生成时间从天/小时级缩短至分钟级。

3. 该系统输出的应急方案，为调度长决策提供了强力支持，最终方案采纳率逾 90%，极大提高了用户体验。

项目二:智能售后知识引擎

- **项目背景**:

随着公司业务规模的持续扩大，项目售后维护与技术支持工作面临日益增长的压力，出现了响应不及时、问题处理效率偏低等问题，导致客户投诉显著增加。为切实提升售后服务质量，公司高度重视该环节，并正式上线了基于 RAG（检索增强生成）技术的智能问答系统，旨在通过智能化手段优化客户支持流程，提高问题响应速度与解决效率。

- **涉及技术**: Python、LangChain、Mysql、Milvus、Redis、BM25、MD5、OCR 等。
- **项目职责**:

1. 设计并实现基于 MySQL + Redis 的 FQA 系统：结合 jieba 分词与 BM25 算法进行问题匹配，通过 Softmax 归一化和阈值判断答案置信度，命中结果自动缓存至 Redis 提升响应速度。

2. 构建端到端 RAG 系统：完成 PDF/PPT/Word 等文档的 OCR 解析、双策略切分（规则+模型）、BGE-M3 向量化（稠密+稀疏），并入库向量数据库。

3. 多策略混合检索模块：基于大模型进行问题分类，针对专业知识动态选择检索策略（如假设、子查询），结合混合检索与重排序以提升召回质量。

4. 多级缓存与降级链路：设计了“缓存 → 规则匹配 → 语义检索 → 大模型生成”的级联服务链路，在保障响应速度的同时，通过自动降级机制确保答案准确性。

5. 向量化存储与管理：为每个文本块生成 MD5 唯一 ID，并关联存储其稠密/稀疏向量元数据，实现了高效且可追溯的向量检索与更新。

- **项目成果:**

1. 项目上线后,售后投诉减少了 80%以上，客户满意度达到 95%以上，提高了售后团队的响应速度和工作效率。

2. 使用多策略加混合检索重排序的优化方法后，召回率达到 89%以上，忠诚度达到 90%。

3. 系统平均响应时间从原来的 15 秒以上降至 5 秒以内，日均处理客户咨询量提升 3 倍以上。

项目三 :制造业设备维修知识库

- **项目背景**:

针对高端制造与能源行业设备维修知识分散、经验依赖性强、平均修复时间长等核心痛点，参与设计并实现了一套基于 RAG 与知识图谱的智能维修辅助系统。该系统深度整合结构化手册、非标经验笔记与历史工单，通过自然语言交互，为现场工程师提供精准的故障排查指引、备件信息与标准化解决方案，旨在将隐性经验转化为可复用的企业数字资产，显著提升维修效率与设备综合利用率。

- **涉及技术**: Python、FastAPI、LangChain、Bge-M3、Milvus、Neo4j、MySQL 等。
- **项目职责**:

1. 系统架构与开发：参与系统技术选型与架构设计，协助实现从文档解析、知识存储、智能检索到服务集成的全链路开发。

2. 数据处理管道开发：开发批流一体的数据处理管道，完成多源维修文档的自动化解析、关键信息抽取与向量化入库。

3. 算法研究与实现：参与文本向量化模型选型与优化、混合检索策略调优，以及基于设备 BOM（物料清单）的知识图谱构建，提升查询准确率与语义理解深度。

4. 负责核心提示词（Prompt）工程，针对故障诊断、步骤查询、报告生成等不同任务，设计了结构化、少样本（Few-Shot）的提示模板，显著降低了模型的幻觉现象。

■ **项目成果:**

1. 运维效率提升：项目上线后，平均故障定位时间缩短约 40%，平均维修作业时间（MTTR）降低约 30%。

2. 知识资产沉淀：累计沉淀标准化维修案例 5000 余条，新工程师独立解决复杂问题的培养周期缩短约 50%。

教育背景

2019-09~2023-06	郑州科技学院	软件工程

自我评价

1. 逻辑思维严谨，思路清晰，善于从复杂问题中提炼关键要素并高效解决。
2. 对数据高度敏感，具备扎实的编码能力与良好的工程习惯，注重代码质量与可维护性。
3. 工作细致认真，具备优秀的团队协作精神和沟通能力，能够高效协同推进项目落地。