求职意向：AI智能体开发工程师

个人简历

姓名：彭潇哲 出生年月：2000.05

性别：男 电话：16670483367

邮箱：[728100660@qq.com](mailto:728100660@qq.com)

**教育背景**

EDUCATION

**2017.09-2021.06 湖南工商大学 物联网工程技术**

**个人技能**

Skills

有过丰富的系统和平台开发经验，掌握大模型开发技术，技能树如下：

**编程语言：**精通 **Python**，熟悉内存结构、垃圾回收机制，具备扎实的面向对象编程能力；了解 Java，有 Java 项目经验。

**大模型开发：**熟练掌握**Dify**框架设计思想及二次开发、**RAG**，能独立完成大模型应用落地；熟悉 **RAGFlow**、**AutoGen**、LangChain、LangGraph、微调（LoRA）、VLLM 部署等技术。

**数据库与检索：**熟练使用 **MySQL**，具备 SQL 调优经验，了解主从复制、读写分离及集群搭建；熟悉 FAISS、向量数据库。

**后端开发：**熟练掌握 Sanic、Flask 等框架；有高并发、异步编程及 WebSocket 开发经验。

**中间件与分布式：**熟悉 **Redis** 核心数据结构及高并发场景应用；熟悉 Kafka、**Celery**分布式任务调度；了解 Tdengine 等时序数据库。

**架构与设计：**熟悉常见设计模式（单例、工厂、代理、发布-订阅等），能基于设计原则实现高复用代码。

**系统运维：**熟悉 Linux 常用命令，能独立进行服务部署与故障排查。

**开发工具：**熟练使用PyCharm、Git、Svn、Navicat、PostMan等开发工具

**工作经历**

work experience

**2025.04-至今 新中大科技 AI应用开发**

1. 参与大模型应用平台 Dify 的定制化开发与优化，聚焦中文环境下的问答检索准确性提升；

2. 参与公司智能问答系统及数据分析类 AI 产品的研发工作，基于大模型完成产品原型设计与实现

3. 探索多轮对话流程中的上下文管理、意图识别、查询修复等问题，提升系统可用性和用户体验

**2022.05-2024.10 清科优能（深圳）技术有限公司 Python工程师**

1. 负责核心后端服务架构设计与开发，维护公司关键系统；

2. 设计数据库结构并进行 SQL 查询性能优化，提高数据处理效率；

3. 主导公司技术框架优化，清理不合理遗留代码，提升系统性能及代码质量；

4. 制定并推广开发标准，提升项目架构一致性与团队开发效率，提高系统扩展性和可维护性。

**2021.07-2022.5 多益网络 Python游戏开发**

1. 负责项目日常维护、热更新、bug 查修，保障项目的持续稳定运行。

2. 利用发布订阅模式、工厂模式、单例模式等各种设计模式完成游戏功能开发。

**项目经历**

PROJECT experience

**六和平台 —— 企业级大模型应用系统**

**项目描述：**该平台面向企业知识问答、结构化数据查询等多场景，构建大模型能力集成中台，支持文档问

答、工作流、智能体等多个智能模块。

**技术栈：Python、Weaviate、Dify、OpenAI API、PostgreSQL、Flask**

**核心工作**：

1. 知识库构建方面：接入父子模式，增强大模型理解能力。提高回答质量；
2. 设计并接入 BM25 中文全文索引增强模块：
   1. 手动分词并存入 segmentedQuestion 字段
   2. 查询时同步对用户提问进行分词，提升关键词召回成功率
   3. 有效解决原生 BM25 中文检索不支持分词问题
3. Web 内容：知识库可见权限、文档上传下载、引入 MongoDB 作为存储引擎等。

**项目亮点：**

1. 全文检索通过 BM25 中文检索准确率提升约 60%，结合 Hybrid 向量搜索策略，进一步提升召回与排序质量；
2. 接入结构化解析拆分文本，支持标题检索，提升特殊情况下的召回率。

**基于 Dify 的多轮对话式 Text2SQL 系统**

**项目描述：**开发了一个基于 Dify 的企业内部数据库问答系统，支持用户通过自然语言提问生成 SQL 并查

询结果。系统可理解多轮上下文，修复错误 SQL，输出可视化数据格式。

**技术栈：Dify、PostgreSQL、Python、LLM（OpenAI/自定义）、Prompt 编排、Redis**

**核心工作：**

1. 计并实现了完整的 Text2SQL 多轮对话流程，包括意图识别、SQL 生成与修复、格式控制、上下文记忆等关键环节；
2. 编写 LLM Prompt 模板，用于识别用户意图（如：直接查询、补充说明、报错反馈、格式要求等）， 实现对不同意图的路由和响应；
3. 构建可配置的工作流节点系统，支持自动记录和追踪用户上一次提问与 SQL，提升对话连续性与准确性；
4. 针对 PostgreSQL 的语法特性，设计语法修复模块，结合错误信息引导大模型重新生成兼容 SQL；
5. 支持多种结果格式输出（如柱状图数据、表格、纯数字摘要），提升数据展示的灵活性和清晰度；
6. 项目已集成至公司内部知识问答平台，用于辅助运营、数据团队快速获取业务数据；

**项目亮点：**

1. 模拟真实用户行为，覆盖多轮补充、错误纠正、结果格式调整等对话路径，强化系统容错能力；
2. 意图识别+上下文追踪+动态 Prompt 生成，构建了一套面向多轮对话的工程化 LLM 应用框架；

**项目名称：微网能管平台开发（EMS + 虚拟电厂系统）**

**项目简介：**

参与公司微网能源系统的核心平台开发，包含 EMS 能量管理系统与虚拟电厂调控系统两个子模块，支撑

电力调度、能耗采集、控制策略下发与收益结算等功能，服务于分布式储能与电力辅助交易场景。该项目

整合了 EMS 能量管理与虚拟电厂调度系统，是公司微网能源平台的重要组成部分。

**技术栈**：Python、Celery、Kafka、Redis、MQTT、TDengine、Sanic、MySQL、装饰器模式、爬虫、

权限认证服务

**核心工作**：

1. 数数据采集与任务调度：使用 Celery + schedule实现定时调度任务，采集 TDengine 中的设 备运行数据，写入业务库，支撑系统可视化与分析模块。
2. 异构数据流管理：基于 Kafka 构建生产测试环境的数据分流系统，保障数据在 Redis（缓存）、 TDengine（持久化）与测试数据库中的同步与一致性。
3. 远程控制与指令下发：使用 MQTT协议实现对储能设备的远程控制（启停、功率调节等），确 保控制命令实时、稳定地发送与响应。
4. 外部数据集成与自动入库：开发爬虫模块，定时抓取碳交易行情、天气数据等外部信息，自动入库用于能耗预测与辅助策略。
5. 高性能 API 与并发优化：引入 Sanic 异步框架 + Redis 缓存机制，显著优化慢查询与高并发接 口响应速度。
6. 统一权限认证服务：抽象出各模块用户登录与鉴权逻辑，独立设计统一认证中心，统一管理登录状态与权限校验，提升系统安全性与可维护性。
7. 数据库性能优化：对高频历史数据查询进行索引与分区表设计，SQL 查询性能提升超过 20%。
8. 低代码组件设计：在 EMS 中构建电气量曲线图模块，结合装饰器模式封装 SQL 逻辑，支持自 动生成前端格式数据，显著提升前后端协作效率。

**AI试题查询系统微调与部署（个人项目）**

**项目描述：**基于大语言模型 Qwen2.5-1.5B 微调并部署的试题问答系统，实现针对题库内容的自然语言查询与回答。

**技术栈：LLaMAFactory、Qwen2.5-1.5B、vLLM、Open WebUI、CUDA、Python**

**核心工作**：

1. 基于 [魔塔社区服务器]（NVIDIA A800）使用 **LLaMAFactory** 对 Qwen2.5-1.5B 模型进行指令微调，构建小规模 AI 试题问答系统；
2. 设计并准备 200 条左右的问答训练数据，控制 loss 收敛至 0.001，模拟真实教学场景；
3. 使用 **vLLM + Open WebUI** 实现模型推理与 Web 页面交互，支持上下文追问和自然语言模糊匹配；
4. 调整 tokenizer、Prompt 模板，优化模型对试题类指令的理解效果。

**自我评价**

Self evaluation

1. **善于沟通，理解能力强**，积极与同事交流，**及时反馈问题**，工作中从未出现过因沟通问题导致开发进
2. 度受阻或项目延期的情况；
3. **热爱科技，拥抱变化**，不局限于现有技术，积极学习 AI 应用开发的相关知识，坚信在 AI 时代，大模型应用才是今后的趋势；
4. **喜欢总结和分享**，对于工作中碰到的问题会整理成文档，共享给同事，提升团队工作效率;技术上的学 习和思考会发到**个人博客**上：<https://blog.csdn.net/xiaoyanboke>

**个人证书**

Self Certificate

1. 英语CET4

2. 网络技术挑战赛A系列国家二等奖