

段沛彤

男 | 21岁 ☎ 13191273507 📩 3276432700@qq.com



工作经历

上海欧冶金诚信息服务股份有限公司 Python/AI应用研发工程师

2025.07-至今

知识管理平台：

设计并实现智能搜索与检索解决方案，覆盖从数据摄取、索引、召回、排序、重排序、结果生成的完整链路。
构建高效的数据管道，支持多类型半结构化数据（会议、消息、文档、视频等）的处理，并实现细粒度权限控制。
研发与迭代搜索相关的机器学习模型（嵌入、语义召回、排序、重排序等），提升搜索相关性与答案质量。
与产品、设计和后端工程师协作，将搜索算法与产品功能深度结合，快速验证并上线搜索功能。
优化搜索系统的性能与稳定性，确保在真实生产环境下的高吞吐量、低延迟与高可用性。
关注用户体验，通过实验、A/B 测试和数据分析持续改进搜索质量与交互体验。

- 端到端落地 RAG/Agent 应用：需求抽象、方案设计、核心模块研发、上线与效果持续优化
- 构建高质量知识与检索体系：文本/表格/多模态数据清洗切片、向量化与索引、混合检索（BM25+Dense/Hybrid/Multi-Vector）、重排与引用溯源
- 设计智能体工作流：工具编排（Function/MCP）、规划/反思/执行、多智能体协同，支撑复杂业务流程自动化
- 打造评测与治理闭环：业务目标驱动评测指标（准确率、成功率、幻觉率、时延/成本）、离线+在线A/B、Badcase 归因与自动化回灌
- AI 工程化与稳定性：vLLM/SGLang/TensorRT-LLM 推理优化、缓存与并发调度、可观测与风控/Guardrails（安全、合规、对齐）
- 平台化沉淀：LangChain/LangGraph/LlamaIndex 等框架最佳实践、Prompt/Tool/Workflow 资产化，形成可复制的行业解决方案

低代码智能体开发：

熟练运用 BISHNEG、DIFY、COZE 实现智能体搭建、功能扩展及性能优化，跟进平台更新并应用新特性。
参与过5个智能体项目开发，深切了解如何通过智能体解决业务问题
集成多种财务相关MCP工具
实现财务部门大部分财务报告自动化，如CRM报告、营销小助手

专业技能

熟练：Python（Django、Flask、FastAPI、Celery、SQLAlchemy等常用组件）、Linux、Docker、Mysql、SQL
掌握：RAG、LangChain、Redis、grpc、网络、数据结构、设计模式
熟悉：Golang（Gin）、CI/CD（GitHub Actions）、PostgreSQL、MongoDB、Vue、向量数据库、算法
了解：Kubernetes、transformers、微服务、分布式、消息队列等原理

熟悉主流的LLM、文生图、图生视频相关技术，如 Stable Diffusion（LoRA / ControlNet / ComfyUI）、DALL-E、Stable Video Diffusion 等；
AI Coding：擅长借助 Cursor、Cloud Code 等 AI 开发工具提升研发效率，积极跟进并应用最新的 AI 辅助编程范式、智能开发技术及工程落地方法。

教育经历

中北大学 本科 软件工程

2021-2025

课程：Python程序设计、计算机网络、数据库原理、Web开发、数据结构与算法、Linux操作系统、自然语言处理
其他：“互联网+”大赛省级一等奖，蓝桥杯三等奖，中北大学创新创业实践和竞赛奖，多次奖学金，英语四级

项目经历

Ami 构思/开发/运营/维护

2025.11-至今

Ami是一个平台级的Character-as-a-Service(CaaS)系统，提供AI角色创建、聊天互动、内容生成等核心功能，支持从个人用户到企业客户的完整服务生态。

Node.js、TypeScript、Comfyui、Prisma、PostgreSQL、Redis、BullMQ、pgvector