

## 周守宇个人简历

syzhou98@foxmail.com 28岁 18701918314



## 教育经历

上海师范大学	地图学与地理信息 (GPA:3.3/4.0) 硕士	2023.09-2026.06
● 专业课程:	计算机系统、数据结构与算法、计算机体系结构、数据库系统、编译原理、计算机网络、操作系统	
● 研究成果		
➢	Ye, Q.,Zhang, X,Zeng, G., Cao,Z., & Zhou,S.(2025). Thespatial distribution and synergistic effect of differentinnovation activities in Chinese cities: An analysis based on technology,design, and market activities.Applied Geography, 176, 103527.	
➢	Yu, C, Xu, D.,Peng, Y., Zhou,S.,& Lu, D.(2025).Analysis ofland surfacetemperature drivers and marginaleffect on the Yangtze River Delta urban agglomeration. Ecological Indicators, 178, 114123.	
➢	参与撰写学术专著。徐迪, <<健康城市:全球视野下的策略与挑战>>, 上海科学技术出版社, 2024.6.	
● 奖项荣誉:	2025全国大学生英语翻译大赛省级一等奖、Kagle, 维苏威火山挑战-表面检测, 银牌	
沈阳航空航天大学	新能源科学与工程 本科	2016.09-2020.06
● 工商管理第二学历		

## 实习经历

科大讯飞有限公司	大模型应用开发工程师	2025.07 - 2025.11
● AI预处理:	针对招投标领域专业文本, 设计多维度数据清洗与去重规则, 构建专业领域数据预处理流程, 剔除无效冗余信息, 将数据纯净度提升, 同时把检索噪声降低18%	
● 向量构建:	结合FAISS向量库搭建私有化向量存储体系, 完成文本分块、Qwen大模型Embedding向量生成、索引结构调优全流程, 将向量检索速度提升22%, 百万级文本召回耗时压缩至0.8秒以内	
● RAG架构:	结合LangChain框架搭建RAG核心架构, 优化检索排序、上下文增强与大模型生成逻辑, 实现问答结果精准可溯源, 系统问答准确率提升42%, 业务信息查询效率明显提升30%	
● Agent闭环:	集成Agent基础决策模块, 搭建任务拆解与工具调用链路, 实现专业问题自主检索、推理、应答闭环, 完善智能问答系统自主处理能力	
● 工程落地:	完成系统全流程调试与部署优化, 解决模型推理卡顿、向量匹配偏差等问题, 保障系统稳定运行, 协助招投标业务人员高效完成专业信息查询与资料梳理	
博世	大模型开发实习生	2024.12-2025.03
● 文档解析:	针对多格式车辆使用手册, 设计层级解析策略, 保留章节结构并构建向量数据库, 为Agent问答提供规范数据支撑	
● 框架搭建:	设计集成 RAG 能力的智能体核心架构, 集成混合检索与LLM推理, 支持自然语言查询复杂问题, 返回带溯源引用可靠答案	
● 功能迭代:	实现多轮对话与思考链展示功能, 增强交互可解释性, 支持用户反馈功能, 为模型RLHF优化提供有效数据	
● 技能深耕:	深入掌握Agent技术与LangChain框架, 熟练运用提示工程、RAG及RLHF训练原理, 提升大模型交互体验与效率	

## 工作经历

新乡航空工业集团有限公司103厂	试验工程师 (试验中心)	2020.07-2023.08
● 研发试验:	主导航空重点产品科研及例行全流程试验验证, 牵头完成某机型配套产品全生命周期试验管控, 全程参与试验方案设计、执行与结果研判, 完善产品研发验证基础, 保障产品研发合规落地	
● 方案优化:	负责重点产品竞标试验板块全流程工作, 结合研发思维优化试验流程与测试方案, 实现试验综合成本直降20%, 助力项目成功中标, 为公司创造千万级年度经济效益	
● 试验优化:	主导编制试验标准、工艺规范及检测报告, 牵头制定《试验标准》, 规范覆盖15+条产品线, 统一试验研发与检测准则, 大幅提升试验精准度, 有效降低试验问题发生率, 整体试验返工率减少20%, 大幅提升研发试验效率, 缩短产品研发验证周期	
● 技术研发:	结合试验数据深度分析, 反向赋能产品研发设计优化, 参与产品技术整改与性能升级, 持续提升产品可靠性	

## 项目经历

C++自定义HttpServer框架及AI服务平台	项目负责人	2025.06-2025.09
● 框架开发:	基于C++实现自定义HttpServer框架, 集成HTTPS、动态路由等核心功能, 通过多级缓存与连接池优化性能, 预留扩展接口, 支撑AI服务稳定运行	
● 服务落地:	搭建Reactor模型HTTP服务器, 设计异步推理架构, 通过消息队列解耦AI请求, 实现图像识别等功能, 支持会话管理与轻量化工具调用扩展	
大语言模型推理框架	项目负责人	2025.10-2026.02
● 框架构建:	利用C++开发可扩展LLM推理框架, 支持主流大模型, 优化KV-Cache等核心技术, CPU/GPU双后端推理吞吐量提升37%	
● 性能优化:	定制CUDA算子并精细化调优, 将目标模型推理速度优化至112Token/s, 结合int8量化技术实现高效模型压缩, 兼顾精度与效率	

## 个人技能

● 语言能力:	大学英语四级 (CET-4)、大学英语六级 (CET-6), 英语听说读写能力良好	
➢	熟练掌握Office工具 (Word/Excel/PPT)、Excel (函数+数据透视表)	
➢	熟悉C/C++、Python,C++封装继承多态, STL常用容器, C++11常用特性(智能指针等)	
➢	熟悉Socket网络编程, 了解并发编程、I/O多路复用原理及使用, 以及MySQL Redis常用基础组件	
➢	熟悉主流大模型架构(Qwen、Llama等)及Transformer推理流程, 具备模型结构分析与性能调优能力	
➢	掌握PyTorch框架, 能编写CUDA算子, 熟悉FlashAttention、RMSNorm等关键kernel优化方法	
➢	熟悉Prompt工程和优化技巧, 能基于Milvus数据库构建RAG知识库, 实现向量检索和查询增强的完整流程	
➢	能利用Cursor/Claude Code/Github Copilot等AI IDE开发全栈项目, 掌握MCP和Agent Skills的开发	
➢	熟练Vibe Coding、Harness Engineering、SSD等多种AI编程范式, 能利用Speckit、OpenSpec等工具完成开发	
➢	了解Agent系统开发, 熟悉LangChain等框架, 能够设计适用于高可用、高扩展性的大模型/Agent工程架构	
➢	具备商业化技能, 可以借助Supabase、Vercel、腾讯云开发、Creem、ShipAny和Stitch等完成低成本的项目开发	

## 个人优势

➢	大模型开发: 熟悉大模型全栈开发与AI应用落地, 熟悉RAG、Agent与LangChain, 可独立完成从数据预处理、向量库构建到模型推理、功能迭代的全流程开发, 熟练掌握Prompt工程、RLHF训练原理与多模态交互优化, 能胜任业务需求设计高可用AI解决方案, 具备扎实的技术功底与工程化落地能力	
➢	AI应用: 具备AI项目全生命周期把控能力, 可完成需求拆解、方案设计、架构搭建与部署优化全流程, 擅长将复杂业务逻辑转化为技术实现, 能高效解决模型推理、系统部署中的各类技术难题, 保障AI系统稳定运行与性能迭代	
➢	个人能力: 能快速掌握不同场景的AI应用需求, 持续追踪大模型与Agent技术前沿, 以创新思维优化技术方案, 具备强学习力、执行力与问题解决能力, 能为AI项目落地提供技术支撑	