

杨一赫

LlamaIndex-Core Contributor 37K stars · LazyLLM collaborator 1k stars

✉ yangchichi@126.com · ☎ (+86) 178-642-60105 · in 主页

🎓 教育背景

华东师范大学, 上海	2025 – 至今
博士研究生拟录取 软件工程。	
北京中关村学院, 北京	2025 – 至今
博士研究生拟录取 传统 AI。	
浙江科技大学, 浙江, 杭州	2022 – 2025
硕士研究生 应用统计, 大数据 (NLP 方向), 2025 年 1 月毕业。	
青岛理工大学, 山东, 青岛	2017 – 2021
学士 经济学, 数据分析与数理统计。	

👤 职业经历

华东师范大学, 无尽前沿实验室	2024 年 12 月 – 至今
研究助理 工作方向: 多媒体 AI 代理、智能量化交易系统开发	

MediaAgent 项目 (2024 年 12 月 – 至今)

- 开发 MediaAgent 系统, 目标实现借助多媒体智能体接管多媒体 (小红书) 账号的完全自主运营。
- 实现自主评论、自主收发信息、代理互动等核心功能, 构建完整的社交媒体 AI 代理生态。
- 通过自然语言处理和多模态理解技术, 实现智能内容生成和用户交互。

btcbot 智能量化机器人 (2024 年 12 月 – 至今)

- 构建自动化比特币交易机器人框架, 实现基于 5 分钟 K 线数据的实时分析和交易决策系统。
- 采用 Prefect 企业级任务驱动架构, 集成 XGBoost、LSTM 等多模型预测算法, 支持自定义交易策略。
- 实现完整的量化交易流程, 包括数据获取、特征工程、模型预测、风险控制和自动执行。

商汤科技, 大装置/训练平台与服务部/AI 专家服务研发组	2023 年 8 月 – 2024 年 8 月
AI 系统开发工程师 工作方向: RAG 标品、LLM 应用编排框架开发、复杂文档解析、信息抽取	

阶段一：RAG 系统开发 (2023 年 8 月 – 2023 年 11 月)

- 开发复杂文档解析器, 基于 LLamaIndex, 实现富文本文档解析。通过模态对齐策略, 构建多模态 RAG 问答系统, 支持对文档图表进行细粒度多轮问答。
- 初期搭建评估框架, 调研并测试了 20+ 算法, 最终收敛到 BM25+Embedding 融合检索策略, 加 Reranker 后处理和 Refine 参考精炼策略, 将检索准确度从 67% 提高到 90% 以上, 满足交付算法精度要求。

阶段二：模型训练与数据处理 (2023 年 12 月 – 2024 年 4 月)

- 完成 170 万篇 Arxiv 论文的数据抓取与目录结构提取, 创建了长文档结构化解析数据集。利用 Rope+Shift-attn 技术训练了 1B 级别的模型, 转化为生成式文档结构化解析器, LLM 特化为语义的长文档结构化解析器。
- 基于给定文档生成蒸馏数据集, 并通过单轮数据增强, 训练 bge-large-zh-1.5 模型, 提高 RAG 系统 10% 的检索准确度。

阶段三：AI 应用编排框架开发 (2024 年 5 月 – 至今)

- 从 0 到 1 参与 AI 应用编排框架 LazyLLM 研发, 负责 LazyLLM Core、LazyLLM Engine 的设计与开发。LazyLLM 创意场景研发。

🔧 IT 技能

编程语言: Python, R, TypeScript 常用开发工具: Docker、Gitlab、Fastapi、Flask、Pytest、Gradio、Streamlit、Redis、Neo4j、MySQL、PDM, React、TailwindCss、Next.js、cookiecutter、MongoDB、Marvin、Prefect、LiteLLM、CrewAI Langchain、Langgraph、Agent-Zero。

常用算法库: Pytorch、Transformers、Pandas、Numpy、Accelerate, Deepspeed

掌握的开源框架:

- [LlamaIndex](#): 熟悉其 Core 中 API, 可进行高度定制化开发。
- [Langchain](#): 熟悉低阶 API, 可进行定制化开发。
- [LLama-Factory](#): 熟悉其模块化设计, 能够快速训练和部署模型。
- [vllm](#): 熟悉其推理部分, 能够用其为特定模型定制 Function Call 能力。

其他技能: 爬虫、selenium 自动化、seaborn、pyechart, Tableau 制图, Python 中级开发

维护的开源项目:

- [LazyLLM](#): 所有框架和模型以同一种方式组织, 训练和推理的统一。
- [llamaindex](#): 大模型的数据框架。
- [sms server](#): 为 SmsForwarder 提供 webhook 通道支持的开源服务端
- [HTIA-mobile-predict](#): 行为轨迹预测算法
- [Xproxy](#): 短效代理 IP 管理框架 [FlowX](#): 重构 langchain [LazyNote](#): Docstrings Manager

📄 其他

论文:

- Mini-batch HTIA-based Seq2Seq Networks –Information Science[中科院 1 区]
- Semantic Alignment in RAG Systems through Generative Semantic Integration –[中科院 3 区在投]

竞赛: 全国大学生数学建模竞赛省一等奖

GitHub: [y1h Github](#)

证书: [ACCA](#)(英国注册会计师中级)、国际人才英语考试初级 (ETIC)