

余非凡

大模型算法工程师 / AI Systems Engineer

个人概述

大模型算法工程师，关注 RAG、Agent、工具调用、评测体系和可上线的 AI 应用系统。擅长把模型能力拆解为数据、流程、权限、工具和工程边界中的可维护系统能力。

技术能力

AI Systems

RAG / Agent Runtime / Tool Calling / Evaluation / Prompt Engineering

Engineering

Python / FastAPI / Java / PostgreSQL / Redis / Playwright

Workflow

MCP / LangGraph / Human-in-loop / Observability / Document Parsing

工作经历

2025 - NOW

Large Language Model Algorithm Engineer

Current / Confidential

- 把自然语言入口连接到可审批、可观测的业务流程。
- 建设面向复杂文档的解析、检索、重排与评测链路。
- 沉淀 Agent Runtime、工具注册、执行日志与人工审核模式。

2024 - 2025

AI Application Engineer

Project / Product Practice

- 围绕网页任务拆解观察、规划、执行和验证流程。
- 关注推理延迟、页面状态、失败归因和评测闭环。

精选项目

Enterprise OA Agent / AI Browser / Intelligent Bidding System / Reusable Agent Runtime / RAG Service

技术写作

RAG 分块中的表格、代码块与图片保护 / Agent Runtime 不只是循环调用模型

教育背景

计算机科学 / AI 工程方向

链接

[Website](#) [GitHub](#) [Email](#)