

刘义轩

Go 后端开发 / AI 应用开发 (2027 届)
成都 • yimolieu@163.com • github.com/baka-world

个人优势

参与企业级 RAG 平台二次开发，具备 Go/Python 后端、文档解析、检索与重排可靠性及容器化交付经验；独立主导个人项目的需求、架构、函数契约、代码审查与测试。

向 GitHub 9.5k+ Star 的开源模型推理平台 XInference 提交的 6 个 PR 已合入 main；在线上实战中定位并修复 FileBrowser 大批量 PDF 同步线上 OOM。

教育经历

成都大学

计算机学院 · 计算机科学与技术 本科

◦ CET-4: 515; 核心课程: 数据结构、操作系统、计算机网络、数据库系统、软件工程

成都

9 月 2023 - 6 月 2027

开源贡献

XInference (开源贡献 · 9.5k+ ★)

8 月 2026

向开源模型推理平台 XInference 提交的 6 个 PR 已合入 main，覆盖 Worker 重连、跨进程 RPC 参数规范化、metrics 启动同步与 OpenAI API 兼容性。

- 将 metrics exporter 启动从 busy-wait 改为带超时的阻塞队列读取，覆盖地址发布、线程提前退出与超时回归；在 Supervisor/Worker 边界规范化 n_worker，避免 JSON/RPC 字符串参数触发 TypeError
- 为 Qwen3.8 将顶层 reasoning_effort 映射到 chat template 参数，补齐冲突与非法值的 HTTP 400 前置校验；标准化 tool-call 参数格式以兼容多种开源模型模板

工作经历

开发实习生 (研发部)

智范科技 (成都) 有限公司

成都

7 月 2026 - 至今

- 定位并修复 FileBrowser 大目录 PDF 同步线上 OOM (服务进程常驻物理内存 (RSS) 曾约 107 GiB，内存与 Swap 耗尽)：将整批文件内容聚合后入库改为逐文件下载、入库与 checkpoint 的流式同步，并限制诊断明细数量；同一问题数据源 10,774 项受控回归中，App 峰值约 3.0 GiB，无 OOM/重启
- 修正 MinerU PDF 准入仅统计外部提交、处理与结果获取阶段，解除本地分块/向量索引对 PDF 队列的阻塞；为远程 Reranker 引入共享熔断、15 秒独立超时与单次 64 文档硬上限

项目经历

ITI (个人项目)

8 月 2026

AI 驱动的 IT 资讯聚合工具：主导需求拆解、模块架构、函数契约、代码审查与测试 (AI 辅助具体实现)，使用 Go 完成多源资讯抓取、SQLite 去重、AI 分类摘要、Rerank 精选与日报/周报推送。

- 独立完成函数级设计文档、9 包模块化架构、94 次 Git 提交与 11 个测试文件的完整迭代，以 TDD 方式推进核心功能
- 实现 RSS/GitHub/LKML 多源抓取与 SQLite 去重入库；OpenAI 兼容客户端采用 SSE 流式与块间空闲超时；dashscope/jina reranker 可插拔精选，失败时按时间排序兜底
- 关键工程点: errgroup 并发单篇 AI 调用 (SetLimit 限流)、超时分层 (120s / 60s / 30s)、预算双键防输出截断、纯 Go SQLite 免交叉编译

专业技能

编程语言: Go (主), Python (辅助)

Go 后端: Gin, GORM, HTTP/REST, SSE, Context, 超时/重试/熔断

AI 应用: RAG 文档解析与检索链路, Embedding/Rerank, Tool Calling, OpenAI-compatible API, 向量数据库

数据与工程化: MySQL, PostgreSQL, Redis, SQLite, Docker/Docker Compose, Linux, Git, Prometheus/Grafana