

熊浩毅

电话: 18779165101 | 邮箱: se_hyxiong@163.com | Github: [Bin-hy](https://github.com/Bin-hy)

求职意向: 后端开发实习生



教育背景

西安电子科技大学 (推免) | 通信工程

2025.09 — 2028.06

江西财经大学 | 软件工程

2021.09 — 2025.06

- 主修课程: Node.js 服务端项目开发(97)、JavaEE 程序设计(100)、Web3.0 大前端基础与实践(100)、Web 前端框架应用与开发 (100)、C++程序设计(95)、高等数学(93)、服务器程序设计(96)
- 在校荣誉: 国家奖学金 (0.1%) 优秀学生一等奖学金 (2%)、国家励志奖学金 (连续两年); 2024 第十七届中国大学生计算机设计大赛 web 开发赛道 **国家级三等奖**; 第五届传智杯全国大学生 IT 技能大赛 **程序设计赛道 B 组国家区域赛二等奖**等

技术技能

- 编程语言: Go、TypeScript / JavaScript、Rust、Python。
- 后端框架: Gin、Express 5、axum
- 数据库: Mysql、PostgreSQL、Redis
- 消息队列/服务发现: Kafka、BullMQ、etcd
- AI/LLM: AnthropicAPI、OpenAI API、阿里云 DashScope (Qwen VL)、RAG 检索增强生成、MCP 协议
- 工具/设施基础: Docker/Docker Compose、gRPC + Protobuf、FFmpeg、MinIO/阿里云 OSS、CI/CD

实习经历

鲨堂 (陕西) 文化传媒有限公司——全栈工程师

2026.04 — 2026.7

- 从 0 到 1 负责电商 AI 视频生成平台 (shatangAI) 的全栈研发, 产品已上线: <https://shatang.top>
- 1. **多源 AI 视频编排**: 集成阿里云 DashScope、豆包 Seedance 等 6+ 个视频服务供应商, 统一 BullMQ 任务队列调度, 支持多段视频一致性合并、分镜合并、首尾帧桥接, 覆盖商品展示、翻拍改编等 10+ 种生成方式
- 2. **任务全生命周期管理**: 设计完整状态机 (pending->processing->done/failed), 实现崩溃恢复 (启动自动扫描异常任务重新入队)、积分冻结日志防重复扣减、引用计数保护
- 3. **全架构落地**: 前端 React 19 + Vite + Zustand, 后端 Express5 + Mysql, 管理后台独立 React 项目 (Recharts 数据可视化), TypeScript 全栈统一类型安全
- 4. **生产级基础设施**: 多环境配置 (本地磁盘 vs 阿里云 OSS)、Docker Compose 一键部署、微信登录+微信支付接入

深圳草莓创新科技有限公司——大模型Rag全栈工程师

2025.01 — 2025.05

- 负责无人机平台的图片 Rag 检索增强的后端与界面开发, 参与需求到上线的完整研发流程。
- 1. 无人机 Rag 检索平台搭建: 开发图片上传管理、图片和元信息处理与召回等策略构建
- 2. 提示词工程: 针对图片检索与问答场景持续优化 Rag Prompt, 提升检索结果相关性与问答准确性。
- 3. 工程实践: 搭建 CI/CD 流水线, 定期 Code Review, 负责测试和 Docker 部署上线

项目经历

一、EasyCoding —— 终端 AI 编程助手 (负责人)

2026.07 — 2026.08

- 从零手写的终端 AI Agent, 类 Claude Code, 全 Go 实现, 支持 Anthropic/OpenAI 双协议、MCP 工具生态、五层权限防御、流式 TUI 交互。源码: <https://github.com/Bin-hy/EasyCoding>
- 关键贡献与技术点:
- 1. **多协议 LLM 引擎**: 协议无关的 Provider 抽象层, 支持 Anthropic Messages API (含扩展思考/Prompt Cache) + OpenAI Chat Completions + 任意兼容端点; 一份配置自由切换, 上层代码对协议无感
- 2. **多轮 ReAct Agent 闭环**: 完整思考→行动→观察→思考循环 (10 轮迭代上限); 只读工具同轮并发执行、副作用工具串行保序; 连续幻觉检测自动停止
- 3. **五层权限防御体系**: ① 危险命令黑名单 → ② 文件操作沙箱 → ③ YAML 规则引擎三层优先级匹配 → ④ 四档模式兜底 (Default/AcceptEdits/Plan/Bypass) → ⑤ 人在回路弹窗, TUI 实时模式切换无需重启

- 4. MCP 协议零配置接入: stdio + HTTP 双传输支持, 配置驱动自动发现, 命名空间隔离 (mcp<server><tool>), 单服务失败隔离无感适配

- 5. 会话持久化与恢复: JSONL 追加写实时 fsync; /resume 浏览历史会话, Token 超限自动压缩; 30 天过期会话后台清理

二、Vistack 视频播放与直播系统 (负责人)

2025.6 —— 2025.10

- 基于云原生架构的流媒体系统, 三角色微服务拆分, 支持实时直播 (Live Streaming) 与视频点播 (VOD)。技术栈: Gin + GORM + PostgreSQL + Redis + gRPC + Kafka + etcd + FFmpeg + MinIO。源码: <https://github.com/Bin-hy/Vistack>
- 关键贡献与技术点:
 - 1. 三角色微服务架构: 将单体拆分为 API (HTTP 服务)、Worker (任务编排)、Transcoder (远程转码) 三个独立进程, 支持独立部署、独立扩容, API 与 Worker 分离解耦。
 - 2. 自适应码率播放 (DASH ABR): FFmpeg 按源分辨率智能选择 240p - 4K 档位 (manifest.mpd + m4s), 前端 dash.js 无缝切换清晰度; MinIO Multipart 分片上传 + MD5 哈希秒传+ 断点续传
 - 3. 远程转码隔离: FFmpeg 独立容器, 通过 gRPC ProcessVideo RPC 调用, 输入/输出均走 MinIO 对象存储, 无状态设计支持并发转码水平扩容
 - 4. 防盗链鉴权: 预签名 URL + STS 临时凭证 + JWT 权限控制

三、BinRag —— 企业级文档知识库 RAG 系统 (负责人)

2026.7 —— 2026.8

- 基于检索增强生成的文档知识库系统, 支持多格式文档上传、混合检索 + LLM 问答, Web / 桌面双形态部署, 内置 OIDC 登录与 MCP 开放接口。源码: <https://github.com/Bin-hy/documentsRag>
- 关键贡献与技术点:
 - 1. 混合检索与精排: 向量语义 + BM25 关键词检索, RRF 融合加权; Cross-encoder 重排序提升召回质量
 - 2. 完整文档处理管道: 多格式加载器 (PDF/DOCX/Excel/HTML) → 三种分块策略 → 异步 Worker 池入库, 失败自动重试
 - 3. LLM 问答链路: Query 改写 → 上下文按 Token 预算组装 → SSE 流式生成回答与引用来源
 - 4. MCP Server 接口: 基于 MCP 协议开放 6 个只读 Tool (问答、检索、知识库管理), 支持外部 Agent 调用