张云帆

现居：瑞典哥德堡 · 邮箱：2529347470@qq.com · 微信：rushac2
GitHub：github.com/44sunsetsf · 个人主页：yunfanteo.world · 求职意向：AI 应用开发 / 后端开发工程师

教育背景

查尔姆斯理工大学（瑞典，QS 2025 #139） — 计算机系统与网络安全 · 硕士 **2025.09 – 2027.06**
获查尔姆斯 IPOET 奖学金（减免 75% 学费）

西北工业大学（985） — 网络空间安全学院 · 信息安全 · 本科 **2020.09 – 2024.06**
GPA 86.4/100，专业排名 10/94（前 11%）；二等奖学金暨优秀大学生；CET-6 510

实习经历

华为技术有限公司 — 终端 BG 软件部 · 软件开发工程师（优秀实习生） **2026.06 – 2026.09**

- 设计 PerfAgent 性能分析智能体的 MCP Server + 声明式 Skill 系统，解决 Trace 数据超出大模型上下文的问题；10 个 MCP Tool 支撑指标查询与工作流编排，免编码完成根因分析，内部用户从 20 人增至 130+ 人。
  技术栈：Python、FastMCP、SQLite、YAML
  - 设计 Skill 执行引擎，参数由框架注入、结论必须通过证据闸门，从机制上降低 Agent 幻觉；上下文以文件形式传递，支撑 SubAgent 并行分析，17 个 Skill 均可断点续跑。
  - 实现统一查询引擎，参数按 5 层优先级合并，降采样时保留尖峰，组合指标按拓扑序计算，覆盖 80+ 原子指标。
  - 设计装饰器零侵入拦截 MCP 调用，识别四类 Agent 失效；构建 27 个黄金样本回归，防止指标静默劣化。
  - 搭建结构化维测，用 session_id 串起 JSON Lines 全链路日志，三级查找可从 Skill 耗时定位到单个算子。
- 定位冰海架构看板首页慢查询，补建复合索引、改写慢 SQL，首页汇总接口耗时由 11.7s 降至 0.93s。
- 实现 OpenHarmony 后台长时任务的通知类型切换（C++），按通知形态分三档用状态机路由，9 种迁移均有单测覆盖。

项目经历

VParser 智能视频解析系统 **2026.03 – 2026.05**
技术栈：*Go*、*Pydantic*、*gRPC*、*MySQL*、*Redis*、*Kafka*、*MinIO*、*Vue* 在线演示：*yunfanteo.world/vparser*
面向长视频内容理解的解析平台，集成用户鉴权、大文件稳定接入、多模态内容解析与 AI 总结问答，已上线。

- 引入 Kafka 分级重试与死信队列，将长耗时任务异步化，提交接口 RT 由 60s+ 降至 50ms 内。
- 设计分布式锁 + MD5 指纹的内容级去重，保证消费幂等，避免相同视频重复解析浪费处理资源。
- 实现分片上传与断点续传，用 Redis 记录分片状态，保障 GB 级视频在弱网环境下稳定上传。
- 实现 Redis 令牌桶限流，分用户级与全局限速，遏制恶意请求带来的高昂 Token 开销。
- 设计 Planner-Executor-Critic 受控 Agent，拆解目标、生成带时间戳结论，代码逐条核验证据，不符则定向补检索重写。

GoEuroOps 留学业务智能运营中枢 **2025.05 – 2025.07**
技术栈：*Python*、*FastAPI*、*Redis*、*ChromaDB*、*Docker*、*GitHub Actions* 在线演示：*yunfanteo.world/goeuroops*
面向留学咨询的多 Agent 客服系统，集成意图路由、知识库问答、服务报价与线索转人工，已上线。

- 设计多 Agent 编排，以角色契约和工具白名单划定各 Agent 职责，经两级路由分派主/协作 Agent，紧急诉求转人工。
- 融合 LLM Few-shot、Embedding 与规则三路信号做意图识别，380 条自建评测准确率 91.7%、宏平均 F1 0.915。
- 设计意图门控的 LLM Wiki 知识库，与意图识别并行预取整页知识，回答完全正确率比向量 RAG 高 7 个百分点。
- 构建统一工具网关，为工具调用统一加 Pydantic 校验、超时、熔断与降级，金额由确定性工具计算，避免模型编造价格。
- 构建 Agentic Memory，参考 A-Mem 生成向量笔记并补全事件时间与链接，结合客户档案与咨询记录实现跨会话服务。
- 设计动态 Skills 注入机制，让不同 Agent 使用各自的业务处理规范，降低越权回复和话术漂移风险。
- 构建 LLM-as-Judge 与 τ-bench 式多轮仿真评测，以 pass^k 衡量稳定性，结合 Trace 与回归检测定位问题。

专业技能

编程语言	熟练掌握 Python、Go，熟悉 C++、SQL、JavaScript
后端与中间件	FastAPI、Gin、GORM、gRPC；MySQL、Redis、Kafka、MinIO；熟悉异步、幂等、限流等设计
Agent / 大模型	MCP Server、Skill 系统设计、多 Agent 编排、Agentic RAG、LLM Wiki、记忆机制、LLM-as-Judge
工程工具	Git、GitHub Actions（CI/CD）、Docker、Linux；熟练使用 Claude Code、OpenCode 等 AI 编程工具