

学术智库报价单

项目名称: 学术智库

致: Xman

联系人: Xman 项目负责人

报价方: 成都追耀信息科技有限公司

项目顾问: 张富意

报价日期: 2025年11月24日

报价版本: V1.0

尊敬的 Xman 项目负责人:

您好!

非常感谢您对成都追耀信息科技有限公司的信任，并选择我们为您规划「学术智库」。

在与您深入沟通并仔细研读了项目功能需求清单后，我们对项目的商业愿景、核心业务流程以及期望达成的市场目标已有了全面且深刻的理解。我们非常荣幸能将我们的技术洞察与您的商业蓝图相结合，为您呈报以下项目开发方案与报价。

一、项目理解与目标

本项目旨在构建一套私有化部署的高精尖「学术智库」AI科研辅助平台。我们了解到，贵方目前的痛点在于通用平台（如Dify）无法满足严格的数据物理隔离需求，且在处理学术文献时，对复杂的图表、曲线理解能力不足。本项目的核心目标是打造一个**数据绝对安全、支持本地大模型**

(Qwen/DeepSeek) 推理、具备深度多模态解析能力 的专属系统。重点在于实现Google/Bing学术搜索的聚合、解决PDF文献中图表数据的精准提取（VLM推理），并确保多租户环境下的会话记录与知识库互不可见，从而为科研人员提供一个高效、精准且私密的学术对话环境。

二、核心功能范围

根据您的需求，本项目将包含以下核心功能模块：

1. 用户端 - 独立鉴权与会话系统：

- 独立账号体系：构建物理隔离的用户数据库，彻底解决Dify记录互看问题，确保Token与数据安全。
- AI对话主页：采用类“知乎直达”的清爽交互界面，支持流式输出、多轮追问及上下文记忆。
- 历史记录管理：实现个人私有会话列表，支持单条删除与管理，严格限制仅当前登录用户可见。

2. 用户端 - 学术增强功能：

- 联网学术搜索：自动聚合Google Scholar与Bing Scholar接口，清洗搜索结果，生成带可信引用的专业回答。
- 多模态文件解析：支持PDF及图片上传，重点开发针对学术文献中复杂图表、曲线的视觉理解能力，将非结构化数据转化为结构化洞察。

3. 后台系统 - 本地模型与推理引擎：

- 本地大模型服务：在4卡GPU环境下部署Qwen或DeepSeek模型（适配36B/72B版本），封装OpenAI兼容格式的推理接口，管理显存资源与推理队列。
- RAG知识引擎：开发高精度PDF OCR解析算法，对文档进行切片、清洗并存入向量数据库。
- 多模态图表推理管道：集成视觉模型（VLM），专门处理文献截图与图表数据，解决传统OCR无法理解曲线趋势的难点。

4. 后台系统 - 权限与代理服务：

- 搜索代理服务：整合第三方学术API，提供稳定的代理访问服务。
- 多租户权限管理：管理员授权机制，确保不同用户的Knowledge Base（知识库）与对话数据在逻辑与物理层面实现隔离。

三、项目开发费用明细

项目开发总价：60,000.00 元

职位/角色	核心工作内容	费用(元)	工时预估(人/天)
UI/UX 设计	学术风格界面设计，对话流式交互视觉设计，多模态上传组件设计，高保真原型图绘制。	6,000	6
前端开发	用户端界面开发（React/Vue），流式 Markdown 渲染，文件上传与预览逻辑，图表数据可视化展示。	12,000	12
后端开发	(核心) 本地大模型 (Qwen/DeepSeek) 部署与调优，RAG 向量检索系统搭建，多模态图表解析管道开发，学术搜索 API 聚合，多租户隔离架构实现。	34,000	34
测试 & 项目管理	AI 回答准确性测试，并发推理性能测试，图表识别率验证，项目进度与风险管控。	8,000	8
合计	-	60,000	60

四、项目周期、技术栈与服务承诺

- 预估周期:** 2.5 个月 (约 75 个自然日)。
- 建议技术栈:** 前端 React/Vue，后端 Python (FastAPI/LangChain) + Java (业务逻辑)，向量数据库 Milvus/PgVector，大模型推理 vLLM/Ollama。
- 源码交付:** 结清全部款项后的 3 个工作日内，我们将交付项目的完整源码及相关技术文档。
- 免费质保:** 系统上线后提供 3 个月 免费质保（包含系统 bug 修复、小型功能优化调整）。

- 系统培训:** 提供 2 次 免费的后台系统操作培训。
- 模型部署支持:** 提供本地服务器环境的大模型部署调试服务，协助配置显卡驱动与推理加速框架。

五、付款方式

为保障项目顺利推进，我们遵循以下三阶段付款方式：

- 启动阶段:** 签订合同后，支付项目总价的 **40%** (即 **24,000.00** 元) 作为项目启动金。
- UI交付阶段:** UI设计稿交付并由您确认后，支付项目总价的 **30%** (即 **18,000.00** 元)。
- 上线阶段:** 项目部署上线并完成测试后，支付剩余 **30%** 尾款 (即 **18,000.00** 元)。

六、报价有效期

本报价方案有效期为 **30天**，自报价日期起算。

七、第三方平台与服务费用明细 (非包含于开发费)

请注意，以上报价为本项目 **软件开发服务费**。平台的长期稳定运营还需依赖一些第三方的基础云服务。鉴于本项目涉及**本地大模型部署**，您可能需要自备高性能GPU服务器或购买云端GPU实例。以下为基础服务及常规云资源的预估费用：

服务项目	价格 (人民币)	说明	官方参考链接
GPU云服务器	需按需评估	特别提示: 运行72B模型通常需要A100/A800或多卡4090配置，成本较高（数千元/月起），建议利用闲置硬件或租赁算力。	阿里云GPU服务器
云服务器ECS (常规)	¥99/年(2核2G)	仅用于部署Web服务与业务逻辑层（不含模型推理），首年新用户价极具优势。	阿里云优惠活动页

域名	¥39 / 年 起	部分后缀或特定活动的首年价格，用于系统访问。	阿里云域名注册
SSL证书	免费版	主流云服务商均提供免费DV型证书，足以满足项目加密传输需求。	阿里云SSL证书
学术API (Google/Bing)	按量付费	调用第三方搜索接口可能产生的费用，通常有免费额度或按次计费。	SerpApi等聚合服务
对象存储OSS	¥40 - ¥100 / 年	用于存储上传的PDF文献、图片等原始文件。	阿里云对象存储

我们坚信，凭借我们专业的技术实力、透明的合作流程以及对您业务的深度理解，能够高效、高质地完成「学术智库」的开发工作，成为您最值得信赖的技术合作伙伴。

再次感谢您的信任，期待与您的进一步沟通与合作！

顺颂商祺！

成都追耀信息科技有限公司

资深项目顾问 & 解决方案架构师

张富意

联系电话: 15281032034

电子邮箱: 3382597124@qq.com

公司地址: 四川省成都市成华区多弗国际中心写字楼1607