

软件需求规格说明书

DevPulse 智能问答助手

版本 V1.0

表 1 文档基本信息

项目	内容
软件名称	DevPulse 智能问答助手
文档名称	软件需求规格说明书
文档版本	V1.0
编写单位	福州大学 2601_FZU_SE 《软件工程实践》课程 DevPulse 团队
编写日期	2026 年 10 月 14 日
适用阶段	需求分析与界面原型设计阶段
文档状态	定稿

表 2 文档修订历史

版本	日期	修订说明	修订人
V0.1	2026-10-10	建立文档框架，完成引言与总体描述	郑宇凌
V0.2	2026-10-11	补充用户场景、类图与数据需求	郑宇凌、郑栋兴
V0.3	2026-10-12	补充界面原型、功能描述与验收标准	郑栋兴、陈仕杰
V1.0	2026-10-14	统一术语与排版，完成全文校对	张幼

目 录

- 1 引言 1
 - 1.1 编写目的 1
 - 1.2 项目背景 1
 - 1.3 术语、缩写与定义 2
 - 1.4 参考资料 3
- 2 总体描述 4
 - 2.1 产品目标 4
 - 2.1.1 开发意图 4
 - 2.1.2 应用目标与作用范围 4
 - 2.1.3 产品前景 5
 - 2.2 用户特征 5
 - 2.3 用户场景与典型故事 6
 - 2.3.1 场景一：新成员了解本地运行方式（SC-01） 6
 - 2.3.2 场景二：开发成员了解历史设计决策（SC-02） 6
 - 2.3.3 场景三：课程助教答复重复提问（SC-03） 7
 - 2.3.4 场景四：管理员处理知识缺口（SC-04） 8
 - 2.4 运行环境与约束 8
 - 2.5 假设与依赖 9
 - 2.6 非目标 9
- 3 具体要求 10
 - 3.1 类图 10
 - 3.2 数据需求 13
 - 3.2.1 实体属性与约束 13
 - 3.2.2 数据保留与清理 14
 - 3.3 接口需求 15
 - 3.3.1 用户界面 15
 - 3.3.2 软件接口 15
 - 3.3.3 通信接口 16
 - 3.4 非功能需求 16
 - 3.4.1 可用性 16

3.4.2 安全性.....	17
3.4.3 性能.....	17
3.4.4 可维护性与可移植性.....	18
4 界面原型.....	18
4.1 信息架构与导航	18
4.2 页面清单	19
4.3 关键交互流程	25
4.4 界面设计规范	28
4.5 状态设计	29
5 功能描述及验收验证标准.....	32
5.1 具体功能描述	32
5.2 输入输出格式	36
5.3 界面验收标准	36
5.4 功能验收标准	37
附录 A 需求追溯矩阵.....	39
附录 B 原型页面索引	39

1 引言

1.1 编写目的

本文档描述 DevPulse 智能问答助手 V1.0 的功能需求与非功能需求，作为后续概要设计、详细设计、编码实现、测试验收和项目评审的共同依据。

文档的预期读者包括四类：项目组成员在设计、开发与测试阶段对照本文档确认实现范围；课程教师与助教据此判断需求是否真实、完整、可验收；评审小组据此对需求分析质量给出评价；后续接手项目的人员据此理解系统要解决的问题与边界。

本文档只描述"系统应当做什么"以及"如何判断做到了"，不描述具体的代码结构、数据库表结构和部署脚本。上述内容属于设计文档的范围。

1.2 项目背景

在课程学习、社团工作和项目协作中，信息通常散落在聊天记录、在线文档、会议纪要、课程资料与网页之中。成员遇到问题时需要在多个平台反复搜索，或者直接询问已经处理过同类问题的人。这一过程带来三个后果：同一个问题被不同成员重复提问，定位成本随资料增加而上升；只给结论而不说明来源，使用者难以判断答案是否可信；问题解决后没有形成可复用的知识资产，人员变动时经验随之流失。

大语言模型可以生成流畅的回答，但如果只依赖模型自身的参数化知识，容易出现事实过期、来源不明与内容幻觉。检索增强生成先把与问题相关的文档检索出来，再让模型基于检索结果组织答案，可以缓解上述问题。DevPulse 的目标不是再做一个通用聊天机器人，而是围绕团队私有资料建立一条可信的问答链路，让答案有出处、可核对、能更新。

项目基本信息如下。

表 3 项目基本信息

项目	内容
软件名称	DevPulse 智能问答助手
项目任务提出者	福州大学 2601_FZU_SE 《软件工程实践》课程组

项目	内容
项目开发者	DevPulse 团队，共 5 名成员
直接用户	课程班级、学生团队、社团、实验室小组、小型项目组
运行形态	浏览器访问的 Web 应用，前后端分离，服务端部署于云服务器
与外部系统的关系	调用大模型服务与向量检索服务；不替代现有在线文档平台，只在其之上提供问答入口

1.3 术语、缩写与定义

本文档出现的缩写与专有术语在首次出现处给出中文解释，并统一列于下表，后续章节不再重复定义。

表 4 术语与缩写表

术语 / 缩写	英文全称	含义
SRS	Software Requirements Specification	软件需求规格说明书，即本文档
LLM	Large Language Model	大语言模型，具备自然语言理解与生成能力
RAG	Retrieval-Augmented Generation	检索增强生成，先检索资料再生成答案的技术路线
知识库	Knowledge Base	一组有序组织的文档与资料的集合，是检索的基本单位
文本片段	Chunk	文档切分后的最小检索单位，携带来源与页码信息
向量检索	Vector Retrieval	将文本与问题映射为向量后按语义相似度查找片段
关键词检索	Keyword Retrieval	基于词项匹配的检索方式，擅长专有名词与精确表达
混合检索	Hybrid Retrieval	同时使用向量检索与关键词检索并合并结果
重排序	Rerank	对初次召回的结果重新排序，把最相关的片段排在前面

术语 / 缩写	英文全称	含义
引用溯源	Citation Tracing	从答案定位到原文片段与所在位置的能力
知识缺口	Knowledge Gap	用户提问在现有知识库中无法被回答的情况
检索命中率	Retrieval Hit Rate	前若干条检索结果中包含正确答案来源的比例
引用准确率	Citation Accuracy	引用片段确实支持对应结论的比例
FR	Functional Requirement	功能需求，编号形如 FR-01
NFR	Non-Functional Requirement	非功能需求，编号形如 NFR-01
MVP	Minimum Viable Product	最小可用产品，指主链路可稳定演示的最小功能集合
OCR	Optical Character Recognition	光学字符识别，用于处理扫描版文档
UAT	User Acceptance Testing	用户验收测试

1.4 参考资料

1. GB/T 9385-2008 《计算机软件需求规格说明规范》
2. GB/T 8567-2006 《计算机软件文档编制规范》
3. 邹欣. 《构建之法》[M]. 北京：人民邮电出版社
4. Lewis P, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks[C]. NeurIPS, 2020
5. LangChain 官方文档. <https://python.langchain.com>
6. Milvus 官方文档. <https://milvus.io/docs>
7. DevPulse 团队. 《DevPulse 智能问答助手选题说明》. 2026
8. DevPulse 团队. 《软件工程实践团队作业——种子队选拔、团队展示及选题》. 2026

2 总体描述

2.1 产品目标

2.1.1 开发意图

本项目希望解决团队协作中知识查找成本高、答案来源不可核对、经验难以沉淀的问题。具体做法是把团队已有的文档、课程资料和常见问答整理为可检索的知识库，用户用自然语言提问后，系统返回基于这些资料生成的答案，并展示答案所依据的原文片段。

从工程实践角度看，本项目还有第二层意图：让团队完整经历一次从需求分析、数据建模、接口设计、前后端实现到测试部署的软件开发过程，并在其中练习如何把模糊的想法转化为可验收的需求。

2.1.2 应用目标与作用范围

系统的应用目标是使用户获取团队内部知识的时间明显下降，并且每次获得的答案都能被独立核对。围绕这一目标，系统的作用范围包括六个方面。

1. **知识入库。** 支持 Markdown、纯文本、PDF、Word 等常用格式的资料导入，完成解析、清洗、切分与向量化，并保留来源与版本信息。
2. **检索与问答。** 支持自然语言提问，通过混合检索与重排序筛选相关片段，由大语言模型基于片段生成答案。
3. **引用溯源。** 答案附带引用卡片，展示文档名称、页码、片段内容与相似度，用户可以据此回到原文核对。
4. **多轮对话。** 在同一会话中保留上下文，支持“这个”“上面提到的”等指代性追问。
5. **反馈闭环。** 用户可以评价回答，系统记录低分回答与未命中问题，形成可被处理的知识缺口清单。
6. **权限与治理。** 区分管理员、普通成员与访客，控制知识库的可访问范围，并保留关键操作的审计记录。

系统不处理的领域在第 2.6 节说明。

2.1.3 产品前景

首版面向课程班级与小型团队，功能范围有意保持收窄，以保证主链路稳定可用。由于系统与具体知识内容解耦，只要能提供文档资料，就可以替换知识库内容服务新的场景，例如实验室的入门资料问答、社团的规章制度问答、课程助教的高频问题问答。

后续版本的扩展方向包括文档版本管理、知识库之间的引用、回答质量自动评估与更细粒度的权限模型。项目采用的模型接口与向量库均通过配置替换，不绑定单一厂商，这为长期维护保留了空间。

2.2 用户特征

系统的使用者可归纳为四类角色，其技术背景与使用诉求差异明显，界面与权限设计需要据此区分。

表 5 用户角色与特征

角色	典型人员	技术水平	使用频率	主要诉求
普通成员	项目组员、社团成员、学生	会使用浏览器与在线文档，不了解检索原理	每周数次	快速得到有出处的答案，不想学习复杂的检索语法
知识库管理员	项目负责人、社团干部、课程助教	熟悉常见办公软件，能判断文档是否准确	每周一至两次	维护文档、处理反馈、发现知识缺口
系统管理员	团队技术人员	具备基本的部署与运维能力	每月一至两次	管理账号与权限、更换模型配置、查看审计日志
访客	被临时授权的同学或外部评审人员	只会浏览	偶尔	在不接触敏感资料的前提下查看公开内容

用户群体的共同特点是了解自己负责的业务，但不熟悉检索与提示词调优，因此系统不能把检索参数暴露给普通成员，默认值必须开箱可用。

2.3 用户场景与典型故事

本节场景来自对 5 位同学与 1 位课程助教的小范围访谈。访谈采用开放式提问，询问受访者最近一次"找不到项目资料"的经历，场景描述保留了受访者的原话表述。

2.3.1 场景一：新成员了解本地运行方式（SC-01）

学号靠后加入团队的组员第一次拿到仓库地址，需要在本机把项目跑起来。部署说明分散在群聊记录、一份在线文档和仓库的 README 中，三处内容版本不一致，其中群聊里提到的一条命令已经过时。该成员在群里提问，等待了四十分钟才得到答复。

表 6 用例 UC-01 查询项目部署方式

项目	内容
用例编号	UC-01
用例名称	查询项目部署方式
参与者	普通成员
对应场景	SC-01
前置条件	用户已登录，且对项目知识库具有查看权限；知识库中已收录最新的部署文档
主流程	1. 用户在会话中输入"这个项目怎么在本地跑起来"；2. 系统检索部署文档并重排序；3. 系统基于命中的片段生成步骤化答案；4. 答案下方展示引用卡片；5. 用户点击引用卡片，查看原文中的完整命令
异常流程	若知识库中只有过时文档，系统仍返回基于该文档的答案，并在引用卡片中标注文档的更新日期，便于用户判断时效性
后置条件	会话记录被保存，用户可继续追问

2.3.2 场景二：开发成员了解历史设计决策（SC-02）

一名成员在修改接口时发现参数命名不符合直觉，想知道当初这样设计的原因。相关讨论只存在于两个月前的会议记录中，参与讨论的成员已经记不清细节。该成员需要确认现状是否有意为之，而不是一次疏忽。

表 7 用例 UC-02 查询历史设计决策

项目	内容
----	----

项目	内容
用例编号	UC-02
用例名称	查询历史设计决策
参与者	普通成员
对应场景	SC-02
前置条件	会议纪要与设计决策记录已录入知识库
主流程	1. 用户输入"某个接口为什么这样设计"; 2. 系统识别问题中的接口名称并执行混合检索; 3. 系统返回包含决策背景、可选方案与最终选择的答案; 4. 引用卡片指向会议纪要的对应段落
异常流程	若检索结果不足, 系统明确提示"知识库中没有找到相关记录", 并建议补充会议纪要, 同时将该问题记入知识缺口
后置条件	未命中问题进入管理员的知识缺口清单

2.3.3 场景三：课程助教答复重复提问（SC-03）

课程助教每周都会收到若干名学生询问作业提交格式的问题。答复内容固定，但每次都需要在一份长文档中找到对应段落并复制，重复劳动较多。助教希望直接得到一段可以转发的答复，并确认这段话确实来自作业要求。

表 8 用例 UC-03 生成可转发的规范答复

项目	内容
用例编号	UC-03
用例名称	生成可转发的规范答复
参与者	普通成员（助教角色）
对应场景	SC-03
前置条件	作业要求文档已录入知识库
主流程	1. 助教输入"作业提交格式是什么"; 2. 系统检索到作业要求文档中的对应段落; 3. 系统生成包含截止时间、提交方式与命名规则的答复; 4. 引用卡片展示原文段落
异常流程	若知识库中的作业要求存在多个版本, 系统优先引用更新时间最新的文档, 并在答案中提示存在多个版本
后置条件	答案可被复制转发, 引用来源可随时复核

2.3.4 场景四：管理员处理知识缺口（SC-04）

管理员发现某一类问题连续一周被反复提问，说明知识库在这部分存在缺失。在缺少汇总机制的情况下，管理员只能凭印象判断要补什么资料。

表 9 用例 UC-04 查看与处理知识缺口

项目	内容
用例编号	UC-04
用例名称	查看与处理知识缺口
参与者	知识库管理员
对应场景	SC-04
前置条件	系统中已积累了一定数量的未命中问题与低分回答
主流程	1. 管理员进入知识缺口分析页；2. 系统按问题出现次数降序展示未命中问题清单；3. 管理员选取其中一个问题，查看相似提问的原始记录；4. 管理员补充或修订对应文档；5. 系统重新解析并更新向量
异常流程	若缺口原因是文档表述不清而非文档缺失，系统支持管理员仅修订原文后重新索引，无需新增文档
后置条件	知识缺口清单中该条目被标记为已处理

2.4 运行环境与约束

表 10 运行环境与约束

类别	要求
服务端运行环境	Linux 服务器，Docker 与 Docker Compose 可用，可访问外部大模型服务接口
客户端运行环境	Windows、macOS 或 Linux 桌面操作系统，具备现代浏览器
浏览器支持	Chrome 100 及以上、Edge 100 及以上、Firefox 100 及以上；不支持 Internet Explorer
网络要求	客户端与服务端之间通过 HTTPS 通信；服务端需要能访问所配置的大模型接口

类别	要求
终端适配	以桌面端为主要使用场景，窗口宽度不小于 1280 像素；窄屏设备仅保证可浏览，不保证完整交互
语言与编码	界面与文档以简体中文为主，文本统一采用 UTF-8 编码
课程约束	项目须在课程周期内完成，部署成本与模型调用成本需控制在课程可接受范围内

2.5 假设与依赖

假设。

1. 用户具备基本的中文输入能力，能够用自然语言表达问题，不需要了解检索语法。
2. 知识库中的文档以中文为主，且多数文档含有可提取的文本层。扫描版文档属于少数情况，可通过 OCR 单独处理。
3. 团队能够获得课程资料、项目文档与常见问答作为初始知识库内容，不需要额外购买数据。
4. 使用者愿意在答案存疑时点击引用核对原文，而不是直接采信模型输出。

依赖。

1. 依赖外部大模型服务提供文本生成能力。若接口不可用，系统需要降级提示而不是返回空白页面。
2. 依赖向量检索组件保存与查询文本片段向量。首版采用可本地运行的轻量方案，后续可替换为服务化方案。
3. 依赖课程提供的服务器或团队自备的云服务器完成部署。
4. 依赖团队成员在课程周期内持续投入，具体分工与进度见项目博客。

2.6 非目标

明确列出不做的事情，可以避免需求在实现阶段不断膨胀。本版本不包含以下内容。

表 11 本版本的非目标

非目标	说明
-----	----

非目标	说明
通用闲聊	系统只回答与知识库内容相关的问题，不承担闲聊、创作等通用对话职能
替代人的判断	系统提供的是带出处的参考信息，高风险结论仍需使用者自行核对原文
全格式文档解析	首版优先支持 Markdown、纯文本、PDF 与 Word，表格、公式与扫描件的支持程度有限
大规模高并发	首版面向课程班级与小型团队，不做高并发与横向扩展设计
互联网全文检索	系统只检索团队自行导入的资料，不主动抓取互联网内容
多租户计费	首版不区分租户，不包含计费、套餐与配额管理
移动端原生应用	不开发 iOS 与 Android 客户端，移动端仅保证页面可浏览

3 具体要求

3.1 类图

类图描述系统的业务实体及其静态关系。图 1 中只包含业务对象，缓存、消息队列、向量库等技术组件不出现在类图中，它们属于接口需求与设计文档的讨论范围。

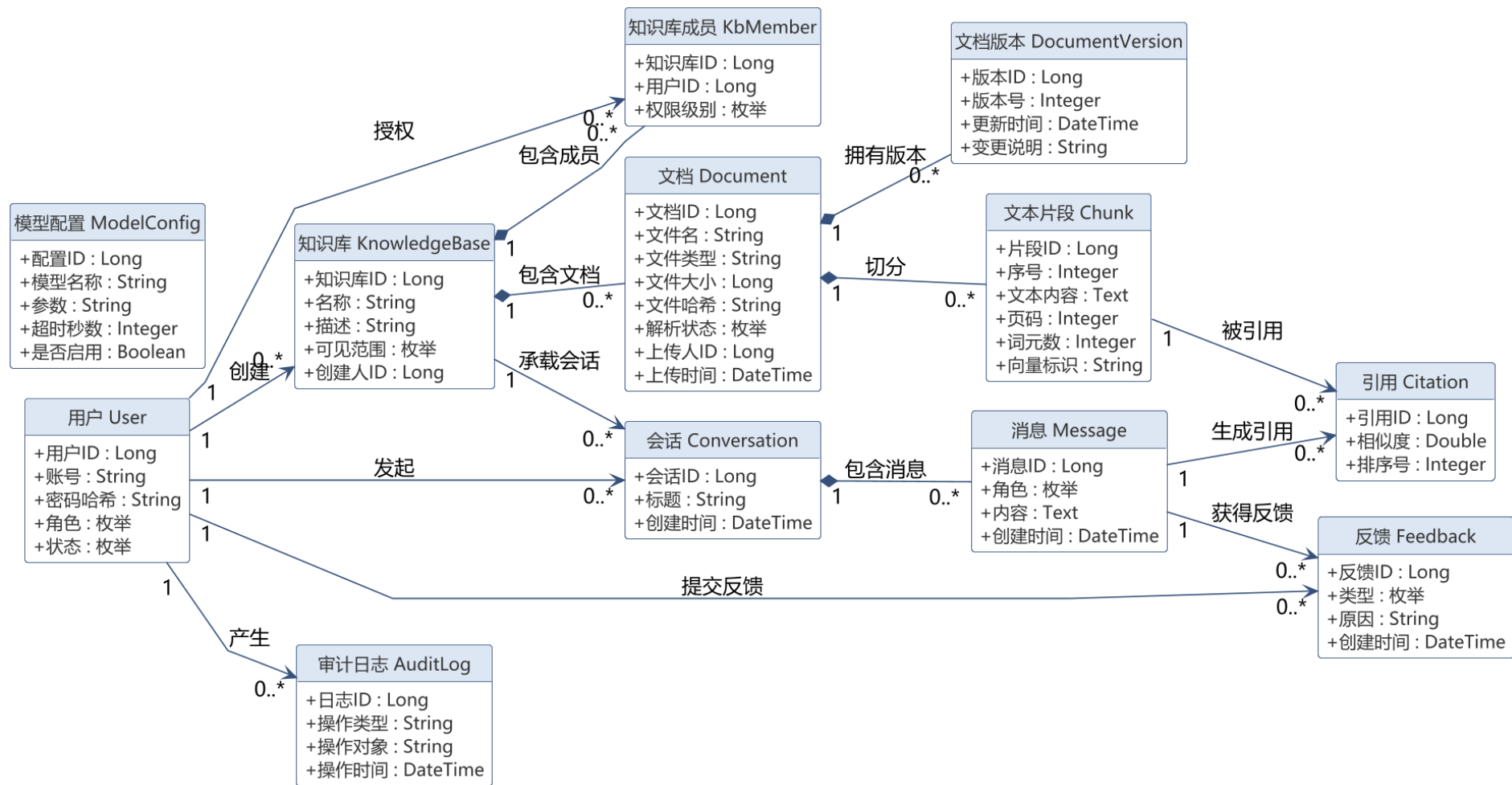


图 1 DevPulse 智能问答助手类图

各类的职责与关键属性说明如下。

表 12 类的职责与关键属性

类	职责	关键属性
用户 User	标识系统中的操作主体，承载身份与角色信息	用户 ID、账号、密码哈希、角色、状态
知识库 KnowledgeBase	组织文档的顶层容器，决定检索范围与可见范围	知识库 ID、名称、描述、可见范围、创建人 ID
知识库成员 KbMember	记录用户与知识库之间的授权关系与权限级别	知识库 ID、用户 ID、权限级别
文档 Document	表示一份被导入的资料，记录来源与解析状态	文档 ID、文件名、文件类型、文件大小、文件哈希、解析状态、上传人 ID、上传时间
文档版本 DocumentVersion	记录文档的更新历史，避免继续使用过期内容	版本 ID、版本号、更新时间、变更说明
文本片段 Chunk	检索的最小单位，承载来源定位信息	片段 ID、序号、文本内容、页码、词元数、向量标识
会话 Conversation	组织一个主题下的连续问答	会话 ID、标题、创建时间
消息 Message	记录一问一答的原始内容	消息 ID、角色、内容、创建时间
引用 Citation	关联答案与支撑答案的片段，支持溯源	引用 ID、相似度、排序号
反馈 Feedback	记录用户对回答的评价，支撑知识缺口分析	反馈 ID、类型、原因、创建时间
模型配置 ModelConfig	保存模型调用参数，支持更换模型供应商	配置 ID、模型名称、参数、超时秒数、是否启用
审计日志 AuditLog	记录关键操作，便于追责与问题排查	日志 ID、操作类型、操作对象、操作时间

类之间的关系说明如下。

- 用户与知识库（1 对多）。** 一个用户可以创建多个知识库，一个知识库只有一个创建人。创建人对该知识库拥有管理员权限。
- 用户与知识库（多对多，通过知识库成员）。** 用户通过授权获得对其他知识库的访问权。引入知识库成员这一关联实体，既表达多对多关系，也承载权限级别属性。

- 3. **知识库与知识库成员（1 对多，组合）。** 知识库被删除时，其成员授权记录随之失效。
- 4. **知识库与文档（1 对多，组合）。** 文档必须归属于一个知识库，知识库被删除时其文档一并删除。
- 5. **文档与文档版本（1 对多，组合）。** 一份文档可以有多个版本，版本记录用于追溯文档的更新过程。
- 6. **文档与文本片段（1 对多，组合）。** 一份文档按语义边界切分为多个片段，片段是向量检索的基本单位。片段被删除时，其对应的向量一并作废。
- 7. **知识库与会话（1 对多）。** 会话在某个知识库范围内进行，检索范围由会话所属知识库决定。
- 8. **用户与会话（1 对多）。** 一个用户可以发起多个会话，会话归属发起人，其他成员不可见。
- 9. **会话与消息（1 对多，组合）。** 会话由多条消息组成，消息按时间顺序排列。
- 10. **消息与引用（1 对多）。** 一条助手消息可以引用多个片段，用于支撑答案的不同部分。
- 11. **片段与引用（1 对多）。** 一个片段可能被多条回答反复引用。消息与片段之间因此通过引用实体构成多对多关系，引用实体上的相似度与排序号用于前端按相关度展示。
- 12. **消息与反馈（1 对多）。** 一条助手消息可以获得多次反馈，但同一用户对同一条消息只保留最新一次评价。
- 13. **用户与反馈（1 对多）。** 反馈归属于提交它的用户，用于区分不同角色的意见权重。
- 14. **用户与审计日志（1 对多）。** 关键操作都会产生审计记录，记录操作主体与操作对象。

3.2 数据需求

3.2.1 实体属性与约束

表 13 主要实体的属性约束

实体	关键属性	主要约束
----	------	------

实体	关键属性	主要约束
用户	用户 ID、账号、密码哈希、角色、状态	账号在系统内唯一；密码以加盐哈希方式保存，不存明文；角色取值限于管理员、普通成员、访客
知识库	知识库 ID、名称、描述、可见范围、创建人 ID	名称在同一创建人名下唯一；可见范围取值为公开或指定成员
知识库成员	知识库 ID、用户 ID、权限级别	知识库 ID 与用户 ID 的组合唯一；权限级别取值为管理、编辑、只读
文档	文档 ID、文件名、文件类型、文件大小、文件哈希、解析状态、上传人 ID、上传时间	文件哈希在同一知识库内唯一，用于识别重复上传；解析状态取值为待解析、解析中、已入库、解析失败
文档版本	版本 ID、文档 ID、版本号、更新时间、变更说明	同一文档的版本号递增且唯一
文本片段	片段 ID、文档 ID、序号、文本内容、页码、词元数、向量标识	同一文档内序号唯一；页码可以为空，表示该片段无法定位到具体页
会话	会话 ID、知识库 ID、用户 ID、标题、创建时间	会话必须归属于某一知识库与某一用户
消息	消息 ID、会话 ID、角色、内容、创建时间	角色取值为用户或助手；内容不为空
引用	引用 ID、消息 ID、片段 ID、相似度、排序号	同一消息内排序号唯一；相似度为 0 至 1 之间的实数
反馈	反馈 ID、消息 ID、用户 ID、类型、原因、创建时间	类型取值为有用或无用；同一用户对同一消息只保留最新记录
模型配置	配置 ID、模型名称、参数、超时秒数、是否启用	同一时刻只允许一条配置处于启用状态
审计日志	日志 ID、操作人 ID、操作类型、操作对象、操作时间	日志只追加，不允许修改或删除

3.2.2 数据保留与清理

1. 会话与消息默认长期保留，用户可主动删除自己的会话记录。
2. 被删除的文档进入回收状态，其片段与向量在索引中立即失效，避免继续被检索到。
3. 审计日志保留时间不短于课程项目周期，用于事后追溯。

4. 模型调用日志只记录调用耗时、状态与词元用量，不记录完整的问题与答案原文，降低隐私风险。

3.3 接口需求

3.3.1 用户界面

用户界面需求在第 4 章界面原型中集中描述。本节只约定界面层需要遵循的三条约束：所有列表类页面必须提供空态提示；所有耗时操作必须提供进度或加载反馈；所有失败操作必须给出可执行的原因说明与重试入口。

3.3.2 软件接口

系统需要与两类外部服务交互，两者都通过配置注入地址与凭据，不在代码中硬编码。

表 14 外部软件接口

接口	用途	交互方式	主要约束
大语言模型服务	基于检索片段生成答案	HTTPS，兼容主流对话补全接口	支持超时与重试；模型名称可配置；失败时降级为提示信息，不返回空白回答
向量检索服务	保存与查询文本片段向量	本地进程或服务化接口	支持按知识库过滤；支持更新与删除；向量维度与所选嵌入模型一致
文档解析组件	从 PDF、Word 等格式中提取文本	进程内调用	解析失败时保留原始文件并记录失败原因

系统对前端暴露的内部接口按资源组织，主要如下。

表 15 主要内部接口

接口	方法	用途
/api/auth/login	POST	用户登录并获取访问凭据
/api/knowledge-bases	GET / POST	查询知识库列表、创建知识库
/api/knowledge-bases/{id}/documents	GET / POST	查询文档列表、上传文档

接口	方法	用途
/api/documents/{id}	DELETE	删除文档并使其片段失效
/api/conversations	GET / POST	查询会话列表、新建会话
/api/conversations/{id}/messages	POST	在会话中提问并获取带引用的回答
/api/messages/{id}/citations	GET	查询某条回答的引用片段详情
/api/messages/{id}/feedback	POST	提交对回答的评价
/api/knowledge-bases/{id}/gaps	GET	查询知识缺口清单
/api/knowledge-bases/{id}/members	GET / POST / DELETE	管理知识库成员与权限

3.3.3 通信接口

1. 客户端与服务端之间采用 HTTPS，请求与响应体使用 JSON 格式，字符编码为 UTF-8。
2. 问答接口采用流式返回，前端逐段渲染，避免长时间等待造成误判。
3. 接口统一返回结构包含状态码、提示信息与数据体三部分，前端据此区分正常结果与业务异常。
4. 单次问答请求的服务端超时时间为 30 秒，超时后向用户返回可重试的提示。

3.4 非功能需求

3.4.1 可用性

表 16 可用性需求

编号	需求描述
NFR-01	新用户在不阅读帮助文档的情况下，应能在 3 分钟内完成一次提问并看到带引用的答案
NFR-02	主链路的关键操作应控制在三次点击以内，例如从知识库列表进入问答
NFR-03	所有耗时操作必须提供进度或加载反馈，超过 2 秒未返回的操作不得出现无反馈的静止界面

编号	需求描述
NFR-04	用户输入错误时，系统应提示具体原因并保留已输入内容，不得清空表单
NFR-05	列表为空时应展示引导性空态，说明下一步可执行的操作
NFR-06	界面文字与背景的对比度应满足可读性要求，关键信息不依赖颜色单独区分

3.4.2 安全性

表 17 安全性需求

编号	需求描述
NFR-07	用户密码以加盐哈希方式存储，任何接口不得返回密码或其哈希
NFR-08	所有业务接口都需要校验访问凭据，未登录请求一律拒绝
NFR-09	知识库内容按授权范围隔离，用户访问未授权知识库时不得泄露文档标题或内容
NFR-10	敏感资料不进入公共知识库，知识库的可见范围在创建时明确指定
NFR-11	上传文件需校验类型与大小，拒绝可执行文件与超大文件，防止恶意文件进入解析流程
NFR-12	关键操作（创建知识库、修改权限、删除文档、变更模型配置）必须写入审计日志
NFR-13	模型调用不发送超出当前问题所需的资料片段，降低信息外泄面

3.4.3 性能

表 18 性能需求

编号	需求描述
NFR-14	在知识库包含 50 篇文档的规模下，问答请求的端到端响应时间不超过 8 秒（百分之九十五分位）
NFR-15	问答首段内容应在 3 秒内开始返回，后续内容以流式方式持续输出
NFR-16	单篇 10 兆字节以内的文档，从上传完成到状态变为已入库的时间不超过 5 分钟

编号	需求描述
NFR-17	在 20 名用户同时在线查询的条件下，系统不出现请求失败，响应时间不超过第 NFR-14 条要求的两倍
NFR-18	检索结果数量可配置，默认单次向模型提交不超过 5 个片段，控制响应时间与调用成本

3.4.4 可维护性与可移植性

表 19 可维护性与可移植性需求

编号	需求描述
NFR-19	模型服务地址、模型名称与调用凭据通过配置文件注入，更换服务商不需要修改业务代码
NFR-20	向量检索组件可替换，替换后的上层检索流程保持不变
NFR-21	文档解析器以插件方式组织，新增一种文件格式不影响既有格式的解析逻辑
NFR-22	检索命中率、引用准确率、响应时间等关键指标应可被采样记录，便于后续调优
NFR-23	系统可通过容器化方式在单机部署，部署步骤记录在项目文档中，具备可复现性

4 界面原型

本章描述界面的组织方式、页面构成与交互流程。原型采用中高保真形式制作，主链路上的五个页面为高保真，管理类页面为中保真。所有页面均按同一套设计规范绘制，规范见第 4.4 节。

4.1 信息架构与导航

系统的导航以知识库为中心组织：用户登录后进入知识库列表，从列表进入具体知识库后，可以在文档管理与智能问答之间切换。管理类功能只在具备相应权限时出现。

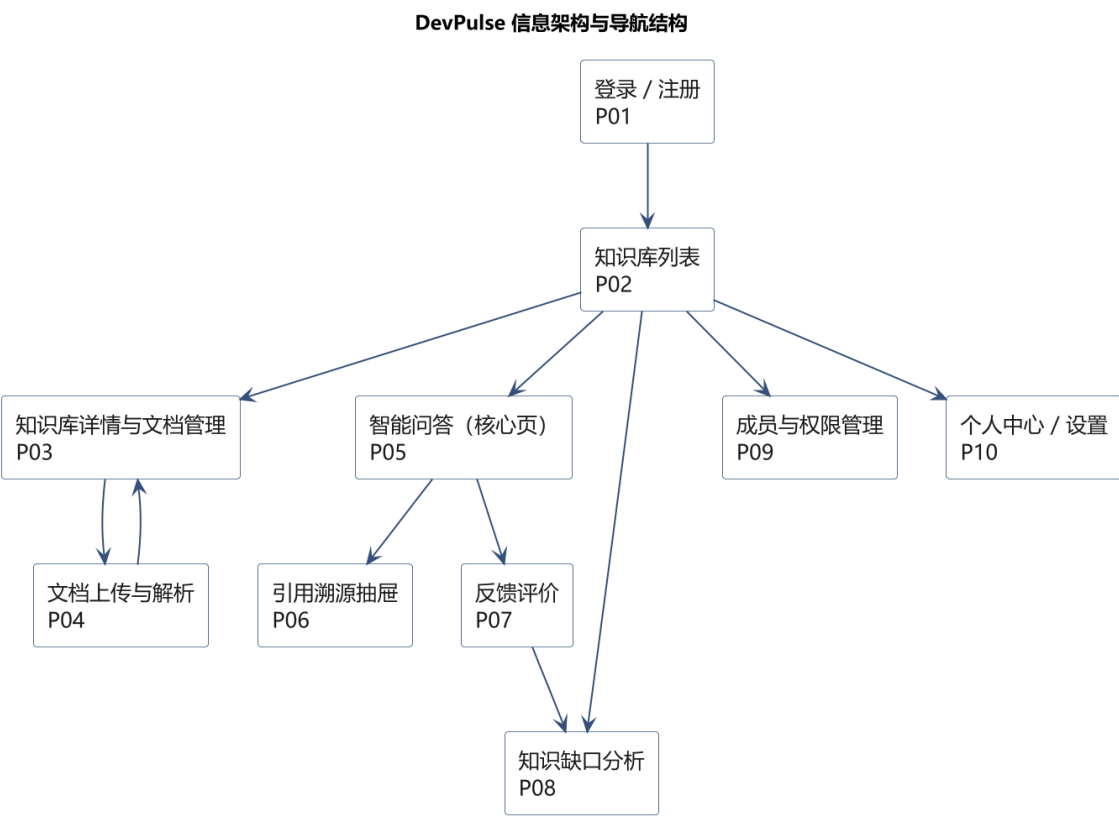


图 2 DevPulse 信息架构与导航结构

导航设计遵循三条原则。第一，全局最多两级导航，避免用户在层级中迷失。第二，当前所处的知识库在页面上方持续可见，用户随时知道自己在哪个知识库中提问。第三，无权限的入口直接不展示，而不是展示后点击报错。

4.2 页面清单

表 20 页面清单

编号	页面名称	保真度	主要功能	可访问角色
P01	登录 / 注册	中保真	账号登录、注册、错误提示	全部
P02	知识库列表	中保真	浏览、搜索、新建知识库	全部
P03	知识库详情与文档管理	中保真	查看文档表格、上传、删除、重新索引	管理员、编辑者
P04	文档上传与解析	中保真	拖拽上传、解析进度、失败重试	管理员、编辑者

编号	页面名称	保真度	主要功能	可访问角色
P05	智能问答	高保真	提问、查看流式回答、查看引用卡片	全部
P06	引用溯源抽屉	高保真	查看原文片段、页码、相似度、跳转原文	全部
P07	反馈评价	中保真	评价回答、选择原因、补充说明	全部
P08	知识缺口分析	中保真	查看未命中问题与低分回答、补充资料	管理员
P09	成员与权限管理	中保真	成员列表、角色设置、移除成员	管理员
P10	个人中心与设置	中保真	会话历史、模型偏好、账号设置	全部

十个页面的原型截图如下，其中 P02 至 P06 构成问答主链路，按高保真绘制，其余管理类页面按中保真绘制。

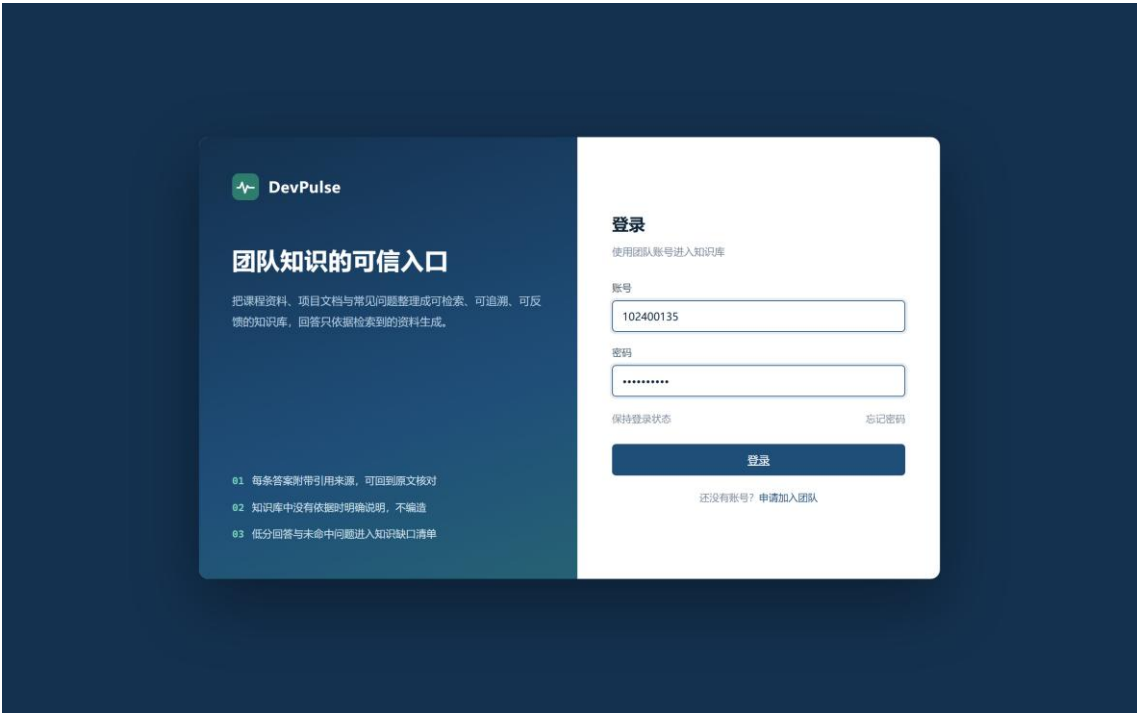


图 3 登录与注册页 P01

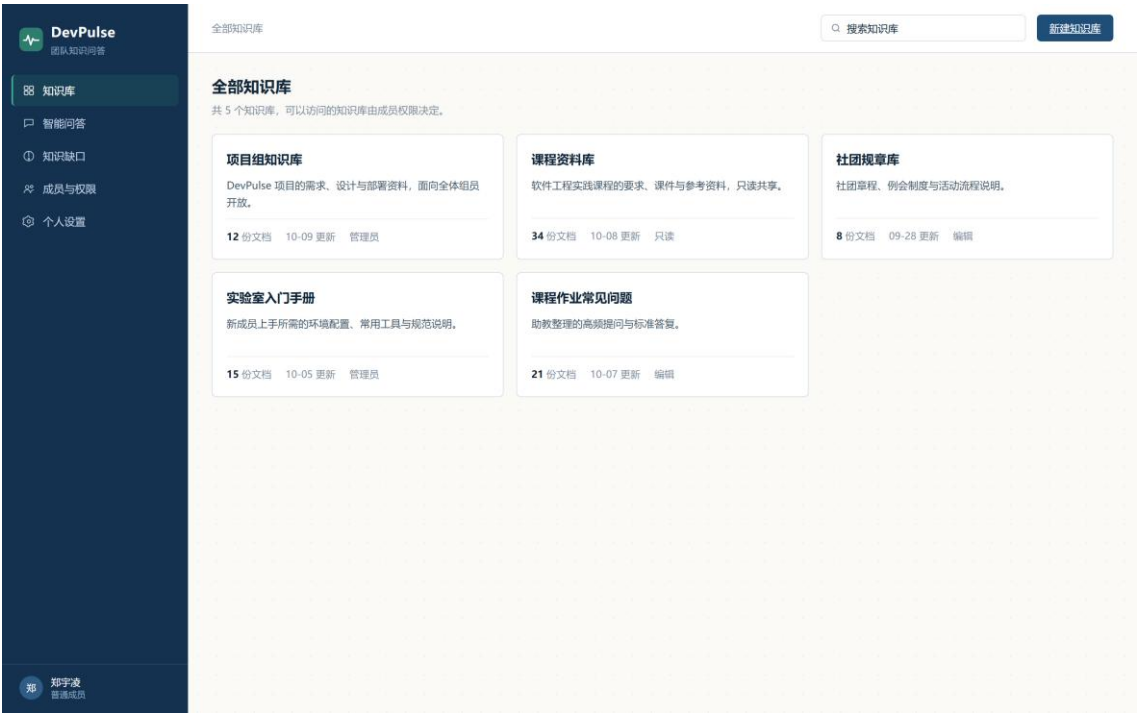


图 4 知识库列表页 P02



图 5 知识库详情与文档管理页 P03



图 6 文档上传与解析页 P04



图 7 智能问答页 P05



图 8 引用溯源抽屉 P06



图 9 反馈评价弹层 P07



图 10 知识缺口分析页 P08



图 11 成员与权限管理页 P09



图 12 个人中心与设置页 P10

4.3 关键交互流程

系统有三条关键交互链路，分别对应知识入库、问答与引用、反馈与知识缺口处理。这三条链路构成系统的完整闭环，也是验收时的重点检查对象。

图 13 说明文档从上传到可被检索的完整过程，其中拦截点位于格式校验环节，避免不可解析的文件进入后续流程。图 14 是最核心的一条链路，检索为空时的处理分支同样重要，系统在无依据时选择明确告知而不是编造答案。图 15 说明反馈如何转化为知识库的改进动作，这条闭环是系统区别于一次性问答工具的关键。

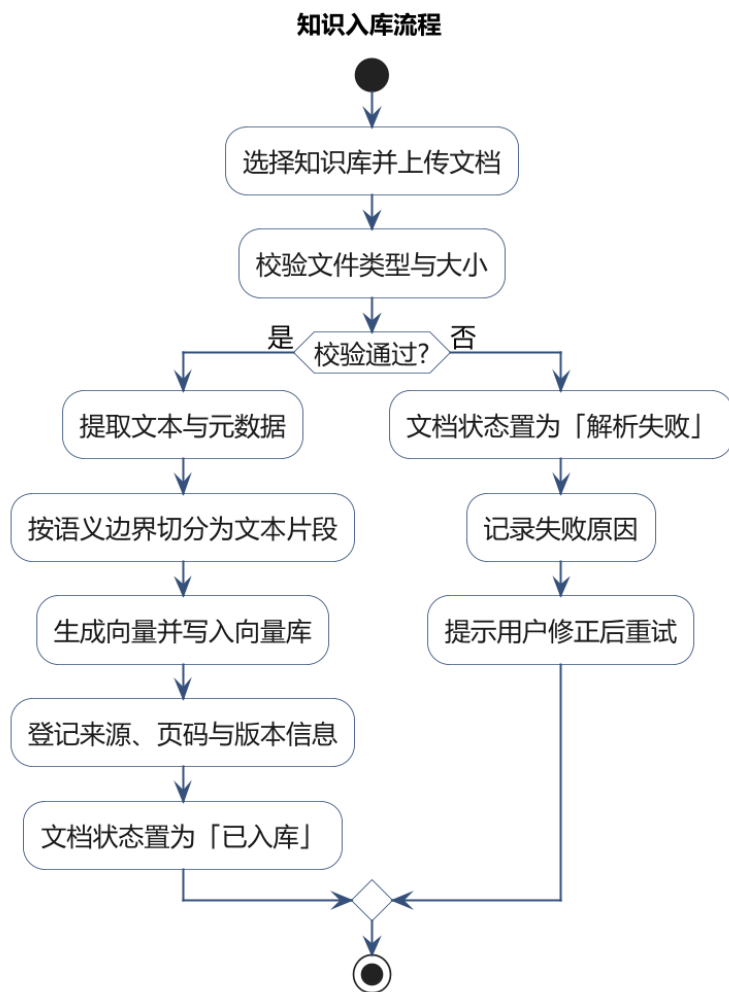
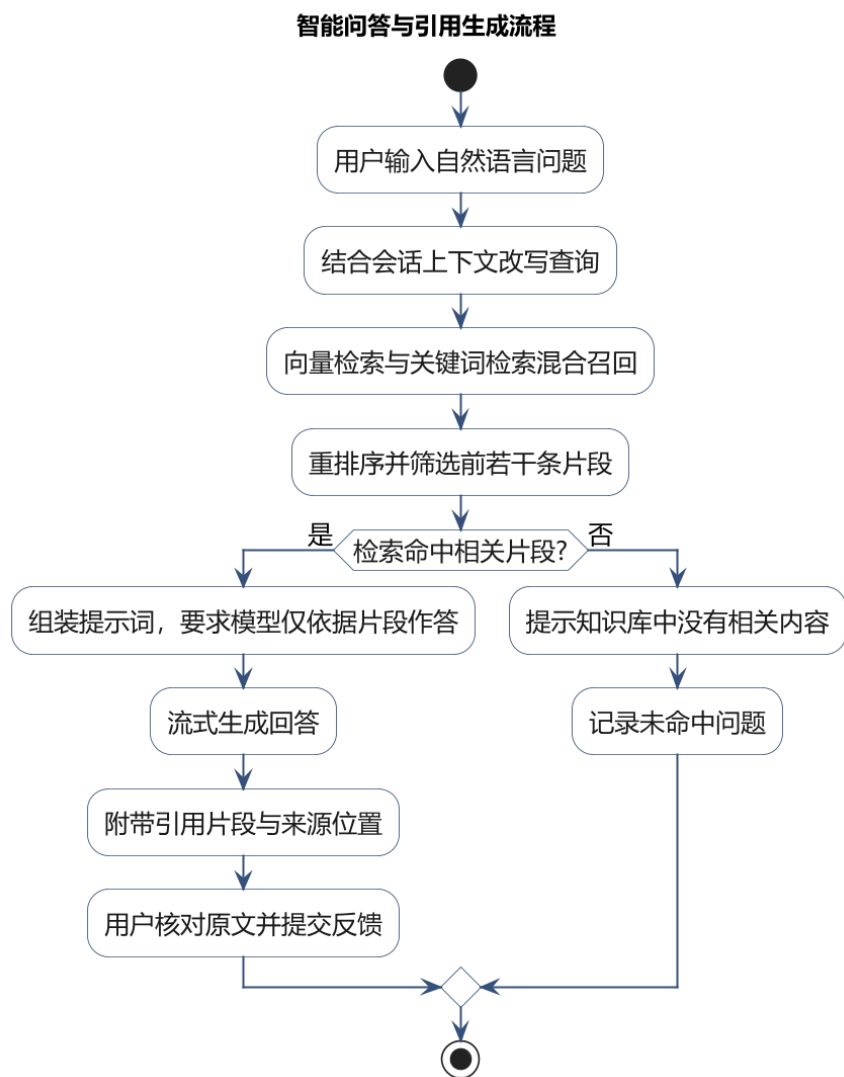


图 13 知识入库流程



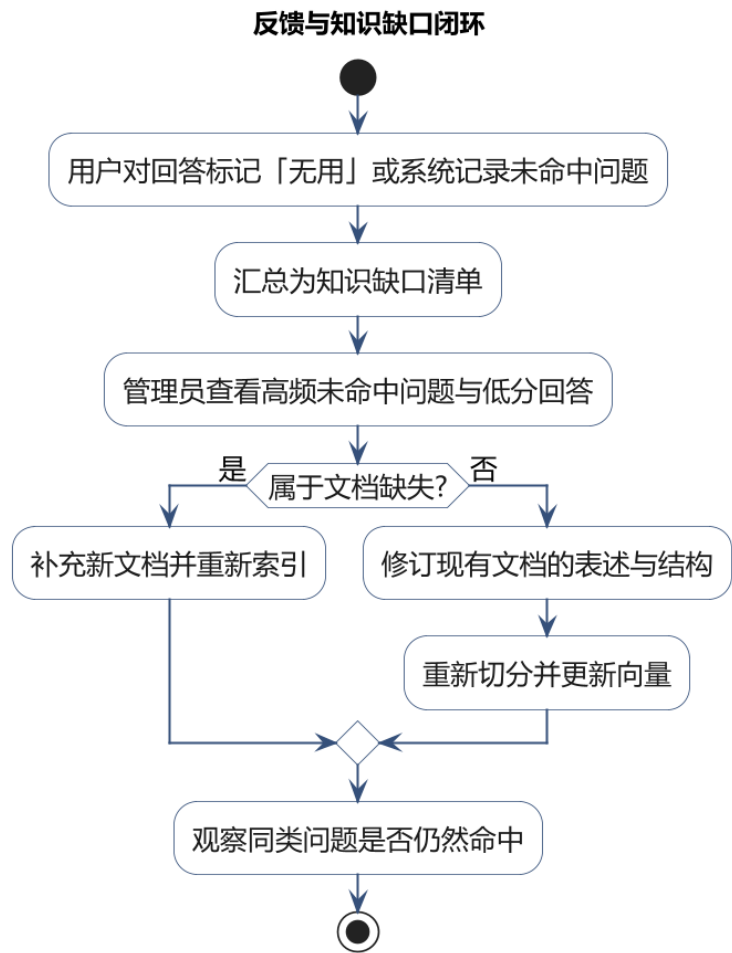


图 15 反馈与知识缺口闭环

4.4 界面设计规范

表 21 界面设计规范

类别	规范
主色	深蓝色，用于主按钮、选中态与关键强调元素
辅助色	灰阶用于建立文本层级；绿色表示成功，橙色表示警告，红色表示错误
字体	全站使用同一无衬线字体，避免中英文混排时字形风格不一致
字号阶梯	页面标题 20 像素，区块标题 16 像素，正文 14 像素，辅助说明 12 像素
间距	采用 8 像素基准栅格，卡片内间距 16 像素，模块之间间距 24 像素

类别	规范
按钮	分为主要、次要、文字三种层级；同一区域最多出现一个主要按钮
输入框	具有默认、聚焦、错误三种状态，错误状态下方给出具体原因
引用卡片	统一为文档名、页码、相似度、片段摘要四段式结构，全站样式一致
反馈组件	以图标加文字的形式呈现，点击后可展开原因选项
表格	表头固定，行高不小于 40 像素，长文本允许换行而不截断
提示信息	采用行内提示或顶部轻提示，避免使用打断操作的模态弹窗

4.5 状态设计

界面的完整程度主要体现在异常状态上。本节对核心页面 P05 明确四种状态的表现要求。

表 22 核心页面的状态设计

状态	触发条件	界面表现
空态	知识库中尚无文档，或用户还没有任何会话	展示引导文案，并提供上传文档与开始提问两个主要操作入口，不出现空白页面
加载态	文档解析中，或模型正在生成回答	文档行显示解析阶段名称与进度；回答区域采用流式输出并提供停止生成按钮
错误态	解析失败、模型超时、网络中断	说明失败原因，提供重试按钮，并保留用户已输入的问题内容
无权限态	访问未获授权的知识库	展示无权限提示与申请访问入口，不显示任何文档标题或内容摘要

四种状态在原型中的具体呈现如下。

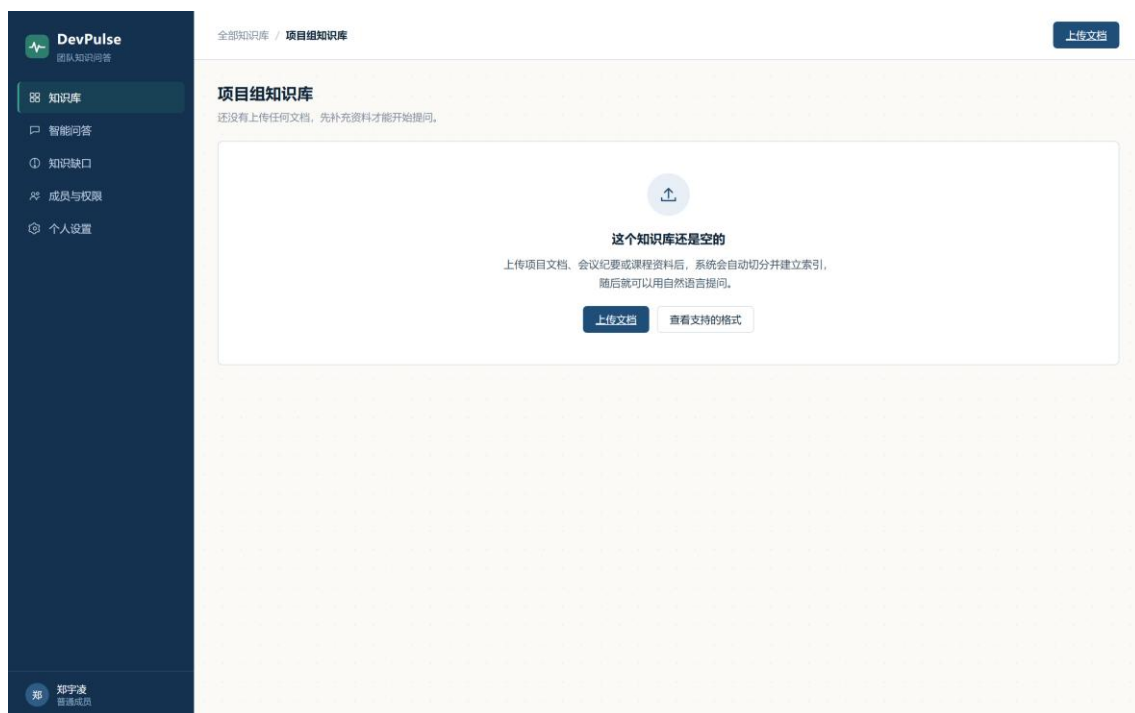


图 16 空态：知识库中还没有文档



图 17 加载态：模型正在流式生成回答

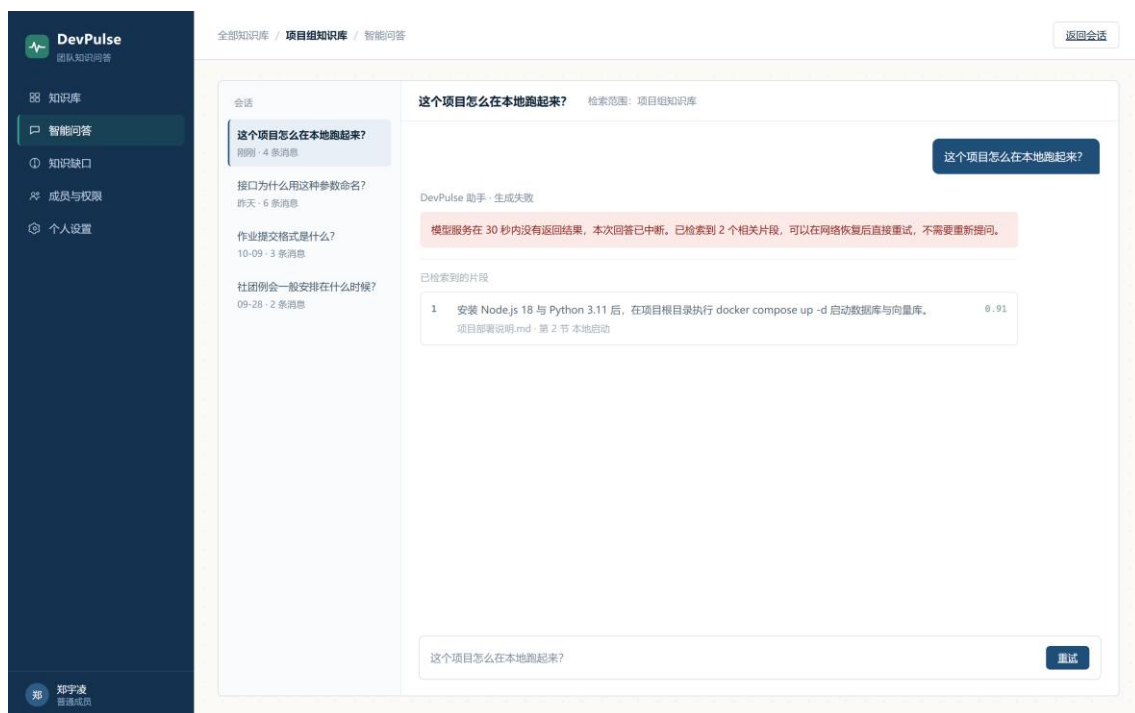


图 18 错误态：模型超时后的提示与重试

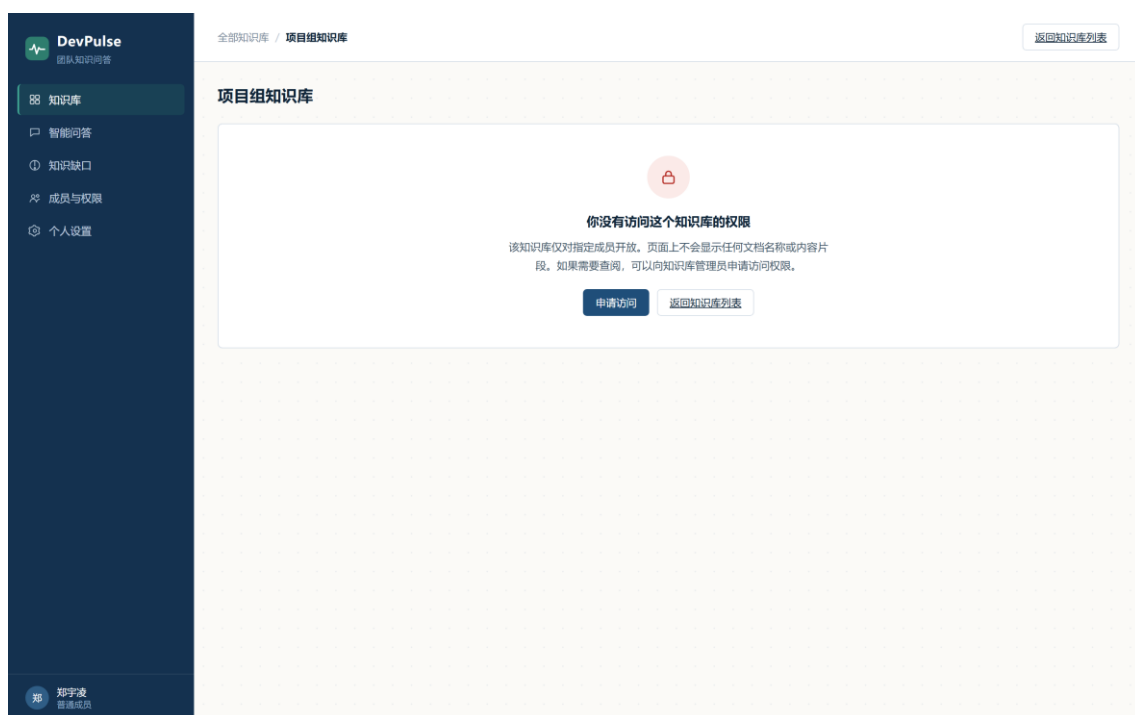


图 19 无权限态：访问未授权知识库

5 功能描述及验收验证标准

5.1 具体功能描述

本节逐条描述功能需求。每条需求包含描述、输入、输出、异常处理、优先级与来源场景，编号在全文范围内唯一，并与第 5.4 节及附录 A 中的条目对应。

表 23 FR-01 用户注册与登录

项目	内容
功能描述	用户通过账号与密码注册和登录，登录后获得访问凭据，系统据此判定其角色与可访问的知识库范围
输入	账号、密码；注册时另有确认密码
输出	登录成功后进入知识库列表页；失败时给出具体原因提示
异常处理	账号已存在、密码不符合复杂度要求、账号被停用、连续多次登录失败时给出不同提示
优先级与来源	高；基础功能，支撑全部场景

表 24 FR-02 知识库创建与管理

项目	内容
功能描述	用户创建知识库并设置名称、描述与可见范围，创建者可修改信息或将知识库归档
输入	知识库名称、描述、可见范围
输出	新知识库出现在列表中，创建者具有管理员权限
异常处理	名称重复时提示修改；无权操作的知识库不展示编辑入口
优先级与来源	高；支撑 SC-01 至 SC-04

表 25 FR-03 文档上传与解析

项目	内容
功能描述	用户上传文档，系统解析文本、清洗内容、切分片段、生成向量并登记来源与版本信息
输入	知识库标识、文档文件（Markdown、纯文本、PDF、Word）

项目	内容
输出	文档出现在管理列表中，状态由待解析、解析中、已入库依次变化；片段携带页码等来源信息
异常处理	文件类型不支持或超过大小限制时拒绝上传并说明原因；解析失败时保留原始文件并允许重试；同一文件重复上传时提示已存在
优先级与来源	高；支撑 SC-01、SC-02、SC-03

表 26 FR-04 文档管理

项目	内容
功能描述	管理员查看文档列表、查看解析状态、删除文档或对文档重新索引
输入	知识库标识、文档标识、操作类型
输出	列表状态同步更新；被删除文档的片段与向量立即失效，不再参与检索
异常处理	删除正在解析的文档时先中断解析任务；重新索引失败时保留原有片段，避免知识库内容丢失
优先级与来源	中；支撑 SC-04

表 27 FR-05 智能问答

项目	内容
功能描述	用户用自然语言提问，系统在知识库范围内检索相关片段，由大语言模型基于片段生成答案，并附带引用来源
输入	知识库标识、问题文本、会话上下文
输出	答案文本、引用片段列表（含文档名称、页码、相似度），以流式方式返回
异常处理	检索为空时明确提示知识库中没有相关内容；模型调用失败或超时时提示重试，不返回空白或部分生成的内容
优先级与来源	高；支撑 SC-01、SC-02、SC-03

表 28 FR-06 引用溯源

项目	内容
----	----

项目	内容
功能描述	用户点击答案下方的引用卡片，查看支撑该结论的原文片段及其所在位置
输入	引用标识
输出	弹出引用抽屉，展示原文片段、所属文档、页码与相似度，并提供跳转原文入口
异常处理	原文已被删除或更新时，提示该引用已失效并说明原因，不展示过期内容
优先级与来源	高；支撑 SC-01、SC-02、SC-03

表 29 FR-07 多轮对话

项目	内容
功能描述	在同一会话中保留上下文，使用户可以针对上一轮回答继续追问
输入	会话标识、追问文本
输出	结合上下文的答案，指代性表达被正确解析
异常处理	会话过长时对历史进行摘要压缩；上下文缺失时提示用户补充完整问题
优先级与来源	中；提升连续使用体验

表 30 FR-08 回答反馈

项目	内容
功能描述	用户对回答标记有用或无用，选择原因并可选填补充说明
输入	消息标识、评价类型、原因、补充说明
输出	反馈入库，管理员端可查看低分回答清单
异常处理	同一用户对同一条消息重复评价时覆盖此前记录；反馈提交失败时保留用户已选内容以便重试
优先级与来源	中；支撑 SC-04

表 31 FR-09 知识缺口分析

项目	内容
----	----

项目	内容
功能描述	管理员查看高频未命中问题与低分回答清单，据此补充或修订知识库文档
输入	知识库标识、统计时间范围
输出	按出现次数降序排列的缺口清单，以及每个问题对应的原始提问记录
异常处理	数据量不足时说明样本不足，不给出可能误导的结论；处理完成后条目状态更新为已处理
优先级与来源	中；支撑 SC-04

表 32 FR-10 成员与权限管理

项目	内容
功能描述	管理员维护知识库成员名单与权限级别，普通用户只能访问被授权的知识库
输入	知识库标识、用户标识、权限级别
输出	成员列表更新，权限变更立即生效
异常处理	移除成员时若该成员有进行中的会话，会话保留但其后续访问被拒绝；不允许移除知识库的唯一管理员
优先级与来源	中；支撑 NFR-09、NFR-10

表 33 FR-11 模型配置管理

项目	内容
功能描述	系统管理员配置模型服务地址、模型名称、超时时间与调用参数，并可在必要时切换模型
输入	模型服务地址、模型名称、参数、超时秒数
输出	配置生效，后续问答使用新配置
异常处理	保存前执行一次连通性校验；校验失败时拒绝保存并说明错误原因，保证原有配置继续可用
优先级与来源	中；支撑 NFR-19

表 34 FR-12 审计日志查询

项目	内容
功能描述	系统管理员按时间范围与操作类型查询关键操作记录
输入	时间范围、操作类型、操作人
输出	按时间倒序排列的日志列表
异常处理	日志只追加不修改；查询范围过大时提示缩小范围，避免一次返回过多数据
优先级与来源	低；支撑 NFR-12

5.2 输入输出格式

表 35 主要输入输出的格式约定

项目	格式	说明
上传文档	.md、.txt、.pdf、.docx	单个文件不超过 10 兆字节；扫描版 PDF 需可提取文本或经 OCR 处理
用户问题	纯文本，不超过 500 字	超出长度时提示截断或拆分提问
回答内容	Markdown 格式文本	支持标题、列表、代码块与表格的渲染
引用信息	结构化数据	包含文档名称、页码、片段文本、相似度与排序号
反馈记录	枚举加文本	类型为有用或无用，原因从预设选项中选择，补充说明不超过 200 字
导出内容	纯文本或 Markdown	会话内容可复制导出，导出内容保留引用来源信息

5.3 界面验收标准

表 36 界面验收标准

编号	验收项	验收标准
UI-01	页面清单完整性	表 20 中列出的十个页面均有对应原型页面，且可以从入口进入

编号	验收项	验收标准
UI-02	导航一致性	全局导航结构一致，任意页面均可在两步内返回知识库列表
UI-03	空态	文档列表为空、会话为空、缺口清单为空时均展示引导文案与下一步操作入口
UI-04	加载态	文档解析与答案生成过程均有可见的进度或流式输出反馈
UI-05	错误提示	表单校验失败、解析失败、超时三类错误均给出具体原因与可执行的重试入口
UI-06	无权限	访问未授权知识库时展示无权限提示，页面不出现文档标题与内容片段
UI-07	引用可见性	每条带引用的回答下方直接展示引用卡片，无需展开二级页面
UI-08	设计规范一致性	主色、字号阶梯、间距与组件样式在全站保持一致

5.4 功能验收标准

表 37 功能验收标准

需求编号	验收项	前置条件	操作步骤	预期结果	判定方式
FR-01	登录成功	存在有效账号	输入正确账号与密码	进入知识库列表页，角色权限正确加载	功能测试
FR-01	登录失败提示	存在有效账号	输入错误密码	提示密码错误，不清空账号输入框	功能测试
FR-02	创建知识库	已登录	输入名称与可见范围后提交	新知识库出现在列表，创建者权限为管理员	功能测试
FR-03	文档上传	已登录且具备编辑权限	上传一份 PDF 文档	文档进入列表，状态依次变为解析中与已入库	功能测试
FR-03	不支持格式拦截	已登录	上传可执行文件	拒绝上传并说明支持的文件类型	功能测试

需求编号	验收项	前置条件	操作步骤	预期结果	判定方式
FR-05	带引用问答	知识库中已有相关文档	输入一个业务问题	返回答案，引用卡片可点击并定位原文片段	功能测试
FR-05	无依据时拒答	知识库中无相关内容	输入超出知识范围的问题	明确提示知识库中没有相关内容，不编造答案	功能测试
FR-06	引用溯源	已产生带引用的回答	点击引用卡片	展示原文片段、文档名称、页码与相似度	功能测试
FR-07	多轮追问	已有一轮对话	输入包含指代性表达的追问	系统结合上下文正确理解追问对象	功能测试
FR-08	反馈记录	已产生至少一条回答	标记无用并选择原因	反馈入库，管理员端低分回答清单可见该条记录	功能测试
FR-09	缺口清单	已积累未命中问题	打开知识缺口分析页	按出现次数降序展示缺口清单	功能测试
FR-10	权限隔离	用户 A 无知识库 B 的权限	直接访问知识库 B 的地址	返回无权限提示，不泄露任何文档内容	安全测试
FR-11	配置校验	具备系统管理员权限	填写不可用的模型地址并保存	拒绝保存并说明原因，原配置继续生效	功能测试
NFR-14	响应时间	知识库含 50 篇文档	连续提问 10 次并记录耗时	百分之九十五的请求端到端响应不超过 8 秒	性能测试
NFR-15	首段返回时间	知识库内容正常	提问并记录首段出现时间	首段内容在 3 秒内开始返回	性能测试
NFR-16	解析耗时	上传一份 10 兆字节以内文档	记录从上传完成到已入库的时间	耗时不超过 5 分钟	功能测试
NFR-17	并发访问	20 名用户同时在线	并发执行查询操作	无请求失败，响应时间不超过 16 秒	性能测试
NFR-22	引用准确率	固定测试集 30 道题	人工核对引用片段是否支持结论	引用支持结论的比例不低于 90%	抽查评审

附录 A 需求追溯矩阵

矩阵用于核对需求来源与验证方式，确保每条需求都能追溯到具体场景，并落到具体的验收用例上。

表 38 需求追溯矩阵

用户场景	功能需求	非功能需求	原型页面	验收条目
SC-01 新成员了解本地运行方式	FR-03、FR-05、FR-06	NFR-01、NFR-03、NFR-15	P03、P04、P05、P06	FR-03 文档上传、FR-05 带引用问答、UI-07
SC-02 开发成员了解历史设计决策	FR-04、FR-05、FR-06	NFR-04、NFR-22	P03、P05、P06	FR-05 无依据时拒答、FR-06 引用溯源、NFR-22
SC-03 助教生成可转发答复	FR-05、FR-06、FR-07	NFR-02、NFR-14	P05、P06	FR-05 带引用问答、FR-07 多轮追问、NFR-14
SC-04 管理员处理知识缺口	FR-08、FR-09、FR-10	NFR-10、NFR-12	P07、P08、P09	FR-08 反馈记录、FR-09 缺口清单、FR-10 权限隔离
全局约束	FR-01、FR-02、FR-11、FR-12	NFR-07 至 NFR-13、NFR-19 至 NFR-23	P01、P02、P09、P10	FR-01 登录、FR-10 权限隔离、FR-11 配置校验

附录 B 原型页面索引

表 39 原型页面与需求的对应关系

页面编号	页面名称	对应功能需求	对应界面验收条目
P01	登录 / 注册	FR-01	UI-01、UI-05
P02	知识库列表	FR-02	UI-01、UI-03
P03	知识库详情与文档管理	FR-03、FR-04	UI-01、UI-03、UI-04
P04	文档上传与解析	FR-03	UI-04、UI-05
P05	智能问答	FR-05、FR-07	UI-01、UI-04、UI-07

页面编号	页面名称	对应功能需求	对应界面验收条目
P06	引用溯源抽屉	FR-06	UI-07
P07	反馈评价	FR-08	UI-01
P08	知识缺口分析	FR-09	UI-01、UI-03
P09	成员与权限管理	FR-10	UI-01、UI-06
P10	个人中心与设置	FR-07、FR-11	UI-01