



湖州学院
HUZHOU COLLEGE

电子信息学院 2025-2026-1 学期
课程设计报告

课程名称 数据库系统原理课程设计

专 业 计算机科学与技术

班 级 计算机 2412

姓 名 沈佳豪

学 号 2024284216

指导教师 施莹

报告成绩

电子信息学院计算机系

二〇二五年十二月二十八日

基于 LangGraph 的法律智能体系统

一、课题背景

自 2022 年底 ChatGPT 横空出世，引发了全球范围内的“百模大战”，人工智能正式迈入大模型时代。经过 2023 年的探索与 2024 年 RAG（检索增强生成）技术的普及，人们逐渐发现单纯的对话模型难以解决复杂的现实问题。2025 年，业界普遍将其定义为“智能体元年”。这一年，AI 的发展重心从单纯的文本生成转向了具备自主感知、规划决策、工具使用能力的“智能体（Agent）”。随着 LangGraph 等图结构编排框架的成熟，构建具备长程记忆、多步推理和自我反思能力的 Agent 已成为人工智能落地垂直领域的关键趋势。

在法律这一高度专业、逻辑严密的垂直领域，通用大模型普遍存在“幻觉”严重、逻辑推理能力不足等问题，难以满足专业法律咨询的需求。然而，社会经济的高速发展使得个人与企业的法律服务需求急剧增加，传统的法律服务模式面临着成本高、效率低、资源分配不均的巨大挑战。因此，利用 2025 年的 Agentic Workflow（代理工作流）技术，结合 LangGraph 的图谱编排能力与 RAG 的精准检索能力，构建一个能够像专业律师一样“思考-行动-反思”的法律智能体系统，不仅是技术发展的必然选择，更是解决法律服务供需矛盾、推动“普惠法治”的重要途径。

二、技术介绍

本系统采用前后端分离的微服务架构，融合了现代 Web 开发技术与人工智能技术，实现了一个智能化的法律咨询服务平台。系统整体架构包括前端展示层、Java 后端服务层、Python AI 智能体层以及云端存储层四个核心组成部分。整体架构图如下图 2-1 所示：

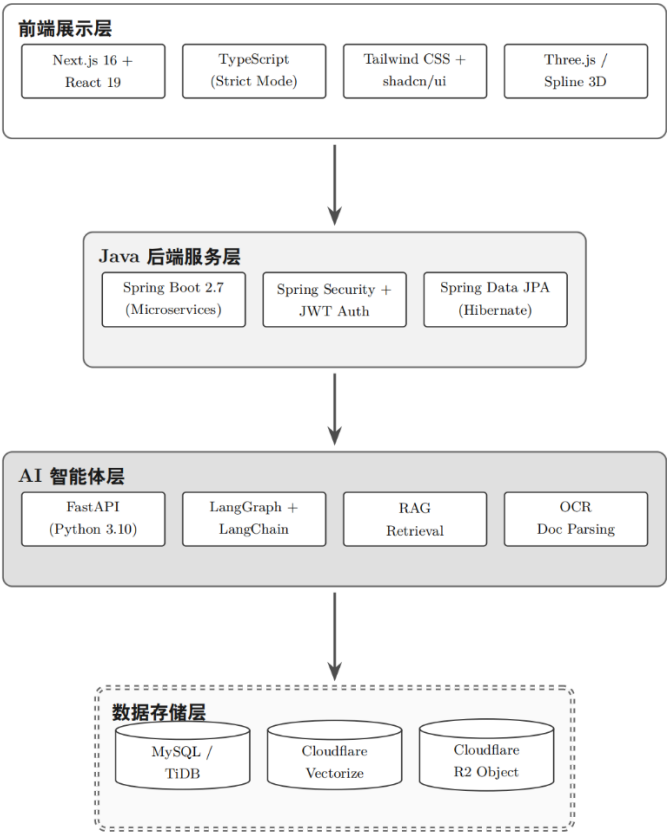


图 2-1 整体架构图

2.1 前端展示层

前端是用户与系统交互的直接窗口，本系统采用了目前 React 生态中的技术栈以确保卓越的用户体验和开发效率。核心框架基于 Next.js 16 和 React 19 构建，利用 Next.js 的服务端渲染特性优化了首屏加载速度和 SEO 表现，同时 React 19 的新特性进一步提升了应用性能。开发语言全线采用 TypeScript，通过静态类型检查在编译阶段捕获错误，显著提高了代码的健壮性和可维护性。在 UI 与交互方面，使用 Tailwind CSS 结合 shadcn/ui 组件库，实现了原子化的样式管理和美观统一的界面风格。此外，为了提升法律服务的科技感与沉浸感，系统引入了 Three.js 和 Spline 3D 技术，在登录页及关键交互区实现了高性能的 3D 场景渲染。

2.2 Java 后端服务层

后端服务层作为系统的业务逻辑中枢，负责处理用户请求、权限验证及核心业务数据的流转，确保系统的稳定性与安全性。核心框架基于 Spring Boot 2.7 微服务框架开发，利用其自动配置和依赖注入特性，快速构建了独立运行、生产级别的 Spring 应用。在安全认证方面，集成了 Spring Security 安全框架，配合 JWT 实现无状态的身份认证与授权机制，有效防御 CSRF 等常见网络攻击，保障用户信息安全。数据持久化层使用 Spring Data JPA 作为 ORM 框架，通过面向对象的方式操作数据库，简化了复杂的 SQL 编写，提高了开发效率。

2.3 AI 智能体层

AI 层是本系统的核心智能化引擎，负责处理复杂的法律咨询推理、文档分析等任务。服务接口使用 Python 3.10 和 FastAPI 框架构建高性能的 AI 服务接口，利用其异步处理能力应对高并发的推理请求。在智能编排方面，引入 LangGraph 和 LangChain 进行大模型的应用编排，通过构建图结构的 Agent 工作流，实现了多轮对话状态管理、意图识别及任务分发。系统的核心 AI 能力主要包括检索增强生成（RAG）和 OCR 文档解析：前者结合外部法律知识库，通过向量检索技术增强大模型的回答准确性，有效解决大模型“幻觉”问题；后者集成 OCR 技术，能够自动提取合同、法律文书中的文本信息，为后续的合同审查功能提供关键数据支撑。

2.4 数据存储层

数据存储层采用了混合云存储方案，针对不同类型的数据选择了最适配的存储介质。关系型数据使用分布式数据库 TiDB 存储用户账号、聊天记录等结构化业务数据，严格保证事务的一致性。向量数据利用 Cloudflare Vectorize 存储法律条文和案例的向量嵌入，为 RAG 检索提供毫秒级的相似度搜索能力。非结构化数据则使用 Cloudflare R2 对象存储服务，安全托管用户上传的合同文件、法律文书图片及系统静态资源，提供高可靠性和低延迟的访问体验。

三、需求分析

本系统的需求分析主要围绕普通用户、系统管理员以及 AI 智能体三个核心角色展开。通过对不同角色的业务场景进行深入梳理，明确了系统的功能边界与核心业务流程，旨在构建一个流程闭环、交互友好的法律 AI 助手系统。用例图如下图 3-1 所示：

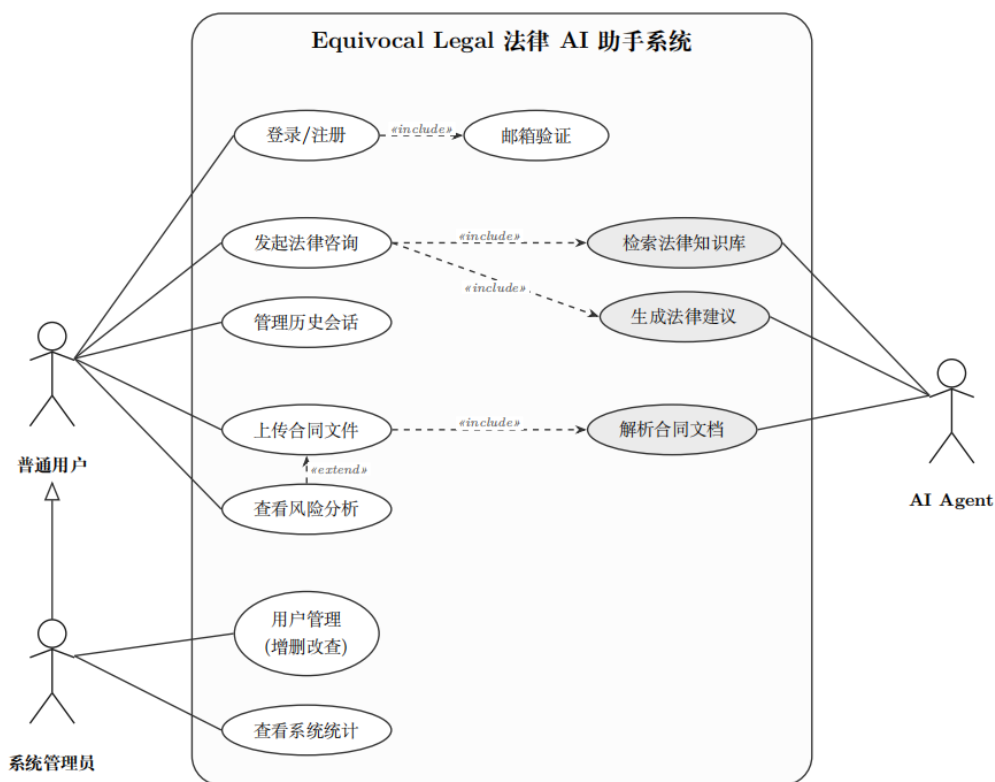


图 3-1 用例图

为了更直观地展示各角色与系统功能之间的交互关系，本系统构建了总体用例图（如图 3-1 所示）。用例图清晰地界定了普通用户、系统管理员及 AI 智能体三个角色的权限边界与协作逻辑。对于普通用户而言，登录注册功能通过包含关系关联了邮箱验证，这表明邮箱验证是身份认证过程中不可或缺的必要环节，确保了用户身份的真实性。发起法律咨询功能内部包含了检索法律知识库与生成法律建议两个子功能，体现了系统在接收用户咨询后，后台自动调用 AI 进行知识检索与回复生成的完整逻辑闭环。上传合同文件功能不仅包含了解析合同文档的基础处理，还通过扩展关系关联了查看风险分析功能，意味着这是一种可选的高级增值服务，用户在上传文件后可根据需求选择是否进行深度的风险评估。系统管理员作为拥有更高权限的角色，继承了普通用户的所有基础功能，并独有用户管理和查看系统统计两项特权，负责系统的日常运维与数据监控。此外，AI 智能体作为系统的后台核心参与者，虽然不直接与用户进行前端交互，但深度参与并支撑了法律知识库检索、法律建议生成及合同文档解析等关键计算任务，是实现系统智能化服务的底层驱动力量。

3.1 普通用户需求

普通用户是本系统的主要服务对象，其核心诉求在于获取便捷、专业的法律咨询服务及智能化的文档审查支持。具体功能需求如下：

首先是身份认证与安全。用户需要通过注册功能创建个人账户，并支持登录系统以访问个性化的历史数据。为了确保账户的真实性与安全性，注册流程必须包含“邮箱验证”环节，该环节作为登录注册功能的必要组成部分，用于防止恶意注册。

其次是智能化法律咨询。这是系统的核心业务功能，用户可向系统发起自然语言形式的法律咨询。该过程不仅是简单的问答，系统在后台自动执行“检索法律知识库”与“生成法律建议”两个关键步骤，即先从海量法律法规中精准定位相关条文，再结合大模型能力生成具

备法律效力的建议。

再次是合同文档审查。用户可以上传合同等法律文件，系统需具备“解析合同文档”的能力，自动提取关键条款。在此基础上，用户可以进一步“查看风险分析”，该功能旨在帮助用户识别合同中的潜在漏洞与法律风险。

最后是历史会话管理。系统需为用户保存所有的咨询记录，用户应能够随时查看、回顾及管理过往的会话历史，以便对法律建议进行回溯。

3.2 系统管理员需求

系统管理员负责平台的日常运维与后台管理，其需求侧重于系统的可控性与数据监控。主要需求包括用户管理，管理员拥有最高权限，能够对平台用户进行增、删、改、查（CRUD）操作，以维护用户数据的规范性与系统的安全性。此外，管理员还需要查看系统统计，通过可视化的方式掌握系统的运行状态、用户活跃度及业务调用量等关键指标，以便对平台运营情况进行整体把控。

四、系统设计

4.1 系统目标

本系统旨在构建一个集智能化、专业化、便捷化于一体的法律咨询服务平台。通过整合前沿的大语言模型技术与专业的法律知识库，致力于解决传统法律咨询服务中存在的成本高、响应慢、专业门槛高等痛点。系统的核心目标包括实现全天候智能咨询，利用 AI 智能体提供 24/7 的法律问答服务，快速响应用户的日常法律疑问；降低法律服务门槛，通过自然语言交互，让普通用户也能轻松获取专业的法律建议和风险提示；提升文档处理效率，自动解析合同与法律文书，辅助用户快速识别潜在的法律风险；保障数据安全隐私，采用多重加密与权限控制机制，确保用户隐私数据与咨询记录的安全性。

4.2 系统功能结构

系统功能结构设计围绕“用户中心”与“智能服务”两大核心展开，主要划分为用户管理、智能咨询、文档审查及历史会话管理四大功能模块。用户管理模块包含用户注册、登录、邮箱验证及个人信息管理，系统支持基于 JWT 的无状态认证，确保用户会话的安全性。智能咨询模块提供基于检索增强生成（RAG）技术的法律问答服务，用户可发起多轮对话，系统自动检索相关法律条文并生成专业建议。文档审查模块支持用户上传合同等法律文件，系统利用 OCR 技术提取文本，并结合 AI 模型进行条款解析与风险评估。历史会话管理模块自动保存用户的咨询记录与文档分析结果，支持用户对历史会话进行查询、回顾与管理。

4.3 概念结构设计

概念结构设计通过实体-关系（E-R）图来抽象表示系统中的核心数据对象及其相互关系。本系统的核心实体主要包括用户、会话及消息。用户是系统的核心主体，拥有唯一的身份标识、邮箱及角色权限。会话代表用户发起的每一次咨询对话，包含标题及创建时间，一个用户可以拥有多个会话，二者呈一对多关系。消息则是会话中的具体交互内容，包含角色、内容及消息类型，一个会话包含多条消息，二者同样呈一对多关系。系统整体 E-R 图，如图 4-1 所示：

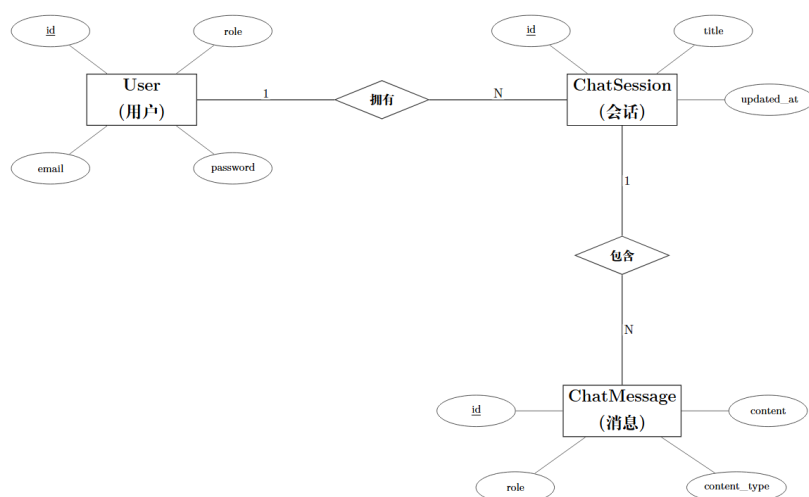


图 4-1 E-R 图

用户 (User) 是系统的核心主体，拥有唯一的身份标识 id，以及用于登录凭证的邮箱 email、登录密码 password 和用于区分权限的角色 role。

会话 (ChatSession) 代表用户发起每一次咨询对话，包含唯一的会话标识 id、用于标识主题的会话标题 title 以及记录最后活动时间的更新时间 updated_at。

消息 (ChatMessage) 则是会话中的具体交互内容，包含唯一的消息标识 id、具体内容 content、消息发送者角色 role (如 user 或 assistant) 以及标识文本或图片等格式的内容类型 content_type。

实体之间的关系如下：用户与会话之间存在一对多 (1:N) 的拥有关系，即一个用户可以创建并拥有多个会话，但每一个会话只能属于一个特定的用户。会话与消息之间存在一对多 (1:N) 的包含关系，即一个会话中可以包含多条消息，但每一条消息只能属于某一个特定的会话。

4.4 数据库设计

(1) 用户表

该表用于存储系统用户的基础信息与认证凭证。其中 id 字段为主键，用于唯一标识用户；email 字段存储用户邮箱，设有唯一约束以防止重复注册；password 字段存储加密后的密码；role 字段用于区分用户角色 (普通用户或管理员)；email_verified 字段标识邮箱是否已验证；created_at 记录账户创建时间；updated_at 记录账户最后更新时间。用户表的具体结构如表 4-1 所示：

字段名	数据类型	约束条件	说明
id	VARCHAR(64)	PK	用户唯一标识（主键）
email	VARCHAR(255)	UNIQUE, NN	用户邮箱（登录凭证）
password	VARCHAR(255)	NN	加密后的密码
role	INT	NN	角色权限 (1: 普通, 10+: 管理)
email_verified	BOOLEAN	DEFAULT 0	邮箱验证状态
created_at	DATETIME	-	账户创建时间
updated_at	DATETIME	-	最后更新时间

注：PK= 主键，NN= 非空，UNIQUE= 唯一索引

表 4-1 用户表结构设计

(2) 会话表

该表用于存储用户发起的咨询会话元数据。id 字段为主键；user_id 为外键，关联用户表，用于标识会话所属用户；title 字段存储会话标题，默认为“新对话”；created_at 和 updated_at 分别记录会话的创建时间和最后更新时间。会话表的具体结构如表 4-2 所示：

字段名	数据类型	约束条件	说明
id	VARCHAR(64)	PK	会话唯一标识
user_id	VARCHAR(64)	FK, NN	关联用户 ID（外键）
title	VARCHAR(255)	-	会话标题（默认为”新对话”）
created_at	DATETIME	-	会话创建时间
updated_at	DATETIME	-	最后活跃时间

注：FK= 外键，关联 users 表 id 字段

表 4-2 会话表结构设计

(3)消息表

该表用于存储会话中的具体对话记录。id 字段为主键，设为自增；session_id 为外键，关联会话表；role 字段标识消息发送者角色（用户或 AI 助手）；content 字段存储消息的具体内容，采用长文本类型；content_type 字段标识内容类型（如文本、图片或文件）；created_at 记录消息发送时间。消息表的具体结构如表 4-3 所示：

字段名	数据类型	约束条件	说明
id	BIGINT	PK, AUTO_INC	消息流水号（自增主键）
session_id	VARCHAR(64)	FK, NN	所属会话 ID
role	VARCHAR(50)	NN	发送角色 (user/assistant)
content	LONGTEXT	NN	消息内容（支持长文本）
content_type	VARCHAR(20)	-	内容类型 (text/image/file)
created_at	DATETIME	-	发送时间

表 4-3 消息表结构设计

(4) 验证码表

该表用于存储系统发送给用户的邮箱验证码，以确保用户身份的真实性与安全性。其中 id 字段为主键（设为自增）；email 字段记录接收验证码的邮箱地址；code 字段存储生成的随机验证码字符；expires_at 字段设定了验证码的有效截止时间，用于防止过期验证码被恶意利用；used 字段标识该验证码是否已被使用，防止重复提交；created_at 记录验证码的发送时间。attempts 字段用于记录验证尝试次数，防止暴力破解。验证码表的具体结构如表 4-4 所示：

字段名	数据类型	约束条件	说明
id	BIGINT	PK, AUTO_INC	唯一标识（自增主键）
email	VARCHAR(255)	NN	关联邮箱
code	VARCHAR(10)	NN	验证码内容
type	VARCHAR(20)	NN	业务类型 (register/login)
expires_at	DATETIME	NN	过期时间
used	BOOLEAN	DEFAULT 0	是否已使用
attempts	INT	DEFAULT 0	验证尝试次数
created_at	DATETIME	-	创建时间

表 4-4 验证码表结构设计

五、系统实现

本章将从普通用户和系统管理员两个角色的视角，详细阐述系统的核心功能模块实现。重点介绍界面的布局结构、关键控件的属性配置以及交互逻辑的实现方法。

5.1 普通用户功能实现

普通用户是系统的主要使用者，其核心功能包含法律咨询与合同审查。

5.1.1 智能法律咨询界面实现

法律咨询界面采用经典的左右分栏布局。左侧为可折叠的“历史会话侧边栏”，用于展示过往咨询记录；右侧为“对话主窗口”。主窗口底部固定悬浮“输入框组件”，支持文本输入与文件上传。法律咨询界面效果图如图 5-1：

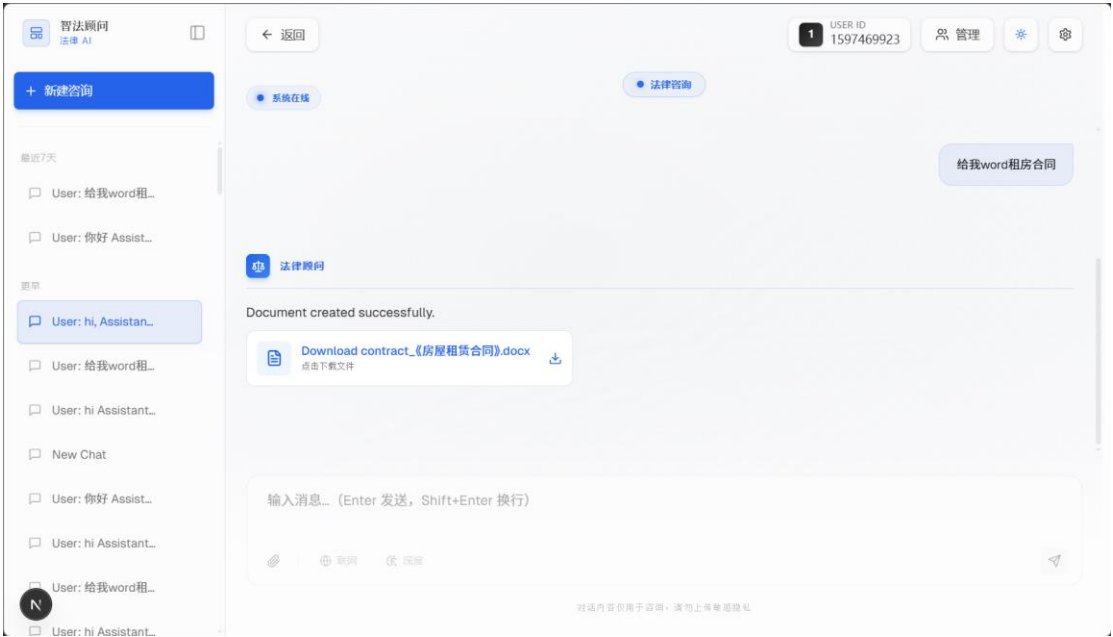


图 5-1 智能法律咨询界面

在关键控件实现方面，流式对话气泡是核心组件。通过将 markdown-render 属性设置为 true，组件能够实时渲染 AI 返回的 Markdown 格式文本，如加粗字体或列表结构。同时，通过 role 属性区分消息来源，用户消息显示为右侧蓝色气泡，AI 消息显示为左侧灰色气泡。在实现上，前端采用了 Server-Sent Events (SSE) 技术接收后端的流式数据。系统维护一个 messages 数组状态，每当接收到新的 token 时，便将其实时追加到最后一条消息的 content 字段中，从而在界面上呈现出流畅的打字机效果。此外，输入框采用了自适应高度设计，设置最小高度为 44px，最大高度为 200px。通过监听 input 事件，每次输入时动态计算 scrollHeight 并更新元素高度，确保输入框随内容自动增高，仅在超过最大高度时才出现滚动条，提升了输入体验。

5.1.2 合同智能审查界面实现

该界面是本系统的核心亮点，采用了“分屏对照”布局。用户上传合同后，界面自动切换为左右两栏：左侧为“文档预览区”，右侧为“风险面板”，方便用户对照查看。合同智能审查界面如图 5-2：

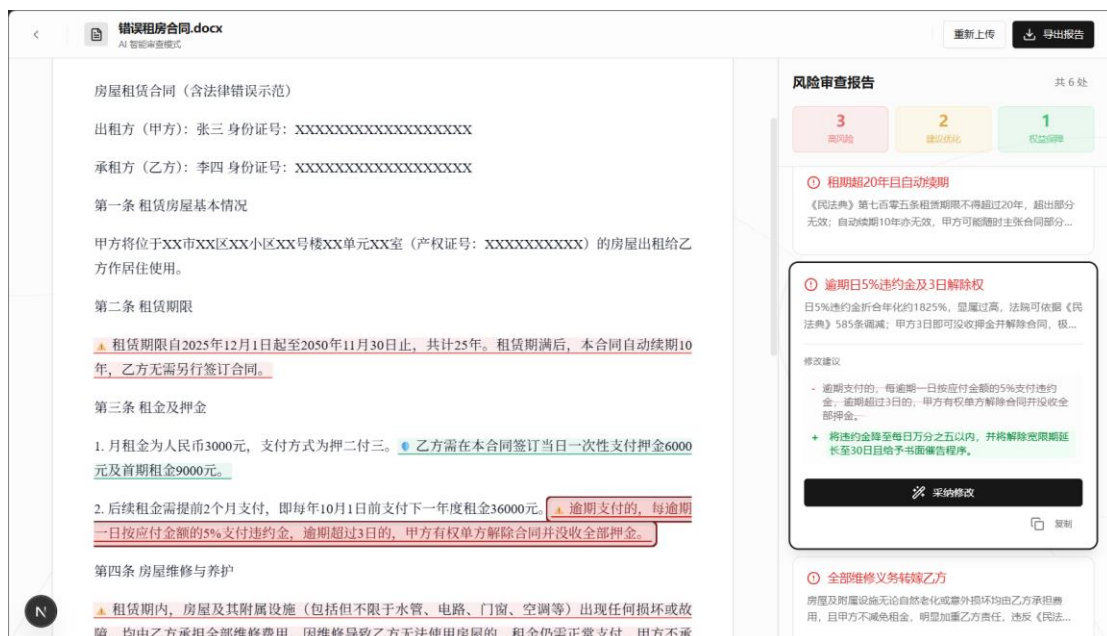


图 5-2 合同智能审查界面

为了实现灵活的分屏体验，系统引入了可拖拽分屏面板。外层容器使用 ResizablePanelGroup，左侧文档面板默认宽度比例设置为 70%，并设置最小宽度为 50%，以保证文档的可读性；右侧风险面板默认占比 30%。通过引入 react-resizable-panels 库并在中间设置 ResizableHandle 把手，系统能够监听鼠标拖拽事件，动态计算并更新左右两个区域的 flex-basis 属性，实现流畅的宽窄调整。

针对合同中的风险点，系统实现了风险详情浮窗。当用户点击文档中的高亮风险点时，浮窗会在附近弹出，显示风险等级及修改建议。为了防止浮窗遮挡关键内容或溢出屏幕，系统内置了一套智能定位算法。首先通过 getBoundingClientRect() 获取目标元素在视口中的坐标，然后计算元素下方和上方的剩余空间。如果下方空间充足，优先显示在下方；否则显示在上方。最后，算法会计算浮窗的水平中心点并进行边界检测，如果超出屏幕左右边界，则强制进行位移修正，确保浮窗内容始终完整可见。

5.2 系统管理员功能实现

管理后台采用了沉浸式仪表盘设计。顶部为“全息数据卡片”，实时展示用户总数及活跃

度；下方为“流体用户列表”，支持按角色筛选和搜索用户。系统管理员界面如图 5-3：



图 5-3 系统管理员界面

在数据展示方面，全息数据卡片开启了数字滚动动画属性。实现上使用了 framer-motion 动画库，当数据从后端 API 返回时，系统不直接显示最终数值，而是利用 useSpring 钩子将数值从 0 平滑过渡到目标值，增强了视觉交互感。在安全性方面，系统实现了严格的权限路由守卫。在页面组件挂载时，useEffect 钩子会调用 useAuth 方法检查当前用户的 role 字段。如果检测到当前用户角色不是管理员，系统会立即执行路由跳转，强制将用户重定向回首页，并显示“权限不足”的提示。在数据加载完成前的间隙，界面会显示全屏的骨架屏动画，避免页面出现空白闪烁，提升了系统的专业感。

六、总结

6.1 系统亮点

本系统成功构建了一个垂直领域的法律 AI 助手，其核心亮点在于深度融合了人工智能技术与法律业务场景。首先，在合同审查方面，系统创新性地采用了“分屏对照”与“风险智能分级”的交互模式，用户不仅能直观地看到合同原文与风险点的对应关系，还能直接获取 AI 生成的修改建议，极大地提升了合同审核的效率与准确性。其次，系统在技术架构上引入了检索增强生成（RAG）技术，通过挂载专业的法律法规向量知识库，有效解决了通用大模型在法律领域容易产生“幻觉”的问题，确保了法律咨询回复的专业性与严谨性。最后，在用户

体验层面，系统采用了现代化的流体动画与沉浸式仪表盘设计，摆脱了传统法律软件刻板、沉闷的界面风格，为用户提供了流畅、愉悦的操作体验。

6.2 存在的不足

尽管系统已实现了核心业务闭环，但在实际应用层面仍存在一定局限性。在文档解析方面，目前系统主要针对标准排版的 PDF 与 Word 文档进行了优化，对于包含复杂嵌套表格、模糊扫描件或手写签名的非结构化文档，OCR 识别的准确率仍有待提高。在交互深度上，目前的合同审查主要是一次性的“全检”模式，缺乏针对特定条款进行深入“多轮辩论”或“模拟谈判”的能力，难以应对极其复杂的商业博弈场景。此外，系统的权限体系目前较为扁平，仅区分了管理员与普通用户，尚未支持企业级的多层级组织架构与审批流管理。

6.3 改进方向

针对上述不足，未来的系统迭代将主要集中在以下三个方向。第一，增强多模态文档处理能力，引入更先进的 OCR 大模型，提升对复杂格式与手写字体的识别精度，确保“看得清”所有合同细节。第二，引入“人机协作”与“专家介入”机制，开发律师端工作台，允许专业律师对 AI 的审查结果进行复核与修正，并将修正数据回流至训练集，通过强化学习不断提升 AI 的判断力。第三，构建个性化知识库功能，支持企业用户上传内部的历史合同模板与合规指引，让 AI 能够学习特定企业的业务偏好，从而提供更加定制化、贴合企业实际的法律服务。