

智谱法章 -基于RAG的民法一体化平台





目录 Content

- ① 项 目 背 景
 Background
- ② 系 统 架 构
 Architecture
- 3 技 术 细 节
 Innovation
- 4 项 目 特 色 Popularize



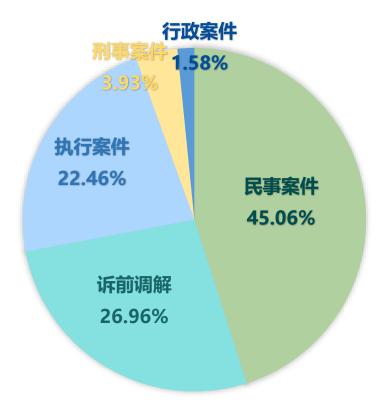








2023年人民法院审判工作主要数据



最高法发布2023年人民法院审判执行工作主要数据

全国法院收案4557.37万件,增长15.62%;

民事案件占比45.06%

收案2004.8万件, 增长10.55%; 结案1997.7万件, 增长8.39%。

基于此背景,我们决定开发一个民事案件智能类案检索与预测一

体化平台,旨在为法律从业者和普通公众提供一个方便、高效、

准确的法律服务工具。

PART 1 论文选题背景

竞品分析



产品	主要功能	特点
法律图书馆	法治新闻动态、阅读法律文书 新法规速递、法律法规检索	与法律数据库交互
正义网法律博客	多用户群体的法律交流平台	互动性好、用户群体广泛
中国裁判文书网	裁判文书查询和下载	检索功能强大
AI民法(本产品)	类案检索与智能推荐 基于RAG-LLM的案件智能审判	专业性强、互动性好 用户群体广泛、智能预判
DISC-LawLLM	法律咨询、法考助手、文书处理	简化文书工作、法考支持

痛点分析



非专业人士法律知识薄弱。 无法清晰陈述案件,难以利用法律专业名词进行检索。

检索难度大

法律领域专业 性强,壁垒高, 用户难以与法 律知识与实际 相联系,缺乏 沟通交流平台。

互动性不足

通用大模型缺乏法律领域专业知识,难以给出准确的答案,答案缺乏可解释性。

LLM幻觉严重

案件陈述文本 较长,且多以 **图片形式呈现。** 用户检索需手

动打字, 耗费

大量时间。

案件长文本



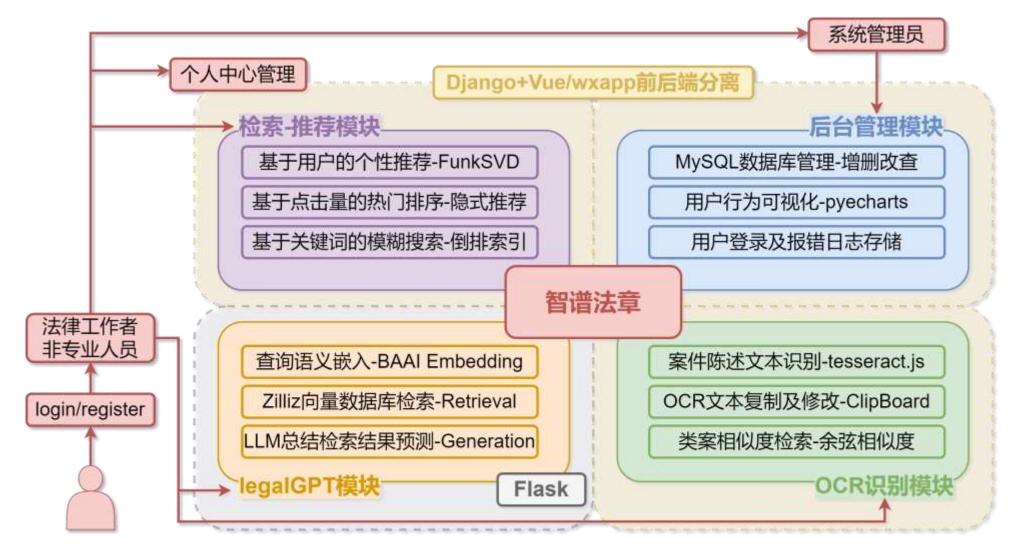




PART 2 系统架构

技术路线

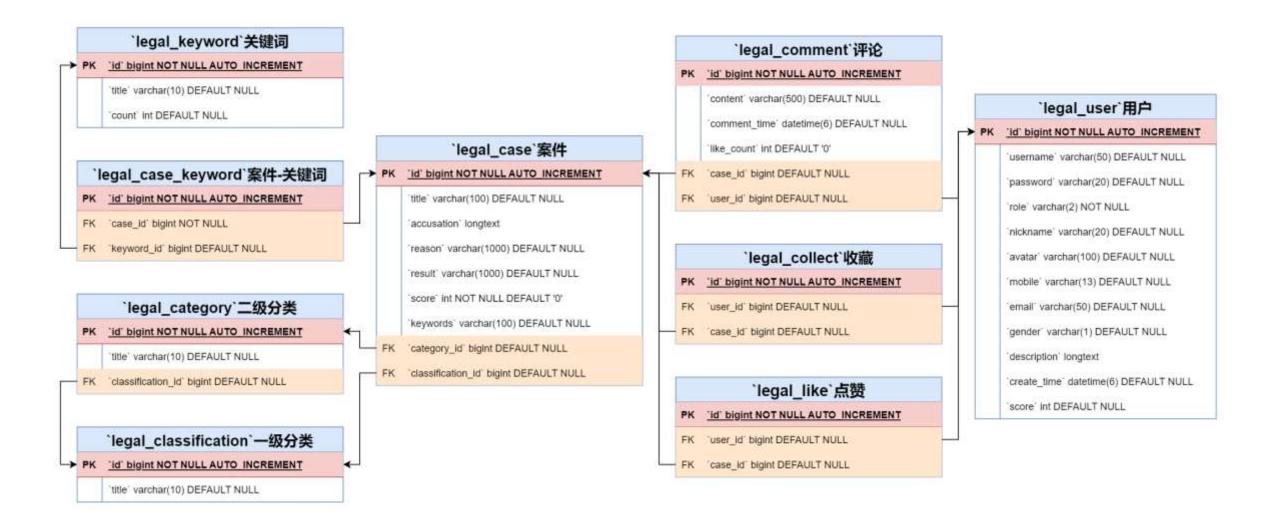




PART 2 系统架构

数据库设计









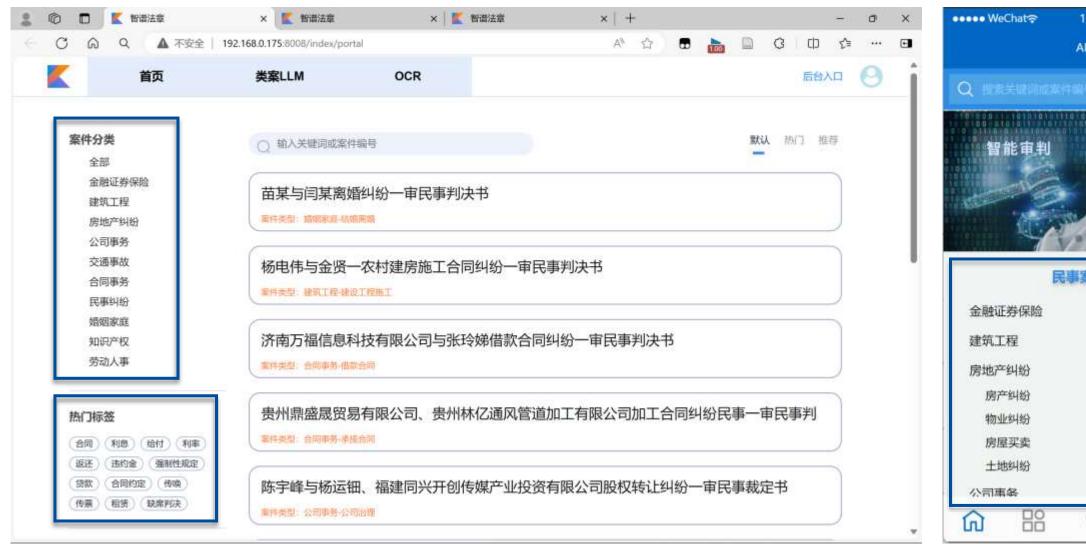






检索-推荐模块





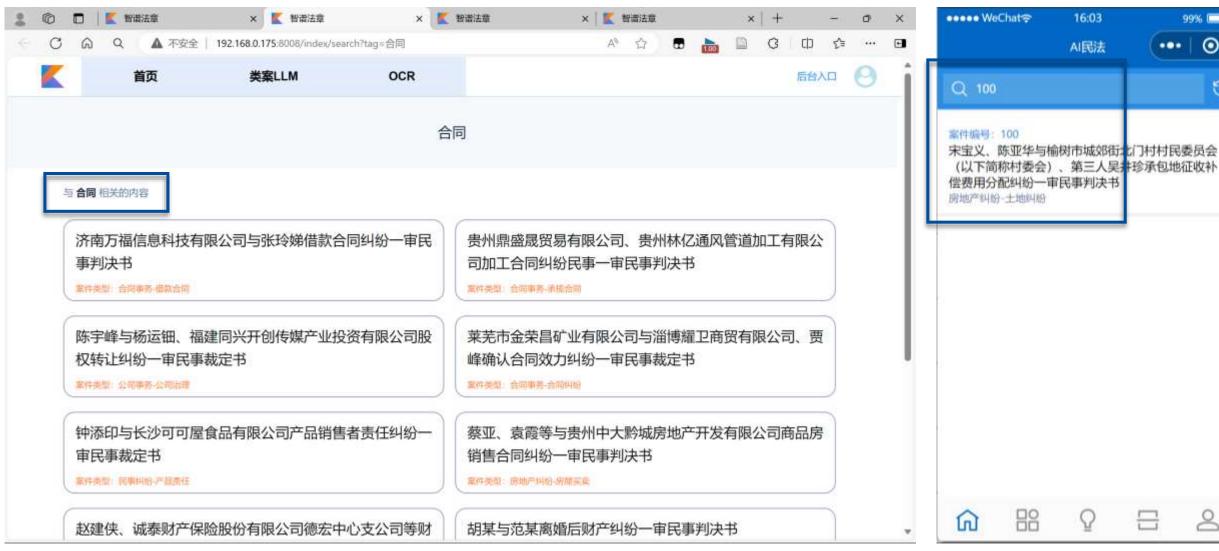




检索-推荐模块



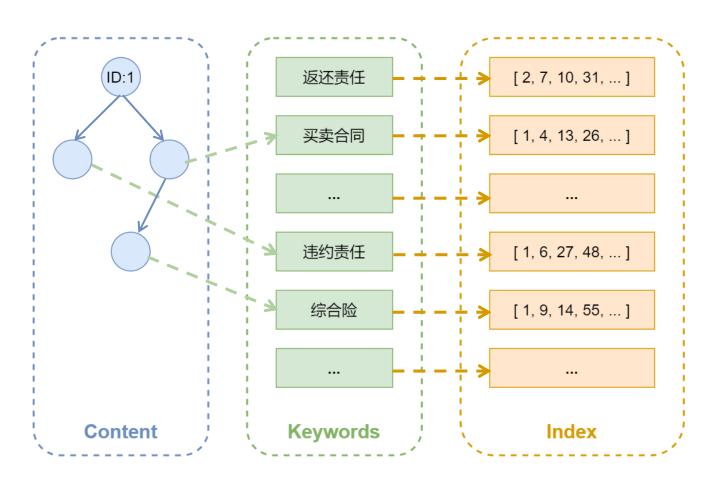
0





倒排索引





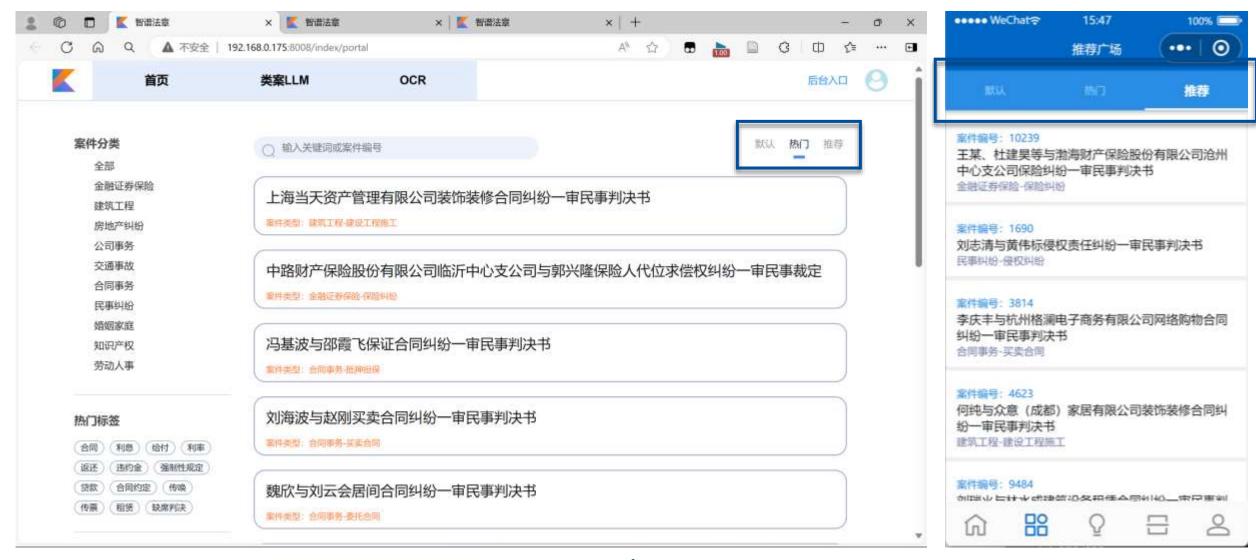
倒排索引: 文档文本提取关键词,建立 从关键词到文本的映射。这些映射关系按 照关键词进行排序,构建出一张**倒排列表**。

查询速度快: 当用户搜索某个关键字时,可以在有序的映射中快速定位到对应的文本,提高搜索效率。而传统索引算法通常需要遍历整个文档,搜索的时间成本较高。

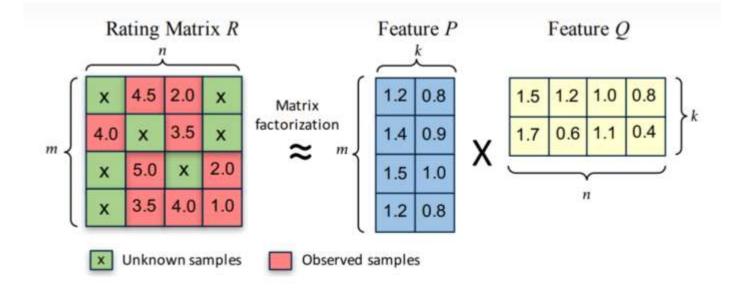
空间开销小: 无需存储每个文档的全部内容, 只保存了关键词与文档的映射关系。

检索-推荐模块





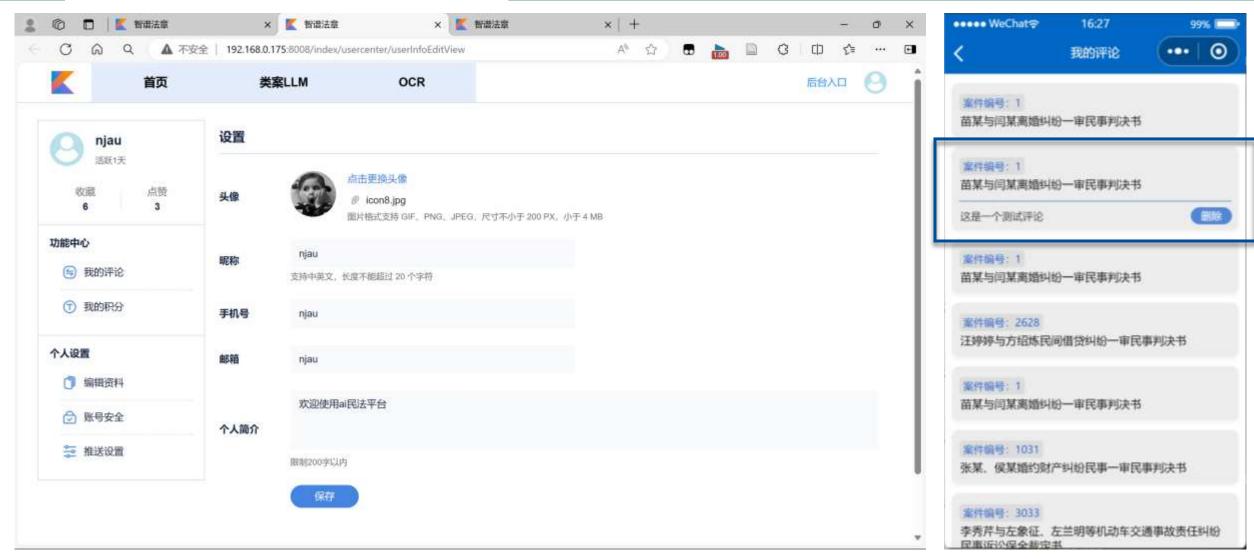




FunkSVD: 改进SVD的计算效率、数据 稀疏问题;将矩阵R分解为两个低维矩阵, 通过重构的低维矩阵预测用户对物品的评 分,目标函数中加入正则,控制模型方差。 核心思想:认为用户的兴趣只受少数几 个因素的影响,因此将稀疏且高维的 User-Item评分矩阵**分解为两个低维矩阵**, 由于用户和物品的特征向量维度比较低, 因而可以通过梯度下降的方法高效地求解。

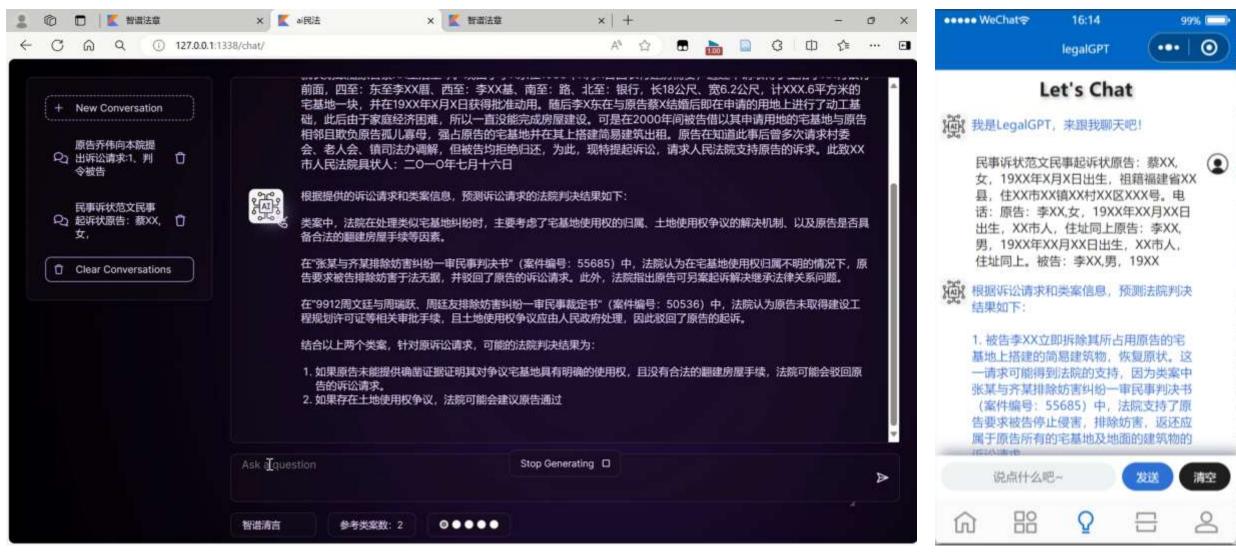
个人中心管理





legalGPT模块







Question A

在中国,结婚的法定年龄是多少岁? What is the legal age for marriage in China?

Chinese LLaMA + SFT

在中国境内,男性必须年满22周岁才能合法地登记结婚。女性则需要达到 21岁以上即可进行婚姻注册手续。

In China, males must be at least 22 years old to legally register for marriage. Females need to be above 21 years old to proceed with the marriage registration process.

Lawyer LLaMA

依据《民法典》第一千零四十七条规定,在中国,男性结婚的法定年龄为22周岁,女性结婚的法定年龄为20周岁。

According to Article 1047 of the Civil Code, in China, the legal age for men to get married is 22 years old, and for women, it is 20 years old...

幻觉 hallucination?

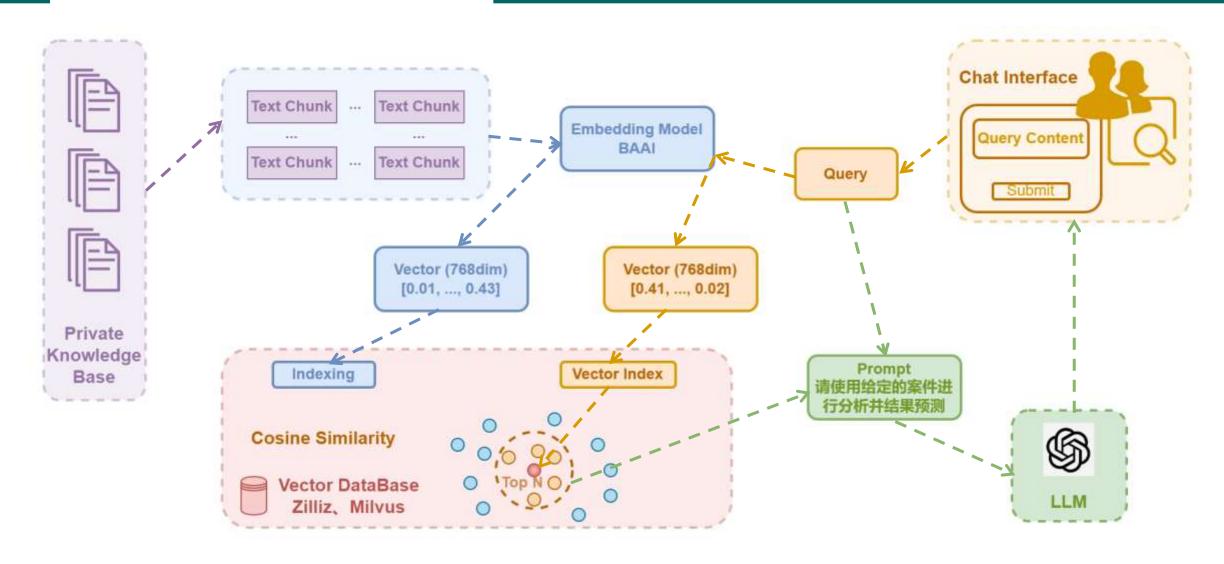
LLMs偶尔会产生偏离用户输入、 与先前生成的上下文相矛盾、与 既定的世界知识相抵触的内容。

H1: LLMs引用了与查询案件无 关的法律条文。

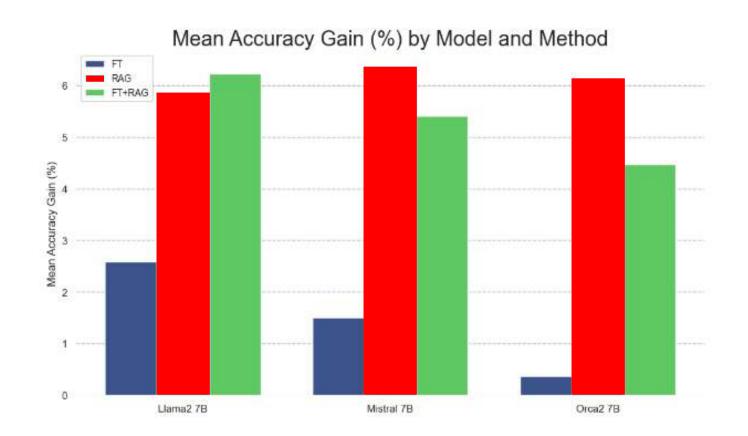
H2: LLMs编造了与现实法律相 违背的虚假条文或案件。

RAG-LLM





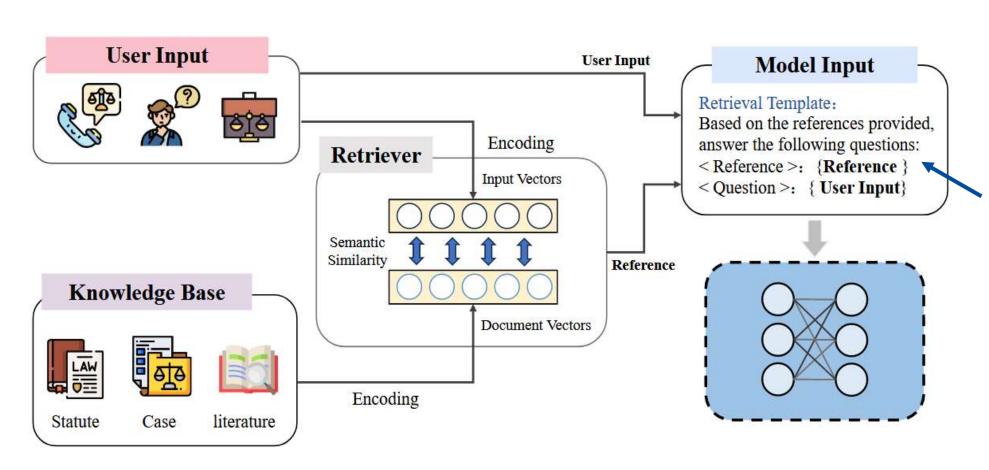




- ➤ RAG不仅为模型添加了知识,还融入了与问题相关的上下文,这是FT不具备的特征。
- ▶ 由于一定程度的灾难性遗忘, FT可能 会影响模型的其他能力。
- ➤ 无监督微调模型可能会受益于SFT或基于RL的微调的进一步对齐。

RAG-Prompt





不是所有检索结 果均为有用信息!

Prompt:

参考案件中可能存在与Query无关的 类案,回复时不要引用这些无关信息

OCR识别模块

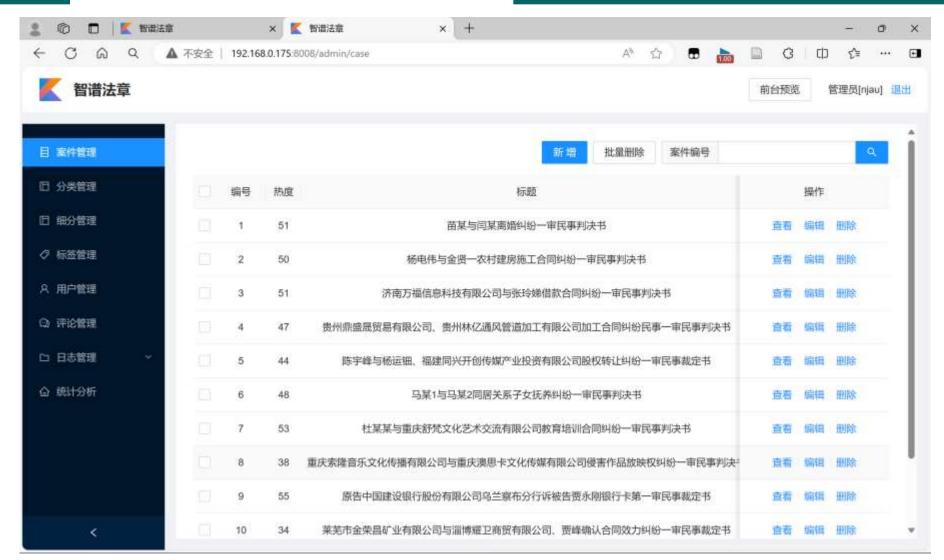


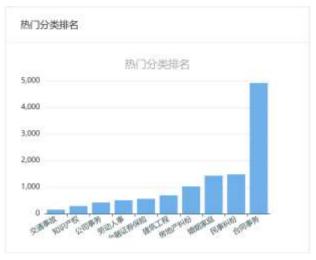


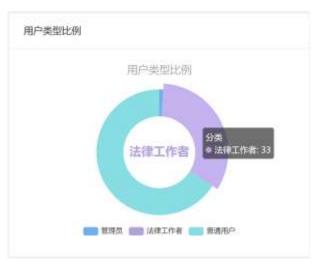


后台管理模块



















创新点与项目特色



关键词模糊搜索

精细化检索

隐式行为记录

个性化智能推荐

检索难度大

案件评论区

热门案件推荐

多途径分享

专业人才互动

互动性不足

类案检索

过滤无关案件

COT提示词

案件结果预测

LLM幻觉严重

合同文字识别

文本复制编辑

类案检索

案件结果预测

案件长文本



感谢各位评委老师聆听



