# 大模型标记模块marker说明文档

## 技术架构

### 依赖库：

json：用于处理 JSON 数据。

dotenv：用于加载环境变量。

openai：用于与 Deepseek API 进行交互。

### 环境变量：

使用 dotenv 加载环境变量，特别是 Deepseek API 的密钥。

### 功能模块：

信息提取：extract\_info\_from\_judgment 函数使用 Deepseek API 从法律文书中提取信息。它构建了一个系统提示和用户提示，并将其发送到 API，解析返回的 JSON 数据。

标记判决：mark\_judgment 函数调用信息提取函数并将提取的信息存储在全局变量 mark\_judgment\_info 中。

## 实现过程

### 开发环境，开发流程

开发环境主要包括以下几个方面：

编程语言：使用 Python 作为主要开发语言。

开发工具：使用 Visual Studio Code作为集成开发环境（IDE），以便于代码编写和调试。

依赖管理：使用 pip 和 requirements.txt 文件来管理项目依赖，确保环境的一致性。

版本控制：使用 Git 进行版本控制，便于团队协作和代码管理。

环境变量管理：使用 dotenv 库来管理敏感信息和配置，如 API 密钥。关键技术解析。

开发流程包括以下步骤：

需求分析：明确项目需求，确定功能模块和技术栈。

环境搭建：配置开发环境，安装所需的库和工具。

模块开发：按照功能模块进行代码编写，逐步实现各个功能。

测试：对每个模块进行单元测试，确保功能正常。

集成：将各个模块集成到主程序中，进行整体测试。

部署：将应用程序部署到服务器，供用户使用。

维护：根据用户反馈和需求变化，进行后续的维护和更新。

### 关键技术解析

OpenAI：本应用程序使用 OpenAI 调用deepseek API 来进行信息提取。通过构建系统提示和用户提示，应用程序能够向 API 发送请求，并获取法律文书中的关键信息。API 的响应以 JSON 格式返回，便于后续处理。

JSON 处理：使用 Python 的 json 库来解析和生成 JSON 数据。提取的信息以字典形式存储，方便后续的 CSV 导出和数据展示。

环境变量管理：使用 dotenv 库来加载环境变量，确保敏感信息（如 API 密钥）不被硬编码在代码中，提高安全性和灵活性。

### 腾讯云Ai代码助手在上述过程中的助力

 **智能代码优化**  
在代码编写和修改过程中，腾讯云 AI 助手能快速分析代码逻辑，提出优化建议，如改进变量命名、简化代码结构和移除冗余代码等，提升代码质量和可维护性。

 **快速问题定位**  
AI 助手可以通过自然语言对话协助排查代码中的逻辑错误、语法错误或依赖问题，减少调试时间。

 **代码补全与生成**  
根据需求快速生成代码片段，例如界面组件（如开关、输入框、按钮）以及复杂的函数逻辑，减少手动编写代码的负担。

 **文档与注释生成**  
自动生成清晰的代码注释，帮助团队成员理解项目逻辑。还可快速创建文档模板，记录接口设计、功能描述等信息。

 **集成调试工具**  
在 VSCode 环境中，AI 助手能够直接集成调试功能，实时检查代码运行状态，协助排查运行时错误。

 **项目加速与创新**  
提供 AI 算法和模型的推荐或实现方案，帮助项目实现复杂任务（如文本标记、数据分析）的快速迭代。

## 使用说明

该项目是一个基于 Gradio 的法律文书信息提取应用，用户可以通过输入判决书文本或上传文件来提取关键信息，并将结果导出为 CSV 文件。以下是详细的使用说明：

1. 环境准备

- 确保已安装 Python 3.6 及以上版本。

-克隆或下载项目代码。

-在项目根目录下创建一个虚拟环境并激活：



- 安装项目依赖：



2. 配置环境变量

在项目根目录下.env 文件中添加 Deepseek API 密钥：



3. 运行应用

在终端中运行以下命令启动应用：



- 应用将启动一个本地服务器，通常在 <http://127.0.0.1:7860>。

4. 使用界面

打开浏览器，访问应用的 URL。

在输入框中输入判决书文本，或通过“上传判决书”按钮上传 Word 文档（.doc 或 .docx 格式）。

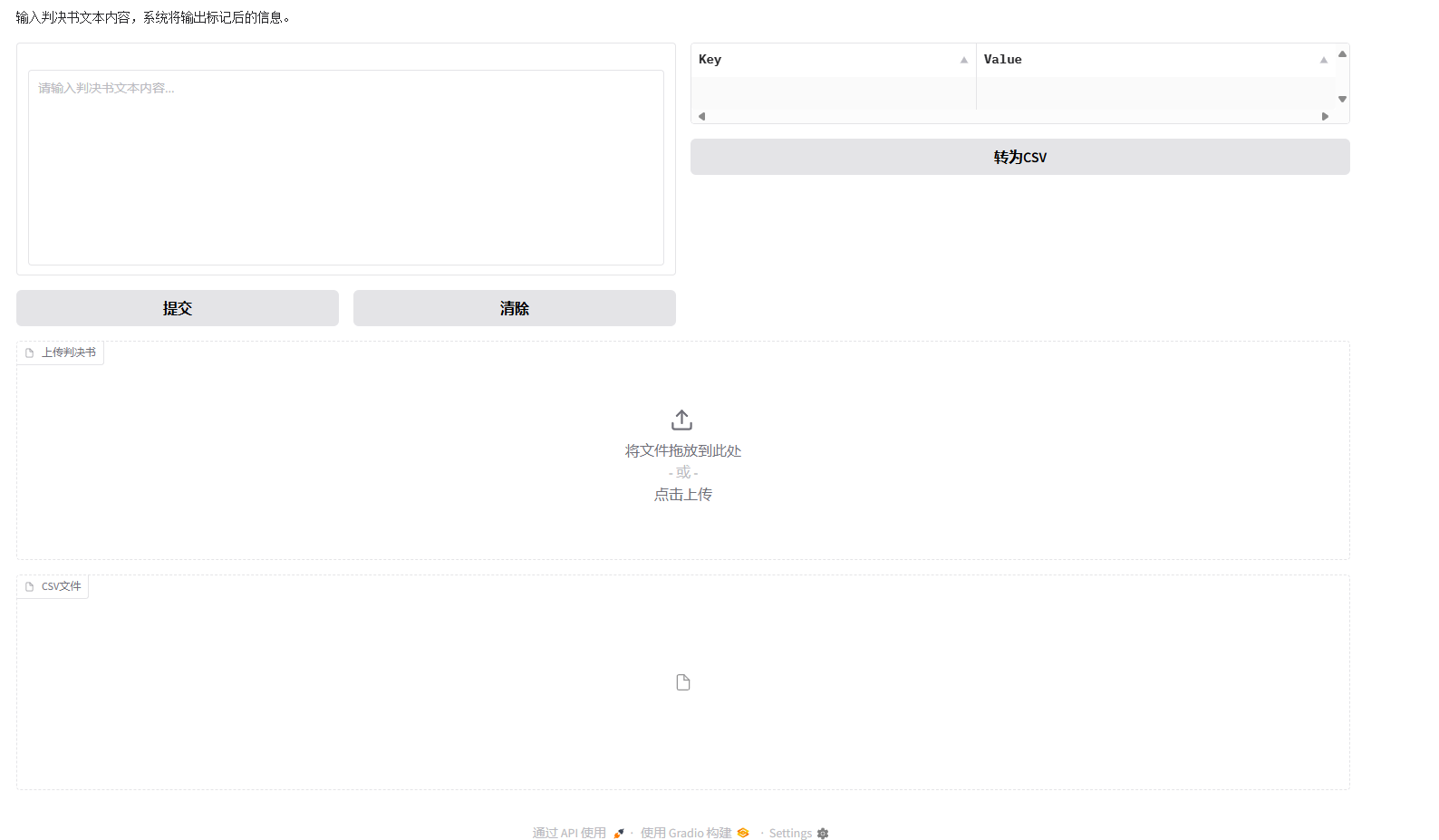
点击“提交”按钮，系统将提取法律文书中的关键信息，并在右方的数据表格中显示结果。

如果需要将提取的信息保存为 CSV 文件，点击“转为CSV”按钮，系统将生成 CSV 文件供下载。

点击“清除”按钮，可以清空输入框和输出结果。

## 效果展示

主界面



输入判决书

