## **一、概述**

### **1.1 目标**

开发一个具备对话功能的Web小界面，以提升用户体验和系统功能性。

### **1.2 背景**

将大模型交互功能作为一个产品，嵌入到CSM系统中，辅助回答用户提问，通过对话功能增强用户体验。该创新点提升我们产品的竞争力，因此需要开发这样一个Web小界面。

## **二、功能需求**

### **2.1 功能设计**

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 功能名称 | 业务功能说明 | 技术实现路线 |
| 1 | 知识查询 | 针对用户提出公司内部相关问题，从公司内部知识库中检索信息辅助大语言模型回答用户问题，生成答案。 | 1. 首先建立公司内部知识库（首次需要建立，后续不用） 2. 用户输入查询语句后，通过向量索引搜索topN个相关文档。 3. 将参考的相关文档内容和用户问题一起输入到大模型进行总结回答。 |
| 2 | 问答助手 | 回答用户关于通用领域的问题，给出答案 | 利用大模型自身的回答问题能力 |
| 3 | 写作助手 | 辅助用户写材料。 | 利用大模型自身的回答问题能力 |
| 4 | 文件解析助手 | 用户上传文件(pdf/docx/ppt)等格式，提取解析文件中的文字内容，利用到OCR文本识别。（仅支持对一个文件回答） | 接收上传的文档，将文档内容解析成文字，返回文本内容 |
| 5 | 文件对话助手 | 将用户上传的文件解析成文本后，辅助大模型回答用户提问的问题。 | 1. 将上传的文件识别，切分，插入到临时知识库。 2. 用户输入查询语句后，通过向量索引搜索topN个相关文档。 3. 将参考的相关文档内容和用户问题一起输入到大模型进行总结回答。 |
| 6 | 知识库管理(不面向用户) | 维护知识库内容，包含文件上传/删除/更新等，维护向量数据库 |  |

注意：第6个功能，不面向用户，面对管理人员维护知识库，知识库管理界面通过内部URL访问。在客户端不显示该URL信息。

### **2.2 基本界面设计**

界面应包括功能选择框、隐藏图标、对话区域、输入框、上传文件按钮、删除文件按钮和发送按钮。

1. 功能选择框

是一个选择框，包含（知识查询/问答助手/写作助手/文件解析助手/文件对话助手），每个选择对应不同的访问链接。(对应图AI助理位置)

1. 隐藏图标

点开后展开，显示 (清空对话/功能描述）

* 1. 清空对话：点击后清空对话区域内容。
  2. 功能描述：一个小界面，显示功能以及功能说明描述

1. 对话区域

位于界面中间。按照时间顺序显示所有消息，最新的消息显示在最下方。

1. 输入框

底部固定位置,有默认提示文字。用户可以在输入框中输入任意文本消息，

点击发送按钮后，消息会出现在对话区域内。

1. 上传文件按钮

点击后选择一个本地文件上传的功能。

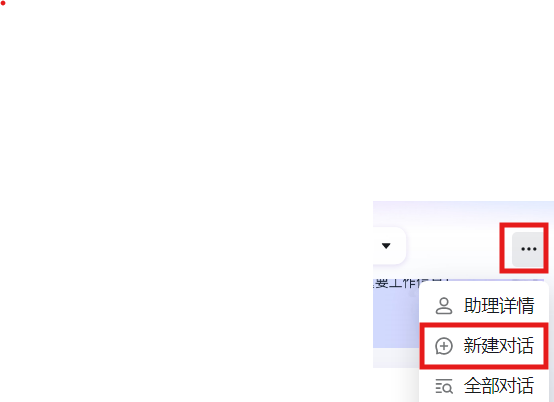
1. 删除文件按钮

针对已经上传的文件，点击X删除已经上传的文件。

1. 发送按钮

在输入框右侧，点击后将输入内容作为消息发送至对话区域。系统接收到用户的消息后，回复一条消息到对话区域。

前端界面示例



**2.3 具体功能界面显示**

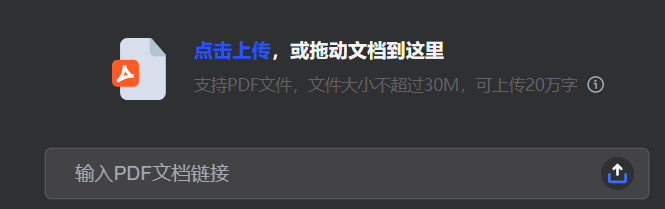
1 知识查询、问答助手、写作助手

其不显示上传文件按钮，点击发送后返回系统回答内容。



2 文件解析助手

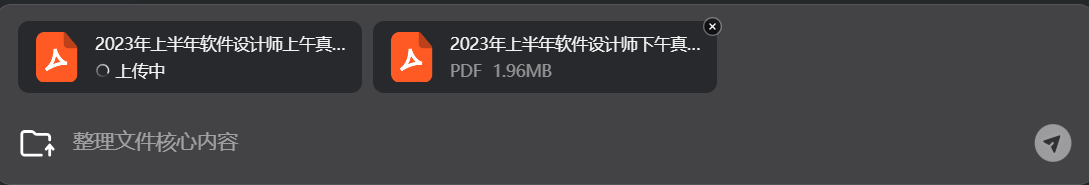
不显示输入框和发送按钮，只显示上传文件按钮，上传文件后直接返回识别内容。其中上传文件设置为单选。



3 文件对话助手

点击上传文件后上传多个文件，可以删除已经上传的文件。

用户在输入框输入问题后，点击发送可以直接回答。



## **三、技术需求**

### **3.1 技术栈**

前端：不限

后端：python、fastchat、huggface、langchain等框架集成

### **3.2 功能实现设计**

设计的5功能模块（知识查询/问答助手/写作助手/文件总结助手），有些模块有复用的情况，总结出有3个核心功能。

1大模型问答功能：利用大模型本身的能力，回答用户问题。

2文件解析功能：将文件解析成文字内容

3文件搜索功能：根据用户输入文本，在知识库中，通过向量索引的方式，搜索出top K个相关的文档。

### **3.3 前端交互基本功能**

1 用户提问回答功能

2 上传文件功能

3 删除文件功能

### **3.4 接口设计**

1 知识查询

|  |  |
| --- | --- |
| 请求方法 | POST |
| url | http://127.0.0.1:7861/chat/knowledge\_base\_chat |
| 参数 | {"query": "详细介绍以下数据安全评估方法",  "knowledge\_base\_name": "laws" //知识库名称} |
| 流式返回 | {"answer": "相关材料"} // 先返回answer, 最后返回docs  {"docs": []} |

2 问答助手

|  |  |
| --- | --- |
| 请求方法 | POST |
| url | http://127.0.0.1:7861/chat/llm\_chat |
| 参数 | {"query": "请用100字左右的文字介绍自己"} |
| 流式返回 | {"text": "信息和支持"} |

3 写作助手

|  |  |
| --- | --- |
| 请求方法 | POST |
| url | http://127.0.0.1:7861/chat/llm\_writer |
| 参数 | {"query": "请用100字左右的文字介绍自己"} |
| 流式返回 | {"text": "信息和支持"} |

4 文件解析助手 - 上传功能

|  |  |
| --- | --- |
| 请求方法 | POST |
| url | http://127.0.0.1:7861/extract/file\_upload |
| 参数 | 1732085999693 |
| 返回 | {"text": "信息和支持"} |

5 文件对话助手

上传文件

|  |  |
| --- | --- |
| 请求方法 | POST |
| url | http://127.0.0.1:7861/file/upload |
| 参数 | 1732085999693 |
| 返回 | {"msg": "上传成功"} |

删除文件

|  |  |
| --- | --- |
| 请求方法 | Del |
| url | http://127.0.0.1:7861/file/delete/{file\_name} |
| 参数 |  |
| 返回 | {"msg": "上传成功"} |

回答内容

|  |  |
| --- | --- |
| 请求方法 | POST |
| url | http://127.0.0.1:7861/chat/llm\_writer |
| 参数 | {"query": "请用100字左右的文字介绍自己",  “knowledge\_id”:} |
| 流式返回 | {"text": "信息和支持"} |