langchain 官网位于 python.langchain.com/en/latest/i… ，它可以简化对各种大语言模型的使用，比如内置有 OpenAI 等。

pdfplumber 顾名思义，是用于读取和处理 PDF 文件的，选择这库是因为今年还在更新，并且对中文的支持还不错。本次项目 demo 会导入自己的 PDF 文件，并将其作为知识库，回答你的提问。

python-dotenv 用于读取.env 文件，本例在该文件中放入 Open AI 平台的 key。

streamlit 用于绘制 UI 界面，当前大多数 ChatGPT 应用都使用它，如大名鼎鼎的 gpt4free，streamlit 默认会收集信息进行分析，可通过配置文件关闭，macOS 和 Linux 位于~/.streamlit/config.toml，Windows 位于%userprofile%/.streamlit/config.toml，添加如下内容即可：

```
[browser]
gatherUsageStats = false
```

其它配置项可通过 streamlit config show 可查看。

faiss-cpu 是 facebook 开源用于相似搜索的库，github.com/facebookres… ，GPU 版本请使用 faiss-gpu。

openai 和 tiktoken 都是调用 ChatGPT 接口时使用的。

保留当前使用版本请使用 pip freeze > requirements.txt。

OpenAI 的注册方式这里就不介绍了，最常用的是在 sms-activate.org/ 上购买服务获取短信验证码完成注册。

提取文本:

接下对文本进行分片，这里每个分片长充为 1000 字符，为保留上下文选择了重叠 200 字符：

接下来配置 embedding，也即将离散值转化为连续向量:

为界面添加一个输入框:

最后完成回复的逻辑: