

1. 绪论

1.1 项目背景、发展现状及意义和目标的理解

本小组致力于开发一款专注于思想政治教育领域的知识问答产品，本产品目前阶段基于本科五本思政教材，以帮助用户提升对思想政治学科等方面知识的理解。在这一领域，知识广泛而复杂，用户通常需要耗费大量时间和精力来获取所需的信息。通过结合先进的技术手段，我们的目标是提供一个高效、便捷的学习方式，使用户能够更轻松地获取思政教育领域的相关知识。

项目背景：

随着社会的不断发展，思想政治教育在培养公民素质、引导社会进步等方面扮演着重要的角色。然而，传统的学习方式可能受到时间和地点的限制，而我们的项目致力于利用先进的技术手段，让用户能够随时随地获取思政教育领域的知识，提高学习的便捷性和灵活性。

发展现状：

当前，虽然存在着大量的思政教育资源，但用户在获取信息时仍然面临一系列挑战，例如信息碎片化、获取途径有限等。想要翻找教材费时费力。我们的项目通过结合先进的大模型技术，借助paddleocr+langchain建立本地知识库，并利用向量相似度进行问题匹配，以提高用户问题获取答案的准确性和效率。同时，我们通过对用户问题建立索引，找到相关书籍和页码，以更全面地满足用户的需求。

意义和目标：

1. **高效获取信息：** 借助我们的产品，用户可以更迅速、方便地获取思政教育领域的相关知识，提高学习效率。
2. **个性化学习：** 我们的系统通过向量相似度匹配，能够根据用户的需求提供个性化的学习内容，增强学习的针对性。
3. **推动思政教育发展：** 通过提供高效的学习方式，我们希望能够促进思政教育领域的现代化发展，推动学科的深入研究和应用。
4. **整合教育资源：** 我们的项目旨在整合思政教育领域的各类资源，为用户提供全面、多样化的学习支持。

预期效果：

通过我们的项目，预计能够为想要学习思想政治的学生、教师和研究者提供一个更加便捷、高效的学习工具，推动思政教育的现代化，促进学科的发展。

1.2 项目实施技术路线及方案

项目灵感来自于：b站技术up主同济子豪兄的思政助手，项目链接如下：<https://github.com/TommyZihao/ChatMarx>。项目视频展示为：<https://www.bilibili.com/video/BV1E94y1v7aa>，在线demo为：<https://chatmarx.zihaoai.cn>。可惜的是这个项目并没有开源代码。本次项目相当于对同济子豪兄的思政助手的一次复现工作。

借助前沿的人工智能大语言模型和中文自然语言处理技术，本项目把大学本科的五本思政教材：《马克思主义原理》、《思想道德修养与法律基础》、《中国近代史纲要》、《习近平新时代中国特色社会主义思想》、《毛泽东思想及中国特色社会主义》，做成了类似ChatGPT的问答AI，可以“又红又专”地回答你的各种问题。

它利用了langchain技术建立本地知识库，通过向量相似度对用户提出的问题进行相似度匹配，从而快速找到用户需要的信息。同时，我们还利用向量相似度对用户提出的问题建立索引，找到最可能的书籍和对应的页码。最后，我们接入大模型，利用知识库匹配用户问的问题，从而找到相似度最高的问题，并且送入大模型中，大模型经过整合从而进行回答。我们的产品主要面向各领域人群，包括教师、学生、爱好者等。我们相信，我们的产品能够帮助他们更加方便和快速地获取他们所需要的信息。

以下是我们产品实现方案的详细说明。

1. 问题搜索：用户可以在搜索框中输入他们感兴趣或者需要了解的问题，系统会自动进行相似度匹配，并且给出相应的答案。
2. 图书检索：用户通过输入问题，系统经过分析后，会检索相关书籍，找到相应页码。
3. LLM答疑：用户输入相关问题，在本应用中，首先会通过向量相似度搜索与问题相似度最高的K个问题，结合用户的问题，一同输入的LLM大模型中，在本应用中使用chatglm-6B模型。

3. 数据预处理

3.1 langchain加载pdf

langchain内置了文档加载器，文档加载器可以从许多不同的来源加载文档。LangChain提供了超过100种不同的文档加载器，以及与空间中其他主要提供商的集成，如AirByte和Unstructured。LangChain提供了与所有类型的文档（HTML、PDF、代码）从各种地方（私有存储、公共网站）加载的集成。

langchain的pdf加载器使用 PyPDF 将 PDF 加载为文档数组，其中每个文档包含页面内容和带有页码、来源的元数据。如下所示：

```
[Document(page_content=' 第三节人民群众在历史发展中的作用 169第三节人民群众在历史发展中的作用唯物史观有时又被称为“群众史观”。把人民群众放在至高无上的地位，无私无我地服务和奉献于人民群众，体现了马克思主义政党和马克思 主义者的情怀和境界。那么，人民群众在社会发展中的作用如何来体现？ 历史人物在社会发展中的作用又如何来理解？一、人民群众是历史的创造者谁是历史的创造者？怎样看待人民群众和个人的历史作用？这是人类 历史上长期困扰人们的难题。唯物史观第一次科学地回答了这些问题。（一）两种历史观在历史创造者问题上的对立唯物史观与唯心史观的对立，在历史创造者问题上表现为群众史观与 英雄史观的对立。马克思主义哲学产生以前，占统治地位的历史观是英雄 史观。这种历史观从社会意识决定社会存在的基本前提出发，否认物质资 料的生产方式是社会发展的决定力量，抹杀人民群众的历史作用，宣扬少 数英雄人物创造历史。英雄史观的产生有其深刻的认识根源、社会历史根源和阶级根源。从 认识根源看，英雄史观之所以产生，主要是因为人们的认识停留在历史现 象的表面，把活跃在历史前台的少数英雄人物的作用尤其是他们的意识的 作用加以夸大并绝对化，而忽视广大人民群众及其历史活动的作用。从社 会历史根源看，英雄史观的产生同社会生产力水平较低，大多数人从事物 质资料的生产活动，少数人从事政治统治、垄断精神文化生活有关。广大 人民群众在私有制社会处于被支配的地位，受剥削、受压迫，其历史创造 性得不到充分发挥和社会应有的承认，而少数剥削阶级则掌握经济、政治 权力，高高在上、独断专行，似乎可以随心所欲地主宰历史。这就为英雄'， metadata={'source': './pdfdata/《马克思主义基本原理》（2023版）.pdf', 'page': 178, 'score': 0.6385535}),
```

使用langchain的pdf加载器可以提取出文档的页码，方便实现本项目的学霸、老师功能。

```
pdfRoot = "./pdfdata"
pdf_files_name = os.listdir(pdfRoot)

# 加载pdf
loaders = []
for pdf_file_name in pdf_files_name:
    loaders.append(PyPDFLoader(pdfRoot+"/"+pdf_file_name))

# 将pdf拆分成每一页，得到pagecontent和metedata
documents = []
for loader in loaders:
    documents += loader.load_and_split()
```

加载得到documents列表，包含有五本思政书的全部内容。是以页号为单位拆分的。

3.2 数据清洗

PyPDF 会为pdf文档的换行自动加上换行符\n，但是\n显然不是原文内容，所以需要被清除掉。在这里使用python提供的replace函数将文本内的\n去除掉

```
# 数据清洗，消除字符串里面的'\n'
for i in range(len(documents)):
    documents[i].page_content = documents[i].page_content.replace('\n', '')
```

3.3 将每一页拆分为chunk

加载了文档后，通常会将加载好的文档进行转换。比如将长文档拆分为较小的块，以适应模型的上下文窗口。LangChain具有许多内置的文档转换器，非常适合进行拆分、合并、过滤以及其他文档操作。本项目只运用了其中的拆分。

文本分割器的工作原理如下：

1. 将文本分割成小的、语义上有意义的块（通常是句子）。
2. 开始将这些小块组合成一个较大的块，直到达到某个大小（由某个函数测量）。
3. 一旦达到该大小，将该块作为自己的文本片段，然后开始创建一个新的文本块，其中包含一些重叠（以保持块之间的上下文）。

本项目使用的是langchain提供的RecursiveCharacterTextSplitter。该文本分割器接受一个字符列表作为参数。它尝试根据第一个字符创建块，但如果某些块太大，它会继续尝试下一个字符，依此类推。本项目设定的块大小为800、重叠字符为100、拆分的字符为常见的中文标点符号：",", "。", "，", "？", "！", "，", "。"。

```
# 将每一份文档拆分成若干个chunk
text_splitter = RecursiveCharacterTextSplitter(
    # Set a really small chunk size, just to show.
    chunk_size = 800,
    chunk_overlap = 100,
    separators=["", "。", "，", "？", "！", "，", "。"]
)

splitted_documents = text_splitter.split_documents(documents)
```

4. text2text文本生成

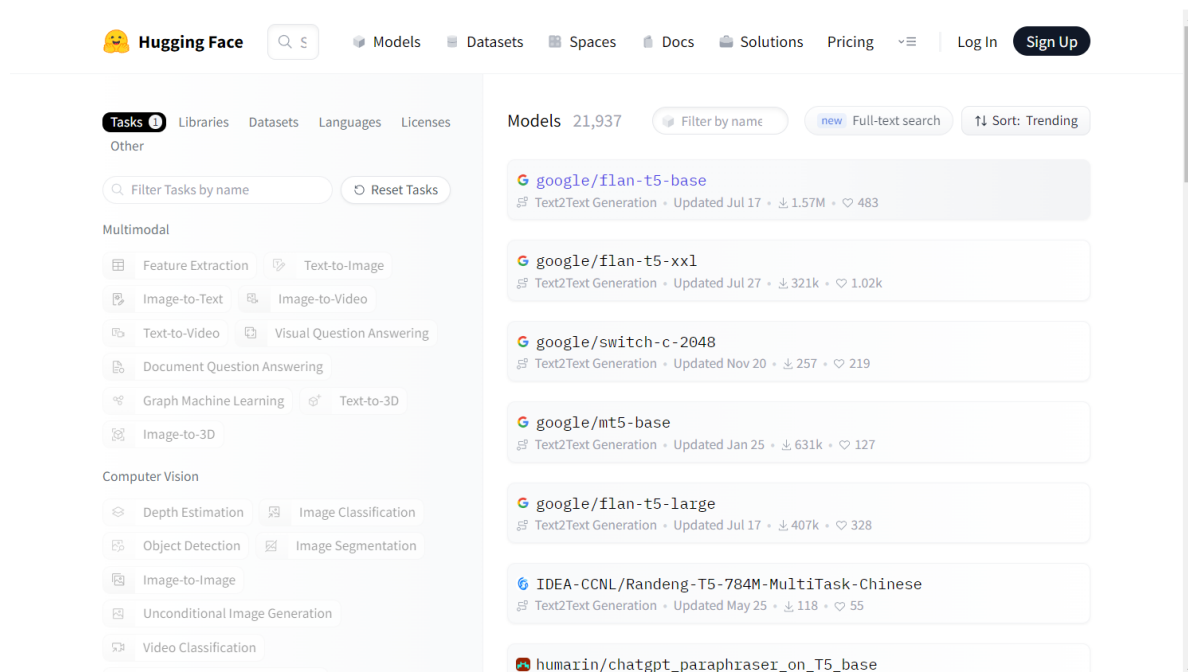
4.1 概述

Text-to-text文本生成是一种自然语言处理（NLP）任务，旨在将输入文本转换为相关的输出文本。与传统的单一目标任务（如机器翻译、文本摘要）不同，text-to-text生成模型具有更广泛的应用范围，因为它们可以适应多种自然语言处理任务。

这种方法的核心思想是采用端到端的学习方式，使用深度神经网络来处理输入文本并生成对应的输出文本。模型的输入和输出都以文本形式呈现，这使得模型在各种任务上的表现更具通用性。text-to-text生成模型通常基于强大的预训练语言模型，如GPT（Generative Pre-trained Transformer）系列。

text-to-text生成可以应用于多个领域，包括机器翻译、文本摘要、问题回答、对话生成等。模型的灵活性和通用性使得其能够适应新任务，只需提供相应的训练数据和适当的目标函数。这种方法的优势在于减少了针对每个任务单独设计模型的复杂性，提高了模型的可迁移性和可扩展性。

这类模型通常训练和推理的计算复杂度大、占用内存、显存、磁盘空间大，因此也被叫做大模型。或者叫做大语言模型（LLM/Large Language Model）。目前开源平台上有大量的模型已发布，截止目前（2023年12月28日），hugging face上共有21937个开源大模型。



根据效果（尤其是在中文文本上的效果）和推理速度以及占用空间综合考虑，本项目运用了清华大学和智谱ai研发的chatglm-6B，其int4量化版本只需要占用6G的显存，在3060显卡上就能跑起来。

4.2 配置chatglm

下载项目

```
git clone https://github.com/THUDM/ChatGLM-6B.git
```

进入项目：

```
cd ChatGLM-6B
```

下载依赖：

```
pip install -r requirements.txt
```

上传模型权重到服务器

本项目选用的是int4量化后的权重：<https://huggingface.co/THUDM/chatglm-6b-int4>

修改api.py

修改模型为int4量化后的版本，再将模型权重修改成本地权重所在的位置：

```
# 修改前
#tokenizer = AutoTokenizer.from_pretrained("THUDM/chatglm-6b",
trust_remote_code=True)
# model = AutoModel.from_pretrained("THUDM/chatglm-6b",
trust_remote_code=True).half().cuda()

#修改后
tokenizer = AutoTokenizer.from_pretrained("THUDM/chatglm-6b",
trust_remote_code=True)
model = AutoModel.from_pretrained("THUDM/chatglm-6b",
trust_remote_code=True, resume_download=True).quantize(4).half().cuda()
```

运行api.py在后台启动chatglm

```
python api.py
```

当看到如下界面时，说明chatglm在后台已启动成功，本机的任何进程都可以通过8000端口调用chatglm的服务。

```
(base) root@08acfb91e9d3:~# python ChatGLM-6B/api.py
Explicitly passing a `revision` is encouraged when loading a model with custom code to ensure no malici
ibuted in a newer revision.
Explicitly passing a `revision` is encouraged when loading a configuration with custom code to ensure n
en contributed in a newer revision.
Explicitly passing a `revision` is encouraged when loading a model with custom code to ensure no malici
ibuted in a newer revision.
No compiled kernel found.
Compiling kernels : /root/.cache/huggingface/modules/transformers_modules/chatglm-weight/quantization_k
Compiling gcc -O3 -fPIC -pthread -fopenmp -std=c99 /root/.cache/huggingface/modules/transformers_module
zation_kernels_parallel.c -shared -o /root/.cache/huggingface/modules/transformers_modules/chatglm-weig
_parallel.so
Load kernel : /root/.cache/huggingface/modules/transformers_modules/chatglm-weight/quantization_kernels
Setting CPU quantization kernel threads to 32
Using quantization cache
Applying quantization to glm layers
No compiled kernel found.
Compiling kernels : /root/.cache/huggingface/modules/transformers_modules/chatglm-weight/quantization_k
Compiling gcc -O3 -fPIC -pthread -fopenmp -std=c99 /root/.cache/huggingface/modules/transformers_module
zation_kernels_parallel.c -shared -o /root/.cache/huggingface/modules/transformers_modules/chatglm-weig
_parallel.so
Load kernel : /root/.cache/huggingface/modules/transformers_modules/chatglm-weight/quantization_kernels
Setting CPU quantization kernel threads to 32
INFO: Started server process [943]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)
```

4.3 langchain接入chatglm

```
from langchain.llms import ChatGLM
llm = ChatGLM(
    endpoint_url='http://127.0.0.1:8000',
    max_token=80000,
    top_p=0.9
)
```

后续可以利用llm这个对象来方便地调用后台运行的chatglm的服务。

5. 文本向量化

5.1 概述

文本向量化是将文本数据转换为数值向量的过程，以便计算机能够理解和处理文本信息。这是自然语言处理（NLP）和机器学习领域中的关键步骤，因为计算机通常需要数值输入来执行各种任务，如文本分类、情感分析、机器翻译等。文本向量化的主要目的是将文本数据转换为向量形式，以便计算机能够利用数学和统计方法进行分析 and 处理。文本向量通常是高维的，其中每个维度对应于语料库中的一个单词或词嵌入模型中的一个特定特征。

文本向量化有多种方法：

- **Bag of Words (BoW)**: 将文本视为一组单词的集合，忽略其顺序，创建一个向量，其中每个元素表示相应单词的出现频率。
- **TF-IDF (Term Frequency-Inverse Document Frequency)**: 结合词频和逆文档频率，强调在当前文档中频繁出现但在整个语料库中罕见的单词。
- **Word Embeddings**: 使用词嵌入模型（如Word2Vec、GloVe、FastText）将单词映射到高维空间中的向量，以捕捉单词之间的语义关系。
- **Doc2Vec**: 扩展Word Embeddings到整个文档级别，为整个文档生成向量表示。

当获得了文本的向量表示后，这意味着我们可以在向量空间中通过简单的线性运算（如计算余弦相似度）来实现复杂的功能，比如诸如语义搜索之类的操作，可以帮助我们在向量空间中寻找最相似的文本片段。

LangChain中的基本Embeddings类提供两种方法：一种用于嵌入文档，另一种用于嵌入查询。前者接受多个文本作为输入，而后者接受单个文本。之所以将它们作为两种不同的方法，是因为一些嵌入提供者对文档（用于搜索）和查询（搜索查询本身）有不同的嵌入方法。

SentenceTransformers是一个集成了计算超过100种语言的句子/文本嵌入的框架，在论文《Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks》中有详细描述。langchain就是基于SentenceTransformers框架实现文本向量化的。该框架支持大量state-of-art text2vec模型，截止目前共支持124个模型。

Models 124

Sort: Recently updated

sentence-transformers/all-MiniLM-L12-v2

Sentence Similarity • Updated Nov 16 • 476k • 81

sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2

Sentence Similarity • Updated Nov 2 • 352k • 337

sentence-transformers/paraphrase-multilingual-mpnet-base-v2

Sentence Similarity • Updated Nov 2 • 272k • 180

sentence-transformers/distiluse-base-multilingual-cased-v2

Sentence Similarity • Updated Nov 2 • 279k • 102

sentence-transformers/all-mpnet-base-v2

Sentence Similarity • Updated Nov 2 • 8.07M • 525

sentence-transformers/msmarco-distilbert-cos-v5

Sentence Similarity • Updated Nov 2 • 3.15k • 7

性能最好的是facebook的llama系列的all-mpnet-base-v2。于是本项目也使用all-mpnet-base-v2完成文本向量化。

All models

Model Name	Performance	Performance	Avg. Performance	Speed	Model Size
	Sentence Embeddings (14 Datasets)	Semantic Search (6 Datasets)			
all-mpnet-base-v2	69.57	57.02	63.30	2800	420 MB
multi-qa-mpnet-base-dot-v1	66.76	57.60	62.18	2800	420 MB
all-distilroberta-v1	68.73	50.94	59.84	4000	290 MB
all-MiniLM-L12-v2	68.70	50.82	59.76	7500	120 MB
multi-qa-distilbert-cos-v1	65.98	52.83	59.41	4000	250 MB
all-MiniLM-L6-v2	68.06	49.54	58.80	14200	80 MB
multi-qa-MiniLM-L6-cos-v1	64.33	51.83	58.08	14200	80 MB

5.2 langchain接入文本向量化模型

首先，下载并解压all-mpnet-base-v2模型权重

```
!wget -O llama_langchain.zip https://bj.bcebos.com/v1/ai-studio-online/5e882eb5d58648658950dd6ac9afb5ced4ca7a30e1f8414ba711967980b05a12?responseContentDisposition=attachment%3B%20filename%3Dllama_langchain%20.zip
!unzip ./llama_langchain.zip

embeddings = HuggingFaceEmbeddings(model_name="./llama_langchain/all-mpnet-base-v2",
                                     model_kwargs={'device':"cuda"})
```

如图所示，all-mpnet-base-v2 将每段文本编码为了一个768维的稠密向量

```
ans=embeddings.embed_documents("其实。现在全专业劝退的原因，很简单。这代学生没生对时间时代已经过饱和了。这")
print(len(ans[0]))
```

768

6. 向量数据库

6.1 概述

向量数据库是一种专为处理向量数据而设计的数据库系统。与传统的关系型数据库不同，向量数据库专注于存储、检索和处理高维向量。这种数据库类型在处理涉及大规模向量数据集的应用中表现出色，如图像识别、自然语言处理、推荐系统等。

以下是向量数据库的一些关键特点和优势：

1. **高维向量存储：** 向量数据库能够高效地存储和管理高维度的向量数据，这对于表示复杂的特征或模式非常重要，比如图像、音频和文本数据。
2. **相似性搜索：** 向量数据库的强项之一是能够执行高效的相似性搜索。这使得用户可以通过查询相似向量找到与给定向量相近的数据，为实时推荐和检索提供了强大的支持。
3. **向量索引：** 向量数据库采用专门设计的索引结构，以支持高效的向量检索。常见的索引技术包括树结构（如KD树、球树）和哈希函数等。
4. **实时性能：** 对于需要快速响应时间的应用，向量数据库通常能够提供实时的查询和检索性能。这对于在线推荐系统和搜索引擎等实时应用非常关键。
5. **可扩展性：** 随着数据量的增长，向量数据库通常具有良好的可扩展性，能够适应大规模向量数据的存储和处理需求。
6. **应用领域：** 向量数据库广泛应用于人工智能和机器学习领域，包括人脸识别、语音识别、文本相似性分析、推荐系统等。这些应用通常涉及大规模的向量数据集。

对于本项目而言，向量数据库是一个存储五本切片后的思政文档的场所，用户询问的query被embedding为向量后，在向量数据库中快速查找到最相似的几个文档，并被返回。

目前市面上有大量开源的向量数据库、langchain支持的有谷歌的chroma、Facebook的Faiss。Facebook AI Similarity Search (Faiss) 是用于高效相似性搜索和密集向量聚类的库。它包含了在任意大小的向量集合中进行搜索的算法，甚至可以处理可能不适合在内存中的向量集。本项目使用Faiss作为向量数据库。

6.2 langchain构建知识库（向量数据库）

```
db = FAISS.from_documents(splitted_documents, embeddings)
```

此时db包含了此前文档转换成的embedding，只是以一种方便检索的方式组织了起来。可以利用它检索与query最相近的几个文档。

6.3 持久化存储

向量数据库可以持久化存储，而不必每次都得加载pdf数据并生成embedding。

```
embeddings = HuggingFaceEmbeddings(model_name="./llama_langchain/all-mpnet-base-v2",
                                     model_kwargs={'device':"cuda"})

if not os.path.exists('./vector_store'):
    db = FAISS.from_documents(splitted_documents, embeddings)
    db.save_local('./vector_store')
else:
    db = FAISS.load_local('./vector_store', embeddings)
```

7. 基于检索增强生成的知识问答

7.1 RAG (Retrieval Augmented Generation, 检索增强生成)

LLM 在回答问题或生成文本时，先会从大量文档中检索出相关的信息，然后基于这些信息生成回答或文本，从而提高预测质量。RAG 方法使得开发者不必为每一个特定的任务重新训练整个大模型，只需要外挂上知识库，即可为模型提供额外的信息输入，提高其回答的准确性。RAG模型尤其适合知识密集型的任务。

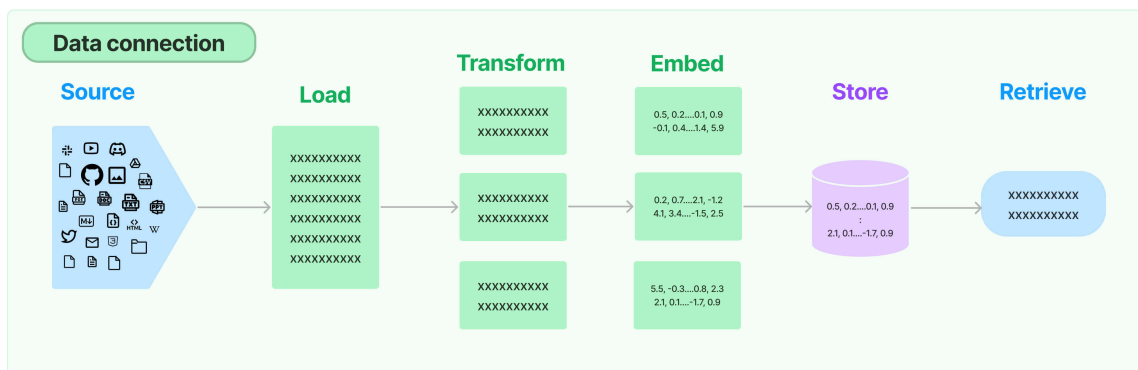
在 LLM 已经具备了较强能力的基础上，仍然需要 RAG，主要有以下几点原因：

- **幻觉问题**：LLM 文本生成的底层原理是基于概率的 token by token 的形式，因此会不可避免地产生“一本正经的胡说八道”的情况。
- **时效性问题**：LLM 的规模越大，大模型训练的成本越高，周期也就越长。那么具有时效性的数据也就无法参与训练，所以也就无法直接回答时效性相关的问题，例如“帮我推荐几部热映的电影？”。
- **数据安全问题**：通用的 LLM 没有企业内部数据和用户数据，那么企业想要在保证安全的前提下使用 LLM，最好的方式就是把数据全部放在本地，企业数据的业务计算全部在本地完成。而在线的大模型仅仅完成一个归纳的功能。

实际上，对于 LLM 存在的上述问题，SFT（supervised Fine-Tuning、有监督微调）是一个最常见最基本的解决办法，也是 LLM 实现应用的基础步骤。但是 sft 对于频繁更改的数据源来说可能不太实用，因为每进行一次微调都需要更改模型权重，这个过程计算复杂度大，且有灾难性遗忘的风险。同时 SFT 就像一个黑匣子，使得响应背后的推理更加不透明。需要准备和整理高质量的训练数据集、定义微调目标以及相应的计算资源。而 RAG 则不存在以上问题，总结一下 RAG 有如下优点：

- **可扩展性(Scalability)**：减少模型大小和训练成本，并允许轻松扩展知识
- **准确性 (Accuracy)**：模型基于事实并减少幻觉
- **可控性 (Controllability)**：允许更新或定制知识
- **可解释性 (Interpretability)**：检索到的项目作为模型预测中来源的参考
- **多功能性 (Versatility)**：RAG 可以针对多种任务进行微调和定制，包括 QA、文本摘要、对话系统等。

这使得 RAG 成为了最适合本项目的方法。



7.2 langchain实现RAG

首先构建prompt：

```
QA_CHAIN_PROMPT = PromptTemplate.from_template("""根据下面的上下文（context）内容回答问题。
如果你不知道答案，就回答不知道，不要试图编造答案。
答案最多3句话，保持答案简介。
总是在答案结束时说”谢谢你的提问！”“
{context}
问题: {question}
""")
```

langchain的向量数据库可以成为retriever，传入llm和prompt可以优雅的实现RAG

```

retriever = vector_store.as_retriever()
qa = RetrievalQA.from_chain_type(
    llm=llm,
    retriever=retriever,
    verbose=True,
    chain_type_kwargs={"prompt": QA_CHAIN_PROMPT}
)

ans = qa.run("什么是社会主义? ")
print(ans)

```

8. 基于gradio的前端快速生成

8.1 gradio概述

Gradio (Gradient Interactive Dashboard) 是一个用于快速创建机器学习模型交互界面的开源Python库。它的设计目标是使机器学习模型的部署和分享变得简单，无需深度的前端开发经验。Gradio的主要优势在于其简单易用的界面，用户可以通过几行代码快速创建一个交互式的界面，用于展示和测试机器学习模型。在以前，ai研究人员要想通过webui的方式展示机器学习模型，需要学习flask作为后端和前端三件套+vue构建前端，通过定义接口实现前后端的通信，这个过程费时费力，偏离了ai研究人员的主线工作。因此本项目使用Gradio构建前端展示界面，而无需写一行前后端代码。

8.2 学霸模式网页生成

ChatInterface制定了聊天窗口样式，其中的参数定义了聊天窗口的标题、描述等界面。

```

import gradio as gr
def good_learner(query):
    ans=vector_store.similarity_search(query)
    return "emmm, 我想了一会, 我感觉这个知识点可能位于教材{}的第{}页,\
我背给你听嗽, 好好记住吧! 下次别问我了。这部分内容是: \n\
【{}】".format(ans[0].metadata['source'].split('/')[0].split('.')[0],ans[0].metadata['page']+1,ans[0].page_content)

def good_learner_chatbot(message, history):
    return good_learner(message)

gr.ChatInterface(
    good_learner_chatbot,
    chatbot=gr.Chatbot(height=300),
    textbox=gr.Textbox(placeholder="我是学霸, 快来问我问题吧!", container=False,
scale=7),
    title="思政大学霸",
    description="快来问学霸问题吧!",
    theme="soft",
    examples=["什么是社会主义?", "马克思、恩格斯", "封建社会"],
    cache_examples=True,
    retry_btn=None,
    undo_btn="Delete Previous",
    clear_btn="Clear",
).launch(share=True)

```

8.3 教师模式网页生成

```
def teacher(query, k=10):
    embedding_vector = embeddings.embed_query(query)
    anss = vector_store.similarity_search_by_vector(embedding_vector, k=k)

    result_string = "同学啊，我刚才翻了翻书，你提到的这个问题与以下内容有关\n\n"

    for index, ans in enumerate(anss):
        result_string += "{}、{}的第{}页：【{}】\n\n".format(index + 1,
                                                                ans.metadata['source'].split('/')[0],
                                                                ans.metadata['page'] + 1,
                                                                ans.page_content)

    result_string += "好好学习，未来是你们的！"

    return result_string

def teacher_chatbot(message, history):
    return teacher(message)

gr.ChatInterface(
    teacher_chatbot,
    chatbot=gr.Chatbot(height=300),
    textbox=gr.Textbox(placeholder="我是老师，快来问我问题吧！", container=False,
scale=7),
    title="思政老师",
    description="快来问老师问题吧！",
    theme="soft",
    examples=["什么是社会主义？", "马克思、恩格斯", "封建社会"],
    cache_examples=True,
    retry_btn=None,
    undo_btn="Delete Previous",
    clear_btn="Clear",
).launch(share=True)
```

8.4 教授模式网页生成

```
def professor(query):
    ans = qa.run(query)
    return ans

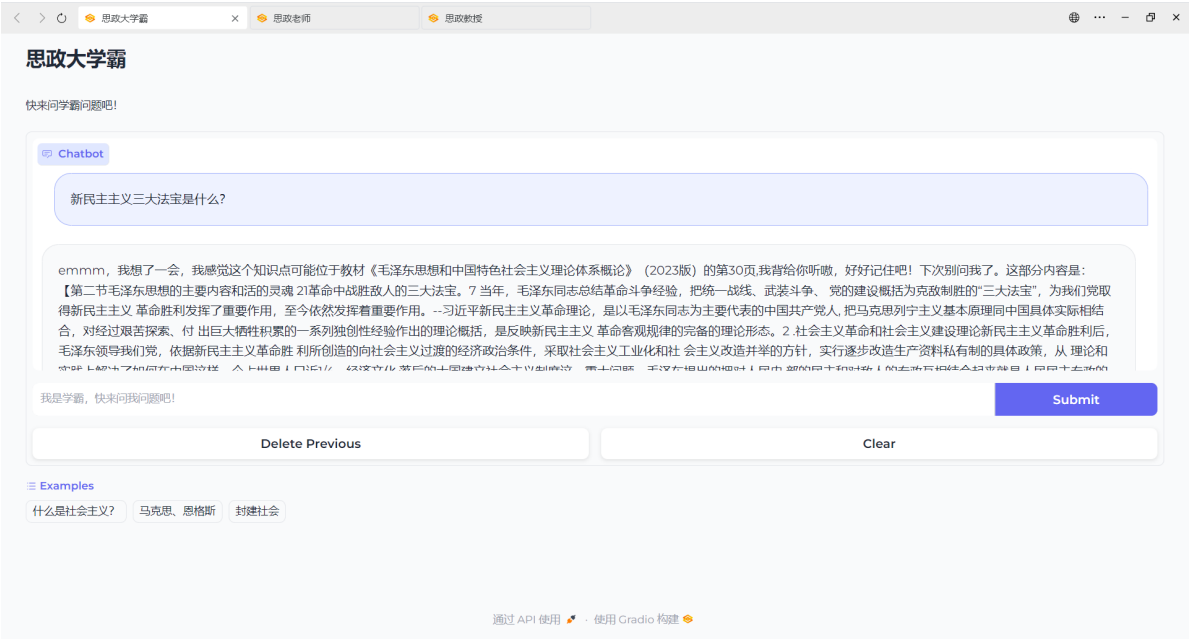
def professor_chatbot(message, history):
    return professor(message)

gr.ChatInterface(
    professor_chatbot,
    chatbot=gr.Chatbot(height=300),
    textbox=gr.Textbox(placeholder="我是教授，快来问我问题吧！", container=False,
scale=7),
```

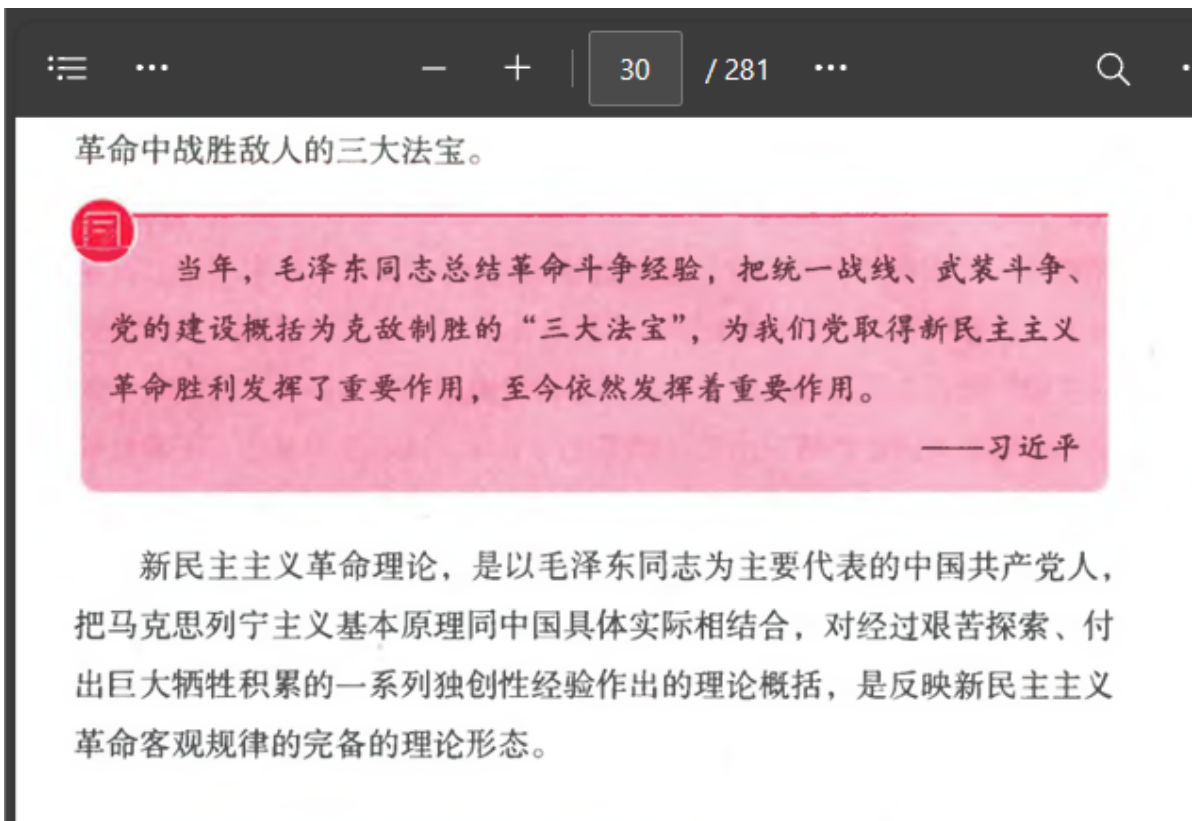
```
title="思政教授",
description="快来问教授问题吧！",
theme="soft",
examples=["什么是社会主义？","马克思、恩格斯","封建社会"],
cache_examples=True,
retry_btn=None,
undo_btn="Delete Previous",
clear_btn="Clear",
).launch(share=True)
```

9. 系统测试

9.1 学霸模式系统测试



查找《毛概》第30页内容：



准确无误

9.2 教师模式系统测试



完整截图如下：

同学啊，我刚才翻了翻书，你提到的这个问题与以下内容有关

1、《毛泽东思想和中国特色社会主义理论体系概论》（2023版）的第30页：【第二节毛泽东思想的主要内容和活的灵魂 21革命中战胜敌人的三大法宝。7 当年，毛泽东同志总结革命斗争经验，把统一战线、武装斗争、党的建设概括为克敌制胜的“三大法宝”，为我们党取得新民主主义革命胜利发挥了重要作用，至今依然发挥着重要作用。--习近平新民主主义革命理论，是以毛泽东同志为主要代表的中国共产党人，把马克思列宁主义基本原理同中国具体实际相结合，对经过艰苦探索、付出巨大牺牲积累的一系列独创性经验作出的理论概括，是反映新民主主义革命客观规律的完备的理论形态。2 社会主义革命和社会主义建设理论新民主主义革命胜利后，毛泽东领导我们党，依据新民主主义革命胜利所创造的向社会主义过渡的经济政治条件，采取社会主义工业化和社会主义改造并举的方针，实行逐步改造生产资料私有制的具体政策，从理论和实践上解决了如何在中国这样一个占世界人口近1/4、经济文化落后的国家建立社会主义制度这一重大问题。毛泽东提出的把对人民内部的民主和对敌人的专政互相结合起来就是人民民主专政的理论，丰富了马克思列宁主义关于无产阶级专政的学说。在社会主义制度建立以后，毛泽东提出把马克思列宁主义基本原理同中国具体实际进行“第二次结合”，领导全党和全国人民积极探索适合中国国情的社会主义建设道路，提出了一系列具有战略意义的正确思想和方针。其中包括：关于社会主义社会是一个很长的历史阶段，可分为不发达的社会主义和比较发达的社会主义两个阶段，后一阶段可能比前一阶段需要更长的时间，社会主义建设具有艰难性、复杂性和长期性的思想；关于社会主义社会仍然存在矛盾，基本的矛盾仍然是生产关系和生产力之间的矛盾、上层建筑和经济基】

2、《马克思主义基本原理》（2023版）的第272页：【，促使资产阶级在吸取和总结社会主义国家经验的基础上对资本主义制度进行改良。例如，重视国家对经济的干预，实行一定的计划化管理；重视职工参与管理，实行经济民主等。最后，主张改良主义的政党对资本主义制度的改革，也对资本主义发生变化起到了重要作用。第二次世界大战结束后，一些改良主义政党在英、法、联邦德国等发达资本主义国家相继获得执政地位。这些政党在执】

3、习近平新时代中国特色社会主义思想概论的第181页：【课后思考 181主、程序民主和实质民主、直接民主和间接民主、人民民主和国家意志相统一。要继续推进全过程人民民主建设，不断健全人民当家作主的制度体系，全面发展协商民主，积极发展基层民主。巩固和发展最广泛的新时代爱国统一战线，铸牢中华民族共同体意识，实现海内外中华儿女大团结大联合。课后思考1.为什么我国不能搬来一座政治制度上的“飞来峰”？2.怎样理解全过程人民民主是最广泛、最真实、最管用的民主？3.如何健全人民当家作主的制度体系？4.如何理解统一战线是凝聚人心、汇聚力量的强大法宝？微信扫描上方二维码，获取更多免费资料】

4、《毛泽东思想和中国特色社会主义理论体系概论》（2023版）的第116页：【，同国民党进行第三次合作的策略。1958年10月，毛泽东表示：台湾如果回归祖国，可以照他们（指蒋介石等）自己的方式生活；台湾的军队也可以保存，我们不压迫他裁兵，不要他搞政，让他搞“三民主义”。此后，党对解决台湾问题又提出了许多重要原则。1960年5月，】

5、《毛泽东思想和中国特色社会主义理论体系概论》（2023版）的第254页：【第二节科学发展观的主要内容 245治文明建设很难搞好，而且会给党和人民事业带来损害，影响社会政治稳定，影响党和国家长治久安。”⑥要充分发挥我国社会主义政治制度优越性，积极借鉴人类政治文明有益成果，绝不照搬西方政治制度模式，不搞多党轮流执政，不搞“三权鼎立”和两院制。以保证人民当家作主为根本，以增强党和国家活力、调动人民积极性为目标，扩大社会主义民主，加快建设社会主义法治国家，发展社会主义政治文明，更加注重改进党的领导方式和执政方式，更加注重健全民主制度、丰富民主形式，更加注重发挥法治在国家治理和社会管理中的重要作用，保证人民依法享有广泛权利和自由。按照建立中国特色社会主义行政体制目标，深入推进政企分开、政资分开、政事分开、政社分开，建设职能科学、结构优化、廉洁高效、人民满意的服务型政府。健全制约和监督体系，坚持用制度管权管事管人，推进权力运行公开化、规范化，让人民监督权力，让权力在阳光下运行。巩固和发展最广泛的爱国统一战线。统一战线是凝聚各方面力量，促进政党关系、民族关系、宗教关系、阶层关系、海内外同胞关系的和谐，夺取中国特色社会主义新胜利的重要法宝。要高举爱国主义、社会主义旗帜，巩固统一战线思想政治基础，正确处理一致性和多样性的关系。全面贯彻党的宗教工作基本方针，依法管理宗教事务，发挥宗教界人士和信教群众在促进经济社会发展中的积极作用。坚持充分尊重、广泛联系、加强团结、热情帮助、积极引导的方针，鼓励和引导新的社会阶层人士为中国特色社会主义事业作出更大贡献。落实党的侨务政策，支持海外侨胞、归侨侨眷关心和参与祖国现代化建设与和平统一大业。三、推进社会主义文化强国建设建设社会主义文化强国，是我们党把握时代和形势发展变化，积极回①《胡锦涛文选》第二卷，人民出版社2016年版，第32页。】

6、《中国近现代史纲要》（2023版）的第330页：【，切实依法履行人大监督职责。认真做好人大讨论决定重大事项工作，更好发挥国家权力机关职能作用。代表工作不断深化和拓展，代表法和有关制度得到更加全面的贯彻落实，社情民意表达和反映渠道更加畅通。社会主义协商民主广泛多层制度化发展。协商民主是中国特色社会主义民主政治的特有形式和独特优势，是切实保障人民当家作主的制度安排。】

7、《毛泽东思想和中国特色社会主义理论体系概论》（2023版）的第75页：【66 第二章新民主主义革命理论矛盾，决定了中国革命的首要任务是推翻帝国主义、封建主义和官僚资本主义的统治，从根本上推翻反动腐朽的政治上层建筑，变革阻碍生产力发展的生产关系。新民主主义革命总路线的内容是无产阶级领导的，人民大众的，反帝、反封建、反官僚资本主义的革命。新民主主义的政治、经济、文化纲领是新民主主义革命总路线的展开和具体化，指明了新民主主义革命的发展方向。中国革命必须走农村包围城市、武装夺取政权的革命道路。统一战线、武装斗争、党的建设是新民主主义革命的三大法宝，是新民主主义革命胜利的基本经验。思1 什么是新民主主义革命的总路线？2 新民主主义基本纲领的主要内容是什么？3 如何认识中国革命走农村包围城市、武装夺取政权道路的必要性及重大意义？4 如何理解新民主主义革命的三大法宝及其相互关系？阅读书目1.毛泽东：《中国社会各阶级的分析》，《毛泽东选集》第一卷，人民出版社1991年版。2.毛泽东：《星星之火，可以燎原》，《毛泽东选集》第一卷，人民出版社1991年版。3.毛泽东：《实践论》，《毛泽东选集》第一卷，人民出版社1991年版。4.毛泽东：《（共产党人）发刊词》，《毛泽东选集》第二卷，人民出版社1991年版。5.毛泽东：《新民主主义论》，《毛泽东选集》第二卷，人民出版社1991年版。6.毛泽东：《论联合政府》，《毛泽东选集》第三卷，人民出版社1991年版。】

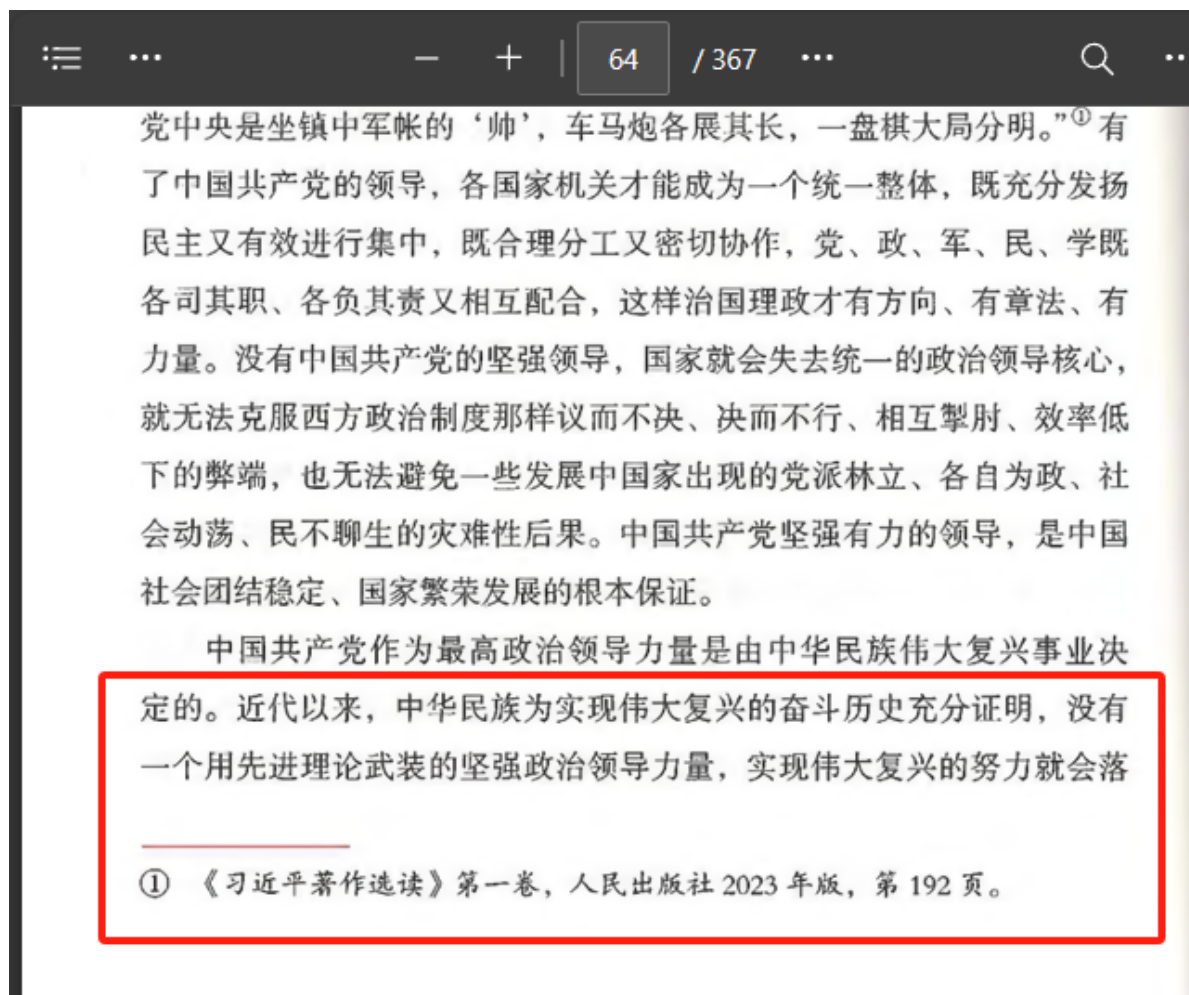
8、习近平新时代中国特色社会主义思想概论的第64页：【，中华民族为实现伟大复兴的奋斗历史充分证明，没有一个用先进理论武装的坚强政治领导力量，实现中华民族伟大复兴的努力就会落④《习近平著作选读》第一卷，人民出版社2023年版，第192页。】

9、《毛泽东思想和中国特色社会主义理论体系概论》（2023版）的第222页：【第二节“三个代表”重要思想的主要内容 213建设社会主义政治文明，必须进行政治体制改革。政治体制改革是社会主义政治制度的自我完善和发展。推进政治体制改革，要有利于增强党和国家的活力，体现和发挥社会主义制度的特点和优势，充分调动人民群众的积极性创造性，维护国家统一、民族团结和社会稳定，促进经济发展和社会全面进步。要坚持从我国国情出发，总结实践经验，同时借鉴人类政治文明的有益成果，绝不照搬西方政治制度的模式。政治体制改革要着重加强制度建设，实现社会主义民主政治的制度化、规范化和程序化。建设社会主义政治文明，必须推进决策的科学化民主化。要改革和完善决策机制，健全深入了解民情、充分反映民意、广泛集中民智、切实珍惜民力的决策机制，推进决策科学化民主化。行政管理体制和机构改革是政治体制改革的重要内容，是推动我国上层建筑更好地适应经济基础的一项重要制度的建设和创新，也是建立和完善社会主义市场经济体制的客观需要。要进一步转变政府职能，改进管理方式，提高行政效率，努力形成行为规范、运转协调、公正透明、廉洁高效的行政管理体制。要深化干部人事制度改革，努力形成广纳群贤、人尽其才、能上能下、充满活力的用人机制，把优秀人才集聚到党和国家的各项事业中来。建设社会主义政治文明，必须尊重和保障人权。新中国成立特别是改革开放以来，我国经济社会发展有力地推动了人权事业的发展。中国所焕发出的巨大活力，是中国人民拥有广泛自由、民主的生动写照。我们尊重国际社会关于人权的普遍性原则，但普遍性原则必须与各国国情相结合。实现人权的根本途径是经济社会发展和社会进步。对于发展中国家，生存权、发展权是最基本最重要的人权。要根据自己的国情把集体人权和个人人权，经济、社会、文化权利和公民政治权利一起来加以推进。人权是一个国家主权范围内的问题。人权要靠主权来保护，不是人权高于主权】

10、《毛泽东思想和中国特色社会主义理论体系概论》（2023版）的第220页：【第二节“三个代表”重要思想的主要内容 211一步走向成熟和完善。建设社会主义政治文明，是我国改革开放和社会主义现代化建设发展的必然要求，是我们党领导人民坚持和发展人民民主长期实践的必然结论，进一步深化了我们党对中国特色社会主义事业的规律性认识。建设社会主义政治文明，发展社会主义民主政治，是社会主义现代化建设的重要目标。必须在坚持四项基本原则的前提下，继续积极稳妥地推进政治体制改革，扩大社会主义民主，健全社会主义法制，建设社会主义法治国家，巩固和发展民主团结、生动活泼、安定和谐的政治局面。建设社会主义政治文明涉及政治思想、政治制度、行政管理、法制建设等方面，是一个系统工程。建设社会主义政治文明，最根本的就是要坚持党的领导、人民当家作主和依法治国的有机统一。这是我们推进政治文明建设必须遵循的基本方针，也是我国社会主义政治文明区别于资本主义政治文明的本质特征。党的领导是人民当家作主和依法治国的根本保证，人民当家作主是社会主义民主政治的本质要求，依法治国是党领导人民治理国家的基本方略。建设社会主义政治文明，必须发展社会主义民主。没有民主就没有社会主义，就没有社会主义的现代化。社会主义愈发展，民主也愈发展。发展社会主义民主政治，就是要健全民主制度，丰富民主形式，扩大公民有序的政治参与，保证人民依法实行民主选举、民主决策、民主管理和民主监督，享有广泛的权利和自由，把广大人民群众的积极性和主动性充分调动起来。建设社会主义政治文明，必须坚持和完善中国特色社会主义政治制度。必须坚持人民代表大会制度，要保证人民通过人民代表大会行使国家权力，保证人民代表大会及其常委会依法履行职能，加强立法和监督工作，使立法和决策更好地体现人民的意志。必须坚持中国共产党领导的多党合作和政治协商制度，要加强同民主党派合作共事】

好好学习，未来是你们的！

对第8个内容进行检验：



在《习思想》第65页确实发现第八个内容

9.3 教授模式系统测试



新民主主义三大法宝确实是统一战线、武装斗争、党的建设，回答正确。

10. 项目心得与收获

