

# Walkthrough of LLM

Code written by Virginia Sun, MD. Last edited on June 10th, 2024

## Import packages

We recommend using the packages listed in `requirements.txt` to ensure that the code runs properly.

```
In [1]: from langchain.callbacks.manager import CallbackManager
from langchain.callbacks.streaming_stdout import StreamingStdOutCallbackHand
from langchain_community.llms import Ollama
import pyreadr
import os
from langchain_community.document_loaders import PyPDFLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain_community.vectorstores import Chroma
from langchain_community.embeddings import LlamaCppEmbeddings
from langchain.chains import LLMChain
from langchain_community.llms import LlamaCpp
from langchain.chains import RetrievalQA
from langchain.embeddings.sentence_transformer import SentenceTransformerEmb
from langchain import hub
import numpy as np
import pandas as pd
from langchain_core.output_parsers import StrOutputParser
from langchain_core.runnables import RunnablePassthrough
from langchain_core.runnables import RunnableParallel
from langchain.prompts import ChatPromptTemplate
from langchain.prompts import PromptTemplate
from langchain.prompts import HumanMessagePromptTemplate
from langchain.schema.messages import SystemMessage
```

## Set GPU (optional)

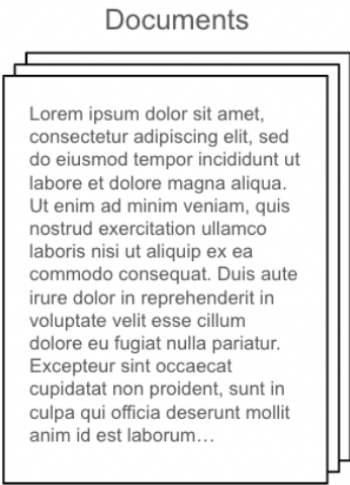
In order for our code to use the correct graphics processing unit (GPU), we ran the below two lines of code, which may require modification depending on your computer system.

```
In [2]: os.environ["CUDA_DEVICE_ORDER"]="PCI_BUS_ID"
os.environ["CUDA_VISIBLE_DEVICES"]="1"
```

## Create Vector Database for a Patient

# Load Documents

Load documents from the RData file into a pandas dataframe using the package PyReadr.



```
In [3]: # load the document
reports = pyreadr.read_r('./demo_reports.rdata')
text = reports['demo_reports']
text["noteID"] = list(range(0, len(text)))
text.head()
```

Out[3]:

	ICI_id	Date	Case	Text	TextType	noteID
0	1	2023-01-13	Hepatitis	***This text report has been converted from th...	prg	0
1	1	2023-01-12	Hepatitis	***This text report has been converted from th...	prg	1
2	1	2023-01-13	Hepatitis	***This text report has been converted from th...	prg	2
3	1	2023-01-13	Hepatitis	***This text report has been converted from th...	prg	3
4	1	2023-01-13	Hepatitis	***This text report has been converted from th...	prg	4

For demonstration purposes, we'll use our LLM on patient #1, who was manually adjudicated to have checkpoint hepatitis.

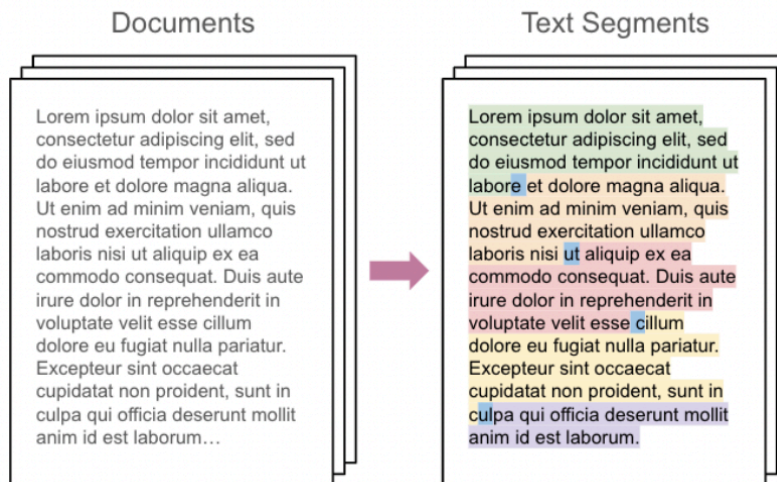
```
In [4]: # Select patient
encounter = text[text["ICI_id"] == "1"]
documents = encounter["Text"]
print("Manual adjudication of patient #1 found to have: ICI-induced " + enc
print("Number of progress and discharge notes written for patient #1: " + st
documents.head()
```

Manual adjudication of patient #1 found to have: ICI-induced Hepatitis  
 Number of progress and discharge notes written for patient #1: 57

```
Out [4]: 0    ***This text report has been converted from th...
        1    ***This text report has been converted from th...
        2    ***This text report has been converted from th...
        3    ***This text report has been converted from th...
        4    ***This text report has been converted from th...
        Name: Text, dtype: object
```

## Segment documents

Each note is segmented to into chunks of 1000 characters, with a 200 character overlap between each segment.



```
In [5]: text_splitter = RecursiveCharacterTextSplitter(chunk_size=1000, chunk_overlap=200)
        all_splits = text_splitter.create_documents(documents)
        print("Number of chunks for patient #1: " + str(len(all_splits)))
        print("Preview of first two chunks: ")
        print("1. " + all_splits[0].page_content)
        print("2. " + all_splits[1].page_content)
```

Number of chunks for patient #1: 425

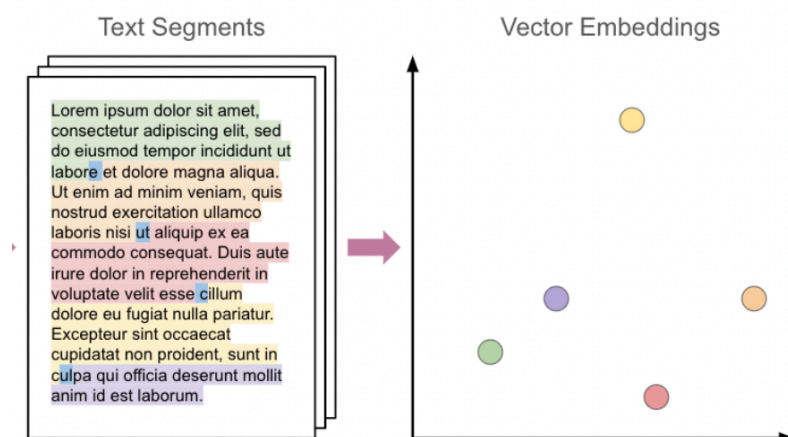
Preview of first two chunks:

1. \*\*\*This text report has been converted from the report, '████████.pdf'. Content may not appear exactly as it appears in the original .pdf. For a download of original content and format (pdf), please contact the RPDR Team at RPDRHelp@partners.org. \*\*\* MGH Main Campus ██████ 55 Fruit St MRN: ██████, DOB: ██████, Sex: M Boston MA 02114-2621 Acct #: ██████ ADM: ██████, D/C: ? Patient Information Patient Name Legal Sex DOB ██████ Male ██████ Progress Notes by ██████, RN at ██████ Author: ██████ Service: Nursing Author Type: Registered Nurse ██████, RN Filed: ██████ Date of Service: ██████ Status: Signed ██████ Editor: ██████, RN (Registered Nurse) Nursing Progress Note Pt A&Ox3. Abdominal pain managed w/ PRN Morphine IR. LFTs continue to downtrend, transitioned to PO Prednisone this shift. PRN senna, miralax, and simethicone administered for discomfort and

2. Note Pt A&Ox3. Abdominal pain managed w/ PRN Morphine IR. LFTs continue to downtrend, transitioned to PO Prednisone this shift. PRN senna, miralax, and simethicone administered for discomfort and bloating from mild constipation. Plan: Monitor LFTs over weekend and ?d/c Monday. Electronically signed by ██████, RN at ██████ [report\_end]

## Transform into embeddings

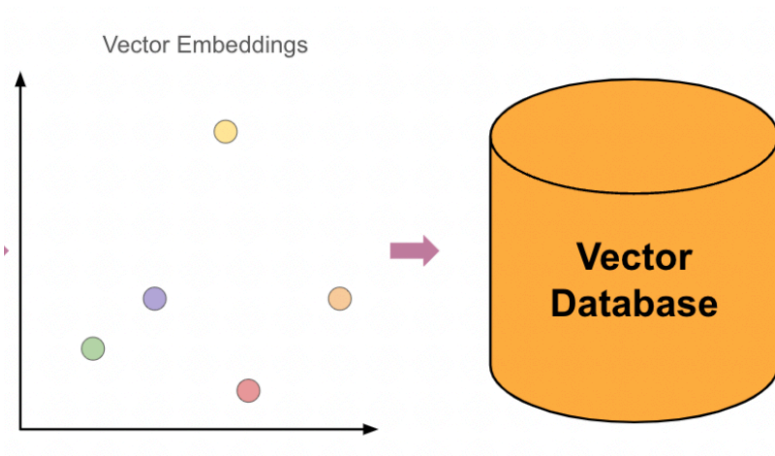
We now convert each text segment into a "vector embedding", aka a numerical representation of the text. In this code example, we used the default embedding function `all-MiniLM-L6-v2`.



```
In [7]: # initialize embedding function
embedding_function = SentenceTransformerEmbeddings(model_name="all-MiniLM-L6-v2")
```

## Store into vector database

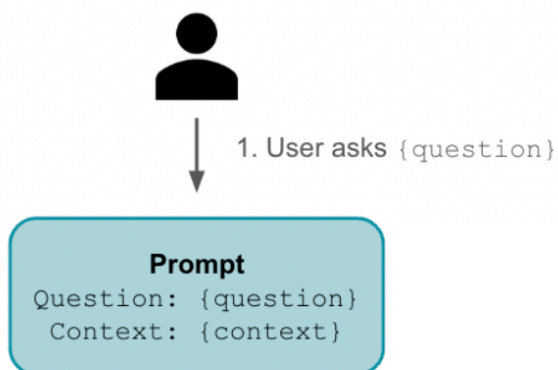
Vectors are stored into a Chroma database for easier retrieval by the LLM later on.



```
In [9]: # store text segments in a vector database
vectorstore = Chroma.from_documents(documents=all_splits, embedding=embedder)
```

## LLM Pipeline

### Create Prompt

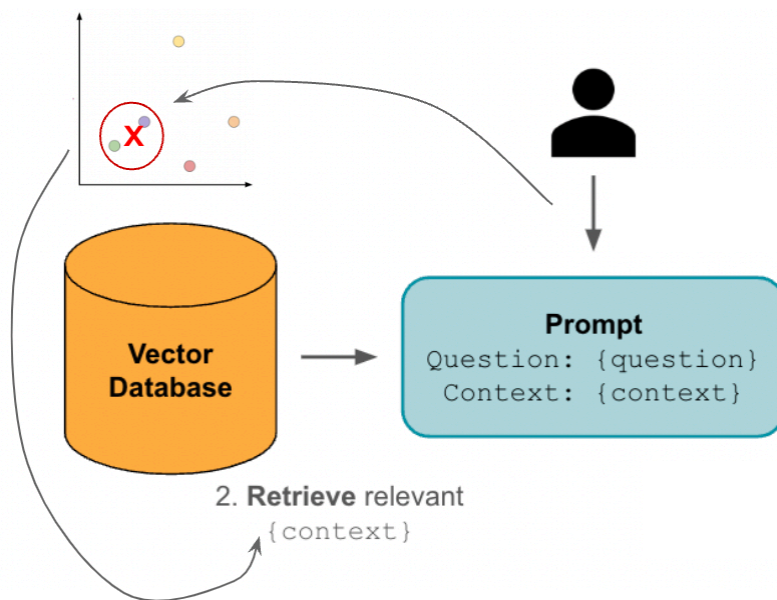


```
In [10]: # initialize prompt
QA_CHAIN_PROMPT = PromptTemplate(input_variables=['context', 'question'], template="""
<|im_start|>system
You are an assistant for question-answering tasks. Use the following pieces
<|im_end|>
<|im_start|>user
Question: {question}
Context: {context}
<|im_end|>
<|im_start|>assistant
""")

question_hepatitis = "Immune checkpoint inhibitor induced hepatitis is an im"
```

### Retrieve relevant context

We now look for the text segments most relevant to the question we asked. You can think of the vector database we created earlier as a "search engine" of all of our patient's medical records that returns the top `k` search results to `question_hepatitis`.



```
In [11]: #Retrieve the top k text segments most similar to the question we asked
docs = vectorstore.similarity_search_with_score(query=question_hepatitis, k=
print("Most relevant text segments: ")
print("1. " + docs[0][0].page_content + "\n\tSimilarity Score: " + str(docs
print("2. " + docs[1][0].page_content + "\n\tSimilarity Score: " + str(docs
```

Most relevant text segments:

1. histologic findings are largely explained by hepatitis B viral infection. However, a contributory component of checkpoint inhibitor-associated injury cannot be excluded, as the exuberant inflammatory injury of the viral infection may be masking a second contributory component. Assessment and Plan [REDACTED] is a [REDACTED] y.o. male with [REDACTED] metastatic lung cancer (s/p VATS and chemoXRT, on nivolumab last dose [REDACTED]), [REDACTED] prior HBV infection (HBV cAb+) who was admitted for incidentally noted LFT elevation (transaminases in the 3-4k range). Workup notable for HBV reactivation. Liver biopsy with panlobular hepatitis and suspicion for a component of both HBV infection and possibly ICI hepatitis. He has been started on entecavir and solumedrol. His steroids were increased to a ICI hepatitis therapeutic dose on [REDACTED] [REDACTED] given increasing LFTs, which appear better

Similarity Score: 0.7363131046295166

2. core antigen is negative. The etiologic differential diagnosis for a panlobular hepatitis is viral infection, drug reaction, and autoimmune hepatitis. The patient is noted to have a hepatitis B viral load of 921,000,000, and is currently on nivolumab. In that context, the histologic findings are largely explained by hepatitis B viral infection. However, a contributory component of checkpoint inhibitor-associated injury cannot be excluded, as the exuberant inflammatory injury of the viral infection may be masking a second contributory component. Assessment and Plan [REDACTED] is a [REDACTED] y.o. male with [REDACTED] metastatic lung cancer (s/p VATS and chemoXRT, on nivolumab last dose [REDACTED]), [REDACTED] prior HBV infection (HBV cAb+) who was admitted for incidentally noted LFT elevation (transaminases in the 3-4k range). Workup notable for HBV reactivation. Liver biopsy

Similarity Score: 0.7454572319984436

## Initialize LLM

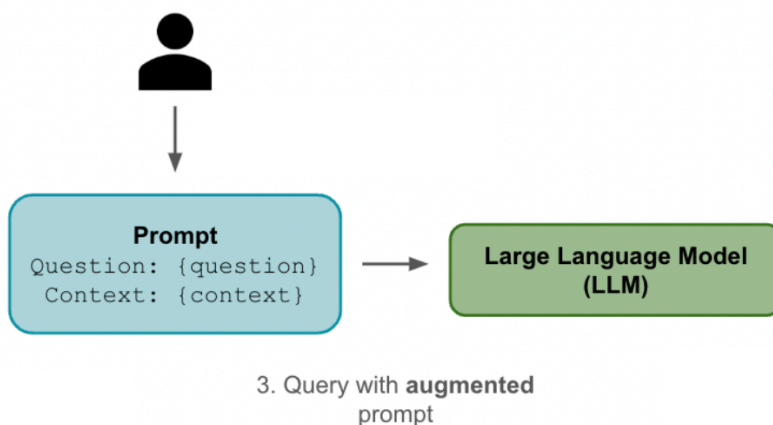
```
In [12]: # initialize LLM
llm = Ollama(
    model="mistral-openorca", # Uses the LLM model, "Mistral-7B-OpenOrca" or
    callback_manager=CallbackManager([StreamingStdOutCallbackHandler()]), #
    temperature = 0, # Sets variability of responses to 0
    num_gpu = 50, # Number of layers to offload onto the GPU
    stop = ["<|im_end|>"] # Ensures that the LLM stops when the token <|im_end|>
)

# QA chain
def format_docs(docs):
    return "\n\n".join(doc.page_content for doc in docs)

rag_chain_from_docs = (
    RunnablePassthrough.assign(context=(lambda x: format_docs(x["context"])))
    | QA_CHAIN_PROMPT
    | llm
    | StrOutputParser() # Formats so that the output is a string
)
```

## Augment LLM with context





```
In [13]: rag_chain_with_source = RunnableParallel(
    {"context": vectorstore.as_retriever(), "question": RunnablePassthrough()}.assign(answer=rag_chain_from_docs)
```

## Generate response



```
In [15]: try:
    response = rag_chain_with_source.invoke(question_hepatitis)
except ValueError:
    # Very occasionally, we receive a ValueError if no data is received from
    # In the event of an error, we can re-run the question on the patient to
    print("No data received from Ollama stream, empty placeholder for now")
    response = {"question": question_hepatitis, "answer": "xxxxx", "context":
```

Answer: Yes

The patient is experiencing immune checkpoint inhibitor induced hepatitis as the histologic findings are largely explained by hepatitis B viral infection, but a contributory component of checkpoint inhibitor-associated injury cannot be excluded. The patient has been started on entecavir and solumedrol to manage this condition.

## Store data and clear vector database

```
In [101]: # Create empty dataframe to store results
results = pd.DataFrame(columns=['ICI_id', 'Adjudicated_Case',
                                'Answer_Hepatitis', 'Source_Hepatitis'])
# Store answer and source documents. You can save this into a CSV for further
results.loc[len(results.index)] = [1, #ICI_id
                                    encounter.iloc[0, encounter.columns.get_loc("Case
                                    response["answer"], response["context"]]

print("Response: \n" + response["answer"])
print("\nContext: \n" + format_docs(response["context"]))
```



```
# Delete vector database  
vectorstore.delete_collection()
```

Response:

Answer: Yes

The patient is experiencing immune checkpoint inhibitor induced hepatitis as the histologic findings are largely explained by hepatitis B viral infection, but a contributory component of checkpoint inhibitor-associated injury cannot be excluded. The patient has been started on entecavir and solumedrol to manage this condition.

Context:

histologic findings are largely explained by hepatitis B viral infection. However, a contributory component of checkpoint inhibitor-associated injury cannot be excluded, as the exuberant inflammatory injury of the viral infection may be masking a second contributory component. Assessment and Plan [REDACTED] is a [REDACTED] y.o. male with [REDACTED] metastatic lung cancer (s/p VATS and chemoXRT, on nivolumab last dose [REDACTED]), [REDACTED] prior HBV infection (HBV cAb+) who was admitted for incidentally noted LFT elevation (transaminases in the 3-4k range). Workup notable for HBV reactivation. Liver biopsy with panlobular hepatitis and suspicion for a component of both HBV infection and possibly ICI hepatitis. He has been started on entecavir and solumedrol. His steroids were increased to a ICI hepatitis therapeutic dose on [REDACTED] given increasing LFTs, which appear better

core antigen is negative. The etiologic differential diagnosis for a panlobular hepatitis is viral infection, drug reaction, and autoimmune hepatitis. The patient is noted to have a hepatitis B viral load of 921,000,000, and is currently on nivolumab. In that context, the histologic findings are largely explained by hepatitis B viral infection. However, a contributory component of checkpoint inhibitor-associated injury cannot be excluded, as the exuberant inflammatory injury of the viral infection may be masking a second contributory component. Assessment and Plan [REDACTED] is a [REDACTED] y.o. male with [REDACTED] metastatic lung cancer (s/p VATS and chemoXRT, on nivolumab last dose [REDACTED]), [REDACTED] prior HBV infection (HBV cAb+) who was admitted for incidentally noted LFT elevation (transaminases in the 3-4k range). Workup not able for HBV reactivation. Liver biopsy

shows staining in rare hepatocytes, but HBV core antigen is negative. The etiologic differential diagnosis for a panlobular hepatitis is viral infection, drug reaction, and autoimmune hepatitis. The patient is noted to have a hepatitis B viral load of 921,000,000, and is currently on nivolumab. In that context, the histologic findings are largely explained by hepatitis B viral infection. However, a contributory component of checkpoint inhibitor-associated injury cannot be excluded, as the exuberant inflammatory injury of the viral infection may be masking a second contributory component. Assessment and Plan [REDACTED] is a [REDACTED] y.o. male with [REDACTED] metastatic lung cancer (s/p VATS and chemoXRT, on nivolumab last dose [REDACTED]), [REDACTED] prior HBV infection (HBV cAb+) who was admitted for incidentally noted LFT elevation (transaminases in the 3-4k range). Workup

hepatitis is viral infection, drug reaction, and autoimmune hepatitis. The patient is noted to have a hepatitis B viral load of 921,000,000, and is currently on nivolumab. In that context, the histologic findings are largely explained by hepatitis B viral infection. However, a contributory component of checkpoint inhibitor-associated injury cannot be excluded, as the exuberant inflammatory injury of the viral infection may be masking a second contributory component. Assessment and Plan [REDACTED] is a [REDACTED] y.o. male with [REDACTED] metastatic lung cancer (s/p VATS and chemoXRT, on nivolumab last dose [REDACTED]), [REDACTED] and prior HBV infection (HBV cAb+) who was admitted for incidentally noted LFT elevation (transaminases in the 3-4k range). Workup not

able for HBV reactivation. Liver biopsy with panlobular hepatitis and suspicion for a component of both HBV infection and

## Fin

Congrats! You are now finished with this Tutorial. Consider running the full code, `demo_LLM_loop.py`, which detects four irAEs, colitis, myocarditis, hepatitis, and pneumonitis, on multiple patients.