

RAG + Conversational Agent = Retrieval Chatbot with OpenAI

Conversational agents can struggle with data freshness, knowledge about specific domains, or accessing internal documentation.

"Naive" retrieval augmentation without the use of an agent is simply context retrieval via query.

Merging these methods gives us the best of both worlds.

We start by doing a `pip install` of all required libraries.

```
!pip install -qU \
  transformers==4.31.0 \
  sentence-transformers==2.2.2 \
  pinecone-client==2.2.2 \
  datasets==2.14.0 \
  accelerate==0.21.0 \
  einops==0.6.1 \
  langchain==0.0.240 \
  xformers==0.0.20 \
  bitsandbytes==0.41.0 \
  openai==0.28.1 \
  tiktoken==0.4.0
```

0:00:00	7.4/7.4 MB	21.3 MB/s	eta
0:00:00	86.0/86.0 kB	9.7 MB/s	eta
0:00:00	etaddata (setup.py) ...		
0:00:00	179.1/179.1 kB	20.8 MB/s	eta 0:00:00
0:00:00	492.2/492.2 kB	39.6 MB/s	eta
0:00:00	244.2/244.2 kB	24.4 MB/s	eta
0:00:00	42.2/42.2 kB	4.5 MB/s	eta
0:00:00	1.4/1.4 MB	74.0 MB/s	eta
0:00:00	109.1/109.1 MB	14.8 MB/s	eta
0:00:00	92.6/92.6 MB	10.2 MB/s	eta
0:00:00	77.0/77.0 kB	7.3 MB/s	eta
0:00:00	1.7/1.7 MB	75.6 MB/s	eta
0:00:00			

0:00:00	7.8/7.8 MB 79.3 MB/s eta
0:00:00	1.3/1.3 MB 69.3 MB/s eta
0:00:00	62.5/62.5 kB 7.3 MB/s eta
0:00:00	300.4/300.4 kB 30.7 MB/s eta
0:00:00	115.3/115.3 kB 13.1 MB/s eta
0:00:00	134.8/134.8 kB 15.1 MB/s eta
0:00:00	48.2/48.2 kB 5.8 MB/s eta
0:00:00	90.0/90.0 kB 10.2 MB/s eta
0:00:00	619.9/619.9 MB 2.6 MB/s eta
0:00:00	21.0/21.0 MB 50.7 MB/s eta
0:00:00	849.3/849.3 kB 55.5 MB/s eta
0:00:00	11.8/11.8 MB 63.2 MB/s eta
0:00:00	557.1/557.1 MB 2.7 MB/s eta
0:00:00	317.1/317.1 MB 3.2 MB/s eta
0:00:00	168.4/168.4 MB 7.6 MB/s eta
0:00:00	54.6/54.6 MB 21.0 MB/s eta
0:00:00	102.6/102.6 MB 10.0 MB/s eta
0:00:00	173.2/173.2 MB 4.3 MB/s eta
0:00:00	177.1/177.1 MB 5.4 MB/s eta
0:00:00	98.6/98.6 kB 11.8 MB/s eta
0:00:00	63.3/63.3 MB 14.9 MB/s eta
0:00:00	153.0/153.0 kB 18.1 MB/s eta
0:00:00	ents to build wheel ... etadata (pyproject.toml) ...
0:00:00	49.4/49.4 kB 5.4 MB/s eta
0:00:00	6.8/6.8 MB 25.6 MB/s eta

```
6.9/6.9 MB 25.5 MB/s eta
0:00:00
6.0/6.0 MB 25.7 MB/s eta
0:00:00
ers (setup.py) ... l) ... ERROR: pip's dependency resolver does not
currently take into account all the packages that are installed. This
behaviour is the source of the following dependency conflicts.
llmx 0.0.15a0 requires cohere, which is not installed.
torchaudio 2.1.0+cud118 requires torch==2.1.0, but you have torch 2.0.1
which is incompatible.
torchdata 0.7.0 requires torch==2.1.0, but you have torch 2.0.1 which
is incompatible.
torchtext 0.16.0 requires torch==2.1.0, but you have torch 2.0.1 which
is incompatible.
```

Initialize the Embedding Model

```
from getpass import getpass
from langchain.embeddings.openai import OpenAIEmbeddings

OPENAI_API_KEY = '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'
model_name = 'text-embedding-ada-002'

embed = OpenAIEmbeddings(
    model=model_name,
    openai_api_key=OPENAI_API_KEY
)
```

Building the Vector Index

We now need to use the embedding pipeline to build our embeddings and store them in a Pinecone vector index. To begin we'll initialize our index, for this we'll need a [free Pinecone API key](#).

```
import os
import pinecone
import pandas as pd
from google.colab import drive

PINECONE_API_KEY = '03355e6b-4a09-4d64-af35-f8e1a9df0ac3'
PINECONE_ENV = 'gcp-starter'

# get API key from app.pinecone.io and environment from console
pinecone.init(
    api_key=PINECONE_API_KEY,
    environment= PINECONE_ENV
)
```

```

/usr/local/lib/python3.10/dist-packages/pinecone/index.py:4:
TqdmExperimentalWarning: Using `tqdm.autonotebook.tqdm` in notebook
mode. Use `tqdm.tqdm` instead to force console mode (e.g. in jupyter
console)
    from tqdm.autonotebook import tqdm

import time

index_name = 'us-visas-db'

if index_name not in pinecone.list_indexes():
    pinecone.create_index(
        index_name,
        dimension=1536,
        metric='dotproduct'
    )
    # wait for index to finish initialization
    while not pinecone.describe_index(index_name).status['ready']:
        time.sleep(1)

index = pinecone.Index('us-visas-db')

```

With our index and embedding process ready we can move onto the indexing process itself. For that, we need the dataset.

Part 1: CFR XML Files

```

# Load data from JSON Lines file to pandas DataFrame
drive.mount('/content/drive')
file_path = '/content/drive/My Drive/Colab
Notebooks/Title-8/chunked/CFR_xml.jsonl'
CFR = pd.read_json(file_path, lines=True)

```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
CFR.head()
```

	Source	Section	Subsection	Chunk ID	\
0	CFR-2023-title8-vol1	sec1	4	0	
1	CFR-2023-title8-vol1	sec1	3	0	
2	CFR-2023-title8-vol1	sec1	2	0	
3	CFR-2023-title8-vol1	sec1	1	0	
4	CFR-2023-title8-vol1	part1		0	

	Text	\
0	8 Aliens and Nationality 1 2023-01-01 2023-01-...	
1	8 Aliens and Nationality 1 2023-01-01 2023-01-...	
2	8 Aliens and Nationality 1 2023-01-01 2023-01-...	
3	8 Aliens and Nationality 1 2023-01-01 2023-01-...	
4	8 Aliens and Nationality 1 2023-01-01 2023-01-...	

	Path
0	/content/drive/My Drive/Colab Notebooks/Title-...
1	/content/drive/My Drive/Colab Notebooks/Title-...
2	/content/drive/My Drive/Colab Notebooks/Title-...
3	/content/drive/My Drive/Colab Notebooks/Title-...
4	/content/drive/My Drive/Colab Notebooks/Title-...

We will embed and index the documents like so:

```
from tqdm import tqdm

batch_size = 64

for i in tqdm(range(0, len(CFR), batch_size)):
    i_end = min(len(CFR), i + batch_size)
    batch = CFR.iloc[i:i_end]

    # Combine 'Source', 'Section', and 'Subsection' columns to create
    unique identifiers
    ids = [f"{row['Source']} - {row['Section']} - {row['Subsection']} -
{row['Chunk ID']}" for _, row in batch.iterrows()]
    texts = batch['Text'].tolist()

    # Assuming you have an embed_model initialized
    embeds = embed_model.embed_documents(texts)

    # Get metadata to store in Pinecone
    metadata = [
        {
            'text': row['Text'],
            'Filepath': row['Path'],
            'Source': row['Source'],
            'Section': row['Section'],
            'Subsection': row['Subsection'],
            'Chunk ID': row['Chunk ID']
        }
        for _, row in batch.iterrows()
    ]

    # Add to Pinecone
    index.upsert(vectors=list(zip(ids, embeds, metadata)))

100%|██████████| 28/28 [00:45<00:00, 1.61s/it]

index.describe_index_stats()

{'dimension': 1536,
 'index_fullness': 0.01599,
```

```
'namespaces': {'': {'vector_count': 1599}},  
'total_vector_count': 1599}
```

Part 2: FAM09 HTML Files

```
# Load data from JSON Lines file to pandas DataFrame  
file_path_FAM09 = '/content/drive/My Drive/Colab  
Notebooks/Title-8/chunked/FAM09_html.jsonl'  
FAM09 = pd.read_json(file_path_FAM09, lines=True)
```

```
FAM09.head()
```

	Source	Section ID	Subsection ID	Chunk ID	\
0	09FAM	0101	01	0	
1	09FAM	0101	02	0	
2	09FAM	0102	01	0	
3	09FAM	0102	01	1	
4	09FAM	0102	01	2	

	Text	\
0	9 FAM 101.1 INTRODUCTION TO 9 FAM UNCLASSIFIED...	
1	9 FAM 101.2 TRAINING AND SUPPORT FOR 9 FAM USE...	
2	9 FAM 102.1 INTRODUCTION TO VISAS (WHAT IS A V...	
3	Sources (CT:VISA-1;00 11-18-2015) A.C.S.A. 41....	
4	sea carrier (airlines or cruise ships), their ...	

	Path
0	/content/drive/My Drive/Colab Notebooks/Title-...
1	/content/drive/My Drive/Colab Notebooks/Title-...
2	/content/drive/My Drive/Colab Notebooks/Title-...
3	/content/drive/My Drive/Colab Notebooks/Title-...
4	/content/drive/My Drive/Colab Notebooks/Title-...

```
from tqdm import tqdm
```

```
batch_size = 16  
batch_size_upsert = 80 # Set your desired batch size for upsert  
rate_limit_delay = 60 / (2* 150000) # Delay to stay within the  
150,000 tokens per minute limit
```

```
for i in tqdm(range(0, len(FAM09), batch_size)):  
    i_end_FAM09 = min(len(FAM09), i + batch_size)  
    batch_FAM09 = FAM09.iloc[i:i_end_FAM09]
```

```
# Combine 'Source', 'Section', and 'Subsection' columns to create  
unique identifiers
```

```
ids_FAM09 = [f"{row['Source']} - {row['Section ID']} -  
{row['Subsection ID']} - {row['Chunk ID']}" for _, row in  
batch_FAM09.iterrows()]  
texts_FAM09 = batch_FAM09['Text'].tolist()
```

```

# Assuming you have an embed_model initialized
embed_FAM09 = embed.embed_documents(texts_FAM09)

# Get metadata to store in Pinecone
metadata_FAM09 = [
    {
        'text': row['Text'],
        'Filepath': row['Path'],
        'Source': row['Source'],
        'Section': row['Section ID'],
        'Subsection': row['Subsection ID'],
        'Chunk ID' : row['Chunk ID']
    }
    for _, row in batch_FAM09.iterrows()
]

# Zip vectors and metadata
vectors_metadata_chunk = list(zip(ids_FAM09, embed_FAM09,
metadata_FAM09))

# Upsert in batches
batch_size_upsert = 100 # Set your desired batch size for upsert
for start in range(0, len(vectors_metadata_chunk),
batch_size_upsert):
    end = min(start + batch_size_upsert, len(vectors_metadata_chunk))
    chunk_to_upsert = vectors_metadata_chunk[start:end]

# Add to Pinecone
index.upsert(vectors=chunk_to_upsert)

# Introduce a delay to stay within the rate limit
time.sleep(rate_limit_delay)

100%|██████████| 76/76 [01:07<00:00, 1.13it/s]

index.describe_index_stats()
{'dimension': 1536,
 'index_fullness': 0.02954,
 'namespaces': {'': {'vector_count': 2954}},
 'total_vector_count': 2954}

```

Creating a Vector Store and Querying

```

from langchain.vectorstores import Pinecone

text_field = "text"

# switch back to normal index for langchain

```

```

index = pinecone.Index(index_name)

vectorstore = Pinecone(
    index, embed.embed_query, text_field
)

```

We can use the `similarity_search` method to do a pure semantic search (without the generation component).

```

query = "Who is qualified to obtain a H-1B visa?"

```

```

vectorstore.similarity_search(
    query, # our search query
    k=3 # return 3 most relevant docs
)

```

[Document(page_content='Nonimmigrants (CT:VISA-1623;00 09-29-2022) a. (U) The H-1B classification\napplies to an applicant who is coming temporarily to the United States to\nperform services in one of the categories described below.0 For information on\nH-1B1 classification for nationals of Chile and Singapore deriving from free\ntrade agreements see 9 FAM\n402.10-5 below. (1)0 (U) Applicants\nin Specialty Occupations: 0 Applicants who are qualified to perform\nservices in a specialty occupation as described in INA 214(i)(1) and (2) (other\nthan agricultural workers, described in INA 101(a)(15)(H)(ii)(A) or applicants\nqualifying under INA 101(a)(15)(O) or (P)) are classifiable as H-1B\nnonimmigrants.0 (a)0 (U) A specialty occupation\nrequires the attainment of a bachelor's or higher degree in the specific\nspecialty (or its equivalent) for entry into the occupation.0 An applicant\nseeking to work in a specialty occupation must have completed such a degree or\nhave experience in the specialty equivalent to the completion of the degree (as\ndetermined by USCIS) and expertise in the specialty through progressively\nresponsible positions relating to the specialty. (b)0 (U) The criteria for\nqualifying as an H-1B physician are found in subparagraph 3 below. (c)0 (U) Before filing a petition\nwith USCIS on behalf of an individual in a\nspecialty occupation, the petitioner must have obtained a certification from\nDOL that it has filed a labor condition application (LCA) as specified in INA\n212(n)(1).0 The filing of an LCA does not constitute a determination that the\noccupation in question is a specialty occupation.0 USCIS is responsible for determining whether the\napplication involves a specialty occupation and whether the individual for whom\nH-1B status is sought qualifies to perform services in that occupation. (2)0 (U) Certain\nFashion Models: 0 H-1B classification may be granted to an applicant who\nis of distinguished merit and ability in the field of fashion modeling.0\n"Distinguished merit and ability" is defined by USCIS as prominence; i.e., the attainment of a high\nlevel of achievement in the field of fashion modeling evidenced by a degree of\nskill and recognition

substantially above that ordinarily encountered to the extent that a person described as prominent is renowned, leading, or well-known in the field. Such an applicant must also be coming to the United States to perform services which require a fashion model of prominence. The petitioner of a fashion model of distinguished merit and ability must file an LCA (see 9 FAM 402.10-6 below) with DOL before filing a petition for the applicant. (3) (U) Physicians: (a) (U) Graduates of Foreign or U.S. Medical Schools: A foreign "graduate of a medical school," as defined in INA 101(a)(41), may enter the United States as an H-1B nonimmigrant to perform services as a member of the medical profession if they have a full and unrestricted license to practice medicine in a foreign state or if they have graduated from medical school in either the United States or in a foreign state. In addition, if they will provide direct patient care, they must generally have a valid medical license in the state of intended employment; however, USCIS may grant a limited-validity petition to allow the beneficiary time to obtain a professional license. An individual involved in a medical residency program, for example, may have an approved H-1B petition, even though they do not yet have a full and unrestricted U.S. medical license. (b) (U) Coming to Teach or Conduct Research: A foreign physician may also be classified as an H-1B nonimmigrant if they are coming to the United States primarily to teach or conduct research, or both, at or for a public or nonprofit private educational or research institution or agency. Such an applicant may only engage in direct patient care that is incidental to their teaching and/or research. (c) (U) Applicants Physicians Not Eligible for H-2B or H-3 Classification: Foreign physicians who are coming to the United States to perform medical services or receive graduate medical training are statutorily ineligible to receive H-2B or H-3 status. (4) (U) Applicants in Department of Defense Cooperative Research and Development or Co-production Projects: Applicants coming to the United States, pursuant to INA 222, to participate in a cooperative research and development project, are eligible for H-1B classification. Metadata: {'Chunk ID': 1.0, 'Filepath': '/content/drive/My Drive/Colab Notebooks/Title-8/FAM state gov downloads/09FAM040210.html', 'Section': '0402', 'Source': '09FAM', 'Subsection': '10'}).

Document(page_content="Only principals are counted against each country's respective numerical limitation. Initial applications for H-1B classification are counted against the H-1B annual numerical limitations, as is each renewed labor condition application (LCA). b. (U) At the end of each fiscal year, unused H-1B numbers will be returned to that year's total H-1B global numerical limit and will be made available to H-1B applicants during the first 45 days of the new fiscal year. c. (U) USCIS is required to keep a numerical count of the H-1B visas issued. The Office of Visa Services (CA/VO) monitors the number used based on workload data. On a periodic basis, CA/VO provides this information to USCIS. 9 FAM 402.10-5(C) (U) Applicants Subject to Labor Condition (CT:VISA-1846; 10-05-2023) (U) Employers must submit a labor condition application (LCA) for

foreign workers of Chilean or Singaporean nationality under the H-1B1 program. If the employee applies for an H-1B1 visa abroad, rather than adjusting status with USCIS, the law requires DOL to certify to the Department that Form ETA-9035, Labor Condition Application for H-1B Nonimmigrants, has been filed with DOL. If certified, the employer transmits a copy of the signed, certified LCA to the applicant together with a written offer of employment. The applicant will present a certified copy of the LCA, clearly annotated by the employer as "H-1B1 Chile" or "H-1B1 Singapore," as proof of filing with the visa application. 9 FAM 402.10-5(D) (U) No Petition Required (CT:VISA-1846; 10-05-2023) (U) An employer of an H-1B1 professional is not required to file a petition with USCIS. Instead, an employee will present evidence for classification directly to you with the visa application. 9 FAM 402.10-5(E) (U) H-1B1 Professionals in Specialty Occupations (CT:VISA-1623; 09-29-2022)

a. (U) The H-1B1 category allows for the entry of nonimmigrant professionals in "specialty occupations." The definition of "specialty occupation" set forth in both FTAs is presently identical to the regulatory definition for H-1Bs; i.e., an occupation that requires: (1) Theoretical and practical application of a body of specialized knowledge; and (2) Attainment of a bachelor's or higher degree in the specific specialty (or its equivalent) as a minimum for entry into the occupation in the United States" (8 CFR 214.2). You should refer to this section for guidance in connection with an applicant's qualifications as an H-1B1 professional.

b. (U) Both agreements allow for alternative credentials for certain professions. The United States has agreed to accept alternative credentials for Chilean and Singaporean nationals in the occupations of Disaster Relief Claims Adjuster and Management Consultant. Disaster Relief Claims Adjusters must have a baccalaureate degree, Licenciatura degree, or Titulo Profesional and completion of training in the appropriate areas of insurance adjustment pertaining to disaster relief claims or three years' experience in claims adjustment and successful completion of training in the appropriate areas of insurance adjustment pertaining to disaster relief claims. Management Consultants must have a baccalaureate degree, Licenciatura degree, or Titulo Profesional, even if in an unrelated discipline. If a Management Consultant has a degree in an unrelated discipline, the applicant additionally must have 3 years of experience in a field or specialty related to the consulting agreement. For Chilean nationals only, Agricultural Managers and Physical Therapists can also qualify with a combination of a post-secondary certificate requiring three years of study in the specialty and three years' experience in lieu of the standard degree requirements. You may accept specified documentary evidence of alternative credentials. 9 FAM 402.10-5(F) (U) Temporary Entry of FTA Professionals (CT:VISA-1846; 10-05-2023) a. (U) Both agreements provide for the temporary entry of professionals into the United States. Temporary entry is defined in both agreements as "an entry into the United States without the

intent to establish permanent residence." You must be",
metadata={'Chunk ID': 6.0, 'Filepath': '/content/drive/My Drive/Colab Notebooks/Title-8/FAM state gov downloads/09FAM040210.html', 'Section': '0402', 'Source': '09FAM', 'Subsection': '10'}),
Document(page_content="physicians who are coming to the United States to perform medical services or receive graduate medical training are statutorily ineligible to receive H-2B or H-3 status. (4) (U) Applicants in Department of Defense Cooperative Research and Development or Co-production Projects: Applicants coming to the United States, pursuant to INA 222, to participate in a cooperative research and development project or a co-production project under a government-to-government agreement administered by DOD are classifiable as H-1B nonimmigrants. Such applicants must perform services of an exceptional nature requiring exceptional merit and ability. For purposes of this classification, services of an exceptional nature must be those which require a bachelor's degree or higher (or its equivalent, as determined by USCIS) to perform the duties. The requirement for filing an LCA with DOL does not apply to petitions involving DOD cooperative research and development or co-production projects. b. (U) General Licensure Requirement: The requirements for classification as an H-1B nonimmigrant professional may or may not include a license because states have different rules in this area. If a state permits applicants to take a licensing exam while in nonimmigrant status, then USCIS will generally require a license before they will approve the H-1B petition. However, some states do not permit applicants to take licensing exams until they enter the United States in H-1B status and obtain a taxpayer identification number. Therefore, a visa must not be denied based solely on the fact that the applicant does not already hold a license to practice in the United States if the applicant intends to work in a state that requires a taxpayer identification number as a prerequisite to licensure. 9 FAM 402.10-4(C) (U) H-1C Nurse in Health Professional Shortage Area (CT:VISA-1474; 02-17-2022) (U) This classification expired as of December 20, 2009. 9 FAM 402.10-4(D) (U) H-2A Nonimmigrants (CT:VISA-1623; 09-29-2022) a. (U) The H-2A classification applies to applicants who are coming temporarily to the United States to perform agricultural work of a temporary or seasonal nature. b. (U) The petitioner must file a temporary agricultural labor certification with DOL before filing a petition with USCIS to classify a beneficiary as an H-2A nonimmigrant. c. (U) Except as noted in 9 FAM 402.10-7(C) below, USCIS generally may only approve a Form I-129, Petition for a Nonimmigrant Worker, filed on behalf of an H-2A worker who is a national of a country designated as an H-2A program eligible country. However, USCIS may still approve H-2A petitions filed for nationals of countries not designated as participating countries, if such an approval is in the U.S. interest, as noted in 9 FAM 402.10-7 (C) paragraph c, below. (1) (U) The designated countries can be found on the USCIS H-2A website. (2) (U) Countries are designated as H-2A program participating countries

based on factors that include the following: (a) (U) The country's cooperation with respect to the issuance of travel documents for citizens, subjects, nationals, and residents of that country who are subject to a final order of removal from the United States; (b) (U) The number of final and unexecuted orders of removal against citizens, subjects, nationals, and residents of that country; (c) (U) The number of orders of removal executed against citizens, subjects, nationals, and residents of that country; and (d) (U) Such other factors as may serve U.S. interest. (3) (U) Posts will be advised when there are changes to the list of participating countries as well as the effective dates for their formal participation in the program. Designations will be valid for one year from the date of publication of the list of eligible countries in the Federal Register. On a case-by-case basis, DHS may allow a worker from a country not on the participating country list to be eligible for the H-2A program if, among other considerations, such participation is in the interest of the United States. (4) (U) Posts recommending that a country obtain, maintain, or lose status as an H-2A program participant", metadata={'Chunk ID': 2.0, 'Filepath': '/content/drive/My Drive/Colab Notebooks/Title-8/FAM state gov downloads/09FAM040210.html', 'Section': '0402', 'Source': '09FAM', 'Subsection': '10'}}]

Building the Chatbot

LangChain library will provide the framework to bring various components together for our chatbot. Component 2: Initializing a ChatOpenAI object with OpenAI API key for naive chatbot.

```
from langchain.chat_models import ChatOpenAI

chat = ChatOpenAI(
    openai_api_key=OPENAI_API_KEY,
    model='gpt-3.5-turbo-16k'
)

from langchain.schema import (
    SystemMessage,
    HumanMessage,
    AIMessage
)

messages = [
    SystemMessage(content="You are a helpful assistant."),
    HumanMessage(content="Hi AI, how are you today?"),
    AIMessage(content="I'm great thank you. How can I help you?"),
]
```

Retrieval Augmented Generation

Now utilizing our developed knowledge base, it's time to connect the knowledge base to the chatbot.

```
def augment_prompt(query: str):  
    # get top 3 results from knowledge base  
    results = vectorstore.similarity_search(query, k=5)  
    # get the text from the results  
    source_knowledge = "\n".join([x.page_content for x in results])  
    # feed into an augmented prompt  
    augmented_prompt = f"""You are an assistant helping the user  
understand about the USA's nonimmigrant visa laws and processes.  
Please answer the query using the provided contexts below.
```

```
    If an answer is contingent on various factors, please list all of  
those factors for the user to consider. If it is unclear whether or  
not the user has considered a factor, ask the user for clarity.
```

```
    Contexts:
```

```
    {source_knowledge}
```

```
    Query: {query}"""
```

```
    return augmented_prompt
```

```
print(augment_prompt("Can you provide a summary of all the inputs  
required to fill out a I-94 form?"))
```

Please answer the query using the provided contexts below.

The assistant is helpful and asks follow up questions whenever possible to make that all various factors are considered by the user.

If an answer is contingent on various factors, please provide a comprehensive list of these factors to the user, then ask them if they have met all the criteria.

Contexts:

8 Aliens and Nationality 1 2023-01-01 2023-01-01 false Definition of Form I-94 1.4 Section 1.4

Aliens and Nationality DEPARTMENT OF HOMELAND SECURITY
GENERAL PROVISIONS DEFINITIONS

§ 1.4 Definition of Form I-94

The term

Form I-94, Form I-94, (a) The terms “annotate,” “note,” “indicate on,” “stamp,” and “endorse,” unless used in part 231 of this chapter, include, but are not limited, to DHS amending, including or completing information in its electronic record of admission, or arrival/departure. For purposes of part 231, the term “endorse” includes but is not limited to the submission of electronic departure

data to CBP. (b) The terms "completed," "completely executed" and "completed and signed" include, but are not limited to, DHS completing its collection of information into its electronic record of admission, or arrival/departure. (c) The terms "issuance" and "given" include, but are not limited to, the creation of an electronic record of admission, or arrival/departure by DHS following an inspection performed by an immigration officer. (d) The term "original I-94" includes, but is not limited to, any printout or electronic transmission of information from DHS systems containing the electronic record of admission or arrival/departure. (e) The terms "present," "presentation," or "submission" of a Form I-94, unless they are used in § 231.1 or § 231.2 of this chapter, include, but are not limited to, providing a printout of information from DHS systems containing an electronic record of admission or arrival/departure. For purposes of § 231.1 of this chapter, the terms "present" or "submission" of the Form I-94 includes ensuring that each passenger presents him/herself to a CBP Officer for inspection at a U.S. port-of-entry. For the purposes of § 231.2 of this chapter, the terms "present," "submit," or "submission" of the Form I-94 includes ensuring that each passenger is available for inspection by a CBP Officer upon request. (f) The term "possession" with respect to a Form I-94 includes, but is not limited to, obtaining a copy or printout of the record of an electronic evidence of admission or arrival/departure from the appropriate CBP systems. (g) The terms "surrendering," "turning in a Form I-94," and "departure I-94" includes, but is not limited to, complying with any departure controls under 8 CFR part 215 that may be prescribed by CBP in addition to the submission of electronic departure data to CBP by a carrier. [78 FR 18472, Mar. 27, 2013]

(vii) Aliens who are the spouse or child of a United States citizen whose visa petition has been approved and

(5) Applicants for asylum under section 208(a) of the Act and applicants for withholding of removal under section 241(b)(3) of the Act or under the Convention Against Torture who have been granted employment authorization, and such applicants under the age of 14 who have had an application pending for at least 180 days.

(b)

Non-issuance of a Notice to Appear and non-enforcement of deportation, exclusion, or removal orders. (1) Issue a Notice to Appear; or (2) Enforce an outstanding order of deportation, exclusion or removal.

§ 1.4 Definition of Form I-94

The term

Form I-94, Form I-94, (a) The terms "annotate," "note," "indicate on," "stamp," and "endorse," unless used in part 231 of this chapter, include, but are not limited, to DHS amending, including or completing information in its electronic record of admission, or arrival/departure. For purposes of part 231, the term "endorse" includes but is not limited to the submission of electronic departure data to CBP. (b) The terms "completed," "completely

executed" and "completed and signed" include, but are not limited to, DHS completing its collection of information into its electronic record of admission, or arrival/departure. (c) The terms "issuance" and "given" include, but are not limited to, the creation of an electronic record of admission, or arrival/departure by DHS following an inspection performed by an immigration officer. (d) The term "original I-94" includes, but is not limited to, any printout or electronic transmission of information from DHS systems containing the electronic record of admission or arrival/departure. (e) The terms "present," "presentation," or "submission" of a Form I-94, unless they are used in § 231.1 or § 231.2 of this chapter, include, but are not limited to, providing a printout of information from DHS systems containing an electronic record of admission or arrival/departure. For purposes of § 231.1 of this chapter, the terms "present" or "submission" of the Form I-94 includes ensuring that each passenger presents him/herself to a CBP Officer for inspection at a U.S. port-of-entry. For the purposes of § 231.2 of this chapter, the terms "present," "submit," or "submission" of the Form I-94 includes ensuring that each passenger is available for inspection by a CBP Officer upon request. (f) The term "possession" with respect to a Form I-94 includes, but is not limited to, obtaining a copy or printout of the record of an electronic evidence of admission or arrival/departure from the appropriate CBP systems. (g) The terms "surrendering," "turning in a Form I-94," and "departure I-94" includes, but is not limited to, complying with any departure controls under 8 CFR part 215 that may be prescribed by CBP in addition to the submission of electronic departure data to CBP by a carrier. [78 FR 18472, Mar. 27, 2013]

requirements in this IV DV Passport Waiver template: (1) Applicant's full name with all aliases; (2) Date and place of birth; (3) Nationality; (4) Flight itinerary including the date and port of expected arrival in the United States; please include the flight reservation as an attachment in the email if you have it. (If the POE is undetermined, please submit a tentative itinerary with the tentative POE and travel itinerary. CBP will not consider the passport waiver request without a POE as it is up to the POE to accept or reject the passport waiver request); (5) Immigrant classification; (6) Documents to be waived; (7) A summary of the of the emergent circumstances surrounding the case which must include information indicating that all the requirements of the subparagraph of 22 CFR 42.2(g) under which the waiver is recommended have been met; (8) Name, address, and telephone number of the person where the applicant intends to reside in the United States. b.

Officers must notify the designated CA/VO/F liaison with any itinerary changes. If the applicant's itinerary changes to a different POE a new waiver of passport request must be submitted. CBP POE concurrence is granted by the Port Director at the POE, therefore a new request must be submitted. UNCLASSIFIED (U) information sheets to CA/VO/F/IE. Although you do not have to receive prior approval to use a nonstandard information sheet, VO wants information sheets to be coordinated and consolidated in the interests of efficiency. Therefore, you must send copies of all new nonstandard forms to NVCPPost@state.gov. b. (U) Do not translate forms used in the IV process into the local language. You may prepare a local language information sheet explaining the forms. 9 FAM 504.4-4 (U) Supporting Documents 9 FAM 504.4-4(A) (U) Basic Document Requirements (CT:VISA-1864; 11-16-2023) a. (U) An applicant applying for an IV must follow all instructions on travel.state.gov for submitting the following documents, if available: (1) (U) Birth Certificates: The applicant must obtain an original or certified copy of their birth certificate and those of any family member immigrating with the applicant. (2) (U) Court and Prison Records: If the applicant has ever been convicted of a crime, they must obtain an original or certified copy of each court and prison record, even if later granted amnesty, a pardon, or other act of clemency. (3) (U) Marriage Records: (a) (U) Marriage Certificates: If the applicant is or has been married, they must obtain an original or certified copy of the marriage certificate of every marriage. (b) (U) Marriage Termination: If the applicant was previously married, they must obtain evidence of the termination of every prior marriage. The evidence must be an original or certified copy of a final legal divorce decree, death certificate, or annulment papers. (4) (U) Military Records: If the applicant has ever served in the military of any country, they must obtain an original or certified copy of their military record. (5) (U) Petitioner Documents: (a) (U) If the applicant is applying for an IR5 or an F4 visa, they must obtain an original or certified

copy of the petitioner's birth certificate. (b) (U) If the applicant is applying for an IR1, CR1, or F2A visa as the spouse of a U.S. citizen or LPR and the petitioning spouse was previously married, they must provide evidence of the termination of every prior marriage of the petitioning spouse. The evidence must be an original or certified copy of a final legal divorce decree, death certificate, or annulment papers. (c) (U) If the applicant is a stepchild and is applying for an IR2, CR2, F2, or any other visa classification requiring that the applicant be a child, they must obtain an original or certified copy of the marriage certificate of every marriage of the petitioner and the applicant's natural parent. If the petitioner or applicant's natural parent was previously married, the applicant must obtain evidence of the termination of every prior marriage. The evidence must be an original or certified copy of a final legal divorce decree, death certificate, or annulment papers. (6) (U) Passport: The applicant, and each family member immigrating with the applicant, must submit a photocopy of the biographic data page of a currently valid passport and provide the passport at the interview. In certain circumstances, this passport requirement may be waived. See 9 FAM 201.2 for details on immigrant travel without a passport. (7) (U) Police Certificates: If the applicant is 16 years of age or older or if there is reason to believe a police record exists, they must provide original or certified copies of police certificates. See 9 FAM 504.4-4(B) below. (8) (U) Adoption Documents: If the applicant is an adopted child and the application to immigrate is based on a parent-child relationship, they must submit: (a) (U) An original or certified copy of the adoption decree; (b) (U) The legal custody decree if custody occurred before adoption; (c) (U) A statement showing dates and places where the applicant resided with the adoptive parent(s); and (d) (U) If the applicant was adopted when ages 16 or 17, evidence that they were adopted with, or

after, the adoption of a natural sibling under the age of 16 by the same adoptive parents. b. (U) Availability paid; (2) (U) The petitioner has submitted a signed Form I-864, Affidavit of Support Under Section 213A of the Act; (3) (U) All traveling applicants have completed Form DS-260, Online Application for Immigrant Visa and Alien Registration; and (4) (U) All required police certificates have been provided. c. (U) If Form I-864 is free of critical errors but the IVIS case lacks supporting financial documentation, NVC will send an Assessment Letter to the petitioner or agent with a list of missing items, instructing the visa applicant to present the required information to the consular officer during the visa interview. The Assessment Letter is included in the case file for your information for IVIS cases. d. (U) Under specific circumstances approved by VO for some IVIS cases and for all PIVOT cases, NVC will not schedule a visa interview until all requested documentation has been provided. NVC refers to this as the Fully Qualified process. e. (U) A case that is required to be Fully Qualified will be reviewed for the following: (1) (U) All required fees are paid; (2) (U) All AOS forms are submitted and void of critical errors; (3) (U) All IV application forms are received and void of critical errors; (4) (U) All civil documents are received and meet the guidelines in the Visa Reciprocity and Country Documents Finder; (5) (U) Any required supporting financial evidence from all sponsors is received. (NVC will send an assessment letter for IVIS cases if all other requirements are met); and (6) (U) The only exception to these rules is if an applicant informs NVC they cannot obtain certain civil documents. f. (U) For cases that are required to be Fully Qualified, there is no limit to the number of requests for information that are sent. NVC will continue to send requests for information until all necessary fees, forms, and documents are received and accurate. If the petitioner or agent does not return the documents within one year

and a visa is available, NVC will initiate the administrative process to begin case termination. 9 FAM 504.4-5(C) (U) Post Document Review, Support, and Scheduling 9 FAM 504.4-5(C)(1) (U) Need for Standard Operating Procedures (SOP) (CT:VISA-1864;00 11-16-2023)

a. (U) Standard Procedures and Forms Have Been Developed and Installed in All IV units to: (1) (U) Ensure uniformity in explaining the requirements of the law to visa applicants; (2) (U) Reduce individual correspondence and possible misunderstandings arising therefrom; and (3) (U) Eliminate needless files and record-keeping by requiring applicants to retain their personal documents until the final step in the processing of the case is reached. b. (U) Document Review Systems Ask Applicants to Obtain Documents Required for IV Interview and Then Submit Them by Upload, Mail, Courier, or Drop Box for IV units to Review: (1) (U) If possible, applicants should not appear in person with these documents until scheduled for interview; and (2) (U) Under such a procedure, applicants are "documentarily complete" only when they have demonstrated that they have in their possession all the documents required. c. (U) It is important in such a prescreening procedure that detailed SOP provide guidance to LE Staff screeners to limit misunderstandings and accusations of impropriety. Specifically, the SOP must provide: (1) (U) Written standards for documents submitted; (2) (U) Escalation procedures in cases when, despite repeated appearances at the consular section, the applicant remains unprepared; (3) (U) Procedures for handling multiple secondary documents submitted in lieu of requested primary documents; (4) (U) Procedures for documenting contacts with applicants or their agents as the documents are submitted and reviewed. (U) Notes recording these contacts should be concise but complete and should utilize the comments feature in the IVO system; and (5) (U) Provisions for regular officer oversight of the process, including regular detailed audits of individual cases and questions to applicants during the interview concerning their experience with the prescreening process. 9 FAM 504.4-5(C)(2) (U)

Flexibility in Determining Whether Applicant Is Documentarily

Query: Can you provide a summary of all the inputs required to fill out a I-94 form?

Now use the template below for follow up questions with the chatbot.

```
prompt = HumanMessage(
    content=augment_prompt(
        "I have an advanced degree in computer science. Am I eligible
for the H-1B visa?"
    )
)
```

```
res = chat(messages + [prompt])
print(res.content)
```

Based on the information provided, individuals with an advanced degree in computer science may be eligible for the H-1B visa. The H-1B visa is available for applicants who are qualified to perform services in a specialty occupation that requires the attainment of a bachelor's degree or higher in a specific specialty.

To qualify for the H-1B visa, you would need to demonstrate that you have completed an advanced degree in computer science or a related field. Additionally, you would need to provide evidence of your eligibility, such as an official academic record showing your degree, letters from current or former employers demonstrating your experience, and any other relevant documentation.

It's important to note that the H-1B visa program has specific requirements and limitations, including the need for an employer to file a labor condition application (LCA) with the Department of Labor (DOL) and the availability of H-1B visa numbers. It is advised to consult with an immigration attorney or review the official USCIS guidelines for the most accurate and up-to-date information regarding H-1B visa eligibility and application processes.

```
prompt = HumanMessage(
    content=augment_prompt(
        "I have a job offer from a US employer, who promises to obtain
a LCA. What are other requirements are set by the USCIS?"
    )
)
```

```
res = chat(messages + [prompt])
print(res.content)
```

In addition to obtaining a Labor Condition Application (LCA) from your potential employer, there are other requirements set by the United

States Citizenship and Immigration Services (USCIS) for nonimmigrant visas. Here are some key requirements to consider:

1. Specialty Occupation: The job you are being offered must qualify as a "specialty occupation." This means that the position requires a theoretical and practical application of specialized knowledge and a minimum of a bachelor's degree or its equivalent in the specific specialty.

2. Qualifications: You must demonstrate that you have the required qualifications for the job, including the necessary education or equivalent experience.

3. Validity of Visa: The validity of the visa should not exceed the validity period of the LCA. The maximum validity period for E-3 visas, for example, is 24 months.

4. Intent to Work Temporarily: You must demonstrate that you do not intend to remain or work permanently in the United States and that you plan to work temporarily in the specific job for which the visa is being sought.

5. Fees: Apart from the normal visa-related Machine Readable Visa (MRV) fees, there are no additional fees associated with the issuance of an E-3 visa.

6. Public Charge Concerns: You need to evaluate potential public charge concerns, particularly if you are planning to work part-time or if you have multiple job offers with different LCAs. Each employment situation must overcome public charge concerns on its own.

7. Change of Plans: If there is a material change in your plans after the visa has been issued, such as a change of employer, type of work, or location of employment, you may need to obtain certification for the new employment or satisfy certification requirements before the visa can be used.

It's important to note that these requirements may vary depending on the specific nonimmigrant visa category you are applying for, so it's advisable to consult the USCIS website or contact an immigration attorney for detailed and up-to-date information.