```
!pip install langchain langchain-community faiss-cpu sentence-transformers
```

Show hidden output

```python
import json
from langchain_core.documents import Document

with open(r"/content/documents_for_embedding.json", "r", encoding="utf-8") as f:
    data = json.load(f)

documents = [
    Document(
        page_content=item["content"],
        metadata=item["metadata"]
    )
    for item in data
]

len(documents)
```

```
117
```

```python
from langchain_core.documents import Document

documents = []

for item in data:
    documents.append(
        Document(
            page_content=item["content"],
            metadata=item["metadata"] | {"id": item["id"]}
        )
    )

print(len(documents))
```

```
117
```

```python
import os
os.environ["OPENAI_API_KEY"] = "sk-proj-QnJsbwskr4ZHXm3tUnXoEAMC1mMBd3oPIzI91uNsaBeNvdzlzhnae8IuCUYgPTLkeBQzrwdwbxT3BlbkF
```

```python
!pip install langchain-openai
from langchain_openai import OpenAIEmbeddings

embeddings = OpenAIEmbeddings(
    model="text-embedding-3-small"
)
```

Show hidden output

```python
from langchain_community.vectorstores import FAISS

vectorstore = FAISS.from_documents(documents, embeddings)
```

```python
vectorstore.save_local("compliance_faiss_index")
```

```python
!pip install langchain langchain-openai faiss-cpu
```

```
Requirement already satisfied: langchain in /usr/local/lib/python3.12/dist-packages (1.2.8)
Requirement already satisfied: langchain-openai in /usr/local/lib/python3.12/dist-packages (1.1.9)
Requirement already satisfied: faiss-cpu in /usr/local/lib/python3.12/dist-packages (1.13.2)
Requirement already satisfied: langchain-core<2.0.0,>=1.2.8 in /usr/local/lib/python3.12/dist-packages (from langchain) (1
Requirement already satisfied: langgraph<1.1.0,>=1.0.7 in /usr/local/lib/python3.12/dist-packages (from langchain) (1.0.7)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in /usr/local/lib/python3.12/dist-packages (from langchain) (2.12.3)
Requirement already satisfied: openai<3.0.0,>=1.109.1 in /usr/local/lib/python3.12/dist-packages (from langchain-openai) (
Requirement already satisfied: tiktoken<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from langchain-openai) (
Requirement already satisfied: numpy<3.0,>=1.25.0 in /usr/local/lib/python3.12/dist-packages (from faiss-cpu) (2.0.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.12/dist-packages (from faiss-cpu) (26.0)
Requirement already satisfied: jsonpatch<2.0.0,>=1.33.0 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.
Requirement already satisfied: langsmith<1.0.0,>=0.3.45 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.
Requirement already satisfied: pyyaml<7.0.0,>=5.3.0 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.0.0,
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in /usr/local/lib/python3.12/dist-packages (from langchain-
Requirement already satisfied: typing-extensions<5.0.0,>=4.7.0 in /usr/local/lib/python3.12/dist-packages (from langchain-
Requirement already satisfied: uuid-utils<1.0,>=0.12.0 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.0
```

```
Requirement already satisfied: langgraph-checkpoint<5.0.0,>=2.1.0 in /usr/local/lib/python3.12/dist-packages (from langgra
Requirement already satisfied: langgraph-prebuilt<1.1.0,>=1.0.7 in /usr/local/lib/python3.12/dist-packages (from langgraph
Requirement already satisfied: langgraph-sdk<0.4.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from langgraph<1.1.
Requirement already satisfied: xxhash>=3.5.0 in /usr/local/lib/python3.12/dist-packages (from langgraph<1.1.0,>=1.0.7->lan
Requirement already satisfied: anyio<5,>=3.5.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->la
Requirement already satisfied: distro<2,>=1.7.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->l
Requirement already satisfied: httpx<1,>=0.23.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->l
Requirement already satisfied: jiter<1,>=0.10.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->l
Requirement already satisfied: sniffio in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->langchain-
Requirement already satisfied: tqdm>4 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->langchain-o
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic<3.0.0,>=2.
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic<3.0.0,>=2.7
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic<3.0.0,>=
Requirement already satisfied: regex>=2022.1.18 in /usr/local/lib/python3.12/dist-packages (from tiktoken<1.0.0,>=0.7.0->l
Requirement already satisfied: requests>=2.26.0 in /usr/local/lib/python3.12/dist-packages (from tiktoken<1.0.0,>=0.7.0->l
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.12/dist-packages (from anyio<5,>=3.5.0->openai<3.0.0,>=
Requirement already satisfied: certifi in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->openai<3.0.0,>=1
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->openai<3.0
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.12/dist-packages (from httpcore==1.*->httpx<1,>=0.23.0-
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.12/dist-packages (from jsonpatch<2.0.0,>=1.33.0-
Requirement already satisfied: ormsgpack>=1.12.0 in /usr/local/lib/python3.12/dist-packages (from langgraph-checkpoint<5.0
Requirement already satisfied: orjson>=3.10.1 in /usr/local/lib/python3.12/dist-packages (from langgraph-sdk<0.4.0,>=0.3.0
Requirement already satisfied: requests-toolbelt>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from langsmith<1.0.0,>
Requirement already satisfied: zstandard>=0.23.0 in /usr/local/lib/python3.12/dist-packages (from langsmith<1.0.0,>=0.3.45
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests>=2.26.0-
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests>=2.26.0->tikto
```

```python
import json
from langchain_core.documents import Document

with open("documents_for_embedding.json", "r", encoding="utf-8") as f:
    raw_docs = json.load(f)

documents = [
    Document(
        page_content=d["content"],
        metadata={**d["metadata"], "id": d["id"]}
    )
    for d in raw_docs
]

len(documents)
```

```
117
```

```python
from langchain_openai import OpenAIEmbeddings
from langchain_community.vectorstores import FAISS

embeddings = OpenAIEmbeddings()

vectorstore = FAISS.from_documents(
    documents,
    embedding=embeddings
)
```

```python
vectorstore.save_local("compliance_faiss_index")
```

```python
from langchain_community.vectorstores import FAISS
from langchain_openai import OpenAIEmbeddings

embeddings = OpenAIEmbeddings()

vectorstore = FAISS.load_local(
    "compliance_faiss_index",
    embeddings,
    allow_dangerous_deserialization=True
)
```

```python
docs = vectorstore.similarity_search(
    "Can a broker accept unredacted Aadhaar?",
    k=3
)

for d in docs:
    print(d.page_content)
```

Where Aadhaar authentication is not required, registered intermediaries shall not accept Aadhaar documents unless the Aadh
Where Aadhaar authentication is not required, an All India Financial Institution shall ensure that the customer redacts or
Registered intermediaries shall not retain or process Aadhaar numbers during digital KYC where Aadhaar authentication is n

```python
from langchain_openai import OpenAIEmbeddings, ChatOpenAI
from langchain_community.vectorstores import FAISS

embeddings = OpenAIEmbeddings()

vectorstore = FAISS.load_local(
    "compliance_faiss_index",
    embeddings,
    allow_dangerous_deserialization=True
)
```

```python
retriever = vectorstore.as_retriever(
    search_type="similarity",
    search_kwargs={"k": 6}
)
```

```python
llm = ChatOpenAI(
    model="gpt-4o-mini",
    temperature=0
)
```

```python
from langchain_core.prompts import PromptTemplate

prompt = PromptTemplate.from_template(
    """
You are a regulatory compliance assistant.
Answer ONLY using the context provided.
If the answer is not found, say: "Not found in regulations."

Context:
{context}

Question:
{question}
"""
)
```

```python
from langchain_core.runnables import RunnablePassthrough
from langchain_core.output_parsers import StrOutputParser
from operator import itemgetter

rag_chain = (
    {
        "context": itemgetter("input") | retriever,
        "question": itemgetter("input")
    }
    | prompt
    | llm
    | StrOutputParser()
)
```

```python
docs = retriever.invoke("Is Aadhaar mandatory for KYC?")
len(docs)
```

```
6
```

```python
for d in docs:
    print(d.page_content)
```

```
Aadhaar number shall not be mandatory for KYC purposes except where benefits under Section 7 of the Aadhaar Act are sought
Registered intermediaries shall not mandate the use of Aadhaar for KYC, as its usage is purely voluntary for the client.
Registered intermediaries shall not insist on In-Person Verification where KYC has been completed using Aadhaar authentica
Where Aadhaar authentication is not required, an All India Financial Institution shall ensure that the customer redacts or
Registered intermediaries shall not accept or rely upon a PAN that has become inoperative due to non-linkage with Aadhaar
An All India Financial Institution shall not allow OTP-based e-KYC accounts to operate beyond one year without full identi
```

```python
!pip install langchain-openai
import os
```

```python
from langchain_openai import OpenAIEmbeddings, ChatOpenAI
from langchain_community.vectorstores import FAISS

# Set the OpenAI API key (replace with your actual API key if running this cell in isolation)
os.environ["OPENAI_API_KEY"] = "sk-proj-QnJsbwskr4ZHXm3tUnXoEAMC1mMBd3oPIzI91uNsaBeNvdzlzhnae8IuCUYgPTLkeBQzrwdwbxT3BlbkF

embeddings = OpenAIEmbeddings()

vectorstore = FAISS.load_local(
    "compliance_faiss_index",
    embeddings,
    allow_dangerous_deserialization=True
)

retriever = vectorstore.as_retriever(
    search_type="similarity",
    search_kwargs={"k": 6}
)

docs = retriever.invoke(
    "Is Aadhaar mandatory for KYC?"
)

for d in docs:
    print(d.metadata["id"])
    print(d.page_content[:300])
    print("-"*50)
```

```
Requirement already satisfied: langchain-openai in /usr/local/lib/python3.12/dist-packages (1.1.9)
Requirement already satisfied: langchain-core<2.0.0,>=1.2.11 in /usr/local/lib/python3.12/dist-packages (from langchain-op
Requirement already satisfied: openai<3.0.0,>=1.109.1 in /usr/local/lib/python3.12/dist-packages (from langchain-openai) (
Requirement already satisfied: tiktoken<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from langchain-openai) (
Requirement already satisfied: jsonpatch<2.0.0,>=1.33.0 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.
Requirement already satisfied: langsmith<1.0.0,>=0.3.45 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.
Requirement already satisfied: packaging>=23.2.0 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.0.0,>=1
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.0.
Requirement already satisfied: pyyaml<7.0.0,>=5.3.0 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.0.0,
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in /usr/local/lib/python3.12/dist-packages (from langchain-
Requirement already satisfied: typing-extensions<5.0.0,>=4.7.0 in /usr/local/lib/python3.12/dist-packages (from langchain-
Requirement already satisfied: uuid-utils<1.0,>=0.12.0 in /usr/local/lib/python3.12/dist-packages (from langchain-core<2.0
Requirement already satisfied: anyio<5,>=3.5.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->la
Requirement already satisfied: distro<2,>=1.7.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->l
Requirement already satisfied: httpx<1,>=0.23.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->l
Requirement already satisfied: jiter<1,>=0.10.0 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->l
Requirement already satisfied: sniffio in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->langchain-
Requirement already satisfied: tqdm>4 in /usr/local/lib/python3.12/dist-packages (from openai<3.0.0,>=1.109.1->langchain-o
Requirement already satisfied: regex>=2022.1.18 in /usr/local/lib/python3.12/dist-packages (from tiktoken<1.0.0,>=0.7.0->l
Requirement already satisfied: requests>=2.26.0 in /usr/local/lib/python3.12/dist-packages (from tiktoken<1.0.0,>=0.7.0->l
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.12/dist-packages (from anyio<5,>=3.5.0->openai<3.0.0,>=
Requirement already satisfied: certifi in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->openai<3.0.0,>=1
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->openai<3.0
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.12/dist-packages (from httpcore==1.*->httpx<1,>=0.23.0-
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.12/dist-packages (from jsonpatch<2.0.0,>=1.33.0-
Requirement already satisfied: orjson>=3.9.14 in /usr/local/lib/python3.12/dist-packages (from langsmith<1.0.0,>=0.3.45->l
Requirement already satisfied: requests-toolbelt>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from langsmith<1.0.0,>
Requirement already satisfied: xxhash>=3.0.0 in /usr/local/lib/python3.12/dist-packages (from langsmith<1.0.0,>=0.3.45->la
Requirement already satisfied: zstandard>=0.23.0 in /usr/local/lib/python3.12/dist-packages (from langsmith<1.0.0,>=0.3.45
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic<3.0.0,>=2.
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic<3.0.0,>=2.7
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic<3.0.0,>=
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests>=2.26.0-
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests>=2.26.0->tikto
RBI_AIFI_KYC_2025_C23_EXP4
Aadhaar number shall not be mandatory for KYC purposes except where benefits under Section 7 of the Aadhaar Act are sought
--------------------------------------------------
SEBI_KYC_MC_2023_C42_AADHAAR_VOLUNTARY
Registered intermediaries shall not mandate the use of Aadhaar for KYC, as its usage is purely voluntary for the client.
--------------------------------------------------
SEBI_KYC_MC_2023_C61_IPV_EXEMPTION
Registered intermediaries shall not insist on In-Person Verification where KYC has been completed using Aadhaar authentica
--------------------------------------------------
RBI_AIFI_KYC_2025_C23_EXP1
Where Aadhaar authentication is not required, an All India Financial Institution shall ensure that the customer redacts or
--------------------------------------------------
SEBI_KYC_MC_2023_C9_INOPERATIVE_PAN
Registered intermediaries shall not accept or rely upon a PAN that has become inoperative due to non-linkage with Aadhaar
--------------------------------------------------
RBI_AIFI_KYC_2025_C25_6
An All India Financial Institution shall not allow OTP-based e-KYC accounts to operate beyond one year without full identi
--------------------------------------------------
```

```python
response = rag_chain.invoke({
    "input": "Is Aadhaar mandatory for KYC under Indian regulations?"
})
```

```
    print(response)
```

```
    Not found in regulations.
```

```python
from langchain_core.prompts import PromptTemplate

prompt = PromptTemplate(
    input_variables=["context", "question"],
    template="""
You are a regulatory compliance assistant.

Answer the question using ONLY the regulatory clauses below.
If the answer is not present, say: "Not found in regulations."

Regulatory clauses:
{context}

Question:
{question}

Answer:
"""
)
```

```python
docs = retriever.invoke("Is Aadhaar mandatory for KYC?")

context = "\n".join([d.page_content for d in docs])

print("DEBUG CONTEXT ↓↓↓")
print(context)
```

```
DEBUG CONTEXT ↓↓↓
Aadhaar number shall not be mandatory for KYC purposes except where benefits under Section 7 of the Aadhaar Act are sought
Registered intermediaries shall not mandate the use of Aadhaar for KYC, as its usage is purely voluntary for the client.
Registered intermediaries shall not insist on In-Person Verification where KYC has been completed using Aadhaar authentica
Where Aadhaar authentication is not required, an All India Financial Institution shall ensure that the customer redacts or
Registered intermediaries shall not accept or rely upon a PAN that has become inoperative due to non-linkage with Aadhaar
An All India Financial Institution shall not allow OTP-based e-KYC accounts to operate beyond one year without full identi
```

```python
prompt = f"""
Based on the provided regulations, answer the question: "Is Aadhaar mandatory for KYC?"

- Use the provided context ONLY.
- If the context provides a conditional answer, explain those conditions.
- If the information is truly missing, say: Not found in regulations.

Context:
{context}
"""
```

```python
response = llm.invoke(prompt)
print(response.content)
```

```
    Aadhaar is not mandatory for KYC purposes except when benefits under Section 7 of the Aadhaar Act are sought. This means t
```