

Code for PDF File Input and Q&A:-

How to use a **local embedding model** from Hugging Face Transformers with **LangChain** and **FAISS** for Reading PDF file.

Required Installations

Make sure you have these packages installed:

```
pip install langchain-core langchain-community langchain-huggingface
sentence-transformers faiss-cpu torch PyPDF2
```

Python code:-

```
from langchain_community.vectorstores import FAISS
from langchain_huggingface import HuggingFaceEmbeddings
from langchain_core.documents import Document
from langchain_community.document_loaders import PyPDFLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter

# 1. Load the local embedding model
embedding = HuggingFaceEmbeddings(model_name="sentence-transformers/all-MiniLM-L6-v2")

# 2. Load text from a PDF file
pdf_path = "my-notes.pdf" # Change to your actual PDF path

loader = PyPDFLoader(pdf_path)
pages = loader.load()

# 3. Split into smaller chunks (important for long PDFs)
text_splitter = RecursiveCharacterTextSplitter(
    chunk_size=500,
    chunk_overlap=50
)
documents = text_splitter.split_documents(pages)

# 4. Create the FAISS vector store
vectorstore = FAISS.from_documents(documents, embedding)

# 5. Ask a question
```

```
query = "What does the document say about machine learning?"

# 6. Do a similarity search
results = vectorstore.similarity_search(query, k=3)

# 7. Print top matching chunks
for i, doc in enumerate(results, 1):
    print(f"Result {i}: {doc.page_content}")
```

my-notes.pdf:-

Machine learning is a field of artificial intelligence that uses statistical techniques to give computer systems the ability to learn from data.

Natural language processing enables computers to understand, interpret, and respond to human language.

Data science combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.

Neural networks are computing systems inspired by the biological neural networks that constitute animal brains.

AI is transforming industries such as healthcare, finance, and transportation through automation and data-driven decisions.

With input data interface :-

Here's a simple **chatbot-style interface** built on top of your PDF+FAISS setup — so you can ask questions interactively from a PDF.

```
from langchain_community.vectorstores import FAISS
from langchain_huggingface import HuggingFaceEmbeddings
from langchain_core.documents import Document
from langchain_community.document_loaders import PyPDFLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter

# 1. Load the local embedding model
embedding = HuggingFaceEmbeddings(model_name="sentence-transformers/all-MiniLM-L6-v2")

# 2. Load and split the PDF
pdf_path = "my-notes.pdf" # 📄 Your PDF file
loader = PyPDFLoader(pdf_path)
pages = loader.load()

text_splitter = RecursiveCharacterTextSplitter(chunk_size=500, chunk_overlap=50)
documents = text_splitter.split_documents(pages)

# 3. Create the FAISS vector store
vectorstore = FAISS.from_documents(documents, embedding)

# 4. Chat loop
print("\n📄 PDF Chatbot Ready! Ask questions (type 'exit' to quit)\n")

while True:
    query = input("👤 You: ")
    if query.lower() in ["exit", "quit"]:
        print("👋 Goodbye!")
        break

    results = vectorstore.similarity_search(query, k=2)

    if results:
        print("📄 Answer:")
        for doc in results:
            print("-", doc.page_content.strip(), "\n")
    else:
        print("👋 Sorry, no relevant information found.\n")
```

🧠 You:what is machine learning

🤖 Answer:

- Om sir is digital marketing trainer and he is also teaching full stack development code.

Machine learning is a field of artificial intelligence that uses statistical techniques to give computer systems the ability to learn from data.

language.

Data science combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.