```
!pip install streamlit pymupdf langchain langchain-community
langchain-huggingface faiss-cpu requests

Collecting streamlit
  Downloading streamlit-1.46.1-py3-none-any.whl.metadata (9.0 kB)
Collecting pymupdf
  Downloading pymupdf-1.26.1-cp39-abi3-
manylinux_2_28_x86_64.whl.metadata (3.4 kB)
Requirement already satisfied: langchain in
/usr/local/lib/python3.11/dist-packages (0.3.26)
Collecting langchain-community
  Downloading langchain_community-0.3.26-py3-none-any.whl.metadata
(2.9 kB)
Collecting langchain-huggingface
  Downloading langchain_huggingface-0.3.0-py3-none-any.whl.metadata
(996 bytes)
Collecting faiss-cpu
  Downloading faiss_cpu-1.11.0-cp311-cp311-
manylinux_2_28_x86_64.whl.metadata (4.8 kB)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (2.32.3)
Requirement already satisfied: altair<6,>=4.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (5.5.0)
Requirement already satisfied: blinker<2,>=1.5.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (1.9.0)
Requirement already satisfied: cachetools<7,>=4.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (5.5.2)
Requirement already satisfied: click<9,>=7.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (8.2.1)
Requirement already satisfied: numpy<3,>=1.23 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (2.0.2)
Requirement already satisfied: packaging<26,>=20 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (24.2)
Requirement already satisfied: pandas<3,>=1.4.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (2.2.2)
Requirement already satisfied: pillow<12,>=7.1.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (11.2.1)
Requirement already satisfied: protobuf<7,>=3.20 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (5.29.5)
Requirement already satisfied: pyarrow>=7.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (18.1.0)
Requirement already satisfied: tenacity<10,>=8.1.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (8.5.0)
Requirement already satisfied: toml<2,>=0.10.1 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (0.10.2)
Requirement already satisfied: typing-extensions<5,>=4.4.0 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (4.14.0)
Collecting watchdog<7,>=2.1.5 (from streamlit)
  Downloading watchdog-6.0.0-py3-none-
manylinux2014_x86_64.whl.metadata (44 kB)
```

```
                                                                 44.3/44.3 kB 3.4 MB/s eta
0:00:00
ent already satisfied: gitpython!=3.1.19,<4,>=3.0.7 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (3.1.44)
Collecting pydeck<1,>=0.8.0b4 (from streamlit)
  Downloading pydeck-0.9.1-py2.py3-none-any.whl.metadata (4.1 kB)
Requirement already satisfied: tornado!=6.5.0,<7,>=6.0.3 in
/usr/local/lib/python3.11/dist-packages (from streamlit) (6.4.2)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.66 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.3.66)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.8
in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.8)
Requirement already satisfied: langsmith>=0.1.17 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.4.1)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in
/usr/local/lib/python3.11/dist-packages (from langchain) (2.11.7)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
/usr/local/lib/python3.11/dist-packages (from langchain) (2.0.41)
Requirement already satisfied: PyYAML>=5.3 in
/usr/local/lib/python3.11/dist-packages (from langchain) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in
/usr/local/lib/python3.11/dist-packages (from langchain-community)
(3.11.15)
Collecting dataclasses-json<0.7,>=0.5.7 (from langchain-community)
  Downloading dataclasses_json-0.6.7-py3-none-any.whl.metadata (25 kB)
Collecting pydantic-settings<3.0.0,>=2.4.0 (from langchain-community)
  Downloading pydantic_settings-2.10.1-py3-none-any.whl.metadata (3.4
kB)
Collecting httpx-sse<1.0.0,>=0.4.0 (from langchain-community)
  Downloading httpx_sse-0.4.1-py3-none-any.whl.metadata (9.4 kB)
Requirement already satisfied: tokenizers>=0.19.1 in
/usr/local/lib/python3.11/dist-packages (from langchain-huggingface)
(0.21.2)
Requirement already satisfied: huggingface-hub>=0.30.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-huggingface)
(0.33.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests) (2025.6.15)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
```

```
>langchain-community) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (1.7.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (6.4.4)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (0.3.2)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (1.20.1)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from altair<6,>=4.0-
>streamlit) (3.1.6)
Requirement already satisfied: jsonschema>=3.0 in
/usr/local/lib/python3.11/dist-packages (from altair<6,>=4.0-
>streamlit) (4.24.0)
Requirement already satisfied: narwhals>=1.14.2 in
/usr/local/lib/python3.11/dist-packages (from altair<6,>=4.0-
>streamlit) (1.44.0)
Collecting marshmallow<4.0.0,>=3.18.0 (from dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Downloading marshmallow-3.26.1-py3-none-any.whl.metadata (7.3 kB)
Collecting typing-inspect<1,>=0.4.0 (from dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Downloading typing_inspect-0.9.0-py3-none-any.whl.metadata (1.5 kB)
Requirement already satisfied: gitdb<5,>=4.0.1 in
/usr/local/lib/python3.11/dist-packages (from gitpython!
=3.1.19,<4,>=3.0.7->streamlit) (4.0.12)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.30.2-
>langchain-huggingface) (3.18.0)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.30.2-
>langchain-huggingface) (2025.3.2)
Requirement already satisfied: tqdm>=4.42.1 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.30.2-
>langchain-huggingface) (4.67.1)
Requirement already satisfied: hf-xet<2.0.0,>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.30.2-
>langchain-huggingface) (1.1.5)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.66->langchain) (1.33)
```

```
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.17-
>langchain) (0.28.1)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.17-
>langchain) (3.10.18)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.17-
>langchain) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.17-
>langchain) (0.23.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas<3,>=1.4.0-
>streamlit) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas<3,>=1.4.0-
>streamlit) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas<3,>=1.4.0-
>streamlit) (2025.2)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4-
>langchain) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.2 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4-
>langchain) (2.33.2)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4-
>langchain) (0.4.1)
Collecting python-dotenv>=0.21.0 (from pydantic-
settings<3.0.0,>=2.4.0->langchain-community)
  Downloading python_dotenv-1.1.1-py3-none-any.whl.metadata (24 kB)
Requirement already satisfied: greenlet>=1 in
/usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4-
>langchain) (3.2.3)
Requirement already satisfied: smmap<6,>=3.0.1 in
/usr/local/lib/python3.11/dist-packages (from gitdb<5,>=4.0.1-
>gitpython!=3.1.19,<4,>=3.0.7->streamlit) (5.0.2)
Requirement already satisfied: anyio in
/usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0-
>langsmith>=0.1.17->langchain) (4.9.0)
Requirement already satisfied: httpcore==1.* in
/usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0-
>langsmith>=0.1.17->langchain) (1.0.9)
Requirement already satisfied: h11>=0.16 in
/usr/local/lib/python3.11/dist-packages (from httpcore==1.*-
>httpx<1,>=0.23.0->langsmith>=0.1.17->langchain) (0.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in
```

```
/usr/local/lib/python3.11/dist-packages (from jinja2->altair<6,>=4.0-
>streamlit) (3.0.2)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33-
>langchain-core<1.0.0,>=0.3.66->langchain) (3.0.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0-
>altair<6,>=4.0->streamlit) (2025.4.1)
Requirement already satisfied: referencing>=0.28.4 in
/usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0-
>altair<6,>=4.0->streamlit) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in
/usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0-
>altair<6,>=4.0->streamlit) (0.25.1)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas<3,>=1.4.0->streamlit) (1.17.0)
Collecting mypy-extensions>=0.3.0 (from typing-inspect<1,>=0.4.0-
>dataclasses-json<0.7,>=0.5.7->langchain-community)
  Downloading mypy_extensions-1.1.0-py3-none-any.whl.metadata (1.1 kB)
Requirement already satisfied: sniffio>=1.1 in
/usr/local/lib/python3.11/dist-packages (from anyio->httpx<1,>=0.23.0-
>langsmith>=0.1.17->langchain) (1.3.1)
Downloading streamlit-1.46.1-py3-none-any.whl (10.1 MB)
                                        ━━━━━━━━━ 10.1/10.1 MB 103.2 MB/s eta
0:00:00
updf-1.26.1-cp39-abi3-manylinux_2_28_x86_64.whl (24.1 MB)
                                        ━━━━━━━━━ 24.1/24.1 MB 83.4 MB/s eta
0:00:00
munity-0.3.26-py3-none-any.whl (2.5 MB)
                                        ━━━━━━━━━ 2.5/2.5 MB 82.7 MB/s eta
0:00:00
anylinux_2_28_x86_64.whl (31.3 MB)
                                        ━━━━━━━━━ 31.3/31.3 MB 17.8 MB/s eta
0:00:00
                                        ━━━━━━━━━ 45.2/45.2 kB 3.7 MB/s eta
0:00:00
                                        ━━━━━━━━━ 6.9/6.9 MB 105.6 MB/s eta
0:00:00
anylinux2014_x86_64.whl (79 kB)
                                        ━━━━━━━━━ 79.1/79.1 kB 7.0 MB/s eta
0:00:00
arshmallow-3.26.1-py3-none-any.whl (50 kB)
                                        ━━━━━━━━━ 50.9/50.9 kB 4.5 MB/s eta
0:00:00
ypy_extensions-1.1.0-py3-none-any.whl (5.0 kB)
Installing collected packages: watchdog, python-dotenv, pymupdf, mypy-
extensions, marshmallow, httpx-sse, faiss-cpu, typing-inspect, pydeck,
pydantic-settings, dataclasses-json, streamlit, langchain-huggingface,
```

```
langchain-community
Successfully installed dataclasses-json-0.6.7 faiss-cpu-1.11.0 httpx-
sse-0.4.1 langchain-community-0.3.26 langchain-huggingface-0.3.0
marshmallow-3.26.1 mypy-extensions-1.1.0 pydantic-settings-2.10.1
pydeck-0.9.1 pymupdf-1.26.1 python-dotenv-1.1.1 streamlit-1.46.1
typing-inspect-0.9.0 watchdog-6.0.0

IBM_API_KEY = "tZMZbILUlguHqSdoXC43U05j3wRtezqHn3wfFFPcvHXm"
PROJECT_ID = "6fb8395d-ad6b-48f6-8eb7-d79b7b198c56"
WATSONX_URL = "https://us-south.ml.cloud.ibm.com"

import fitz  # PyMuPDF
import requests
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain_community.vectorstores import FAISS
from langchain_huggingface import HuggingFaceEmbeddings

# KEYS
IBM_API_KEY = "tZMZbILUlguHqSdoXC43U05j3wRtezqHn3wfFFPcvHXm"
PROJECT_ID = "6fb8395d-ad6b-48f6-8eb7-d79b7b198c56"
WATSONX_URL = "https://us-south.ml.cloud.ibm.com"

def extract_text(pdf_file):
    text = ""
    with fitz.open(stream=pdf_file.read(), filetype="pdf") as doc:
        for page in doc:
            text += page.get_text()
    return text

def get_chunks(text):
    splitter = RecursiveCharacterTextSplitter(chunk_size=500,
chunk_overlap=100)
    return splitter.create_documents([text])

def create_vector_store(docs):
    embeddings = HuggingFaceEmbeddings(model_name="all-MiniLM-L6-v2")
    return FAISS.from_documents(docs, embeddings)

def get_iam_token(api_key):
    url = "https://iam.cloud.ibm.com/identity/token"
    headers = {"Content-Type": "application/x-www-form-urlencoded"}
    data = f"apikey={api_key}&grant_type=urn:ibm:params:oauth:grant-
type:apikey"
    response = requests.post(url, headers=headers, data=data)
    return response.json()["access_token"]

def ask_watsonx(prompt, token):
    url = f"{WATSONX_URL}/ml/v1/text/generation?version=2024-05-01"
    headers = {
        "Authorization": f"Bearer {token}",
```

```python
        "Content-Type": "application/json"
    }
    payload = {
        "model_id": "ibm/granite-3-2-8b-instruct",
        "input": prompt,
        "project_id": PROJECT_ID,
        "parameters": {
            "decoding_method": "greedy",
            "max_new_tokens": 300
        }
    }
    response = requests.post(url, headers=headers, json=payload)
    return response.json()['results'][0]['generated_text']

def generate_answer(vectorstore, question, token):
    docs = vectorstore.similarity_search(question, k=3)
    context = "\n".join(doc.page_content for doc in docs)

    prompt = f"""
You are a helpful AI assistant. Use the context below to answer the
user's question accurately.

Context:
{context}

Question:
{question}

Answer only using the provided context. If the answer is not available
in the provided context, say:
"The answer is not available in the provided context."
    """.strip()

    return ask_watsonx(prompt, token)

!pip install pdfminer.six
```

```
Collecting pdfminer.six
  Downloading pdfminer_six-20250506-py3-none-any.whl.metadata (4.2 kB)
Requirement already satisfied: charset-normalizer>=2.0.0 in
/usr/local/lib/python3.11/dist-packages (from pdfminer.six) (3.4.2)
Requirement already satisfied: cryptography>=36.0.0 in
/usr/local/lib/python3.11/dist-packages (from pdfminer.six) (43.0.3)
Requirement already satisfied: cffi>=1.12 in
/usr/local/lib/python3.11/dist-packages (from cryptography>=36.0.0-
>pdfminer.six) (1.17.1)
Requirement already satisfied: pycparser in
/usr/local/lib/python3.11/dist-packages (from cffi>=1.12-
>cryptography>=36.0.0->pdfminer.six) (2.22)
Downloading pdfminer_six-20250506-py3-none-any.whl (5.6 MB)
```

```
━━━━━━━━━━━━━━━━━━━━━━━━━━ 0.0/5.6 MB ? eta -:--:--
━━━━━━━━━━━━━━━━━━━━━━━━━ 5.5/5.6 MB 164.7 MB/s eta
0:00:01 ━━━━━━━━━━━━━━━━━━━━━━━━━ 5.6/5.6 MB 90.5 MB/s
eta 0:00:00
iner.six
Successfully installed pdfminer.six-20250506
```

```python
from pdfminer.high_level import extract_text

from IPython.display import display
import ipywidgets as widgets
import io
from pdfminer.high_level import extract_text

# Create upload widget
upload = widgets.FileUpload(accept='.pdf', multiple=False)
display(upload)
```

{"model_id":"404271b744614561a5f15112cf871e83","version_major":2,"version_minor":0}

```python
if upload.value:
    uploaded_filename = list(upload.value.keys())[0]
    content = upload.value[uploaded_filename]['content']

    pdf_file = io.BytesIO(content)
    text = extract_text(pdf_file)

    print(f"Extracted text length: {len(text)}")
    print(f"Preview:\n{text[:500]}")

    chunks = get_chunks(text)
    print(f"Number of chunks created: {len(chunks)}")
    if len(chunks) > 0:
        print(f"First chunk preview: {chunks[0].page_content[:200]}")
    else:
        print("⚠ No chunks created!")

    vectorstore = create_vector_store(chunks)  # Pass Document list as
is
    token = get_iam_token(IBM_API_KEY)

    question = "What is this PDF about?"
    answer = generate_answer(vectorstore, question, token)

    print("☻ Answer:", answer)
else:
    print("⬚ Please upload a PDF file first.")
```

```
Extracted text length: 17591
Preview:
```

1.Definition and Evolution of entrepreneurship.

Definition of Entrepreneurship:

Entrepreneurship is the act of initiating, developing, organizing, and managing a new business
venture to make a profit, while bearing most of the risks. It involves identifying opportunities,
mobilizing resources, and creating value through innovation, creativity, and leadership.

Entrepreneurs are individuals who drive this process by recognizing market needs and taking
the initiative to fulfill them, ofte
Number of chunks created: 45
First chunk preview: 1.Definition and Evolution of entrepreneurship.

Definition of Entrepreneurship:

Entrepreneurship is the act of initiating, developing, organizing, and managing a new business
venture to make a p
◉ Answer:

This PDF appears to be a document outlining various factors that contribute to entrepreneurship and business growth. It is divided into several sections, each focusing on a different aspect of this topic.

The first section, "Observability," discusses how the benefits of entrepreneurial activities can be made visible to others, with applications in marketing, public health, and technology.

The second section, "Idea Generation Techniques in Entrepreneurship," highlights three key techniques: Resource Mobilization (efficient use of financial, human, and technical resources), Value Creation (creating economic or social value through innovative solutions), and Risk and Uncertainty Management (navigating uncertain environments and financial risks).

The third section, "Economic Contribution," explains how entrepreneurship can generate employment, promote competition, and boost GDP.

The fourth section, "Growth Orientation," describes entrepreneurship as being driven by the potential for scaling and expansion.

The fifth section, "Political Stability," notes that a stable environment encourages investment and business growth.

The sixth section, "Technological Factors," identifies three key

factors: Access to Technology (availability and affordability of digital tools and platforms), the Innovation Ecosystem (research institutions, tech hubs, and patent systems),

```python
question = input("Ask a question about the PDF: ")
answer = generate_answer(vectorstore, question, token)
print("Answer:", answer)
```

Ask a question about the PDF: what is this pdf about
Answer:

This PDF appears to be a guide or outline on various aspects of entrepreneurship. It covers topics such as observability, idea generation techniques, resource mobilization, value creation, risk and uncertainty management, economic contribution, growth orientation, observation and trend analysis, customer feedback and surveys, and reverse thinking. Each topic is briefly described, providing a high-level overview of its relevance and application in entrepreneurship. For instance, observability is discussed in the context of marketing, public health, and technology, while idea generation techniques involve resource mobilization, value creation, and risk management. The guide also emphasizes the importance of understanding customer needs through observation, trend analysis, and direct feedback.

```python
question = input("Ask a question about the PDF: ")
answer = generate_answer(vectorstore, question, token)
print("Answer:", answer)
```

Ask a question about the PDF: state the types of entrepuerners
Answer:

1. Startup founders like Jeff Bezos or Elon Musk.
2. Individuals who start and own their own business.
3. Those who use personal resources and capital, taking on full financial risk.
4. People who work independently, outside an existing organization.
5. Entrepreneurs who make foundational decisions and set the overall direction of the business.
6. Individuals driven by vision, creativity, and a willingness to take risks.