

```
!pip install -r requirements.txt
```

```
Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.11/dist-packages (from rich<14,>=10.14.0->streamlit->r
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.11/dist-packages (from rich<14,>=10.14.0->streamlit->
Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4->langchain->-r req
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.11/dist-packages (from beautifulsoup4->unstructured->-r require
Requirement already satisfied: six>=1.9 in /usr/local/lib/python3.11/dist-packages (from html5lib->unstructured->-r requirements.txt (
Requirement already satisfied: webencodings in /usr/local/lib/python3.11/dist-packages (from html5lib->unstructured->-r requirements.t
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk->unstructured->-r requirements.txt (line 1
Requirement already satisfied: olefile in /usr/local/lib/python3.11/dist-packages (from python-oxmsg->unstructured->-r requirements.tx
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->sentence-transformer
Requirement already satisfied: aiofiles>=24.1.0 in /usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured->-r
Requirement already satisfied: cryptography>=3.1 in /usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured->-
Requirement already satisfied: eval-type-backport<0.3.0,>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from unstructured-client->
Requirement already satisfied: jsonpath-python<2.0.0,>=1.0.6 in /usr/local/lib/python3.11/dist-packages (from unstructured-client->uns
Requirement already satisfied: nest-asyncio>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured-
Requirement already satisfied: pypdf>=4.0 in /usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured->-r requi
Requirement already satisfied: pyproject_hooks in /usr/local/lib/python3.11/dist-packages (from build>=1.0.3->chromadb!=0.5.10,!0.5.1
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.11/dist-packages (from cryptography>=3.1->unstructured-client->uns
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,
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx>=0.27.0->chromadb!=0.5.10,!0.5.11,!0.5.1
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx>=0.27.0->chromadb!=0.5.10,!0.5.11
Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx>=0.27.0->chromadb
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-core<
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0->
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0->altair<6,>=4.0->s
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0->altair<6,>=4.0->stream
Requirement already satisfied: google-auth>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.1
Requirement already satisfied: websocket-client!=0.40.0,!0.41.*,!0.42.*,>=0.32.0 in /usr/local/lib/python3.11/dist-packages (from ku
Requirement already satisfied: requests-oauthlib in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.10
Requirement already satisfied: oauthlib>=3.2.2 in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.10,!
Requirement already satisfied: durationpy>=0.7 in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.10,!
Requirement already satisfied: mdurl~0.1 in /usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.2.0->rich<14,>=10.14.0->s
Requirement already satisfied: coloredlogs in /usr/local/lib/python3.11/dist-packages (from onnxruntime>=1.14.1->chromadb!=0.5.10,!0.
Requirement already satisfied: flatbuffers in /usr/local/lib/python3.11/dist-packages (from onnxruntime>=1.14.1->chromadb!=0.5.10,!0.
Requirement already satisfied: deprecated>=1.2.6 in /usr/local/lib/python3.11/dist-packages (from opentelemetry-api>=1.2.0->chromadb!=
Requirement already satisfied: importlib-metadata<=8.5.0,>=6.0 in /usr/local/lib/python3.11/dist-packages (from opentelemetry-api=1.2
Requirement already satisfied: googleapis-common-protos~=1.52 in /usr/local/lib/python3.11/dist-packages (from opentelemetry-exporter-
Requirement already satisfied: opentelemetry-exporter-otlp-proto-common==1.29.0 in /usr/local/lib/python3.11/dist-packages (from opent
Requirement already satisfied: opentelemetry-proto==1.29.0 in /usr/local/lib/python3.11/dist-packages (from opentelemetry-exporter-otl
Requirement already satisfied: opentelemetry-instrumentation-asyncio==0.50b0 in /usr/local/lib/python3.11/dist-packages (from opentelemet
Requirement already satisfied: opentelemetry-instrumentation==0.50b0 in /usr/local/lib/python3.11/dist-packages (from opentelemetry-in
Requirement already satisfied: opentelemetry-semantic-conventions==0.50b0 in /usr/local/lib/python3.11/dist-packages (from opentelemet
Requirement already satisfied: opentelemetry-util-http==0.50b0 in /usr/local/lib/python3.11/dist-packages (from opentelemetry-instrume
Requirement already satisfied: asgiref~=3.0 in /usr/local/lib/python3.11/dist-packages (from opentelemetry-instrumentation-asyncio==0.50b
Requirement already satisfied: monotonic>=1.5 in /usr/local/lib/python3.11/dist-packages (from posthog>=2.4.0->chromadb!=0.5.10,!0.5.
Requirement already satisfied: shellingham>=1.3.0 in /usr/local/lib/python3.11/dist-packages (from typer>=0.9.0->chromadb!=0.5.10,!0.
Requirement already satisfied: mypy-extensions>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from typing-inspect<1,>=0.4.0->datac
Requirement already satisfied: httpxtools>=0.6.3 in /usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3->chromadb!=
Requirement already satisfied: uvloop!=0.15.0,!0.15.1,>=0.14.0 in /usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.
Requirement already satisfied: watchfiles>=0.13 in /usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3->chromadb!=
Requirement already satisfied: websockets>=10.4 in /usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3->chromadb!=
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx>=0.27.0->chromadb!=0.5.10,!
Requirement already satisfied: pycparser in /usr/local/lib/python3.11/dist-packages (from cffi>=1.12->cryptography>=3.1->unstructured-
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.11/dist-packages (from google-auth>=1.0.1->kubernetes>=
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.11/dist-packages (from google-auth>=1.0.1->kubernetes>=28.1.0->
Requirement already satisfied: zipp>=3.20 in /usr/local/lib/python3.11/dist-packages (from importlib-metadata<=8.5.0,>=6.0->openteleme
Requirement already satisfied: humanfriendly>=9.1 in /usr/local/lib/python3.11/dist-packages (from coloredlogs->onnxruntime>=1.14.1->c
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in /usr/local/lib/python3.11/dist-packages (from pyasn1-modules>=0.2.1->google-aut
```

```
from langchain_community.document_loaders import UnstructuredURLLoader
urls = ['https://www.livemint.com/economy/budget-2024-key-highlights-live-updates-nirmala-sitharaman-infrastructure-defence-income-tax-modi-
'https://cleartax.in/s/budget-2024-highlights',
# 'https://www.hindustantimes.com/budget',
'https://economictimes.indiatimes.com/news/economy/policy/budget-2024-highlights-india-nirmala-sitharaman-capex-fiscal-deficit-tax-s
loader = UnstructuredURLLoader(urls=urls)
data = loader.load()
```

```
len(data)
```

```
3
```

```
data[2]
```

Shrimps, and Fish Feed: Duty reduced to 5%. \n\nAmmonium Nitrate: Duty raised to 10%. \n\nNon-Biodegradable Plastics: Duty increased to 25%. \n\nMobiles: Duty reduced to 15% on mobile phones, mobile PCBA, and chargers. \n\nCritical Minerals: Duties exempted for 25 critical minerals. \n\nPlastic Products: Increased customs duty on plastic products. \n\nMethylene Diphenyl Diisocyanate (MDI): Duty reduced from 7.5% to 5% for spandex yarn production. \n\nX-Ray Tubes and Flat Panel Detectors: Duty changes proposed under the Phased Manufacturing Programme. \n\n2024 Budget Highlights: Employer NPS deduction increased from 10% to 14%. \n\nBudget 2024 Highlights: A new solution for NPS (Central Government employees) will be developed based on a review committee's recommendations. \n\nCorporate Tax Rate: The tax rate for foreign companies has been reduced from 40% to 35%. \n\nAssessment Rules: Reopening and reassessment rules have been relaxed. Assessments can now be reopened beyond three years only if undisclosed income exceeds Rs 50 lakh, with a maximum reopening period of five years from the end of the assessment year. \n\nIndian Professionals: Indian professionals working for multinational companies will no longer face penalties for not reporting movable foreign assets, such as ESOPs, if their value is up to Rs 20 lakh. \n\n2024 Budget Capex outlay retained at Rs 11.1 lakh crore \n\nUnion Budget 2024 Highlights: Capital expenditure for infrastructure is set at Rs 11.1 lakh crore, maintaining the interim budget target. \n\nBudget 2024 Railway: For the fiscal year 2024-25, the government has allocated a historic Rs 2,62,200 crore for railway capital expenditure, with gross budgetary support amounting to Rs 2,52,200 crore for the same period. \n\nBudget 2024 Fiscal Deficit \n\nUnion Budget 2024 Highlights: The fiscal deficit for the current fiscal year is projected to be 4.9% of GDP, with a target to lower it to 4.5% by the end of the next fiscal year. \n\nBudget 2024 Highlights: The government has reduced its planned gross market borrowing for FY 2025 by Rs 12,000 crore, setting a revised target of Rs 14.01 lakh crore, down from the Rs 14.13 lakh crore initially announced. \n\nIndia Budget 2024 Highlights: The Pradhan Mantri Janatiya Unnat Gram Abhiyan has been launched to benefit 5 crore tribal people in 63,000 tribal villages, aiming for full coverage in tribal-majority villages and aspirational districts. \n\nBudget 2024 Highlights: The central government will encourage states to lower stamp duty and is considering additional reductions for properties bought by women. \n\n2024 Budget Highlights: The Centre plans to set up 100 branches of India Post Payments Bank in the North East region. \n\nUnion Budget 2024 Highlights: Twelve new industrial parks will be established under the National Industrial Corridor Development Programme. \n\nBudget 2024 Highlights: Rs 10 lakh crore has been allocated to provide 1 crore houses for the urban poor and middle class. \n\n2024 Budget Highlights: Finance Minister Sitharaman announced a critical mineral mission focusing on domestic production, recycling, and overseas acquisition of critical mineral assets. The government will also auction its first offshore mining blocks. \n\nUnion Budget 2024 Highlights: A rental housing scheme for the working class will be introduced, providing dormitory-style rental housing for industrial workers through public-private partnerships. \n\nBudget 2024 Highlights: The government will amend the Insolvency and Bankruptcy Code (IBC) and enhance tribunals. An Integrated Technology Platform will be created to improve IBC processes. \n\nBudget Announcement for MSMEs \n\nBudget 2024 Highlights: The eligibility limit for Mudra loans will be increased to Rs 20 lakh from Rs 10 lakh for those who have successfully repaid loans under the TARUN category. \n\n2024 Budget Highlights: To help MSMEs unlock working capital by converting trade receivables into cash, the turnover threshold for mandatory onboarding on the TReDS platform will be reduced from Rs 500 crore to Rs 250 crore. This change is expected to include 22 additional CPSEs and 7,000 more companies on the platform. \n\nIndia Budget 2024 Highlights: New credit guarantee schemes for MSMEs will be introduced, providing a guarantee cover of Rs 100 crore for larger loan amounts. \n\nUnion Budget 2024 Highlights: SIDBI plans to open 24 new branches over the next three years to broaden its reach. \n\nBudget 2024 Highlights: Financial support will be provided for establishing 50 multi-product food irradiation units and 100 food quality and safety testing labs. \n\n2024 Budget Highlights: E-Commerce Export Hubs will be created through public-private partnerships (PPP). \n\nBudget Announcement for Andhra Pradesh \n\n2024 Budget Highlights: Social and infra funds announced for Andhra Pradesh. Backward region grant will be provided to three districts of Andhra Pradesh. \n\n2024 Budget Highlights: Package announced for Vizag-Chennai Industrial corridor \n\nBudget 2024 Highlights: Government to provide Rs 15,000 crore financial assistance for Andhra Pradesh for development of Amaravati. Additional amounts promised for future years \n\nUnion Budget 2024 Highlights: Centre to finance and early completion of Polavaram Irrigation project, considered as a lifeline for Andhra Pradesh and its farmers. \n\nIndia Budget Highlights: Under the AP Reorganisation Act, funds will be provided for essential infrastructure such as water, power, railways, and roads and also grants for backward regions of Rayalaseema, Prakasam North coastal Andhra Pradesh, as stated in the Act will be provided. \n\nBudget Announcement for Bihar \n\nUnion Budget 2024 Highlights: Road and expressways bonanza announced for Bihar \n\nIndia Budget 2024 Highlights: Industrial node at Gaya on Amritsar Kolkata Industrial Corridor to be developed. Road connectivity projects to be developed at the cost of Rs 26,000 crore: \n\nPatna-Purnia Expressway \n\nBuxar-Bhagalpur Expressway \n\nBodhgaya, Rajgir, Vaishali and Darbhanga Spur \n\nAdditional 2-lane bridge over river Ganga at Buzar \n\nBudget announcement for Bihar: Funds to develop temples in Bihar and Odisha \n\nUnion Budget 2024 announcement for Bihar: 2400 MW power projects to be set up at Pirpainti with an investment of Rs 21,400 crore. New airports, medical colleges and sports infrastructure in Bihar will be constructed. \n\n2024 Budget announcement for Bihar: An additional allocation to support capital investments will be provided through external assistance from multilateral development banks. \n\nIndia Budget 2024 announcement: Govt to provide assistance to flood-affected states like Himachal, Uttarakhand, Sikkim and Assam \n\nBudget 2024 Announcement for Employment and Skill Development \n\nBudget announcement for jobs: Three employment-linked schemes to be launched based on enrolment in the EPFO. Budget

```
from langchain.text_splitter import RecursiveCharacterTextSplitter
```

```
# split data
text_splitter = RecursiveCharacterTextSplitter(chunk_size=1000)
docs = text_splitter.split_documents(data)
```

```
print("Total number of documents: ",len(docs))
```

➡ Total number of documents: 173

docs[7]

Document(metadata={'source': '<https://www.livemint.com/economy/budget-2024-key-highlights-live-updates-nirmala-sitharaman-infrastructure-defence-income-tax-modi-budget-23-july-11721654502862.html>'}, page_content='This initiative, alongside the comprehensive development plans for Rajgir and Nalanda and the support for Odisha's diverse attractions, reflects a visionary approach to tourism. Promoting these regions will not only enhance India's cultural heritage but also drive substantial economic growth. The expected increase in tourism will create numerous jobs and stimulate growth in the hospitality, infrastructure, technology, and local craft sectors.\n\nThis holistic strategy positions India as a leading global destination, highlighting its rich cultural legacy while promoting sustainable economic development.'\n\n24 Jul 2024, 09:45:48 PM IST\n\nBudget 2024 Key Highlights Live Updates: Rohan Bhargava, Co-Founder of CashKaro, said')

```
# Embedding models: https://python.langchain.com/v0.1/docs/integrations/text\_embedding/
# Let's load the Hugging Face Embedding class. sentence_transformers
from langchain_community.embeddings import HuggingFaceEmbeddings
embeddings = HuggingFaceEmbeddings(
    model_name="sentence-transformers/all-mpnet-base-v2", model_kwargs={"device": "cpu"}
)
```

```

)
# embeddings = HuggingFaceEmbeddings(model_name="sentence-transformers/all-mpnet-base-v2")
vector = embeddings.embed_query("hello")
vector[:3]

<ipython-input-7-94531d9e87ae>:4: LangChainDeprecationWarning: The class `HuggingFaceEmbeddings` was deprecated in LangChain 0.2.2 and w
embeddings = HuggingFaceEmbeddings(
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secre
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
warnings.warn(
model.safetensors: 100% 438M/438M [00:04<00:00, 99.3MB/s]

tokenizer_config.json: 100% 363/363 [00:00<00:00, 13.9kB/s]

vocab.txt: 100% 232k/232k [00:00<00:00, 2.34MB/s]

tokenizer.json: 100% 466k/466k [00:00<00:00, 7.34MB/s]

special_tokens_map.json: 100% 239/239 [00:00<00:00, 9.86kB/s]

1_Pooling/config.json: 100% 190/190 [00:00<00:00, 7.91kB/s]
[0.030639857053756714, -0.0062302048318088055, -0.0021215202286839485]

from langchain_chroma import Chroma
from langchain_community.embeddings import HuggingFaceEmbeddings

vectorstore = Chroma.from_documents(
    documents=docs,
    # embedding=HuggingFaceEmbeddings(model_name="sentence-transformers/all-MiniLM-L6-v2", model_kwargs={"device": "cpu"})
    embedding=HuggingFaceEmbeddings(model_name="sentence-transformers/all-mpnet-base-v2", model_kwargs={"device": "cpu"})
)

retriever = vectorstore.as_retriever(search_type="similarity", search_kwargs={"k": 3})

retrieved_docs = retriever.invoke("Budget highlights")

len(retrieved_docs)

3

print(retrieved_docs[0].page_content)

24 Jul 2024, 02:01:06 PM IST

Budget 2024 Key Highlights Live Updates: Manish Shah praises balanced budget supporting MSMEs and urban housing

Manish Shah, MD & CEO at Godrej Capital, commended the 2024 budget for striking a perfect balance between short-term needs and long-term

!pip install huggingface_hub

Requirement already satisfied: huggingface_hub in /usr/local/lib/python3.11/dist-packages (0.27.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from huggingface_hub) (3.17.0)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface_hub) (2024.10.0)
Requirement already satisfied: packaging>=20.9 in /usr/local/lib/python3.11/dist-packages (from huggingface_hub) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from huggingface_hub) (6.0.2)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from huggingface_hub) (2.32.3)
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.11/dist-packages (from huggingface_hub) (4.67.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface_hub) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface_hub) (3.4)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface_hub) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface_hub) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface_hub) (2024.12.1)

from langchain_huggingface import HuggingFacePipeline
from langchain.prompts import PromptTemplate
from transformers import pipeline
from langchain_core.output_parsers import StrOutputParser
import torch
from transformers import AutoModelForCausalLM, AutoTokenizer
from huggingface_hub import login

```

```
# model_id = "meta-llama/Meta-Llama-3-8B"
model_id = "distilgpt2"
login(token='') # Token to be added here

text_generation_pipeline = pipeline(
    "text-generation", model=model_id, model_kwargs={"torch_dtype": torch.bfloat16}, max_new_tokens=400, device=-1)

# torch.cuda.empty_cache()
llm = HuggingFacePipeline(pipeline=text_generation_pipeline)

prompt_template = """
<|system|>
Answer the question based on your knowledge. Use the following context to help:

{context}

</s>
<|user|>
{question}
</s>
<|assistant|>

"""

prompt = PromptTemplate(
    input_variables=["context", "question"],
    template=prompt_template,
)

llm_chain = prompt | llm | StrOutputParser()
```

```
🔄 config.json: 100% 762/762 [00:00<00:00, 11.9kB/s]
model.safetensors: 100% 353M/353M [00:05<00:00, 52.0MB/s]
generation_config.json: 100% 124/124 [00:00<00:00, 2.10kB/s]
tokenizer_config.json: 100% 26.0/26.0 [00:00<00:00, 324B/s]
vocab.json: 100% 1.04M/1.04M [00:00<00:00, 11.6MB/s]
merges.txt: 100% 456k/456k [00:00<00:00, 11.6MB/s]
tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 18.6MB/s]
Device set to use cpu
```

```
!pip install transformers datasets
```

```
🔄 Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-packages (4.47.1)
Collecting datasets
  Downloading datasets-3.2.0-py3-none-any.whl.metadata (20 kB)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from transformers) (3.17.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.24.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.27.1)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (1.26.4)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2024.11.6)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.21.0)
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.5.2)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (17.0.0)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess<0.70.17 (from datasets)
  Downloading multiprocess-0.70.16-py311-none-any.whl.metadata (7.2 kB)
Collecting fsspec<=2024.9.0,>=2023.1.0 (from fsspec[http]<=2024.9.0,>=2023.1.0->datasets)
  Downloading fsspec-2024.9.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.11/dist-packages (from datasets) (3.11.11)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (2.4.4)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (24.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (6.1.0)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (0.2.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.18.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.24.1)
```

```

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.4.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2024.12.14)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas->datasets) (1.
Downloading datasets-3.2.0-py3-none-any.whl (480 kB)
480.6/480.6 kB 9.0 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
116.3/116.3 kB 7.3 MB/s eta 0:00:00
Downloading fsspec-2024.9.0-py3-none-any.whl (179 kB)
179.3/179.3 kB 9.0 MB/s eta 0:00:00
Downloading multiprocess-0.70.16-py311-none-any.whl (143 kB)
143.5/143.5 kB 8.2 MB/s eta 0:00:00
Downloading xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
194.8/194.8 kB 11.9 MB/s eta 0:00:00
Installing collected packages: xxhash, fsspec, dill, multiprocess, datasets
Attempting uninstall: fsspec
Found existing installation: fsspec 2024.10.0
Uninstalling fsspec-2024.10.0:
Successfully uninstalled fsspec-2024.10.0
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
ERROR: gcsfs 2024.10.0 requires fsspec==2024.10.0, but you have fsspec 2024.9.0 which is incompatible.

```

```

from langchain_huggingface import HuggingFacePipeline
from langchain.prompts import PromptTemplate
from transformers import pipeline, AutoTokenizer
from langchain_core.output_parsers import StrOutputParser
import torch
from transformers import AutoModelForCausalLM
from huggingface_hub import login

# model_id = "meta-llama/Meta-Llama-3-8B"
model_id = "distilgpt2"
login(token='hf_rgtFgJmTsuypFkirZffouWSiguSnISuPf')

# Load the tokenizer to ensure consistent tokenization
tokenizer = AutoTokenizer.from_pretrained(model_id)

text_generation_pipeline = pipeline(
    "text-generation",
    model=model_id,
    tokenizer=tokenizer, # Pass the tokenizer explicitly
    model_kwargs={"torch_dtype": torch.bfloat16},
    max_new_tokens=400,
    device=-1
)

# torch.cuda.empty_cache()
llm = HuggingFacePipeline(pipeline=text_generation_pipeline)

# Modify the prompt template to remove special tokens that distilgpt2 may not understand.
prompt_template = """
Answer the question based on your knowledge. Use the following context to help:

{context}

Question: {question}

Answer:
"""

prompt = PromptTemplate(
    input_variables=["context", "question"],
    template=prompt_template,
)

llm_chain = prompt | llm | StrOutputParser()

🔄 Device set to use cpu

from langchain_core.runnables import RunnablePassthrough

rag_chain = {"context": retriever, "question": RunnablePassthrough()} | llm_chain

```

```
question = "2024 Budget Highlights"
```

```
rag_chain.invoke(question)
```

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

Answer the question based on your knowledge. Use the following context to help:

[Document(id='f3aade14-2d82-4ed1-acd5-040186f273da', metadata={'source': 'https://economictimes.indiatimes.com/news/economy/policy/budget-2024-highlights-india-nirmala-sitharaman-capex-fiscal-deficit-tax-slab-key-announcement-in-union-budget-2024-25/articleshow/111942707.cms?from=mdr'}, page_content='2024 Budget Announcement for Space Industry\\n\\n> Budget 2024 Highlights: A venture capital fund of Rs 1,000 crore will be set up. > 2024 India Budget Highlights: Land records in urban areas will be digitized with GIS mapping. An IT based system for property record administration, updating, and tax administration will be established. > Union Budget 2024 Highlights: Sitharaman said that the Budget is focused on poor, women, youth and farmer. > India Budget 2024 Highlights: Global uncertainty poses a downside for growth. Despite these headwinds

```
question = "What is the Union Budget?"
```

```
response = rag_chain.invoke(question)
```

```
# Making the response readable
```

```
response = response.replace("</s>", "").strip()
```

```
print("Response:", response)
```

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

This is a friendly reminder - the current text generation call will exceed the model's predefined maximum length (1024). Depending on the Response: Answer the question based on your knowledge. Use the following context to help:

[Document(id='7f522229-556c-4130-8914-892d915e14c6', metadata={'source': 'https://economictimes.indiatimes.com/news/economy/policy/budget-2024-highlights-india-nirmala-sitharaman-capex-fiscal-deficit-tax-slab-key-announcement-in-union-budget-2024-25/articleshow/111942707.cms?from=mdr'}, page_content='2024 Budget Announcement for Space Industry\\n\\n> Budget 2024 Highlights: A venture capital fund of Rs 1,000 crore will be set up. > 2024 India Budget Highlights: Land records in urban areas will be digitized with GIS mapping. An IT based system for property record administration, updating, and tax administration will be established. > Union Budget 2024 Highlights: Sitharaman said that the Budget is focused on poor, women, youth and farmer. > India Budget 2024 Highlights: Global uncertainty poses a downside for growth. Despite these headwinds

Question: What is the Union Budget?

Answer:

- First, there is no point in comparing government finances to the fiscal targets in the last few months. The last thing we want to do is
- The budget level is very similar to fiscal targets, but there are different fiscal parameters by some of the most important dimensions
 - The budget level is very similar to fiscal targets, so there are different fiscal parameters by some of the most important dimensions:
 - The budget level is very similar to fiscal targets, but there are different fiscal parameters by some of the most important features:
 - The budget level is very similar to fiscal targets, but there are different fiscal parameters by some of the most important aspects, i
 - The budget level is very similar to fiscal targets, but there are different fiscal parameters by some of the most important aspects, i
 - The budget level is very similar to fiscal targets, but there are different fiscal parameters by some of the most important aspects of

```
from langchain_huggingface import HuggingFacePipeline
from langchain.prompts import PromptTemplate
from transformers import pipeline
from langchain_core.output_parsers import StrOutputParser
import torch
from transformers import AutoModelForCausalLM, AutoTokenizer
```

```
model_name = "EleutherAI/gpt-neo-1.3B"
```

```
model = AutoModelForCausalLM.from_pretrained(model_name)
```

```
tokenizer = AutoTokenizer.from_pretrained(model_name)
```

```
text_generation_pipeline = pipeline(
    model=model,
    tokenizer=tokenizer,
    task="text-generation",
    temperature=0.1,
    do_sample=True,
    repetition_penalty=1.1,
    return_full_text=True,
    max_new_tokens=100,
)
llm = HuggingFacePipeline(pipeline=text_generation_pipeline)
```



```

prompt_template = """
<|system|>
Answer the question based on your knowledge. Use the following context to help:

{context}

</s>
<|user|>
{question}
</s>
<|assistant|>

"""

prompt = PromptTemplate(
    input_variables=["context", "question"],
    template=prompt_template,
)

llm_chain = prompt | llm | StrOutputParser()

from langchain_core.runnables import RunnablePassthrough
rag_chain = {"context": retriever, "question": RunnablePassthrough()} | llm_chain

question = "What is the Union Budget?"

response = rag_chain.invoke(question)

# Making the response readable
response = response.replace("</s>", "").strip()
print("Response:", response)

```

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

Response: <|system|>
 Answer the question based on your knowledge. Use the following context to help:

[Document(id='7f522229-556c-4130-8914-892d915e14c6', metadata={'source': 'https://economictimes.indiatimes.com/news/economy/policy/budg

<|user|>
 What is the Union Budget?

<|assistant|>

Start coding or [generate](#) with AI.