

Chatpdf - lets chat with our documents

February 23, 2024

```
[1]: !pip install pypdf
      !pip install python-dotenv
```

Requirement already satisfied: pypdf in /usr/local/lib/python3.10/dist-packages (4.0.2)

Requirement already satisfied: python-dotenv in /usr/local/lib/python3.10/dist-packages (1.0.1)

```
[2]: !pip install -q transformers einops accelerate langchain bitsandbytes
      ↪sentence_transformers
```

```
[3]: !pip install llama-index-core llama-index-readers-file llama-index-llms-ollama
      ↪llama-index-embeddings-huggingface llama-index-llms-replicate
      ↪llama-index-llms-huggingface
      # !pip install llama-index==0.10.12
```

Requirement already satisfied: llama-index-core in /usr/local/lib/python3.10/dist-packages (0.10.12)

Requirement already satisfied: llama-index-readers-file in /usr/local/lib/python3.10/dist-packages (0.1.5)

Requirement already satisfied: llama-index-llms-ollama in /usr/local/lib/python3.10/dist-packages (0.1.2)

Requirement already satisfied: llama-index-embeddings-huggingface in /usr/local/lib/python3.10/dist-packages (0.1.3)

Requirement already satisfied: llama-index-llms-replicate in /usr/local/lib/python3.10/dist-packages (0.1.3)

Requirement already satisfied: llama-index-llms-huggingface in /usr/local/lib/python3.10/dist-packages (0.1.3)

Requirement already satisfied: PyYAML>=6.0.1 in /usr/local/lib/python3.10/dist-packages (from llama-index-core) (6.0.1)

Requirement already satisfied: SQLAlchemy[asyncio]>=1.4.49 in /usr/local/lib/python3.10/dist-packages (from llama-index-core) (2.0.27)

Requirement already satisfied: aiohttp<4.0.0,>=3.8.6 in /usr/local/lib/python3.10/dist-packages (from llama-index-core) (3.9.3)

Requirement already satisfied: dataclasses-json in /usr/local/lib/python3.10/dist-packages (from llama-index-core) (0.6.4)

Requirement already satisfied: deprecated>=1.2.9.3 in /usr/local/lib/python3.10/dist-packages (from llama-index-core) (1.2.14)

```
jinja2->torch<3.0.0,>=2.1.2->llama-index-embeddings-huggingface) (2.1.5)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-
packages (from sympy->torch<3.0.0,>=2.1.2->llama-index-embeddings-huggingface)
(1.3.0)
```

```
[ ]: # !pip install torch-scatter -f https://data.pyg.org/whl/torch-1.13.1+cu117.html
```

```
[4]: import logging
import sys
import transformers

logging.basicConfig(stream=sys.stdout, level=logging.INFO)
logging.getLogger().addHandler(logging.StreamHandler(stream=sys.stdout))

from llama_index.core import VectorStoreIndex, SimpleDirectoryReader, ServiceContext, Settings, PromptTemplate
from llama_index.llms.huggingface import HuggingFaceLLM
# from transformers.models.llama import modeling_llama
```

```
[5]: document = SimpleDirectoryReader(input_files=['/content/drive/MyDrive/Colab_
↳Notebooks/NLP/Lllamaindex/data/unique_document.pdf']).load_data()
```

```
[6]: query_wrapper_prompt = PromptTemplate(
    "Below is an instruction that describes a task. "
    "Write a response that appropriately completes the request.\n\n"
    "### Instruction:\n{query_str}\n\n### Response:"
)

query_wrapper_prompt = PromptTemplate("<|USER|>{query_str}<|ASSISTANT|>")
```

```
[7]: !huggingface-cli login
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85															

A token is already saved on your machine. Run ``huggingface-cli whoami`` to get more information or ``huggingface-cli logout`` if you want to log out.

Setting a new token will erase the existing one.

To login, ``huggingface_hub`` requires a token generated from <https://huggingface.co/settings/tokens> .

Token:

Add token as git credential? (Y/n) y
 Token is valid (permission: read).
 Cannot authenticate through git-credential as no helper is defined on
 your machine.
 You might have to re-authenticate when pushing to the Hugging Face Hub.
 Run the following command in your terminal in case you want to set the 'store'
 credential helper as default.
 git config --global credential.helper store
 Read <https://git-scm.com/book/en/v2/Git-Tools-Credential-Storage> for more
 details.
 Token has not been saved to git credential helper.
 Your token has been saved to /root/.cache/huggingface/token
 Login successful

```
[8]: import torch

llm = HuggingFaceLLM(
    context_window=2048,
    max_new_tokens=256,
    generate_kwargs={"temperature": 0.25, "do_sample": False},
    query_wrapper_prompt=query_wrapper_prompt,
    tokenizer_name="meta-llama/Llama-2-7b-chat-hf",
    model_name="meta-llama/Llama-2-7b-chat-hf",
    device_map="auto",
    tokenizer_kwargs={"max_length": 2048},
    # uncomment this if using CUDA to reduce memory usage
    model_kwargs={"torch_dtype": torch.float16, "load_in_8bit": True}
)

Settings.chunk_size = 512
Settings.llm = llm
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88:
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 secret in your Google Colab
and restart your session.
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.index.json: 0%|          | 0.00/26.8k [00:00<?, ?B/s]
```

```

Downloading shards: 0%|          | 0/2 [00:00<?, ?it/s]
model-00001-of-00002.safetensors: 0%|          | 0.00/9.98G [00:00<?, ?B/s]
model-00002-of-00002.safetensors: 0%|          | 0.00/3.50G [00:00<?, ?B/s]
Loading checkpoint shards: 0%|          | 0/2 [00:00<?, ?it/s]
generation_config.json: 0%|          | 0.00/188 [00:00<?, ?B/s]
tokenizer_config.json: 0%|          | 0.00/1.62k [00:00<?, ?B/s]
tokenizer.model: 0%|          | 0.00/500k [00:00<?, ?B/s]
tokenizer.json: 0%|          | 0.00/1.84M [00:00<?, ?B/s]
special_tokens_map.json: 0%|          | 0.00/414 [00:00<?, ?B/s]

```

[12]: `!pip install llama-index-embeddings-langchain`

```

Collecting llama-index-embeddings-langchain
  Downloading llama_index_embeddings_langchain-0.1.2-py3-none-any.whl (2.5 kB)
Requirement already satisfied: llama-index-core<0.11.0,>=0.10.1 in
/usr/local/lib/python3.10/dist-packages (from llama-index-embeddings-langchain)
(0.10.12)
Requirement already satisfied: PyYAML>=6.0.1 in /usr/local/lib/python3.10/dist-
packages (from llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-
langchain) (6.0.1)
Requirement already satisfied: SQLAlchemy[asyncio]>=1.4.49 in
/usr/local/lib/python3.10/dist-packages (from llama-index-
core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (2.0.27)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.6 in
/usr/local/lib/python3.10/dist-packages (from llama-index-
core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (3.9.3)
Requirement already satisfied: dataclasses-json in
/usr/local/lib/python3.10/dist-packages (from llama-index-
core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (0.6.4)
Requirement already satisfied: deprecated>=1.2.9.3 in
/usr/local/lib/python3.10/dist-packages (from llama-index-
core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (1.2.14)
Requirement already satisfied: dirtyjson<2.0.0,>=1.0.8 in
/usr/local/lib/python3.10/dist-packages (from llama-index-
core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (1.0.8)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.10/dist-packages (from llama-index-
core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (2023.6.0)
Requirement already satisfied: httpx in /usr/local/lib/python3.10/dist-packages
(from llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain)
(0.27.0)
Requirement already satisfied: llamaindex-py-client<0.2.0,>=0.1.13 in
/usr/local/lib/python3.10/dist-packages (from llama-index-
core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (0.1.13)

```

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.31.0->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (2.0.7)

Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.10/dist-packages (from SQLAlchemy[asyncio]>=1.4.49->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (3.0.3)

Requirement already satisfied: mypy-extensions>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from typing-inspect>=0.8.0->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (1.0.0)

Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in /usr/local/lib/python3.10/dist-packages (from dataclasses-json->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (3.20.2)

Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (2.8.2)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (2023.4)

Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anyio->httpx->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (1.2.0)

Requirement already satisfied: packaging>=17.0 in /usr/local/lib/python3.10/dist-packages (from marshmallow<4.0.0,>=3.18.0->dataclasses-json->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (23.2)

Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic>=1.10->llama-index-py-client<0.2.0,>=0.1.13->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (0.6.0)

Requirement already satisfied: pydantic-core==2.16.2 in /usr/local/lib/python3.10/dist-packages (from pydantic>=1.10->llama-index-py-client<0.2.0,>=0.1.13->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (2.16.2)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas->llama-index-core<0.11.0,>=0.10.1->llama-index-embeddings-langchain) (1.16.0)

Installing collected packages: llama-index-embeddings-langchain

Successfully installed llama-index-embeddings-langchain-0.1.2

```
[13]: from langchain.embeddings.huggingface import HuggingFaceEmbeddings
# from llama_index import ServiceContext
from llama_index.embeddings.langchain import LangchainEmbedding

embed_model=LangchainEmbedding(
    HuggingFaceEmbeddings(model_name="sentence-transformers/all-mpnet-base-v2"))
```

modules.json: 0% | 0.00/349 [00:00<?, ?B/s]

```

config_sentence_transformers.json: 0%|          | 0.00/116 [00:00<?, ?B/s]
README.md: 0%|          | 0.00/10.6k [00:00<?, ?B/s]
sentence_bert_config.json: 0%|          | 0.00/53.0 [00:00<?, ?B/s]
config.json: 0%|          | 0.00/571 [00:00<?, ?B/s]
pytorch_model.bin: 0%|          | 0.00/438M [00:00<?, ?B/s]
tokenizer_config.json: 0%|          | 0.00/363 [00:00<?, ?B/s]
vocab.txt: 0%|          | 0.00/232k [00:00<?, ?B/s]
tokenizer.json: 0%|          | 0.00/466k [00:00<?, ?B/s]
special_tokens_map.json: 0%|          | 0.00/239 [00:00<?, ?B/s]
1_Pooling/config.json: 0%|          | 0.00/190 [00:00<?, ?B/s]

```

```

[14]: service_context=ServiceContext.from_defaults(
        chunk_size=1024,
        llm=llm,
        embed_model=embed_model
    )
    service_context

```

```

<ipython-input-14-efaa4bbc4276>:1: DeprecationWarning: Call to deprecated class
method from_defaults. (ServiceContext is deprecated, please use
`llama_index.settings.Settings` instead.) -- Deprecated since version 0.10.0.
    service_context=ServiceContext.from_defaults(

```

```

[14]: ServiceContext(llm_predictor=LLMPredictor(system_prompt=None,
query_wrapper_prompt=None, pydantic_program_mode=<PydanticProgramMode.DEFAULT:
'default'>), prompt_helper=PromptHelper(context_window=2048, num_output=256,
chunk_overlap_ratio=0.1, chunk_size_limit=None, separator=' '),
embed_model=LangchainEmbedding(model_name='sentence-transformers/all-mpnet-
base-v2', embed_batch_size=10,
callback_manager=<llama_index.core.callbacks.base.CallbackManager object at
0x7e2f72a308b0>), transformations=[SentenceSplitter(include_metadata=True,
include_prev_next_rel=True,
callback_manager=<llama_index.core.callbacks.base.CallbackManager object at
0x7e2f72a308b0>, id_func=<function default_id_func at 0x7e2fa8015e10>,
chunk_size=1024, chunk_overlap=200, separator=' ', paragraph_separator='\n\n\n',
secondary_chunking_regex='[^\.; ]+[.,; ]?')],
llama_logger=<llama_index.core.service_context_elements.llama_logger.LlamaLogger
object at 0x7e2f72a301f0>,
callback_manager=<llama_index.core.callbacks.base.CallbackManager object at
0x7e2f72a308b0>)

```

```

[15]: index=VectorStoreIndex.from_documents(document,service_context=service_context)
    query_engine=index.as_query_engine()

```

```
response=query_engine.query("what is the document about of?")
print(response)
```

```
/usr/local/lib/python3.10/dist-
packages/transformers/generation/configuration_utils.py:392: UserWarning:
`do_sample` is set to `False`. However, `temperature` is set to `0.25` -- this
flag is only used in sample-based generation modes. You should set
`do_sample=True` or unset `temperature`.
  warnings.warn(
/usr/local/lib/python3.10/dist-
packages/transformers/generation/configuration_utils.py:397: UserWarning:
`do_sample` is set to `False`. However, `top_p` is set to `0.9` -- this flag is
only used in sample-based generation modes. You should set `do_sample=True` or
unset `top_p`.
  warnings.warn(
```

The document appears to be about an AI Research Engineer named Rohit Chauhan, highlighting his skills, experience, and achievements in the field of AI and ML. The document provides information about his educational background, certifications, and professional experience, showcasing his expertise in various areas such as Generative AI, NLP, CV, Time Series Analysis, and more. Based on the information provided, it can be inferred that the document is likely related to Rohit's professional portfolio or resume.

```
[21]: chat={}
while True:
    chat=chat
    query=input()
    print('user question:',query)
    response=query_engine.query(query)
    print('Answer from the document:',response)
    chat[query]=response
```

```
count of the page in the document
user question: count of the page in the document
Answer from the document: The count of pages in the document is 2.
what is the document about?
user question: what is the document about?
Answer from the document: The document appears to be a resume or portfolio of an
AI/ML engineer named Rohit Chauhan. The document highlights his educational
background, skills, and professional experience in developing innovative AI/ML
solutions for various industries. The document also mentions his leadership
abilities and passion for mentoring team members. Based on the information
provided, it can be inferred that the document is about Rohit Chauhan's
professional experience and skills in the field of AI/ML.
mention me candidates skill if he have any?
user question: mention me candidates skill if he have any?
Answer from the document: Based on the provided context information, Rohit
```

Chauhan has the following skills:

1. Programming languages: Python
2. Scraping: Selenium, BeautifulSoup, PowerAutomate
3. Machine Learning (ML) and Deep Learning (DL): Scikit-Learn, Supervised and Unsupervised learning, Feature Engineering, Neural Networks, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Transfer Learning, Transformer
4. Natural Language Processing (NLP): NLTK (Natural Language ToolKit), Gensim, spaCy, googletrans, Sequence Classification, Sentiment Analysis, Sequence2Sequence
5. Computer Vision (CV): OpenCV, Easyocr, PaddleOCR, PIL
6. Time Series Analysis (Future Forecasting): ARIMA, ARIMAX, Long Short-Term Memory (LSTM), Gated Recurrent units (GRU), Exponential Smoothing (ETS), Seasonal -Trend Decomposition using LOESS (STL)
7. Data Modelling: Linear Regression, Logistic Regression, Dec

can you mention candidates last academic?

user question: can you mention candidates last academic?

Answer from the document: Yes, Rohit Chauhan's last academic qualification is Postgraduate Diploma in Data Science and AI from IT-Vedant Education Pvt. Ltd, which he completed in Nov 2020 - Dec 2022.

can you describe me candidates something about professional history?

user question: can you describe me candidates something about professional history?

Answer from the document: Certainly! Based on the provided context information, we can see that Rohit Chauhan is an AI-ML Engineer with experience in developing innovative solutions using advanced AI/ML techniques, particularly in Generative AI. He has a proven track record in optimizing and enhancing existing AI models for improved performance, scalability, and efficiency. Rohit has also demonstrated leadership abilities and a passion for mentoring team members and delivering end-user training. Additionally, he has experience in coordinating with cross-functional teams to align AI projects with business objectives and providing technical support to ensure seamless integration of AI solutions.

In terms of professional experience, Rohit worked at Cloudstrats Technologies Private Limited from July 2021 to January 2024 as an Executive - AI and Analytics. During his tenure, he spearheaded the development of multiple ML models for various industries, including Manufacturing, Transportation, BFSI, and more. He also engineered a robust scraper model that boosted data extraction and automated the process efficiency by 70%. Furthermore, Rohit innovated specialized

is this candidate is fit for any AI job?

user question: is this candidate is fit for any AI job?

Answer from the document: Yes, this candidate is fit for any AI job.

Reason:

1. Programming Language: The candidate has experience in programming languages

such as Python, which is a fundamental skill for any AI job.

2. Scraping: The candidate has experience in web scraping using Selenium, BeautifulSoup, and PowerAutomate, which is a valuable skill in AI.

3. Machine Learning and Deep Learning: The candidate has experience in various machine learning and deep learning techniques, including supervised and unsupervised learning, feature engineering, neural networks, and transfer learning.

4. Natural Language Processing: The candidate has experience in NLP techniques such as text summarization, sentiment analysis, emotion detection, and translating models, which is a crucial skill in AI.

5. Computer Vision: The candidate has experience in computer vision techniques such as OpenCV, Easyocr, and PaddleOCR, which is a valuable skill in AI.

6. Time Series Analysis: The candidate has experience in time series analysis using ARIMA, ARIMAX, LSTM, GRU, ETS, and STL, which is a crucial skill in AI.