Project_Chatbot

December 9, 2024

1 Nestle HR ChatBot

Author : Manohari Wijesooriya

Date : Dec 08, 2024

2 Project Goal

To create an intelligent HR Chatbot that provides accurate and user-friendly responses to HR-related inquiries based on Nestlé's HR policies and procedures.

3 Scope and Requirements

- Identify the key HR-related topics (e.g., leave policies, benefits, onboarding, grievance redressal).
- Understand the chatbot's user base (employees, HR team, new hires, etc.).
- Define the chatbot's tone (formal, friendly, etc.).

```
[1]: !pip install tiktoken PyPDF2 openai gradio chromadb pysqlite3-binary pypdf␣
      ↪langchain langchain-community --upgrade --use-feature=fast-deps
```

Defaulting to user installation because normal site-packages is not writeable
WARNING: pip is using lazily downloaded wheels using HTTP range requests to

obtain dependency information. This experimental feature is enabled through

--use-feature=fast-deps and it is not ready for production.

Requirement already satisfied: tiktoken in
/voc/work/.local/lib/python3.10/site-packages (0.8.0)
Requirement already satisfied: PyPDF2 in /voc/work/.local/lib/python3.10/site-
packages (3.0.1)
Requirement already satisfied: openai in /voc/work/.local/lib/python3.10/site-
packages (1.57.0)
Requirement already satisfied: gradio in /voc/work/.local/lib/python3.10/site-
packages (5.8.0)
Requirement already satisfied: chromadb in /voc/work/.local/lib/python3.10/site-

packages (0.5.23)
Requirement already satisfied: pysqlite3-binary in
/voc/work/.local/lib/python3.10/site-packages (0.5.4)
Requirement already satisfied: pypdf in /voc/work/.local/lib/python3.10/site-
packages (5.1.0)
Requirement already satisfied: langchain in
/voc/work/.local/lib/python3.10/site-packages (0.3.10)
Requirement already satisfied: langchain-community in
/voc/work/.local/lib/python3.10/site-packages (0.3.10)
Requirement already satisfied: regex>=2022.1.18 in
/usr/local/lib/python3.10/site-packages (from tiktoken) (2023.10.3)
Requirement already satisfied: requests>=2.26.0 in
/usr/local/lib/python3.10/site-packages (from tiktoken) (2.31.0)
Requirement already satisfied: anyio<5,>=3.5.0 in
/usr/local/lib/python3.10/site-packages (from openai) (3.7.1)
Requirement already satisfied: distro<2,>=1.7.0 in
/usr/local/lib/python3.10/site-packages (from openai) (1.7.0)
Requirement already satisfied: httpx<1,>=0.23.0 in
/voc/work/.local/lib/python3.10/site-packages (from openai) (0.28.1)
Requirement already satisfied: jiter<1,>=0.4.0 in
/voc/work/.local/lib/python3.10/site-packages (from openai) (0.7.1)
Requirement already satisfied: pydantic<3,>=1.9.0 in
/voc/work/.local/lib/python3.10/site-packages (from openai) (2.10.3)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/site-
packages (from openai) (1.3.0)
Requirement already satisfied: tqdm>4 in /usr/local/lib/python3.10/site-packages
(from openai) (4.66.1)
Requirement already satisfied: typing-extensions<5,>=4.11 in
/voc/work/.local/lib/python3.10/site-packages (from openai) (4.12.2)
Requirement already satisfied: aiofiles<24.0,>=22.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (23.2.1)
Requirement already satisfied: fastapi<1.0,>=0.115.2 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (0.115.6)
Requirement already satisfied: ffmpy in /usr/local/lib/python3.10/site-packages
(from gradio) (0.3.1)
Requirement already satisfied: gradio-client==1.5.1 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (1.5.1)
Requirement already satisfied: huggingface-hub>=0.25.1 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (0.26.2)
Requirement already satisfied: jinja2<4.0 in /usr/local/lib/python3.10/site-
packages (from gradio) (3.1.2)
Requirement already satisfied: markupsafe~=2.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (2.1.3)
Requirement already satisfied: numpy<3.0,>=1.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (1.26.2)
Requirement already satisfied: orjson~=3.0 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (3.10.10)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/site-

packages (from gradio) (23.2)
Requirement already satisfied: pandas<3.0,>=1.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (2.1.3)
Requirement already satisfied: pillow<12.0,>=8.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (10.1.0)
Requirement already satisfied: pydub in /usr/local/lib/python3.10/site-packages
(from gradio) (0.25.1)
Requirement already satisfied: python-multipart>=0.0.18 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (0.0.19)
Requirement already satisfied: pyyaml<7.0,>=5.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (6.0.1)
Requirement already satisfied: ruff>=0.2.2 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (0.8.2)
Requirement already satisfied: safehttpx<0.2.0,>=0.1.6 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (0.1.6)
Requirement already satisfied: semantic-version~=2.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (2.10.0)
Requirement already satisfied: starlette<1.0,>=0.40.0 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (0.41.3)
Requirement already satisfied: tomlkit<0.14.0,>=0.12.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (0.12.0)
Requirement already satisfied: typer<1.0,>=0.12 in
/voc/work/.local/lib/python3.10/site-packages (from gradio) (0.15.1)
Requirement already satisfied: uvicorn>=0.14.0 in
/usr/local/lib/python3.10/site-packages (from gradio) (0.24.0.post1)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/site-packages
(from gradio-client==1.5.1->gradio) (2023.10.0)
Requirement already satisfied: websockets<15.0,>=10.0 in
/usr/local/lib/python3.10/site-packages (from gradio-client==1.5.1->gradio)
(11.0.3)
Requirement already satisfied: build>=1.0.3 in
/voc/work/.local/lib/python3.10/site-packages (from chromadb) (1.2.2.post1)
Requirement already satisfied: chroma-hnswlib==0.7.6 in
/voc/work/.local/lib/python3.10/site-packages (from chromadb) (0.7.6)
Requirement already satisfied: posthog>=2.4.0 in /usr/local/lib/python3.10/site-
packages (from chromadb) (3.0.2)
Requirement already satisfied: onnxruntime>=1.14.1 in
/usr/local/lib/python3.10/site-packages (from chromadb) (1.16.3)
Requirement already satisfied: opentelemetry-api>=1.2.0 in
/usr/local/lib/python3.10/site-packages (from chromadb) (1.21.0)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-grpc>=1.2.0 in
/usr/local/lib/python3.10/site-packages (from chromadb) (1.21.0)
Requirement already satisfied: opentelemetry-instrumentation-fastapi>=0.41b0 in
/usr/local/lib/python3.10/site-packages (from chromadb) (0.42b0)
Requirement already satisfied: opentelemetry-sdk>=1.2.0 in
/usr/local/lib/python3.10/site-packages (from chromadb) (1.21.0)
Requirement already satisfied: tokenizers<=0.20.3,>=0.13.2 in
/voc/work/.local/lib/python3.10/site-packages (from chromadb) (0.20.1)

Requirement already satisfied: pypika>=0.48.9 in /usr/local/lib/python3.10/site-packages (from chromadb) (0.48.9)
Requirement already satisfied: overrides>=7.3.1 in /usr/local/lib/python3.10/site-packages (from chromadb) (7.3.1)
Requirement already satisfied: importlib-resources in /usr/local/lib/python3.10/site-packages (from chromadb) (6.1.1)
Requirement already satisfied: grpcio>=1.58.0 in /usr/local/lib/python3.10/site-packages (from chromadb) (1.59.3)
Requirement already satisfied: bcrypt>=4.0.1 in /usr/local/lib/python3.10/site-packages (from chromadb) (4.1.1)
Requirement already satisfied: kubernetes>=28.1.0 in /usr/local/lib/python3.10/site-packages (from chromadb) (28.1.0)
Requirement already satisfied: tenacity>=8.2.3 in /usr/local/lib/python3.10/site-packages (from chromadb) (8.2.3)
Requirement already satisfied: mmh3>=4.0.1 in /usr/local/lib/python3.10/site-packages (from chromadb) (4.0.1)
Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.10/site-packages (from chromadb) (13.5.3)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.10/site-packages (from langchain) (1.4.20)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.10/site-packages (from langchain) (3.9.1)
Requirement already satisfied: async-timeout<5.0.0,>=4.0.0 in /usr/local/lib/python3.10/site-packages (from langchain) (4.0.3)
Requirement already satisfied: langchain-core<0.4.0,>=0.3.22 in /voc/work/.local/lib/python3.10/site-packages (from langchain) (0.3.22)
Requirement already satisfied: langchain-text-splitters<0.4.0,>=0.3.0 in /voc/work/.local/lib/python3.10/site-packages (from langchain) (0.3.0)
Requirement already satisfied: langsmith<0.2.0,>=0.1.17 in /voc/work/.local/lib/python3.10/site-packages (from langchain) (0.1.137)
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in /usr/local/lib/python3.10/site-packages (from langchain-community) (0.5.13)
Requirement already satisfied: httpx-sse<0.5.0,>=0.4.0 in /voc/work/.local/lib/python3.10/site-packages (from langchain-community) (0.4.0)
Requirement already satisfied: pydantic-settings<3.0.0,>=2.4.0 in /voc/work/.local/lib/python3.10/site-packages (from langchain-community) (2.6.1)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/site-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (23.1.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/site-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (6.0.4)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/site-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.9.3)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/site-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.4.0)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/site-packages (from aiohttp<4.0.0,>=3.8.3->langchain)

(1.3.1)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.10/site-packages (from anyio<5,>=3.5.0->openai) (3.6)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/site-packages (from anyio<5,>=3.5.0->openai) (1.2.0)
Requirement already satisfied: pyproject_hooks in /voc/work/.local/lib/python3.10/site-packages (from build>=1.0.3->chromadb) (1.2.0)
Requirement already satisfied: tomli>=1.1.0 in /usr/local/lib/python3.10/site-packages (from build>=1.0.3->chromadb) (2.0.1)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in /usr/local/lib/python3.10/site-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community) (3.20.1)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in /usr/local/lib/python3.10/site-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community) (0.9.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/site-packages (from httpx<1,>=0.23.0->openai) (2023.11.17)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.10/site-packages (from httpx<1,>=0.23.0->openai) (1.0.2)
Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.10/site-packages (from httpcore==1.*->httpx<1,>=0.23.0->openai) (0.14.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/site-packages (from huggingface-hub>=0.25.1->gradio) (3.13.1)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.10/site-packages (from kubernetes>=28.1.0->chromadb) (1.16.0)
Requirement already satisfied: python-dateutil>=2.5.3 in /usr/local/lib/python3.10/site-packages (from kubernetes>=28.1.0->chromadb) (2.8.2)
Requirement already satisfied: google-auth>=1.0.1 in /usr/local/lib/python3.10/site-packages (from kubernetes>=28.1.0->chromadb) (2.23.4)
Requirement already satisfied: websocket-client!=0.40.0,!=0.41.*,!=0.42.*,>=0.32.0 in /usr/local/lib/python3.10/site-packages (from kubernetes>=28.1.0->chromadb) (1.3.3)
Requirement already satisfied: requests-oauthlib in /usr/local/lib/python3.10/site-packages (from kubernetes>=28.1.0->chromadb) (1.3.1)
Requirement already satisfied: oauthlib>=3.2.2 in /usr/local/lib/python3.10/site-packages (from kubernetes>=28.1.0->chromadb) (3.2.2)
Requirement already satisfied: urllib3<2.0,>=1.24.2 in /usr/local/lib/python3.10/site-packages (from kubernetes>=28.1.0->chromadb) (1.26.18)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.10/site-packages (from langchain-core<0.4.0,>=0.3.22->langchain) (1.33)

```
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/voc/work/.local/lib/python3.10/site-packages (from
langsmith<0.2.0,>=0.1.17->langchain) (1.0.0)
Requirement already satisfied: coloredlogs in /usr/local/lib/python3.10/site-
packages (from onnxruntime>=1.14.1->chromadb) (15.0.1)
Requirement already satisfied: flatbuffers in /usr/local/lib/python3.10/site-
packages (from onnxruntime>=1.14.1->chromadb) (23.5.26)
Requirement already satisfied: protobuf in /usr/local/lib/python3.10/site-
packages (from onnxruntime>=1.14.1->chromadb) (4.25.1)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/site-packages
(from onnxruntime>=1.14.1->chromadb) (1.12)
Requirement already satisfied: deprecated>=1.2.6 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-
api>=1.2.0->chromadb) (1.2.13)
Requirement already satisfied: importlib-metadata<7.0,>=6.0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-
api>=1.2.0->chromadb) (6.8.0)
Requirement already satisfied: backoff<3.0.0,>=1.10.0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0->chromadb) (2.2.1)
Requirement already satisfied: googleapis-common-protos~=1.52 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0->chromadb) (1.61.0)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-common==1.21.0
in /usr/local/lib/python3.10/site-packages (from opentelemetry-exporter-otlp-
proto-grpc>=1.2.0->chromadb) (1.21.0)
Requirement already satisfied: opentelemetry-proto==1.21.0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0->chromadb) (1.21.0)
Requirement already satisfied: opentelemetry-instrumentation-asgi==0.42b0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-instrumentation-
fastapi>=0.41b0->chromadb) (0.42b0)
Requirement already satisfied: opentelemetry-instrumentation==0.42b0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-instrumentation-
fastapi>=0.41b0->chromadb) (0.42b0)
Requirement already satisfied: opentelemetry-semantic-conventions==0.42b0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-instrumentation-
fastapi>=0.41b0->chromadb) (0.42b0)
Requirement already satisfied: opentelemetry-util-http==0.42b0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-instrumentation-
fastapi>=0.41b0->chromadb) (0.42b0)
Requirement already satisfied: setuptools>=16.0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-
instrumentation==0.42b0->opentelemetry-instrumentation-
fastapi>=0.41b0->chromadb) (69.0.2)
Requirement already satisfied: wrapt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.10/site-packages (from opentelemetry-
instrumentation==0.42b0->opentelemetry-instrumentation-
```

```
fastapi>=0.41b0->chromadb) (1.14.1)
Requirement already satisfied: asgiref~=3.0 in /usr/local/lib/python3.10/site-
packages (from opentelemetry-instrumentation-asgi==0.42b0->opentelemetry-
instrumentation-fastapi>=0.41b0->chromadb) (3.7.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/site-
packages (from pandas<3.0,>=1.0->gradio) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/site-
packages (from pandas<3.0,>=1.0->gradio) (2023.3)
Requirement already satisfied: monotonic>=1.5 in /usr/local/lib/python3.10/site-
packages (from posthog>=2.4.0->chromadb) (1.6)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.10/site-packages (from pydantic<3,>=1.9.0->openai)
(0.6.0)
Requirement already satisfied: pydantic-core==2.27.1 in
/voc/work/.local/lib/python3.10/site-packages (from pydantic<3,>=1.9.0->openai)
(2.27.1)
Requirement already satisfied: python-dotenv>=0.21.0 in
/usr/local/lib/python3.10/site-packages (from pydantic-
settings<3.0.0,>=2.4.0->langchain-community) (1.0.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/site-packages (from requests>=2.26.0->tiktoken)
(3.3.2)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.10/site-packages (from rich>=10.11.0->chromadb) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.10/site-packages (from rich>=10.11.0->chromadb) (2.16.1)
Requirement already satisfied: greenlet!=0.4.17 in
/usr/local/lib/python3.10/site-packages (from SQLAlchemy<3,>=1.4->langchain)
(1.1.2)
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.10/site-
packages (from typer<1.0,>=0.12->gradio) (8.1.7)
Requirement already satisfied: shellingham>=1.3.0 in
/usr/local/lib/python3.10/site-packages (from typer<1.0,>=0.12->gradio) (1.5.4)
Requirement already satisfied: httptools>=0.5.0 in
/usr/local/lib/python3.10/site-packages (from
uvicorn[standard]>=0.18.3->chromadb) (0.6.1)
Requirement already satisfied: uvloop!=0.15.0,!=0.15.1,>=0.14.0 in
/usr/local/lib/python3.10/site-packages (from
uvicorn[standard]>=0.18.3->chromadb) (0.19.0)
Requirement already satisfied: watchfiles>=0.13 in
/usr/local/lib/python3.10/site-packages (from
uvicorn[standard]>=0.18.3->chromadb) (0.21.0)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/usr/local/lib/python3.10/site-packages (from google-
auth>=1.0.1->kubernetes>=28.1.0->chromadb) (5.3.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/usr/local/lib/python3.10/site-packages (from google-
auth>=1.0.1->kubernetes>=28.1.0->chromadb) (0.3.0)
```

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/site-packages (from google-auth>=1.0.1->kubernetes>=28.1.0->chromadb) (4.9)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.10/site-packages (from importlib-metadata<7.0,>=6.0->opentelemetry-api>=1.2.0->chromadb) (3.17.0)
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.10/site-packages (from jsonpatch<2.0,>=1.33->langchain-core<0.4.0,>=0.3.22->langchain) (2.3)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/site-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->chromadb) (0.1.1)
Requirement already satisfied: mypy-extensions>=0.3.0 in /usr/local/lib/python3.10/site-packages (from typing-inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain-community) (0.4.3)
Requirement already satisfied: humanfriendly>=9.1 in /usr/local/lib/python3.10/site-packages (from coloredlogs->onnxruntime>=1.14.1->chromadb) (10.0)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/site-packages (from sympy->onnxruntime>=1.14.1->chromadb) (1.3.0)
Requirement already satisfied: pyasn1<0.6.0,>=0.4.6 in /usr/local/lib/python3.10/site-packages (from pyasn1-modules>=0.2.1->google-auth>=1.0.1->kubernetes>=28.1.0->chromadb) (0.5.1)

```python
[2]: from langchain.document_loaders import TextLoader, PyPDFLoader
     from langchain.text_splitter import RecursiveCharacterTextSplitter
     from langchain.embeddings import HuggingFaceEmbeddings, OpenAIEmbeddings
     from langchain.vectorstores import Chroma
     import pysqlite3
     import sys
     sys.modules["sqlite3"] = sys.modules.pop("pysqlite3")
     import chromadb
     import openai
     import os
     from langchain.chains.question_answering import load_qa_chain
     from langchain.prompts import PromptTemplate
     from langchain.chat_models import ChatOpenAI
     import openai
     import gradio as gr
```

```python
[12]: #import openai
      #openai.api_key = "voc-730917260112272230113166a5077cac4479.92242159"
      #print(openai.api_key)
```

voc-730917260112272230113166a5077cac4479.92242159

4 Data Preparation

- Collect Nestlé's HR policy documents, FAQs, and other relevant resources.

- Ensure documents are clean, well-structured, and up-to-date.

```
[3]: #import os
     # Get the current working directory
     current_path = os.getcwd()
     print("Current working directory:", current_path)
```

Current working directory: /voc/work/Project1

```
[12]: #Reading HR policy document
      loader = PyPDFLoader(r"/voc/work/Project1/
       ↪1728286846_the_nestle_hr_policy_pdf_2012.pdf")
```

5 Data Processing

- Import necessary libraries and tools (e.g., PyPDFLoader for PDFs, OpenAI API for embeddings).
- Preprocess and split the documents into smaller, meaningful chunks for embedding and retrieval.

```
[13]: # Split document into chunks
      pages = loader.load_and_split()
      len(pages)
```

```
[13]: 8
```

```
[14]: # Print the last chunk (if chunks are used)
      print(f"Last chunk of text:\n{pages[7].page_content}")
```

```
Last chunk of text:
The Nestlé Human Resources Policy
5
Since its founding, Nestlé has built a culture
based on values of trust, mutual respect and
dialogue. Nestlé management and employees all
over the world work daily to create and maintain
positive individual and collective relationships,
and are expected to do so as a core part of their
job.
Nestlé not only upholds the freedom of
association of its employees and the effective
recognition of the right to collective bargaining,
but also ensures that direct and frequent
communication is established in the workplace.
While dialogue with trade unions is essential, it
does not replace the close relationship that our
management maintains with all employees.
In the spirit of continuous improvement, we
```

encourage two-way dialogue with our employees
that goes beyond the traditional aspects of
collective bargaining in order to share knowledge
and to jointly find opportunities related to
important matters such as Creating Shared Value,
the health and safety in the workplace and our
concern for the environment.
The Company and employee representatives
are expected to make all necessary efforts
to develop fair and constructive dialogues,
overcome the difficulties that they might
encounter, reach sustainable agreements and
implement them.
Corporate policy:
The Nestlé Employee Relations Policy
 Employee relations
The Human Resources management described
in this document requires and supports an
organisation "on the move".
Nestlé is committed to continue the journey
to establishing flat and flexible structures with
minimal levels of management and broad spans
of control, which enable people development,
increase efficiency, and ease implementation
of our "Nestlé Management and Leadership
Principles".
Less hierarchical layers call for increased
cooperation between colleagues. This is what
will make the organisation more flexible and
more accountable. Indeed, it supports today's
and tomorrow's business requirements for an
agile and innovative company working with ever
competitive intensity.
These simple beliefs have inspired us to create
an environment that puts the emphasis not just on
individual responsibility and autonomy, but also
on a strong willingness to support others, to work
in multi-skilled teams, and to cooperate rather
than to compete internally.
A dynamic organisation creates a climate
of innovation and allows people to think from
different perspectives. At Nestlé we encourage
our people to take risks. Mistakes may be made
but there is always a willingness to correct and
learn from them.
We combine the scope and brand strength
of a global company with the creativity and
knowledge of a local business. As a result, people

can have far-reaching influence every day and
explore their full long-term potential, propelled by
continual support and a collaborative approach by
line managers and employees.
Corporate policy:
Nestlé on the Move
 A flexible and dynamic organisation

```python
[15]: # Use a text splitter for better chunking
      text_splitter = RecursiveCharacterTextSplitter(chunk_size=1000,␣
       ↪chunk_overlap=100)
      documents = text_splitter.split_documents(pages)

      # Display total chunks
      print(f"Total chunks: {len(documents)}")
```

Total chunks: 21

```python
[16]: # Initialize ChromaDB for persistent storage
      client = chromadb.PersistentClient(path=r"/voc/work/Project1/
       ↪ChromaDB\Nestle_Chatbot")
```

6 Embeddings and Vector Store

- Create vector representations of the HR documents using OpenAI's embeddings.
- Store embeddings in a vector database like ChromaDB or Pinecone for efficient retrieval.

```python
[17]: !pip install --upgrade openai langchain-openai
      from langchain_openai import OpenAIEmbeddings
```

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: openai in /voc/work/.local/lib/python3.10/site-
packages (1.57.0)
Requirement already satisfied: langchain-openai in
/voc/work/.local/lib/python3.10/site-packages (0.2.11)
Requirement already satisfied: anyio<5,>=3.5.0 in
/usr/local/lib/python3.10/site-packages (from openai) (3.7.1)
Requirement already satisfied: distro<2,>=1.7.0 in
/usr/local/lib/python3.10/site-packages (from openai) (1.7.0)
Requirement already satisfied: httpx<1,>=0.23.0 in
/voc/work/.local/lib/python3.10/site-packages (from openai) (0.28.1)
Requirement already satisfied: jiter<1,>=0.4.0 in
/voc/work/.local/lib/python3.10/site-packages (from openai) (0.7.1)
Requirement already satisfied: pydantic<3,>=1.9.0 in
/voc/work/.local/lib/python3.10/site-packages (from openai) (2.10.3)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/site-
packages (from openai) (1.3.0)

Requirement already satisfied: tqdm>4 in /usr/local/lib/python3.10/site-packages (from openai) (4.66.1)
Requirement already satisfied: typing-extensions<5,>=4.11 in /voc/work/.local/lib/python3.10/site-packages (from openai) (4.12.2)
Requirement already satisfied: langchain-core<0.4.0,>=0.3.21 in /voc/work/.local/lib/python3.10/site-packages (from langchain-openai) (0.3.22)
Requirement already satisfied: tiktoken<1,>=0.7 in /voc/work/.local/lib/python3.10/site-packages (from langchain-openai) (0.8.0)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.10/site-packages (from anyio<5,>=3.5.0->openai) (3.6)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/site-packages (from anyio<5,>=3.5.0->openai) (1.2.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/site-packages (from httpx<1,>=0.23.0->openai) (2023.11.17)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.10/site-packages (from httpx<1,>=0.23.0->openai) (1.0.2)
Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.10/site-packages (from httpcore==1.*->httpx<1,>=0.23.0->openai) (0.14.0)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.10/site-packages (from langchain-core<0.4.0,>=0.3.21->langchain-openai) (6.0.1)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.10/site-packages (from langchain-core<0.4.0,>=0.3.21->langchain-openai) (1.33)
Requirement already satisfied: langsmith<0.2.0,>=0.1.125 in /voc/work/.local/lib/python3.10/site-packages (from langchain-core<0.4.0,>=0.3.21->langchain-openai) (0.1.137)
Requirement already satisfied: packaging<25,>=23.2 in /usr/local/lib/python3.10/site-packages (from langchain-core<0.4.0,>=0.3.21->langchain-openai) (23.2)
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in /usr/local/lib/python3.10/site-packages (from langchain-core<0.4.0,>=0.3.21->langchain-openai) (8.2.3)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.10/site-packages (from pydantic<3,>=1.9.0->openai) (0.6.0)
Requirement already satisfied: pydantic-core==2.27.1 in /voc/work/.local/lib/python3.10/site-packages (from pydantic<3,>=1.9.0->openai) (2.27.1)
Requirement already satisfied: regex>=2022.1.18 in /usr/local/lib/python3.10/site-packages (from tiktoken<1,>=0.7->langchain-openai) (2023.10.3)
Requirement already satisfied: requests>=2.26.0 in /usr/local/lib/python3.10/site-packages (from tiktoken<1,>=0.7->langchain-openai) (2.31.0)
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.10/site-packages (from jsonpatch<2.0,>=1.33->langchain-core<0.4.0,>=0.3.21->langchain-openai) (2.3)

Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/voc/work/.local/lib/python3.10/site-packages (from
langsmith<0.2.0,>=0.1.125->langchain-core<0.4.0,>=0.3.21->langchain-openai)
(3.10.10)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/voc/work/.local/lib/python3.10/site-packages (from
langsmith<0.2.0,>=0.1.125->langchain-core<0.4.0,>=0.3.21->langchain-openai)
(1.0.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/site-packages (from
requests>=2.26.0->tiktoken<1,>=0.7->langchain-openai) (3.3.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/site-packages (from
requests>=2.26.0->tiktoken<1,>=0.7->langchain-openai) (1.26.18)

```python
[10]: import os
      #from langchain.embeddings.openai import OpenAIEmbeddings


      os.environ["OPENAI_API_KEY"] = "voc-730917260112272230113166a5077cac4479.
       ↪92242159"
```

```python
[18]: #Create Vector Store for Text Embeddings
      # Initialize OpenAI embeddings
      embeddings = OpenAIEmbeddings()

      # Use Chroma for storing embeddings
      chroma_db = Chroma.from_documents(documents, embeddings) #vectorstore
```

```python
[23]: search_result = chroma_db.similarity_search_with_score("What is the candidate's
       ↪skill sets?", k=2)
      print(search_result)  # Prints the top 2 most similar documents to the query
```

[(Document(metadata={'page': 3, 'source':
'/voc/work/Project1/1728286846_the_nestle_hr_policy_pdf_2012.pdf'},
page_content='Only relevant skills and experience and \nadherence to the Nestlé
principles will \nbe considered in employing a person. No \nconsideration will
be given to a candidate's origin, \nnationality, religion, race, gender,
disability, sexual \norientation or age.\nWhilst adequate recruitment tools may
\nimprove the hiring process, the decision to hire a \ncandidate remains in the
hands of the responsible \nmanager, supported by the HR team. \n Joining
Nestlé'), 0.39040276408195496), (Document(metadata={'page': 3, 'source':
'/voc/work/Project1/1728286846_the_nestle_hr_policy_pdf_2012.pdf'},
page_content='Only relevant skills and experience and \nadherence to the Nestlé
principles will \nbe considered in employing a person. No \nconsideration will
be given to a candidate's origin, \nnationality, religion, race, gender,
disability, sexual \norientation or age.\nWhilst adequate recruitment tools may
\nimprove the hiring process, the decision to hire a \ncandidate remains in the
hands of the responsible \nmanager, supported by the HR team. \n Joining

```
Nestlé'), 0.39040276408195496)]
```

```python
[26]:  # Load GPT-3.5 Turbo model
       llm = ChatOpenAI(model="gpt-3.5-turbo", temperature=0)

       # Define a custom prompt template
       prompt_template = PromptTemplate(
           input_variables=["context", "question"],
           template="""
       You are an HR chatbot for Nestlé. Use the provided context to answer the␣
        ↪question accurately and succinctly.

       Context: {context}

       Question: {question}

       Answer:"""
       )

       # Update the QA chain to use the custom prompt
       qa_chain = load_qa_chain(llm, chain_type="stuff", prompt=prompt_template)
```

```python
[38]:  def chatbot(query):
           try:
               # Search for relevant chunks in the vector store
               docs = chroma_db.similarity_search(query, k=4)

               # Combine chunks into context
               context = " ".join([doc.page_content for doc in docs])

               # Perform similarity search and get the answer
               answer = qa_chain.run({"input_documents": docs, "question": query})
               return answer
           except Exception as e:
               return f"An error occurred: {str(e)}"
```

```python
[ ]:
```

```python
[39]:  # Build Gradio interface
       with gr.Blocks() as demo:
           gr.Markdown("### Nestlé HR Chatbot")
           query = gr.Textbox(label="Ask your HR question")
           output = gr.Textbox(label="Chatbot Response")
           submit_btn = gr.Button("Submit")
           submit_btn.click(fn=chatbot, inputs=query, outputs=output)

       # Launch the Gradio app
```

```
demo.launch(share=True)
```

```
* Running on local URL:  http://127.0.0.1:7864
* Running on public URL: https://e9f8432e015033a5ce.gradio.live
```

This share link expires in 72 hours. For free permanent hosting and GPU upgrades, run `gradio deploy` from the terminal in the working directory to deploy to Hugging Face Spaces (https://huggingface.co/spaces)

`<IPython.core.display.HTML object>`

[39]:

Questions

What is Nestle's sick leave policy?

What strategies does Nestlé implement to nurture and develop employee talent?

How does Nestle in the Market process streamline HR functions, and what are the three areas of specilization within the HR structure?

What criterea does Nestle consider when hiring new employees?

How many company holidays are there in 2012?

[40]:
```
demo.launch(share=True)
```

```
Rerunning server… use `close()` to stop if you need to change `launch()`
parameters.
----
* Running on public URL: https://e9f8432e015033a5ce.gradio.live
```

This share link expires in 72 hours. For free permanent hosting and GPU upgrades, run `gradio deploy` from the terminal in the working directory to deploy to Hugging Face Spaces (https://huggingface.co/spaces)

`<IPython.core.display.HTML object>`

[40]:

[ ]: