News Summarization and Text-to-Speech Application

Documentation

1. Project Setup

Prerequisites:

- Python 3.8+
- Pip installed
- Virtual environment (optional but recommended)

Installation Steps:

1.Clone the repository:

```
git clone https://huggingface.co/spaces/sivapurush/sankaripugal_App
cd sankaripugal_App
```

2.Create and activate virtual environment (optional):

```
python -m venv venv
source venv/bin/activate  # On Windows use: venv\Scripts\activate
```

3.Install dependencies:

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

4.Run the application:

```
python app.py
```

This will start the FastAPI backend and Gradio UI on `https://huggingface.co/spaces/sivapurush/sankaripugal_App`.

2. Project Structure

The repository includes the following key files:

- **app.py**: Main script that integrates all functionalities, including API endpoints (No separate `api.py`, all API logic is integrated into `app.py`).
- **utils.py**: Contains utility functions for processing news, sentiment, and TTS.
- **requirements.txt**: Lists all necessary dependencies.
- **README.md**: Provides setup and usage instructions.

3. Model Details

News Summarization:

- Extracts key information from news articles.
- Uses regex-based keyword extraction.
- Limits summary to 20 words for concise output.

Sentiment Analysis:

- Uses `nlptown/bert-base-multilingual-uncased-sentiment` model.
- Analyzes sentiment based on article title and summary.
- Labels sentiment as **Positive, Negative, or Neutral**.

Text-to-Speech (TTS):

- Uses **Google Text-to-Speech (gTTS)**.
- Converts final sentiment summary into Hindi.
- Saves and serves MP3 files for playback.

4. API Development

The application is built using **FastAPI** for backend processing and **Gradio** for the UI.

Endpoints:

1. **Get News Analysis**

- **Endpoint:** `GET /news`
- **Query Parameter:** `company` (str) - Company name for news analysis.
- **Response:** JSON containing sentiment analysis, coverage comparison, and audio link.

Example Request:

```
curl -X GET
"https://huggingface.co/spaces/sivapurush/sankaripugal_App/news?company=Tata"
```

2. **Get Audio File**

- **Endpoint:** `GET /get_audio`
- **Query Parameter:** `filename` (str) - Audio file name.
- **Response:** Serves the MP3 file.

Example Request:

```
curl -X GET
"https://huggingface.co/spaces/sivapurush/sankaripugal_App/get_audio?filename=abc123.mp3"
```

Swagger UI Implementation:

FastAPI provides interactive API documentation through Swagger UI.

- **Access Swagger UI:**
  - Open `https://huggingface.co/spaces/sivapurush/sankaripugal_App/docs` in your browser.
  - This provides an interactive way to test API endpoints.

- **Alternative Redoc UI:**
  - Available at `https://huggingface.co/spaces/sivapurush/sankaripugal_App/redoc`.
  - Provides a clean and structured API documentation.

Using Postman:

1. Open Postman and create a new `GET` request.
2. Enter API URL: `https://huggingface.co/spaces/sivapurush/sankaripugal_App/news?company=Tata`.
3. Click **Send** and check the response JSON.

5. API Usage - Third-party Integrations

Service	Purpose
1.Bing News RSS	**Fetches news articles based on company name.**
2.Hugging Face Transformers	**Provides sentiment analysis model.**
3.Google Translator API	**Translates English text into Hindi.**
4.Google Text-to-Speech (gTTS)	**Converts Hindi text into speech**

6. Assumptions & Limitations

Assumptions:

- News articles are fetched only from **Bing News RSS**.
- Sentiment is based on article **title and summary**, not full content.
- Hindi translation accuracy depends on Google Translator.

Limitations:

- **Article Limit:** Maximum **10 articles** are processed per request.
- **Sentiment Model:** May misclassify complex or mixed sentiments.
- **TTS Output:** Audio quality depends on gTTS capabilities.
- **Language Support:** Hindi is the only supported language for audio output.

7. Dependencies (requirements.txt)

The following libraries are used in this project:

numpy
pandas
torch
torchvision
torchaudio
tensorflow==2.12.0
scikit-learn
matplotlib
seaborn
gradio
streamlit
requests
beautifulsoup4
huggingface_hub
transformers
sentencepiece
protobuf
pydantic
pytz
python-dotenv

soundfile
pydub
gtts
pillow
pyarrow
tqdm
plotly
scipy
deep-translator
feedparser