Project: Israel-Hamas War Question Answering System

Objective

The goal of this project is to create a rudimentary Question Answering (QA) system capable of answering questions related to the Israel-Hamas war using a dataset of news articles. The system should be able to filter relevant articles, clean the data, and provide answers to specific questions about the conflict.

Dataset

The dataset is a JSON file containing 37,000 news articles from various sources, reporting on various topics including the Israel-Hamas war. The articles span from October 2023 to March 2024.

Approach

The solution can be divided into the following steps:

1. **Loading and Exploring Data**: Read the JSON file and load the data into a pandas Data Frame.
2. **Filtering Relevant Articles**: Filter articles that specifically mention "Israel" and "Hamas" in their content.
3. **Data Cleaning**: Clean the text data by removing unnecessary spaces, punctuation, and special characters.
4. **Setting Up a QA Model**: Utilize a pre-trained BERT model from Hugging Face's transformers library for the QA task.
5. **Implementing the QA System**: Create functions to process the input question and return relevant answers from the context provided by the filtered articles.

Two main functions are defined:

- **`answer_question`:** Uses the QA pipeline to get the top 3 answers for a given question and context.
- **`ask_question`:** Aggregates the cleaned article bodies into a single context and passes the question to `answer_question`.

```
question = "What happened at the Al-Shifa Hospital?"
answers = ask_question(question, israel_hamas_articles)
for i, answer in enumerate(answers, 1):
    print(f"Q: {question}\nA{i}: {answer}")
```

```
/usr/local/lib/python3.10/dist-packages/transformers/pipel
  warnings.warn("topk parameter is deprecated, use top_k i
Q: What happened at the Al-Shifa Hospital?
A1: bombed
Q: What happened at the Al-Shifa Hospital?
A2: Hospitals are bombed
Q: What happened at the Al-Shifa Hospital?
A3: Hospitals are bombed
```