

# Problem:

Create a rudimentary Question Answering system that can answer questions related to the Israel Hamas war.

## Steps Followed in the code

### Step 1: Loading and Preprocessing the JSON Data

Simple Explanation: Preprocess and clean news data to ensure it's accurate, relevant, and easy to analyze. This process removes errors, duplicates, and irrelevant information, making the data consistent and suitable for generating meaningful insights and summaries.

Note: Here very small subset of the original data is taken as the transformer model was taking very huge time to generate the summaries.

### Step 2: Taking Input and Generating Keywords

Simple Explanation:

To ask the user which event he/ she would like view like "Israel Hamas War" and generate the keywords so that we can retrieve the more relevant information related to the given prompt or user input.

Example: You tell me you want to know about the "Israel-Hamas War". I think of important words like "Israel", "Hamas", "war", "conflict", etc.

### Step 3: Filter relevant summarized events using extracted keywords

Simple Explanation:

Articles should be related to the text like "Israel-Hamas War" as other will not be of any use and might give poor results.

### Step 4: Prepare the context

Simple Explanation:

Make a big document to give all the information as a context to answer the more accurately question.

### Steps 5: Result

Simple Explanation:

Pass the context and query the answer from the “deepset/roberta-base-squad2” model and tokenizer.

Improvements:

Build a rag by using the langchain and openAI and one vector database.

My work:

<https://github.com/aneeshgadhwal/Rag>