**Step-I: Data Preprocessing:-**

* The JSON file was stored using a Pandas DataFrame
* Only the columns named ‘articles’ and ‘title’ was used for solving the assignment
* First, we cleaned the data of punctuation and other things such as special chareters etc
* After this step, the relevant data was filtered out of the corpus.
* Only a portion and not the whole dataset of the dataset was actually relevant with the Israel-Hamas conflict
* This was done in three steps:-
  + Some words (must words) if present, are included as relevant
  + Some other words (exclude words) if present in the sentences, were removed and thus not included
  + Finally, words with less impact: If their no exceeded a certain threshold, then they are also included

**Step-II: Fitting into the Model:-**

In this part of the code, a pre-trained language model is initialized for generating embeddings. The model chosen is "sentence-transformers/all-mpnet-base-v2", which specializes in converting text sentences into numerical representations (embeddings). Keyword arguments are provided for model initialization, specifying that the model should run on CPU. Additional parameters for encoding the input are also specified, indicating that the embeddings should be normalized. This step ensures that the language model is set up and ready to convert textual input into numerical representations suitable for further processing. Finally, the question-answering bot is made using this model.