**Steps performed in the challenge:**

The dataset is first preprocessed using NLP techniques. The developed code trains the model to predict the intent based on the message provided by the user. Using the intent, the sentences in the dataset which have the closest sentence embedding to the provided message, the therapist’s answer is selected and provided as an empathetic response.

**Model performance evaluation:**

The model can be evaluated on the BLEU metrics to see the closeness of the model’s predicted response to the actual therapist’s response.

**Approaches taken/Challenge faced while developing the challenge:**

1. Certain therapist’s response did not have the associated ‘questionText’, so I considered the ‘questionTitle’ for such cases.
2. I tried training a seq2seq model to predict an empathetic response, but due to the lack of variations in the dataset, the model was not able to output good empathetic reflections.
3. I also tried looking at a way to paraphrase the client’s message, so that an empathetic reflection could be built using the essential keywords from the client’s message but was the model did not perform well.
4. So, I built a model which using the client’s message predicts the intent, and then using the intent, finds the closest sentence in the dataset with the same intent that matches the client’s message and return the therapist’s response.

**Next steps that I would take to improve the model:**

1. I would first try to gather a larger dataset with more variations to a given client question.
2. The vocabulary of my model is currently fixed to the words that have occurred in the dataset, I would improve the model’s vocabulary by initializing it with Glove embeddings.
3. Currently I have used a basic neural network to get the intent, I would replace it with a RNN/Bi-LSTM model to use the context of the client’s message to predict the intent.
4. I would then as a different approach, train my seq2seq model on the dataset to get the empathetic reflections.