# REPORT

- Firstly we convert the data into a format that could be fine tuned. I converted the paragraphs and questions into two lists and the answer with the starting index in a list of dictionaries.
- Because BERT needs the ending index too, we find the ending index of the word and store it as answer_end along with answer_text and answer_start in the dict.
- Now the data can be passed to the tokenizer. Here we use distilbert tokenizer. We select the VERT-base pretrained model "bert-base-uncased".
- Using the data, we make the encodings from the tokenizer. We convert start-end pos to token start end pos.
- SquadDataset Class helps to train and convert encodings to dataset.
- We select BertforQuestionAnswering and use a pretrained model to initialize weights.
- The model is then trained.
- The results passes a score, start,end and the word.  A threshold score can be used to answer Problem 1 and start,end and the word can answer problem 2. A similar result can be found using a pretrained model mentioned in the code using `"deepset/roberta-base-squad2"`
- There were some errors in code which I'm trying to correct and will push my changes to the code with the accuracy results on the testing data.
- Transformers can be build from scratch and a deeper implementation of Bert can be done for this problem statement.