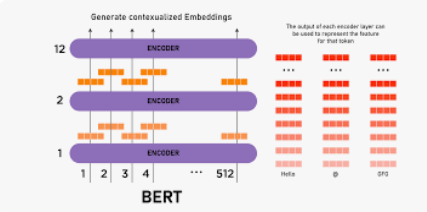
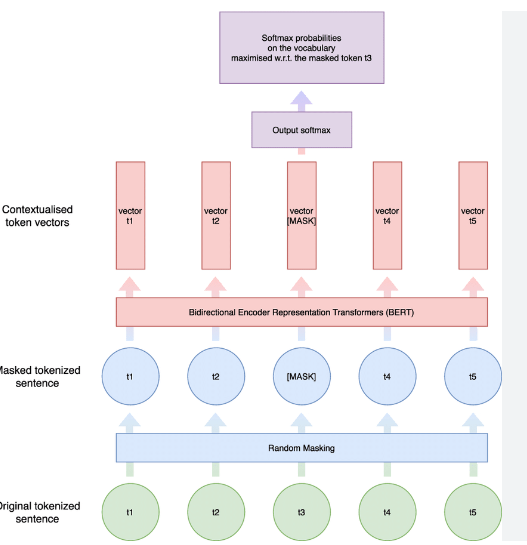
1. Explain the architecture of BERT  
   
2. Explain Masked Language Modeling (MLM)  
   
3. Explain Next Sentence Prediction (NSP)

In the BERT training process, the model receives pairs of sentences as input and learns to predict if the second sentence in the pair is the subsequent sentence in the original document.

1. What is Matthews evaluation?  
   MCC is a statistical tool used for model evaluation. Its job is to gauge or measure the difference between the predicted values and actual values and is equivalent to chi-square statistics for a 2 x 2 contingency table.
2. What is Matthews Correlation Coefficient (MCC)?  
   Matthew's correlation coefficient, also abbreviated as MCC was invented by Brian Matthews in 1975. MCC is a statistical tool used for model evaluation. Its job is to gauge or measure the difference between the predicted values and actual values and is equivalent to chi-square statistics for a 2 x 2 contingency table.
3. Explain Semantic Role Labeling  
   In natural language processing, semantic role labeling (also called shallow semantic parsing or slot-filling) is the process that assigns labels to words or phrases in a sentence that indicates their semantic role in the sentence, such as that of an agent, goal, or result. It serves to find the meaning of the sentence.
4. Why Fine-tuning a BERT model takes less time than pretraining  
   BERT takes between 1-25mins on a single Cloud TPU or between 1-130mins on a single GPU. During pre-training, the model is trained on unlabeled data over different pre-training tasks. For fine-tuning, the BERT model is first initialized with the pre-trained parameters, and all of the parameters are fine-tuned using labeled data from the downstream tasks.
5. Recognizing Textual Entailment (RTE)  
   Textual entailment recognition is the task of deciding, given two text fragments, whether the meaning of one text is entailed (can be inferred) from another text. This task captures generically a broad range of inferences that are relevant for multiple applications.
6. Explain the decoder stack of GPT models.

GPT generates one token at a time just like decoder of transformer and has causal language modeling so it is strictly decoder only model.

