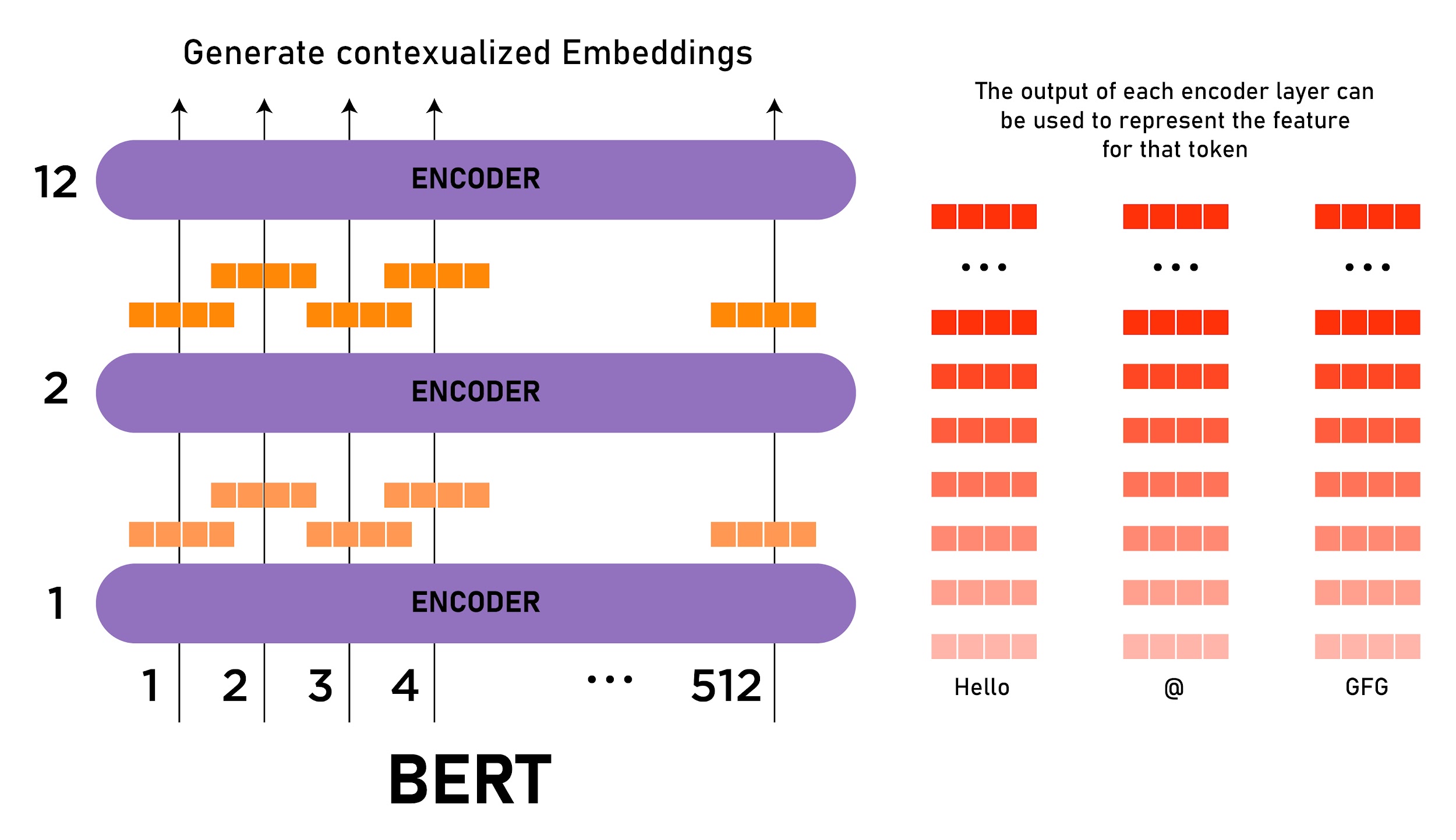
1.Explain the architecture of BERT.

Ans : BERT is basically an Encoder stack of transformer architecture. A transformer architecture is an encoder-decoder network that uses self-attention on the encoder side and attention on the decoder side. BERTbase has 12 layers in the Encoder stack while BERTlarge has 24 layers in the Encoder stack.



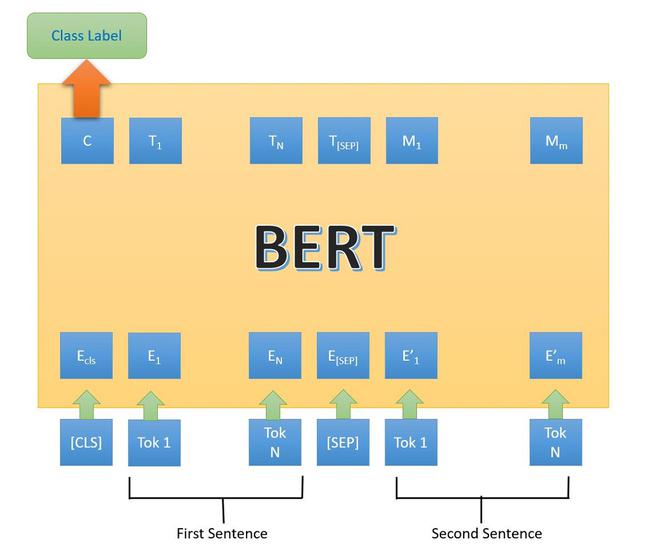
BERT, which stands for Bidirectional Encoder Representations from Transformers, is based on Transformers, a deep learning model in which every output element is connected to every input element, and the weightings between them are dynamically calculated based upon their connection.

2.Explain masked language modeling(MLM)

Ans : Masking is a way to tell sequence-processing layers that certain timesteps in an input are missing.Masked language modeling is one of such interesting applications of natural language processing. Masked image modeling is a way to perform word prediction that was originally hidden intentionally in a sentence.

3.Explain Next sentence prediction(NSP)

Ans : next sentence prediction is 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.



4.what is mattews evaluation ?

Ans : The Matthews correlation coefficient (MCC), instead, is a more reliable statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories true positives, false negatives, true negatives, and false positive.

5.what is mattews correlation coefficient(MCC) ?

Ans : The Matthews correlation coefficient (MCC), instead, is a more reliable statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories true positives, false negatives, true negatives, and false positive.Matthews defined VM, known as the Matthews coefficient, as the crystal volume per unit of protein molecular weight, and showed that VM bears a straightforward relationship to the fractional volume of solvent in the crystal.

6 Explain semantic role Labeling.

Ans : Semantic roles known as thematic roles or theta roles attempt to capture similarities and differences in verb meaning that are reflected in argument expression, with that will contribute to the mapping from semantics to syntax.

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.&

semantic roles have 8 types: agent, patient, theme, location, experiencer, instruments, goal, and source.

7.why fine tuning a BERT model takes less time than pretraining

Ans : BERT Takes time 24 min with BERT-base or 68 min with BERT-large on a single tesla V100 16GB.

8.Recognizing Textual Entailment(RTE)

Ans : Textual entailment recognition is the task of deciding, given two text fragments, whether the meaning of one text is entailed from another text .

9.Explain the decoder stack of GPT models.

Ans : GPT models it was made of decoders stacked on top of 12 decoders

GPT does not require the encoder part of the original transformer architecture as it is decoder only and there are no encoder attention blocks.GPT-3 Pre-trained transformer is a neural network machine learning model trained using internet data to generate any type of text.