1. **Explain the architecture of BERT**

**Answer:-**

BERT (Bidirectional Encoder Representations from Transformers) is a neural network architecture that was introduced by Google AI in 2018. BERT is a transformer-based model, which means that it uses attention mechanisms to learn long-range dependencies in the input text. BERT is trained on a massive dataset of text and code, and it can be used for a variety of natural language processing tasks, such as question answering, sentiment analysis, and text classification.

The architecture of BERT is composed of two main components: a masked language modeling (MLM) task and a next sentence prediction (NSP) task. The MLM task is used to train the BERT model to predict the masked words in a sentence. The NSP task is used to train the BERT model to predict whether two sentences are consecutive or not.

1. **Explain Masked Language Modeling (MLM)**

**Answer:-**

Masked language modeling (MLM) is a technique that is used to train language models. In MLM, a portion of the input text is masked, and the language model is then asked to predict the masked words. This helps the language model to learn the context of the words in the input text.

1. **Explain Next Sentence Prediction (NSP)**

**Answer:-**

Next sentence prediction (NSP) is a technique that is used to train language models. In NSP, two sentences are presented to the language model, and the language model is then asked to predict whether the two sentences are consecutive or not. This helps the language model to learn the relationships between sentences.

1. **What is Matthews evaluation?**

**Answer:-**

Matthews evaluation is a metric that is used to evaluate the performance of binary classification models. Matthews evaluation takes into account both the precision and the recall of the model, and it is often considered to be a more accurate measure of performance than accuracy.

1. **What is Matthews Correlation Coefficient (MCC)?**

**Answer:-**

Matthews correlation coefficient (MCC) is a metric that is used to evaluate the performance of binary classification models. MCC is a correlation coefficient that measures the agreement between the predicted labels and the ground truth labels. MCC is often considered to be a more accurate measure of performance than accuracy.

1. **Explain Semantic Role Labeling**

**Answer:-**

Semantic role labeling (SRL) is a task in natural language processing that involves identifying the semantic roles of the words in a sentence. The semantic roles of a word are the relationships that the word has with other words in the sentence. For example, the word "ate" in the sentence "John ate the apple" has the semantic role of "agent".

1. **Why Fine-tuning a BERT model takes less time than pretraining**

**Answer:-**

Fine-tuning a BERT model takes less time than pretraining because the BERT model has already been trained on a massive dataset of text and code. When you fine-tune a BERT model, you are only training the model on a smaller dataset that is specific to your task. This means that the fine-tuning process can be completed much more quickly.

1. **Recognizing Textual Entailment (RTE)**

**Answer:-**

Recognizing Textual Entailment (RTE) is a task in natural language processing that involves determining whether one sentence entails another sentence. Entailment is a relationship between two sentences where the meaning of one sentence implies the meaning of the other sentence. For example, the sentence "John ate the apple" entails the sentence "John had an apple".

1. **Explain the decoder stack of GPT models.**

**Answer:-**

The decoder stack of GPT models is a stack of self-attention layers that are used to generate text. The decoder stack is responsible for taking the input text and generating the output text. The self-attention layers in the decoder stack allow the model to attend to different parts of the input text when generating the output text.