**Transformers:**

It is a neural network architecture that was discovered in 2017 in the paper, Attention is all you need.

It is made up of two parts, an encoder and decoder.

It uses self-attention to weigh in the importance of different words

It processes sequences(documents) in parallel as opposed to previous structures like LSTMs or RNNs.

**LLMs:**

LLMs are large neural networks trained in large corpuses of data using the transformer architecture.

It pretrained using unsupervised learning.

It can replicate human language. It performs text generation, classification, NER and many more tasks.

Examples are: BERT, GPT-2

**BERT** (Bidirectional Encoder Representations from Transformers):

It is an encoder only model.

It reads sentences bidirectionally. This means that it considers the weight of the word both before and after the current word to generate context.

It is trained using MLM and NSP

**RoBERT**(A Robustly Optimized BERT):

A variant of BERT that is trained on more data. Performs better than BERT on most NLP tasks.