1. Explain the architecture of BERT
2. Explain Masked Language Modeling (MLM)
3. Explain Next Sentence Prediction (NSP)
4. What is Matthews evaluation?
5. What is Matthews Correlation Coefficient (MCC)?
6. Explain Semantic Role Labeling
7. Why Fine-tuning a BERT model takes less time than pretraining
8. Recognizing Textual Entailment (RTE)
9. Explain the decoder stack of GPT models.