Q1 What is the importance of adding special tokens like [CLS], [PAD], [SEP], and [MASK] during the preprocessing of text data for transformer models.
Q2 What is a Retrieval Augmented Generator (RAG) system, and how does it help in reducing the context size for generation tasks?
Q3 Explain the concept of "Zero Shot Prompting" and provide an example of a zero-shot prompt scenario.

Q4 What is the purpose of using Term Frequency – Inverse Document Frequency (TFIDF) in NLP?
- To count occurrences of words in a document
- To normalize over documents and make embeddings that facilitate document distinction
- To calculate the number of documents a term appears in
- To determine the total number of terms in a corpus

Q1 Explain how one-hot encoding works and provide an example using a simple sentence.
Q2 How do transformers solve the problem of parallelization in sequence-to-sequence models and why is this significant for NLP tasks?
Q3 What is the primary purpose of training language models with human feedback?

Q4 Discuss the techniques used for handling unknown words in statistical language models.
- To make the models align with their users' preferences
- To reduce the cost of manual data labeling
- To increase the complexity of large prompt datasets
- To automate the process of fine-tuning language models

Q1 How does prefix tuning differ from parameter-efficient fine-tuning methods like LoRa, in terms of their approach to updating parameters while maintaining model efficiency?
Q2 What is the intuition behind the smoothing techniques in statistical language modeling, and how do they help with the sparsity issue of n-gram models?
Q3 Why would a term with an inverse document frequency of 0 not be useful for distinguishing between documents in this context?

Q4 What is the role of the Byte Pair Encoding (BPE) token learner algorithm in text preprocessing for NLP?
- It separates punctuation from words.
- It normalizes case folding, making everything lowercase or uppercase as needed.
- It learns subword tokens that can represent frequent word subparts and are often morphemes
- It performs full stemming of the words in a corpus.

asdasdasdsadsad