Attention based Models and Transfer Learning Assignment

1 What is BERT and how does it work

BERT is a pre-trained transformer-based model that uses bidirectional context for understanding language.

2 What are the main advantages of using the attention mechanism in neural networks

Attention mechanisms help models focus on important parts of input sequences, improving performance in tasks like translation and summarization.

3 How does the self-attention mechanism differ from traditional attention mechanisms

Self-attention evaluates relationships between words within a sequence, while traditional attention typically uses an external context.

4 What is the role of the decoder in a Seq2Seq model

The decoder generates the output sequence from the encoded input sequence.

5 What is the difference between GPT-2 and BERT models

GPT-2 is autoregressive, generating text step-by-step, while BERT is bidirectional and focuses on understanding context.

6 Why is the Transformer model considered more efficient than RNNs and LSTMs

Transformers process input sequences in parallel, unlike RNNs/LSTMs which process sequentially, improving efficiency.

7 Explain how the attention mechanism works in a Transformer model

It computes a weighted sum of input elements, allowing the model to focus on relevant parts of the sequence.

8 What is the difference between an encoder and a decoder in a Seq2Seq model

The encoder processes the input sequence, and the decoder generates the output sequence from the encoded information.

9 What is the primary purpose of using the self-attention mechanism in transformers
To capture dependencies between all words in a sequence, allowing better contextual understanding.

10 How does the GPT-2 model generate text
GPT-2 generates text by predicting the next word in the sequence based on prior words.

11 What is the main difference between the encoder-decoder architecture and a simple neural network
Encoder-decoder handles sequence-to-sequence tasks, while a simple neural network processes fixed inputs/outputs.

12 Explain the concept of “fine-tuning” in BERT
Fine-tuning adapts a pre-trained BERT model to a specific task by training it on task-specific data.

13 How does the attention mechanism handle long-range dependencies in sequences
It directly connects every token to every other token, allowing it to capture long-range dependencies.

14 What is the core principle behind the Transformer architecture
The core principle is using self-attention to capture relationships within a sequence and process it in parallel.

15 What is the role of the "position encoding" in a Transformer model
Position encoding provides information about the order of tokens, which transformers don't inherently capture.

16 How do Transformers use multiple layers of attention
Multiple attention layers allow the model to capture increasingly complex relationships and patterns in the data.

17 What does it mean when a model is described as “autoregressive” like GPT-2
Autoregressive models generate output step-by-step, where each step uses the previous output as input.

18 How does BERT's bidirectional training improve its performance
BERT considers context from both directions, enhancing its understanding of word meaning in sentences.

19 What are the advantages of using the Transformer over RNN-based models in NLP

Transformers capture long-range dependencies better and process sequences in parallel, making them faster and more accurate.

20 What is the attention mechanism's impact on the performance of models like BERT and GPT-2

It allows these models to focus on key parts of input sequences, improving accuracy and context understanding.