Large Language Model (Winter 2024) Quiz - 1

Date of Examination: 06.02.2024 Duration: 30 mins Total Marks: 20 marks

Multiple Choice Questions : [8 Marks]

1.What is “few-shot learning” in the context of LLMs?

a. Learning with only a small amount of data
b. Learning with a small number of input features
c. Learning with a small learning rate
d. A technique for fine-tuning pre-trained models

2. What is the primary purpose of specifying a “temperature” parameter when generating text with a large language model like GPT-3?

a. Controlling the speed of text generation
b. Adjusting the model's verbosity and creativity
c. Fine-tuning the model on specific tasks
d. Managing the model's memory consumption

3. What is the significance of the “pre-training” and “fine-tuning” approach used in Large Language Models?

a. Pre-training refers to training a model from scratch, while fine-tuning adapts it to specific tasks
b. Pre-training involves optimizing the model for a specific task, while fine-tuning is a general training phase
c. Pre-training is the process of compressing the model, while fine-tuning is for model expansion
d. Pre-training and fine-tuning are synonymous and used interchangeably

4. What is the significance of the “attention mechanism” in transformer-based LLMs?

a. It controls the learning rate during training
b. It determines the number of layers in the network
c. It decides the batch size for training
d. It allows the model to focus on specific parts of the input sequence when making predictions

5. How does GPT-3 typically generate text output?

a. By memorizing and regurgitating predefined responses
b. By applying deterministic rules and heuristics
c. By sampling from a probability distribution over words, using a combination of learned patterns and context to generate text
d. By performing keyword-based searches on the internet

6. What is “knowledge distillation” in the context of Large Language Models ?

a. The process of extracting knowledge from a human teacher and transferring it to an LLM
b. A technique for increasing the model's capacity by adding more layers
c. The process of converting LLM-generated text into a distilled, shorter form
d. A method for training smaller models to mimic the behavior of a larger pre-trained LLM by transferring its knowledge

7. In the context of LLMs, what is “zero-shot learning”?

a. Training the model with zero data samples
b. Learning without any teacher or supervision
c. A learning technique that starts from zero knowledge
d. The ability to perform tasks for which the model has not been explicitly trained, with zero examples provided

8 . In transformer-based models, what is the “position encoding” used for?

a. To encode the geographical positions of data samples
b. To indicate the order of words in a sequence, as the transformer architecture does not have built-in positional information
c. To encode the position of the model's parameters
d. To control the learning rate

Fill In the Blanks Questions : [4 Marks]

1. _____ of the Transformer model acts as a gating mechanism, similar to that found in GRUs and LSTMs.

2. BERT is an auto-encoding transformer model, while GPT is an _____ transformer model

3. _____ mechanism is often used in sequence-to-sequence tasks to align the elements of the source and target sequences during translation.

4. _____ is caused by the mismatch between the LLM’s internal knowledge and the labeler’s internal knowledge.

Descriptive Questions and Answers :

1. How next sentence prediction (NSP) is used in language modeling? **[2 Marks]**

2. Explain Difference in the training strategy between BERT and GPT ? **[2 Marks]**

3. How can bias be mitigated with human-in-the-loop approaches when developing LLMs? **[4 Marks]**