| | |
|---|---|
| **Started on** | Wednesday, 5 February 2025, 12:16 PM |
| **State** | Finished |
| **Completed on** | Wednesday, 5 February 2025, 12:18 PM |
| **Time taken** | 1 min 48 secs |
| **Marks** | 10.00/10.00 |
| **Grade** | **100.00** out of 100.00 |

**Question 1**

Complete

Mark 1.00 out of 1.00

What is the primary reason Transformers outperform RNNs in NLP tasks?

- a. They rely on recurrence
- b. They require fewer parameters
- c. They handle long-range dependencies efficiently
- d. They use convolutions

**Question 2**

Complete

Mark 1.00 out of 1.00

What is a common application of the Transformer model?

- a. Object detection
- b. Image segmentation
- c. Speech synthesis
- d. Machine translation

**Question 3**

Complete

Mark 1.00 out of 1.00

What is the purpose of the softmax function in self-attention?

- a. To reduce computational complexity
- b. To activate neurons
- c. To normalize attention scores
- d. To update model weights

**Question 4**

Complete

Mark 1.00 out of 1.00

Who introduced the Transformer model in the paper "Attention Is All You Need"?

- a. Vaswani et al.
- b. Geoffrey Hinton
- c. Andrew Ng
- d. Yann LeCun

**Question 5**

Complete

Mark 1.00 out of 1.00

What is a major advantage of pre-trained Transformer models?

- a. They do not need large datasets
- b. They require no fine-tuning
- c. They are computationally inexpensive
- d. They generalize well to new tasks

**Question 6**

Complete

Mark 1.00 out of 1.00

Which model is based on the Transformer architecture?

- a. LSTM
- b. CNN
- c. ResNet
- d. BERT

**Question 7**

Complete

Mark 1.00 out of 1.00

Why do Transformers use positional encodings?

- a. To reduce overfitting
- b. To inject the order of words into the model
- c. To increase model depth
- d. To improve the efficiency of training

**Question 8**

Complete

Mark 1.00 out of 1.00

How does the attention mechanism compute relevance scores?

- a. Using recurrent units
- b. Using dropout
- c. Using max pooling
- d. Using dot-product similarity

**Question 9**

Complete

Mark 1.00 out of 1.00

How does Transformer differ from CNNs in feature extraction?

- a. Transformers use pooling layers
- b. Transformers use stride-based filters
- c. Transformers use weight sharing
- d. Transformers use self-attention instead of convolutions

**Question 10**

Complete

Mark 1.00 out of 1.00

Which of the following is NOT a part of the Transformer architecture?

- a. Decoder
- b. Recurrent unit
- c. Encoder
- d. Self-attention