

<b>Started on</b>	Tuesday, 4 February 2025, 11:51 AM
<b>State</b>	Finished
<b>Completed on</b>	Tuesday, 4 February 2025, 11:56 AM
<b>Time taken</b>	4 mins 17 secs
<b>Marks</b>	9.00/10.00
<b>Grade</b>	<b>90.00</b> out of 100.00

### Question 1

Complete

Mark 0.00 out of 1.00

What is a common application of the Transformer model?

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

### Question 2

Complete

Mark 1.00 out of 1.00

How does Transformer differ from CNNs in feature extraction?

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

### Question 3

Complete

Mark 1.00 out of 1.00

Which model is based on the Transformer architecture?

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

**Question 4**

Complete

Mark 1.00 out of 1.00

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

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

**Question 5**

Complete

Mark 1.00 out of 1.00

Why do Transformers use positional encodings?

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

**Question 6**

Complete

Mark 1.00 out of 1.00

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

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

**Question 7**

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 dot-product similarity
- ☐ d. Using max pooling

**Question 8**

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 9**

Complete

Mark 1.00 out of 1.00

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

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

**Question 10**

Complete

Mark 1.00 out of 1.00

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

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