

MCQ Assessment - Answers Included

Question 1: Which type of neural networks have been established as state-of-the-art for sequence modeling?

- a: Recurrent neural networks
- b: Convolutional neural networks
- c: Feedforward neural networks
- d: Bayesian neural networks

Correct Answer: a

Question 2: Who proposed replacing RNNs with self-attention in the Transformer model?

- a: Jakob
- b: Ashish
- c: Noam
- d: Llion

Correct Answer: a

Question 3: Which of the following was not a feature proposed by Noam in the Transformer model?

- a: Scaled dot-product attention
- b: Multi-head attention
- c: Positional encoding
- d: Recurrent computation

Correct Answer: d

Question 4: Who designed and implemented the first Transformer models?

- a: Llion
- b: Ashish
- c: Niki
- d: Jakob

Correct Answer: b

Question 5: What is one of the inherent limitations of recurrent models?

- a: Limited parallelization
- b: High computational cost
- c: Poor generalization ability
- d: Inability to model long-range dependencies

Correct Answer: a

Question 6: How do attention mechanisms allow for more efficient modeling of dependencies in sequence modeling?

- a: By allowing dependencies to be modeled regardless of distance
- b: By reducing the number of parameters required
- c: By parallelizing the computation of hidden states
- d: By introducing a hierarchical structure to the model

Correct Answer: a

Question 7: What is a key difference between the Transformer model and previous attention-based models?

- a: The Transformer relies solely on attention mechanisms

- b: The Transformer uses a different attention mechanism
- c: The Transformer combines attention with recurrence
- d: The Transformer does not use attention mechanisms

Correct Answer: a

Question 8: Which of the following was NOT a significant contributor to the development of the Transformer model?

- a: Jakob
- b: Ashish
- c: Noam
- d: Lukasz

Correct Answer: d

Question 9: What type of GPUs were used to train the Transformer model?

- a: P100
- b: V100
- c: RTX 3090
- d: Titan Xp

Correct Answer: a

Question 10: What was the approximate training time for the Transformer model?

- a: 6 hours
- b: 12 hours
- c: 24 hours
- d: 48 hours

Correct Answer: b

Question 11: What was one of the benefits of the factorization tricks and conditional computation used in recent recurrent models?

- a: Improved computational efficiency
- b: Enhanced model performance
- c: Reduced memory consumption
- d: Faster convergence

Correct Answer: a

Question 12: Who was responsible for our initial codebase, efficient inference, and visualizations in the Transformer model?

- a: Llion
- b: Ashish
- c: Niki
- d: Jakob

Correct Answer: a

Question 13: What did Lukasz and Aidan contribute to the Transformer model?

- a: Designing and implementing tensor2tensor
- b: Proposing the attention mechanism
- c: Tuning and evaluating model variants
- d: Visualizing the attention patterns

Correct Answer: a

Question 14: Who is not listed as a contributor to the Transformer model?

- a: Jakob
- b: Ashish
- c: Niki
- d: David

Correct Answer: d

Question 15: What is one of the limitations of recurrent models that the Transformer model addresses?

- a: Sequential computation
- b: High parameter count
- c: Limited memory capacity
- d: Inability to learn long-term dependencies

Correct Answer: a

Question 16: What was one of the major advantages of the Transformer model?

- a: Significantly more parallelization
- b: Elimination of recurrent connections
- c: Improved generalization ability
- d: All of the above

Correct Answer: d