Name: **Matriculation No.:**

1. In the Transformer architecture, what is the correct formula for the scaled-dot-product attention calculation given $\mathbf{Q} \in \mathbb{R}^{l_q \times d_q}, \quad \mathbf{K} \in \mathbb{R}^{l_k \times d_k} \quad \text{and} \quad \mathbf{V} \in \mathbb{R}^{l_v \times d_v}$? (2pt)

   $Att(Q, K, V) =$

2. In the transformer architecture, the embedding dimension of the value vectors $\mathbf{V}$ must be equal to which of the following? (1pt)

   ☐ The embedding dimension of the keys $\mathbf{K}$.

   ☐ The embedding dimension of the queries $\mathbf{Q}$.

   ☐ Both the keys and the queries.

   ☐ None of the above.

3. The output sequence length of the attention mechanism is equal to the sequence length of which component? (1pt)

   ☐ Keys.

   ☐ Queries.

   ☐ Values.

   ☐ More information needed.

4. **Statement**: In the vanilla transformer model, each encoder layer output is passed to the same level decoder layer? (1pt)

   ☐ True ☐ False

5. Given an input sequence length of 4 and an embedding dimension of 5, generate the look-ahead mask? (2pts)

6. What is the reason for usage of the positional encoding in the Transformer? (1pts)

7. What is the purpose of the residual connections around each sub-layer (e.g., attention, feedforward) in the Transformer? (1pts)

8. Cross the **TRUE** statement (only one is true)? (1pts)

☐ The number of encoder layers must equal the number of decoder layers in a transformer model.

☐ The time complexity for computing the queries, keys, and values in the Transformer model is $O(log(n) \cdot d^3)$. (Assume: sequence length n and embedding dimension d)

☐ In a translation setting, it is necessary to use the same tokenizer for both the source and target languages.

☐ The attention mechanism does not enhance the performance of Recurrent Neural Networks (RNNs) when dealing with long sequences.

☐ In the attention mechanism, the scaling of the dot product of queries and keys by $\sqrt{d_k}$ is done to prevent extremely small gradients for most predictions.