Q1 Instructions
1 Point

Open notes: The quiz is open notes. You are free to use any content from the course website or from your own personal notes.

No communication: ANY communication with other students about the quiz content is strictly forbidden and will result in a failing grade for the whole class (not just this quiz).

No partial credit: Every question is all or nothing credit. Thus, you must get the answer exactly right to get credit for the question (including SELECT ALL questions). No partial credit will be given on quizzes.

Number Format: When giving numbers as short answers, please give in standard decimal notation with preceding "0." for decimals if needed but no trailing 0s (e.g., "0.15", "2.9", "0.001", "100" but NOT "0.15000" NOR ".15" NOR ".001" NOR "6.0").

Honor Pledge: I assert that I have not received any information about this quiz and will not share any quiz content with anyone else. I understand that any violation of this will result in a failing grade for the whole class (not just this quiz).

Yes, I understand the policies above and assert the honor pledge.

No

Q2
1 Point

The self-attention mechanism in the Transformer model allows a token in the input sequence to attend to all other tokens in the same sequence, including itself.

True

False

Q3
1 Point

Self-attention is a sequential process that attends to one token at a time in the input sequence.

True

False

Q4
1 Point

Transformers are a type of recurrent neural network (RNN) architecture that is particularly suited for sequential data processing tasks.

True

False

Q5
1 Point

One nice property of attention is to visualize the attention map. Assuming that we have an attention model trained for English-French translation, and the input is "He has an adorable dog" and the output is "Il a un chien adorable", what is the size of the cross-attention map? Note that an <EOS> token is used.

6 by 6

5 by 5

7 by 7

Q6
1 Point

When we compute self-attention, Query, Key and Value are obtained by forwarding the input representation into the corresponding weight matrix,

i.e., $Q = XW_Q, K = XW_K, V = XW_V$. Which pair must have the same shape?

Q, K

Q, V

K, V

Q7
1 Point

Assuming the input of self attention is $[1, 3, 2, 7]^T$ and the output is $[6, 8, 0, 4]^T$. If we permute the input to be $[1, 2, 7, 3]^T$, what is the output?

$[0, 4, 6, 8]^T$

$[6, 0, 4, 8]^T$

$[4, 0, 6, 8]^T$

$[8, 4, 6, 0]^T$

Unable to be determined by information given.

Q8
1 Point

Suppose we have an attention **weights** (after softmax) $M = \sigma(QK^T) \in \mathbb{R}^{10\times10}$. What is the sum over all the elements in the matrix, i.e., $\sum_{i,j} M_{i,j}$?

100

Q9
1 Point

We have learned that positional encoding is important in Transformers. Suppose we have the input embedding of size (L, D_{in}), the output embedding of size $(L, 1)$, the hidden embedding of size (L, D_h). What is the size of the positional encoding?

Scalar

$(L, 1)$

(L, D_{in})

(L, D_h)

Q10
1 Point

For multi-headed attention, suppose the output dimension of every head is $D_{head} = 8$ and there are 4 heads and the final linear layer matrix of multiheaded attention is $W_H \in \mathbb{R}^{C \times D_{out}}$, what is C?

32

Q11
1 Point

Which of the following best describes the role of attention in Transformers?

- ✓ A technique used to increase the computational efficiency of Transformers by reducing the number of computations required.
- ✓ A mechanism that allows Transformers to dynamically weight the importance of different tokens in a sequence for processing.
- ✓ A model architecture that enables the decoder to use all encoder outputs efficiently

A mechanism that controls the rate at which the Transformer model updates its parameters during optimization.

Answer

A mechanism that allows Transformers to dynamically weight the importance of different tokens in a sequence for processing., A model architecture that enables the decoder to use all encoder outputs efficiently

Quiz 11

● Graded

Student
Paloma Arellano

Total Points
8 / 11 pts

Question 1 Instructions	1 / 1 pt
Question 2 (no title)	1 / 1 pt
Question 3 (no title)	1 / 1 pt
Question 4 (no title)	1 / 1 pt
Question 5 (no title)	1 / 1 pt
Question 6 (no title)	1 / 1 pt
Question 7 (no title)	1 / 1 pt
Question 8 (no title)	0 / 1 pt

Question 9
(no title) 0 / 1 pt

Question 10
(no title) 1 / 1 pt

Question 11
(no title) 0 / 1 pt

+ 1 pt Correct

✔ **+ 0 pts** Incorrect