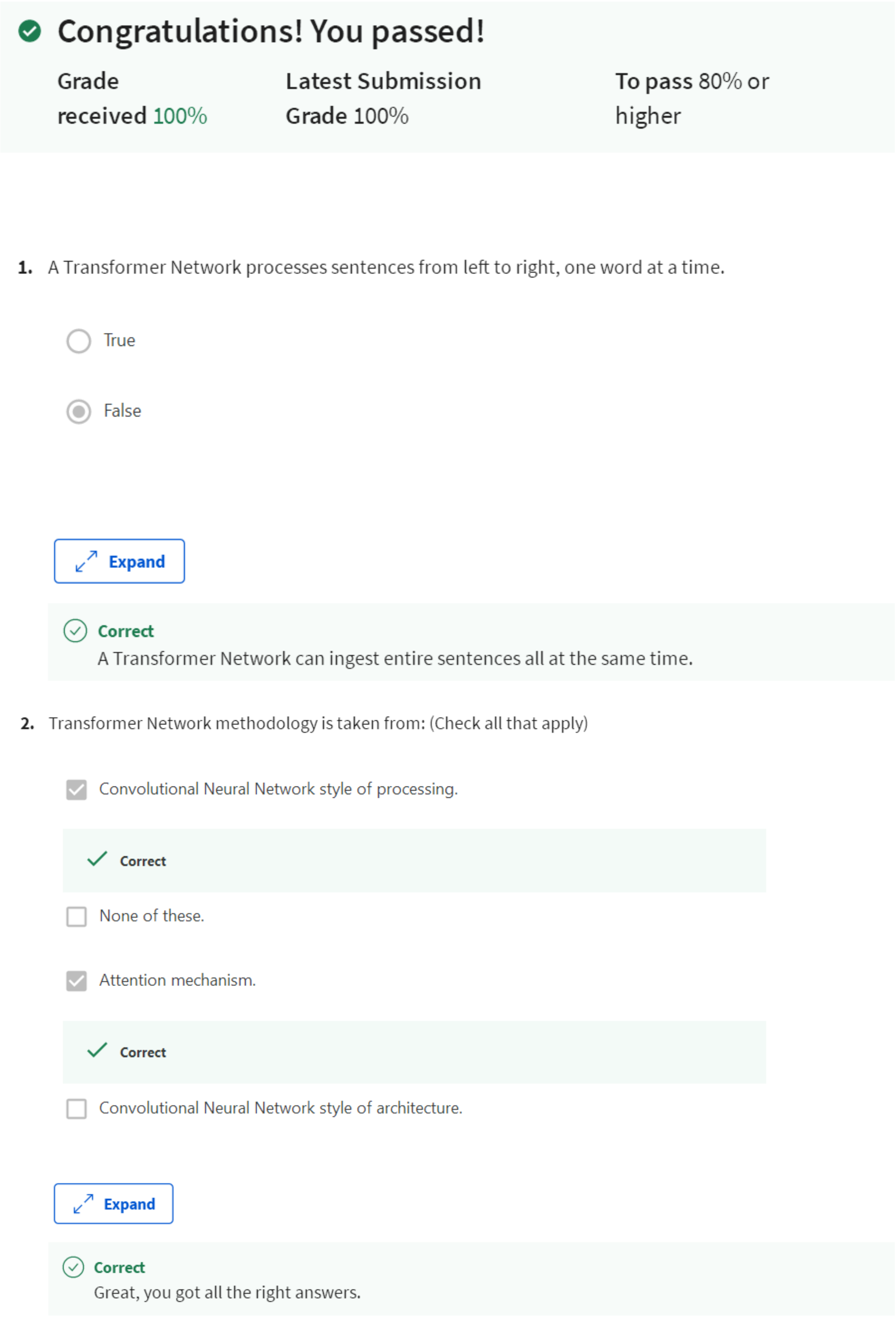
DLS/ C5/ W4 / quiz

Transformers



Congratulations! You passed!

Grade received 100%

Latest Submission Grade 100%

To pass 80% or higher

1. A Transformer Network processes sentences from left to right, one word at a time.

○ True

● False

Correct
A Transformer Network can ingest entire sentences all at the same time.

2. Transformer Network methodology is taken from: (Check all that apply)

☑ Convolutional Neural Network style of processing.

✓ **Correct**

□ None of these.

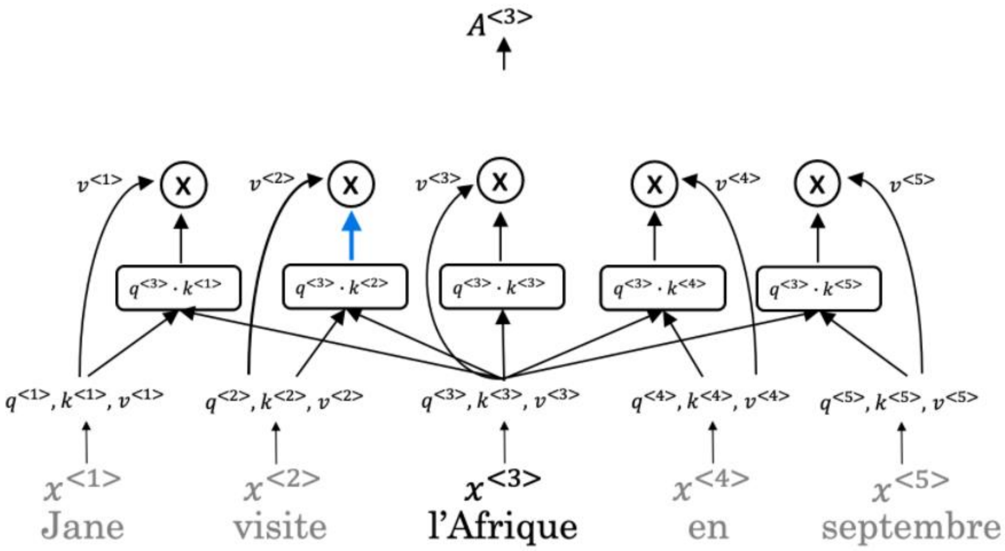
☑ Attention mechanism.

✓ **Correct**

□ Convolutional Neural Network style of architecture.

Correct
Great, you got all the right answers.

3. The concept of Self-Attention is that:



● Given a word, its neighbouring words are used to compute its context by summing up the word values to map the Attention related to that given word.

○ Given a word, its neighbouring words are used to compute its context by selecting the lowest of those word values to map the Attention related to that given word.

○ Given a word, its neighbouring words are used to compute its context by taking the average of those word values to map the Attention related to that given word.

○ Given a word, its neighbouring words are used to compute its context by selecting the highest of those word values to map the Attention related to that given word.

Expand

✓ **Correct**

4. Which of the following correctly represents Attention?

○ $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q * k^{j})} {{\sum}_j\exp(q * k^{j})})* {\sum}_i{v}^{i}$$

○ $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q * v^{j})} {{\sum}_j\exp(q * v^{j})})* K^{j}$$

● $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q * k^{j})} {{\sum}_j\exp(q * k^{j})})* V^{j}$$

○ $${A(Q,K,V)} = (\frac{\exp(q * k^{j})} {\exp(q * k^{j})})* V^{j}$$

Expand

✓ **Correct**

This is the correct Attention formula.

5. Which of the following statements represents Key (K) as used in the self-attention calculation?

○ K = interesting questions about the words in a sentence

● K = qualities of words given a Q

○ K = the order of the words in a sentence

○ K = specific representations of words given a Q

Expand

✓ **Correct**

The qualities of words given a Q are represented by Key (K).

6. $\boldsymbol{Attention(W_i^Q Q, W_i^K K, W_i^V V)}$

i here represents the computed attention weight matrix associated with the ith “word” in a sentence.

○ True

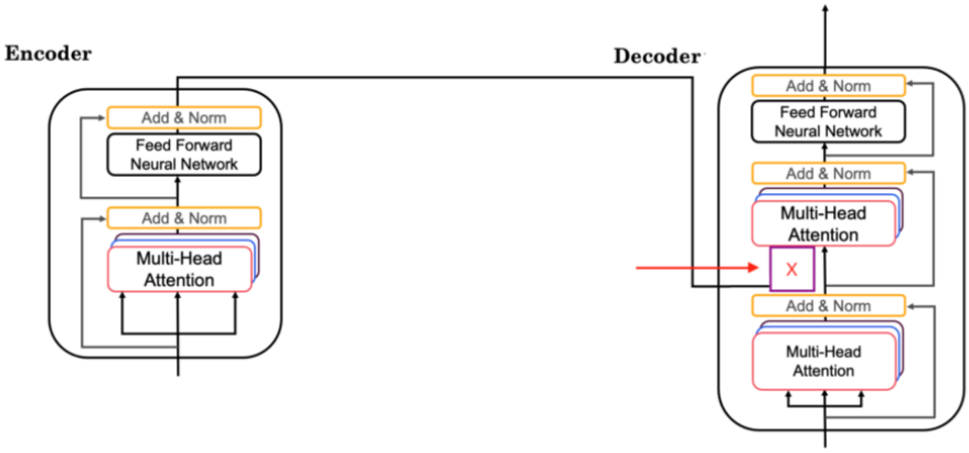
● False

Expand

✓ **Correct**

Correct! i here represents the computed attention weight matrix associated with the ith “head” (sequence).

7. Following is the architecture within a Transformer Network **(without displaying positional encoding and output layers(s)).**



What information does the Decodertake from the Encoder for its second block of Multi-Head Attention ? (Marked X, pointed by the independent arrow)

(Check all that apply)

□ Q

☑ V

✓ Correct

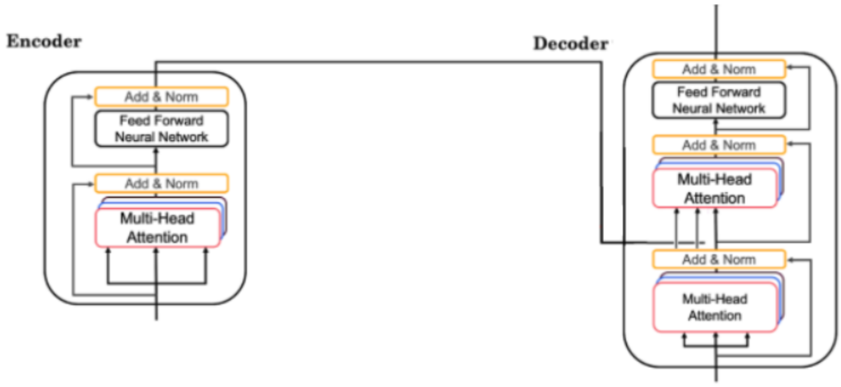
☑ K

✓ Correct

Expand

Correct
Great, you got all the right answers.

8. Following is the architecture within a Transformer Network **(without displaying positional encoding and output layers(s)).**



What does the output of the encoder block contain?

● Contextual semantic embedding and positional encoding information

○ Softmax layer followed by a linear layer.

○ Linear layer followed by a softmax layer.

○ Prediction of the next word.

Expand

✓ **Correct**
The output of the encoder block contains contextual semantic embedding and positional encoding information.

9. Why is positional encoding important in the translation process? (Check all that apply)

☑ Position and word order are essential in sentence construction of any language.

✓ **Correct**

□ It helps to locate every word within a sentence.

□ It is used in CNN and works well there.

☑ Providing extra information to our model.

✓ **Correct**

Expand

✓ **Correct**
Great, you got all the right answers.

10. Which of these is a good criterion for a good positional encoding algorithm?

☑ It should output a unique encoding for each time-step (word's position in a sentence).

✓ **Correct**

☑ Distance between any two time-steps should be consistent for all sentence lengths.

✓ **Correct**

☑ The algorithm should be able to generalize to longer sentences.

✓ **Correct**

□ None of these.

Expand

✓ **Correct**
Great, you got all the right answers.