Congratulations! You passed!

Grade received 100% Latest Submission Grade 100% To pass 80% or higher

Go to next item

1. A Transformer Network, like its predecessors RNNs, GRUs and LSTMs, can process information one word at a time. (Sequential architecture). **1 / 1 point**

- (•) False
- ( ) True

Expand

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

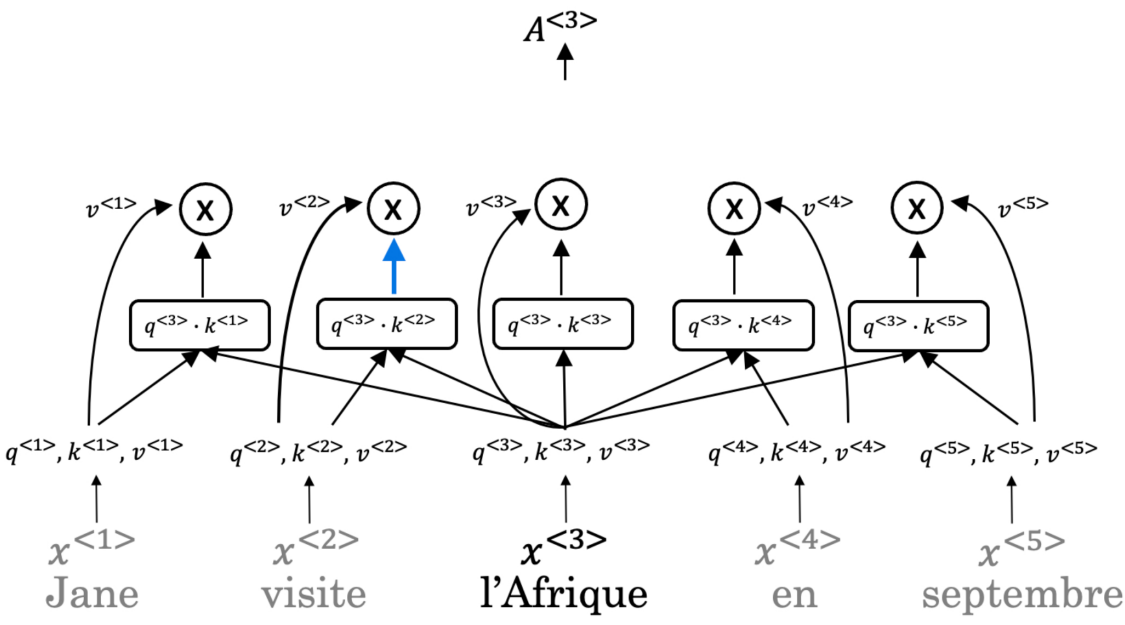
2. The major innovation of the transformer architecture is combining the use of LSTMs and RNN sequential processing. **1 / 1 point**

- (•) False
- ( ) True

Expand

Correct
The major innovation of the transformer architecture is combining the use of attention based representations and a CNN convolutional neural network style of processing.

3. What are the key inputs to computing the attention value for each word? **1 / 1 point**



- ( ) The key inputs to computing the attention value for each word are called the query, knowledge, and vector.

- ○ The key inputs to computing the attention value for each word are called the quotation, knowledge, and value.
- ◉ The key inputs to computing the attention value for each word are called the query, key, and value.
- ○ The key inputs to computing the attention value for each word are called the quotation, key, and vector.

Expand

✓ Correct
The key inputs to computing the attention value for each word are called the query, key, and value.

4. Which of the following correctly represents *Attention ?* **1 / 1 point**

- ◉ $Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$
- ○ $Attention(Q, K, V) = min(\frac{QK^T}{\sqrt{d_k}})V$

$Attention(Q, K, V) = min(\frac{QK^T}{\_\_})V$

Expand

✓ Correct

5. Which of the following statements represents Key (K) as used in the self-attention calculation? **1 / 1 point**

- ○ K = interesting questions about the words in a sentence
- ○ K = the order of the words in a sentence
- ◉ K = qualities of words given a Q
- ○ K = specific representations of words given a Q

Expand

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

6. $\boldsymbol{Attention(W_i^Q Q, W_i^K K, W_i^V V)}$ **1 / 1 point**

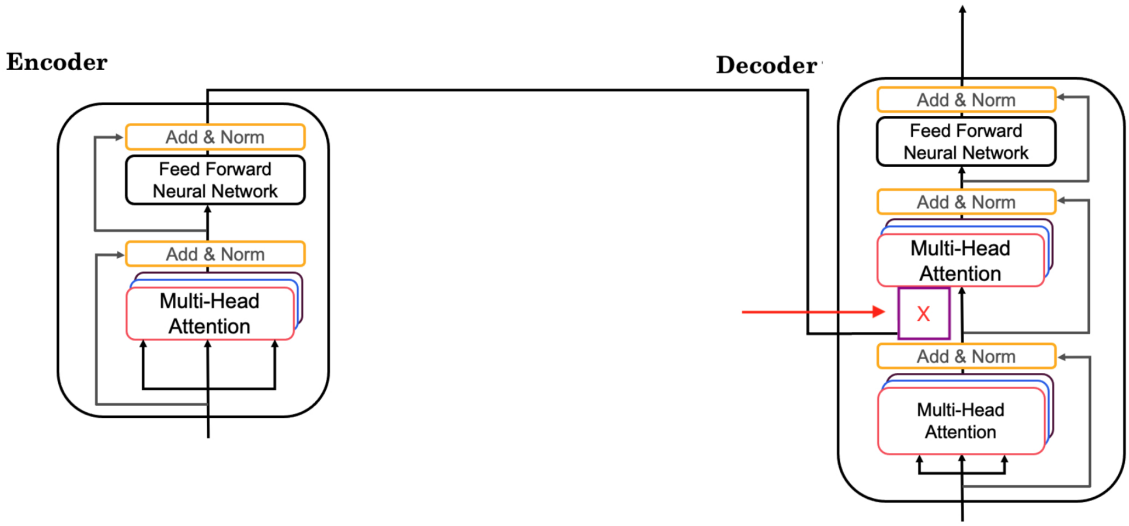
i here represents the computed attention weight matrix associated with the ith "word" in a sentence.

- ○ 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)).*** **1 / 1 point**



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

(Check all that apply)

- [ ] Q
- [x] K

✓ Correct

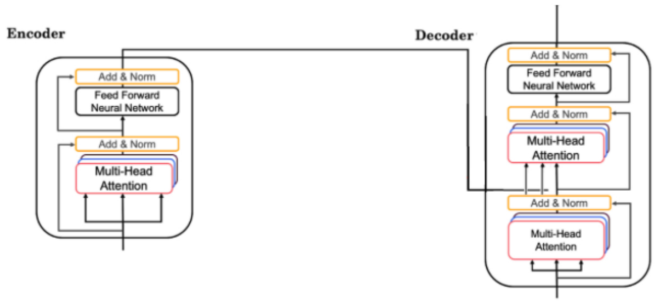
- [x] V

✓ 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)).*** 1 / 1 point



The output of the decoder block contains a softmax layer followed by a linear layer to predict the next word one word at a time.

- (•) False
- ( ) True

Expand

✓ Correct
The output of the decoder block contains a linear layer followed by a softmax layer to predict the next word one word at a time.

9. Why is positional encoding important in the translation process? (Check all that apply) **1 / 1 point**

- [x] 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.
- [x] 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? **1 / 1 point**

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

  ✓ **Correct**

- [x] Distance between any two time-steps should be consistent for all sentence lengths.

  ✓ **Correct**

- [x] The algorithm should be able to generalize to longer sentences.

  ✓ **Correct**

- [ ] None of these.

Expand

Correct
Great, you got all the right answers.