

✔

Congratulations! You passed!

Grade
received 80%

Latest Submission
Grade 80%

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. Transformer Network methodology is taken from:

1 / 1 point

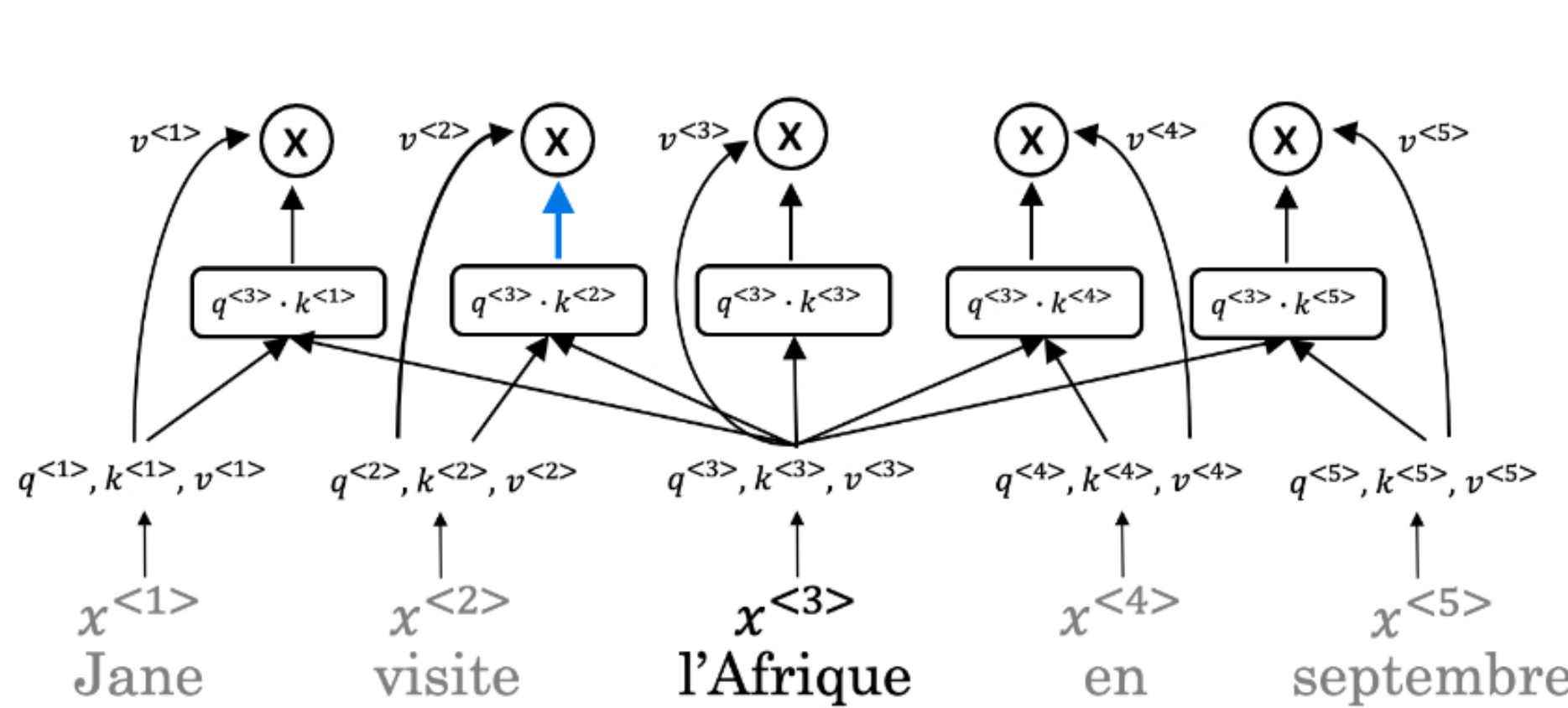
- ☐ Attention Mechanism and RNN style of processing.
- ☐ RNN and LSTMs
- ☐ GRUs and LSTMs
- ☒ Attention Mechanism and CNN style of processing.

Expand

✔ Correct
Transformer architecture combines the use of attention based representations and a CNN convolutional neural network style of processing.

3. The concept of *Self-Attention* is that:

1 / 1 point



- ☐ 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.
- ☐ 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 summing up the word values to map the Attention related to that given word.

Expand

✔ Correct

4. What letter does the "2" represent in the following representation of *Attention*?

1 / 1 point

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

- ☐ v
- ☐ q
- ☒ k
- ☐ t

Expand

✔ Correct
k is represented by the \$K\$ in the representation.

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

0 / 1 point

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

Expand

✘ Incorrect
To revise the concept watch the lecture *Self-Attention*; Q = interesting questions about the words in a sentence

6. $Attention(W_i^Q Q, W_i^K K, W_i^V V)$

1 / 1 point

i here represents the computed attention weight matrix associated with the *i*th "head" (sequence).

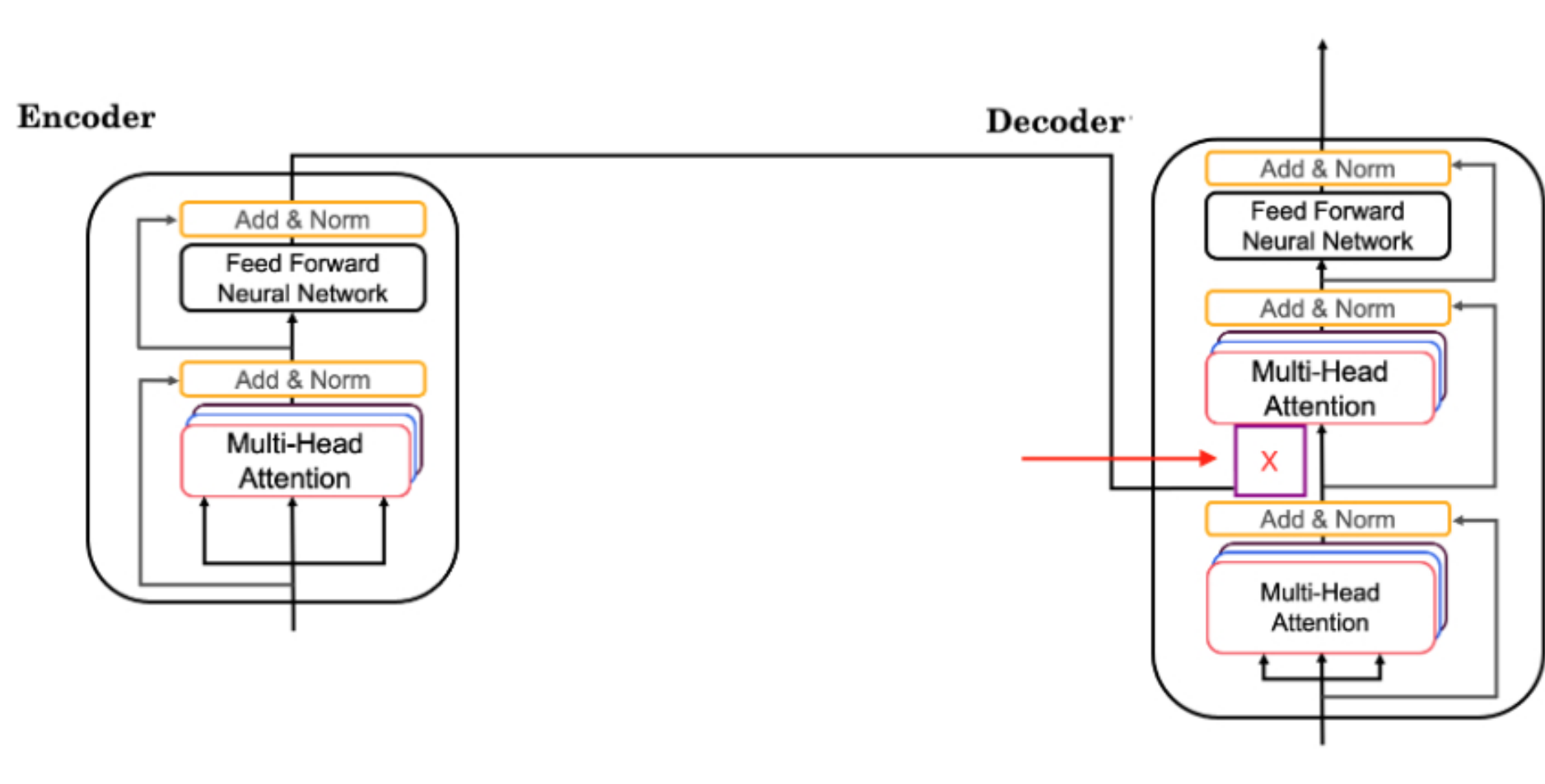
- ☐ False
- ☒ True

Expand

✔ Correct
\$i\$ here represents the computed attention weight matrix associated with the *i*th "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
- ☒ V

✔ Correct

✔ K

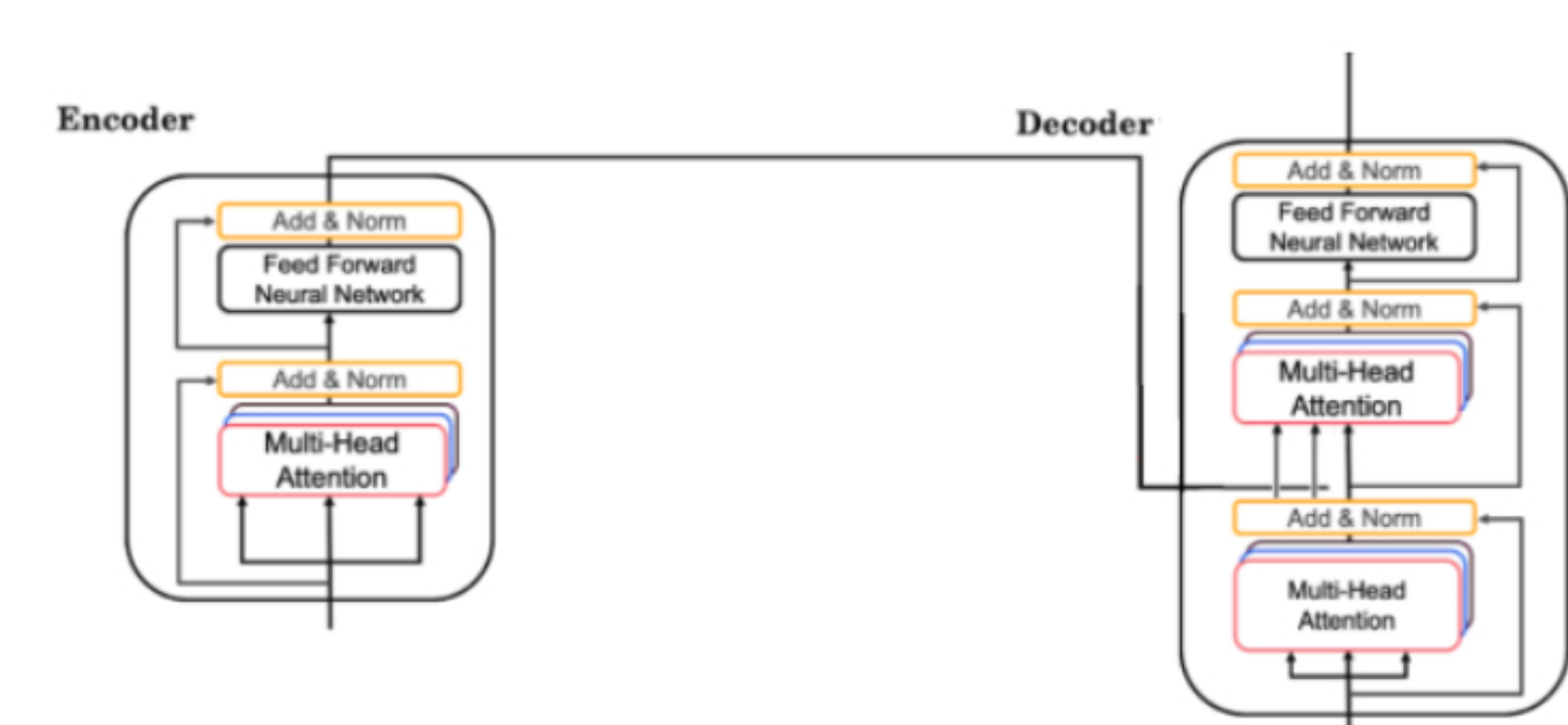
✔ Correct

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)*).

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. Which of the following statements is true about positional encoding? Select all that apply.

1 / 1 point

- ☐ Positional encoding is used in the transformer network and the attention model.
- ☒ Positional encoding is important because position and word order are essential in sentence construction of any language.

✔ Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

✔ Positional encoding uses a combination of sine and cosine equations.

✔ Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

✔ Positional encoding provides extra information to our model.

✔ Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

Expand

✔ Correct
Great, you got all the right answers.

10. Which of these is *not* a good criterion for a good positional encoding algorithm?

0 / 1 point

- ☒ It must be deterministic.
- ☐ Distance between any two time-steps should be consistent for all sentence lengths.
- ☐ The algorithm should be able to generalize to longer sentences.
- ☐ It should output a common encoding for each time-step (word's position in a sentence).

Expand

✘ Incorrect
This is a good criterion for a good positional encoding algorithm.