**[Skip to Main Content](https://www.coursera.org/learn/nlp-sequence-models/exam/0s68u/transformers/view-attempt" \l "main)**

[[](https://www.coursera.org/)](https://www.coursera.org/)

Explore

Top of Form



Bottom of Form

* **1**
* Vijay Misra

1. [Sequence Models](https://www.coursera.org/learn/nlp-sequence-models/home/welcome)
2. [Week 4](https://www.coursera.org/learn/nlp-sequence-models/home/week/4)
3. Transformers

[**Previous**](https://www.coursera.org/learn/nlp-sequence-models/supplement/M7RxM/lectures-in-pdf)[**Next**](https://www.coursera.org/learn/nlp-sequence-models/programming/roP5y/transformers-architecture-with-tensorflow)

* **Transformers**
* **Lecture Notes (Optional)**
* **Quiz**

**[Quiz:](https://www.coursera.org/learn/nlp-sequence-models/exam/0s68u/transformers)**[Transformers](https://www.coursera.org/learn/nlp-sequence-models/exam/0s68u/transformers)

[10 questions](https://www.coursera.org/learn/nlp-sequence-models/exam/0s68u/transformers)

* **Programming Assignment**
* **Transformer Applications - Ungraded Labs**
* **Conclusion**
* **References & Acknowledgments**

**QUIZQuiz • 30 MIN30 minutes**

**Transformers**

**Submit your assignment**

**DUE DATE**Jul 11, 11:59 PM PDTJuly 11, 11:59 PM PDT

Resume

**Receive grade**

**TO PASS**80% or higher

**Grade**

82.50%

View Feedback

We keep your highest score

Transformers

Graded Quiz • 30 min

**Due** Jul 11, 11:59 PM PDT

**Congratulations! You passed!**

**TO PASS**80% or higher

Keep Learning

**GRADE**

82.50%

**Transformers**

**LATEST SUBMISSION GRADE**

82.5%

1.

Question 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

**Correct**

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

2.

Question 2

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

**1 / 1 point**



Convolutional Neural Network style of processing.

**Correct**



Attention mechanism.

**Correct**



None of these.

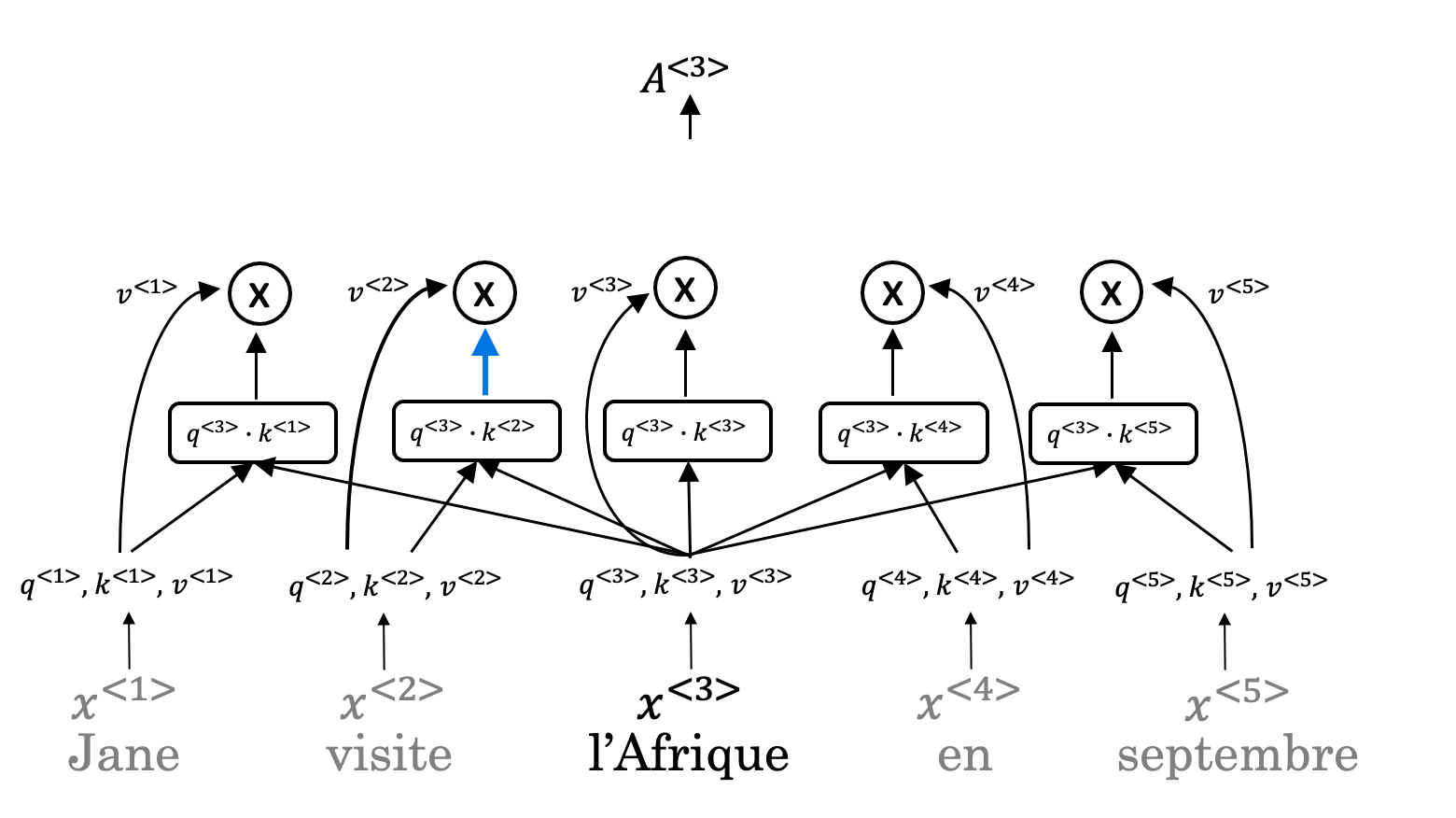


Convolutional Neural Network style of architecture.

3.

Question 3

The concept of *Self-Attention* is that:



**1 / 1 point**



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 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 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 selecting the highest of those word values to map the Attention related to that given word.

**Correct**

4.

Question 4

Which of the following correctly represents *Attention ?*

**1 / 1 point**



Attention(Q, K, V) = softmax(\frac{QV^T}{\sqrt{d\_k}})K*Attention*(*Q*,*K*,*V*)=*softmax*(*dk*​​*QVT*​)*K*



Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d\_k}})V*Attention*(*Q*,*K*,*V*)=*softmax*(*dk*​​*QKT*​)*V*



Attention(Q, K, V) = min(\frac{QV^T}{\sqrt{d\_k}})K*Attention*(*Q*,*K*,*V*)=*min*(*dk*​​*QVT*​)*K*



Attention(Q, K, V) = min(\frac{QK^T}{\sqrt{d\_k}})V*Attention*(*Q*,*K*,*V*)=*min*(*dk*​​*QKT*​)*V*

**Correct**

5.

Question 5

Are the following statements true regarding Query (Q), Key (K) and Value (V) ?

Q = interesting questions about the words in a sentence

K = specific representations of words given a Q

V = qualities of words given a Q

**1 / 1 point**



False



True

**Correct**

Correct! Q = interesting questions about the words in a sentence, K = qualities of words given a Q, V = specific representations of words given a Q

6.

Question 6



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

**1 / 1 point**



True



False

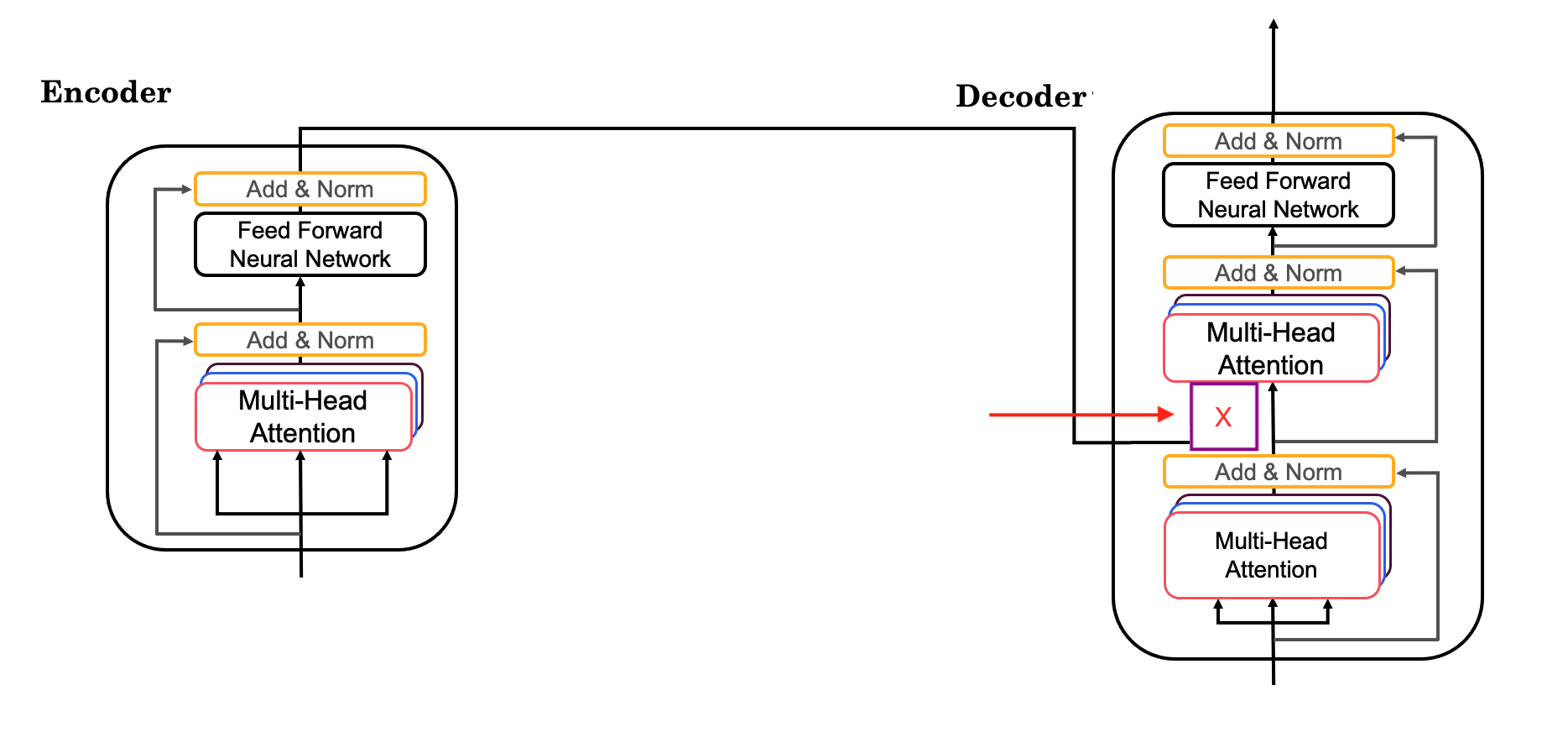
**Correct**

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

7.

Question 7

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



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

(Check all that apply)

**1 / 1 point**



K

**Correct**



Q



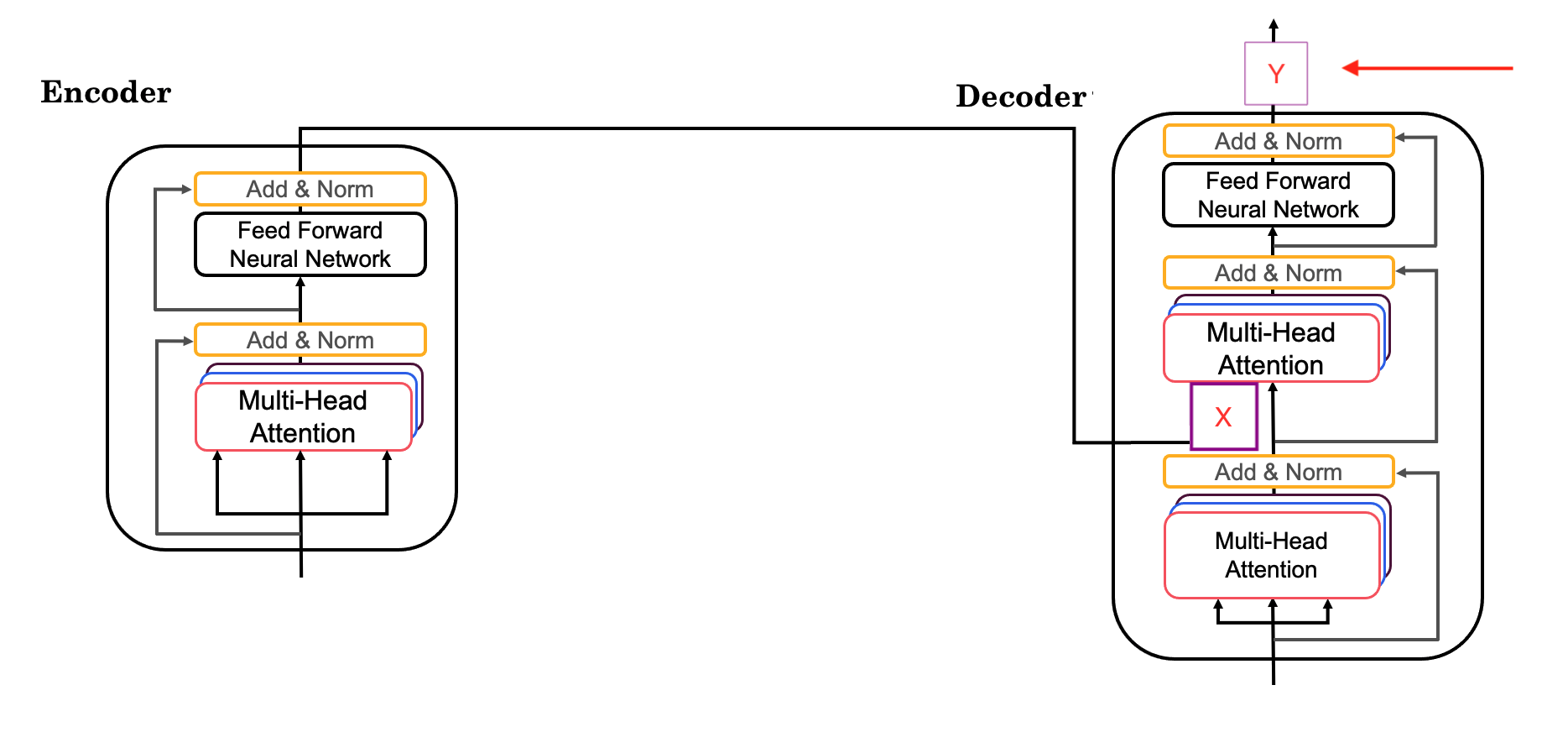
V

**Correct**

8.

Question 8

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



What is the output layer(s) of the *Decoder* ? (Marked Y*Y*, pointed by the independent arrow)

**0 / 1 point**



Linear layer followed by a softmax layer.



Softmax layer followed by a linear layer.



Softmax layer



Linear layer

**Incorrect**

9.

Question 9

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

**0.75 / 1 point**



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.

You didn’t select all the correct answers

10.

Question 10

Which of these is a good criteria for a good positionial encoding algorithm?

**0.5 / 1 point**



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



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



The algorithm should be able to generalize to longer sentences.

**Correct**



None of the these.

You didn’t select all the correct answers