

Copyright Notice

These slides are distributed under the Creative Commons License.

[DeepLearning.AI](#) makes these slides available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite [DeepLearning.AI](#) as the source of the slides.

For the rest of the details of the license, see <https://creativecommons.org/licenses/by-sa/2.0/legalcode>



deeplearning.ai

Sequence to
sequence models

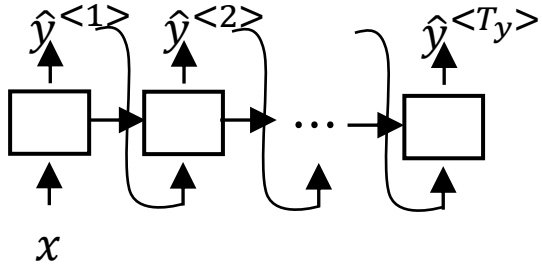
Transformers
Intuition

Transformers Motivation

In that these all intermediate feels like a bottleneck to the flow

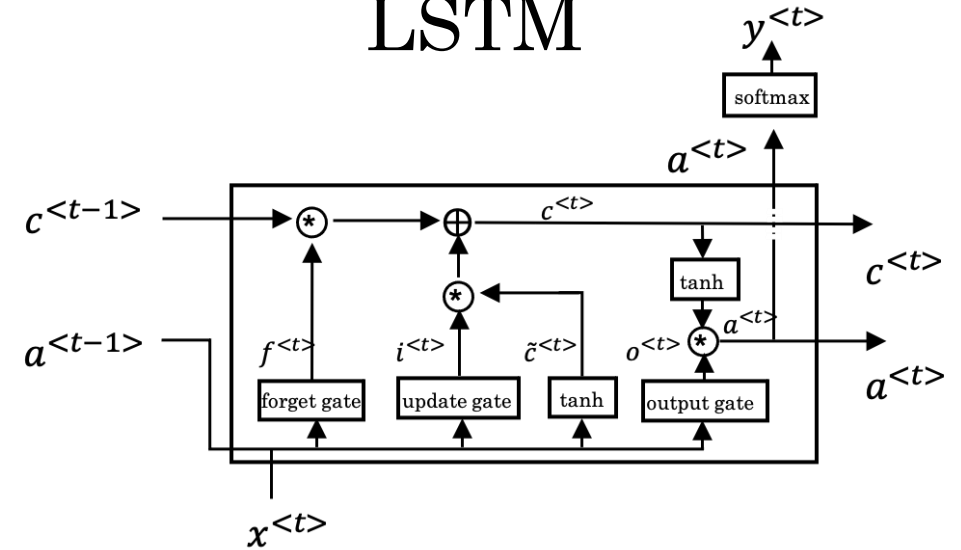
Increased complexity,
sequential

RNN



GRU

LSTM



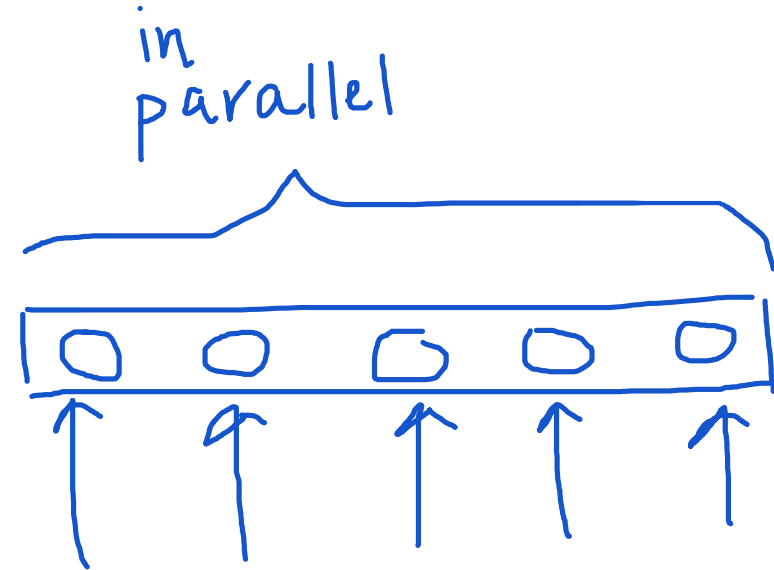
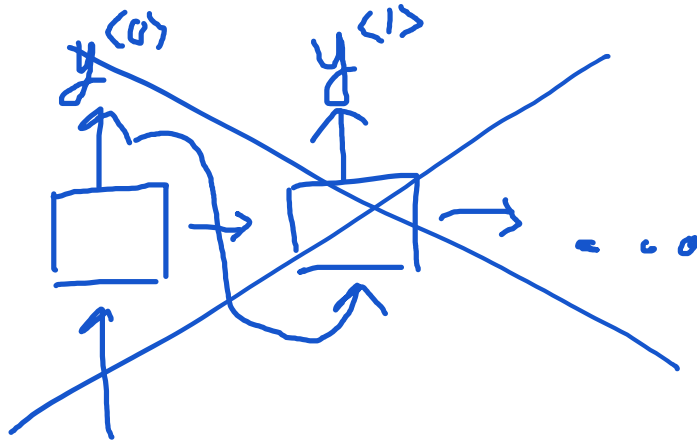
Transformers Intuition

- Attention + CNN

- Self-Attention

The goal of self-attention is for a sentence of five words it will end up computing 5 representation for these 5

- Multi-Head Attention





deeplearning.ai

Sequence to sequence models

Self-Attention

Self-Attention Intuition

$A(q, K, V)$ = attention-based vector representation of a word

↪ calculate for each word

Calculate that A for each of the word. In our case 5 words A1, A2, ,A3, A4, A5. Like

RNN Attention

$$\alpha^{<t, t'>} = \frac{\exp(e^{<t, t'>})}{\sum_{t'=1}^T x \exp(e^{<t, t'>})}$$

Transformers Attention

$$A(q, K, V) = \sum_i \frac{\exp(e^{<q \cdot k^{<i>}>})}{\sum_j \exp(e^{<q \cdot k^{<j>}>})} v^{<i>}$$

Query Key and Value.

$x^{<1>}$ Jane $x^{<2>}$ visite $x^{<3>}$ l'Afrique

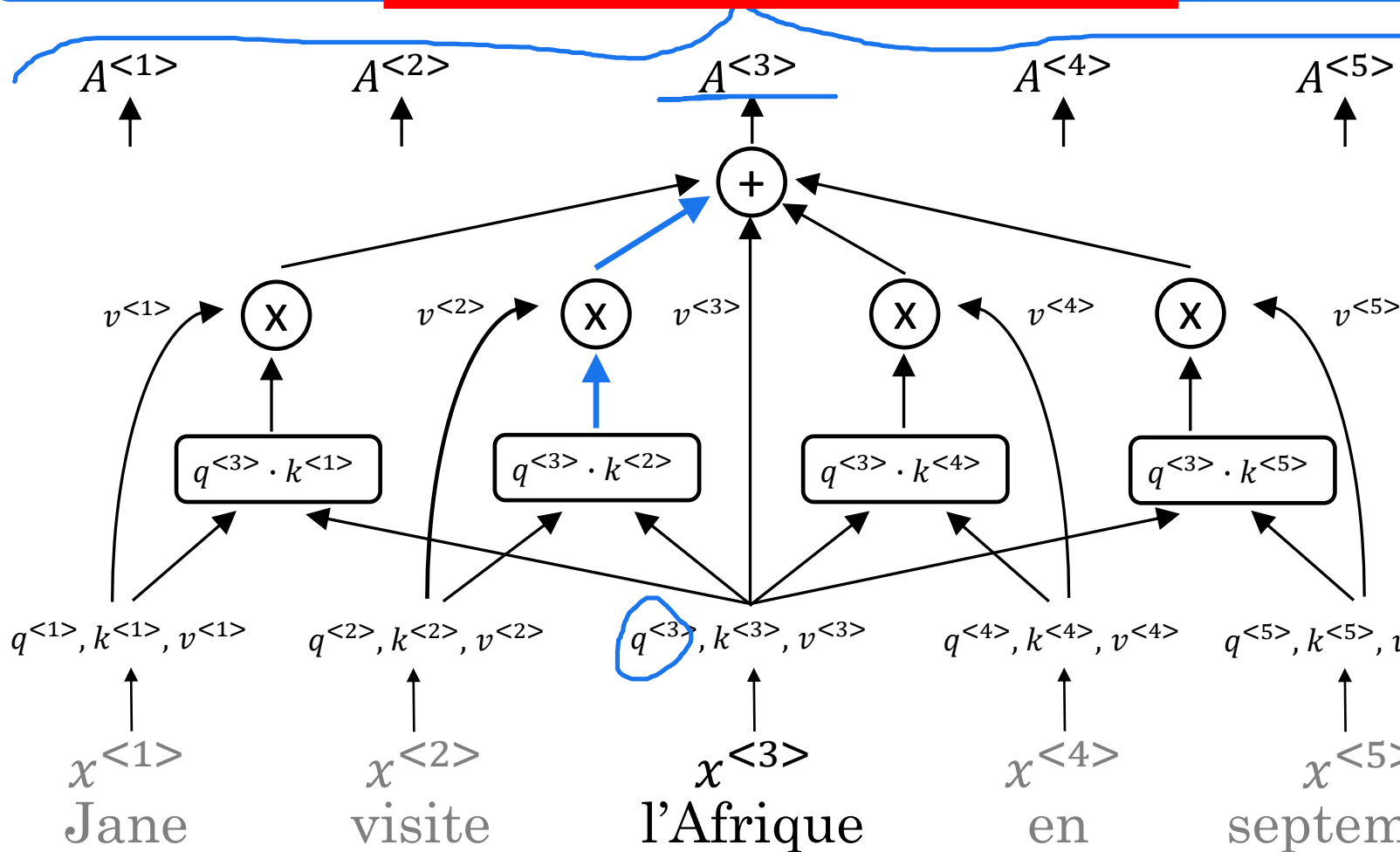
$x^{<4>}$ en $x^{<5>}$ septembre

Self-Attention

$$A(q, K, V) = \sum_i \frac{\exp(e^{q \cdot k^{<i>}})}{\sum_j \exp(e^{q \cdot k^{<j>}})} v^{<i>}$$

softmax

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



We are going to associate each of the word with 3 values called query, key, and value.

Query (Q)	Key (K)	Value (V)
$q^{<1>}$	$k^{<1>}$ person	$v^{<1>}$ Jane
$q^{<2>}$	$k^{<2>}$ action	$v^{<2>}$ visit
$q^{<3>}$ What's happening there?	$k^{<3>}$	$v^{<3>}$
$q^{<4>}$	$k^{<4>}$	$v^{<4>}$
$q^{<5>}$	$k^{<5>}$	$v^{<5>}$

W^Q, W^K, W^V

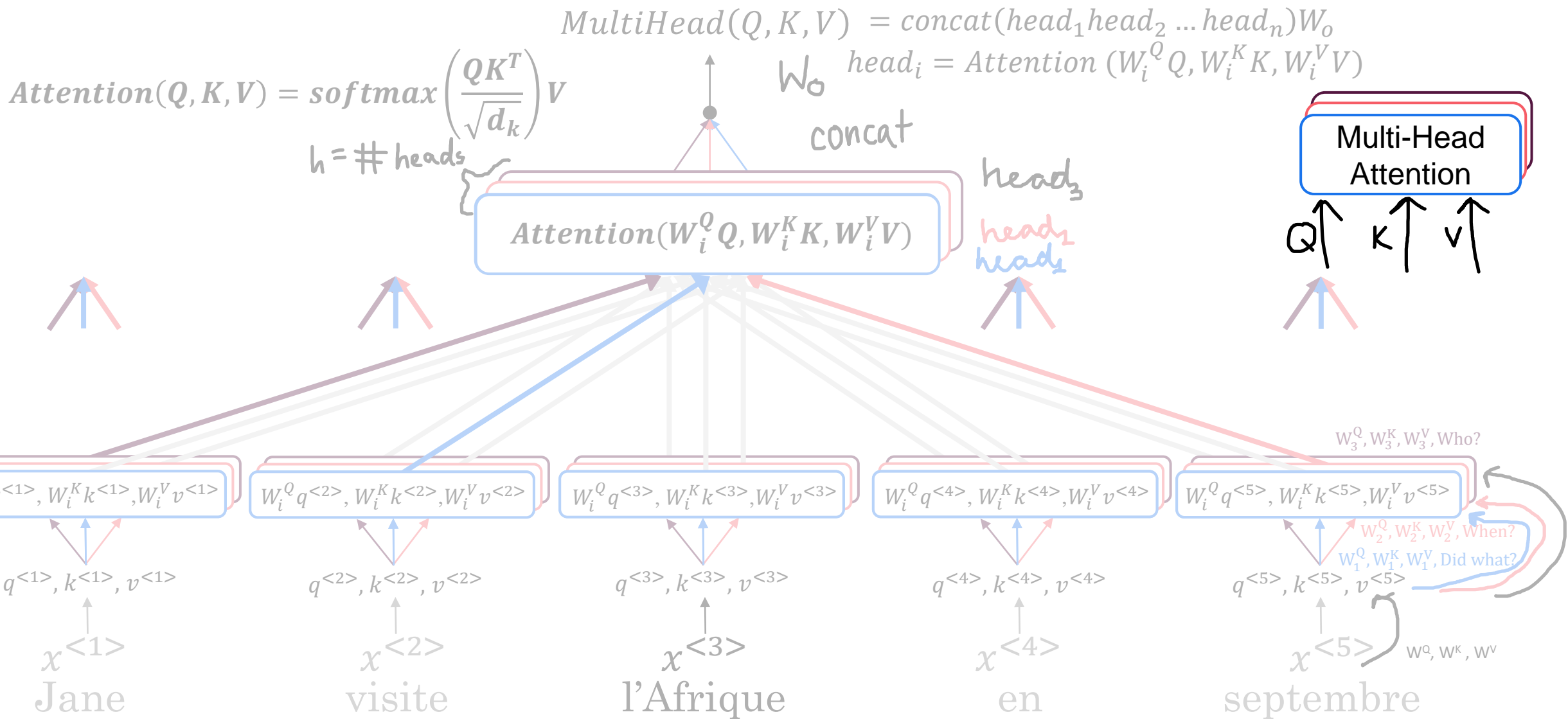


deeplearning.ai

Sequence to sequence models

Multi-Head Attention

Multi-Head Attention





deeplearning.ai

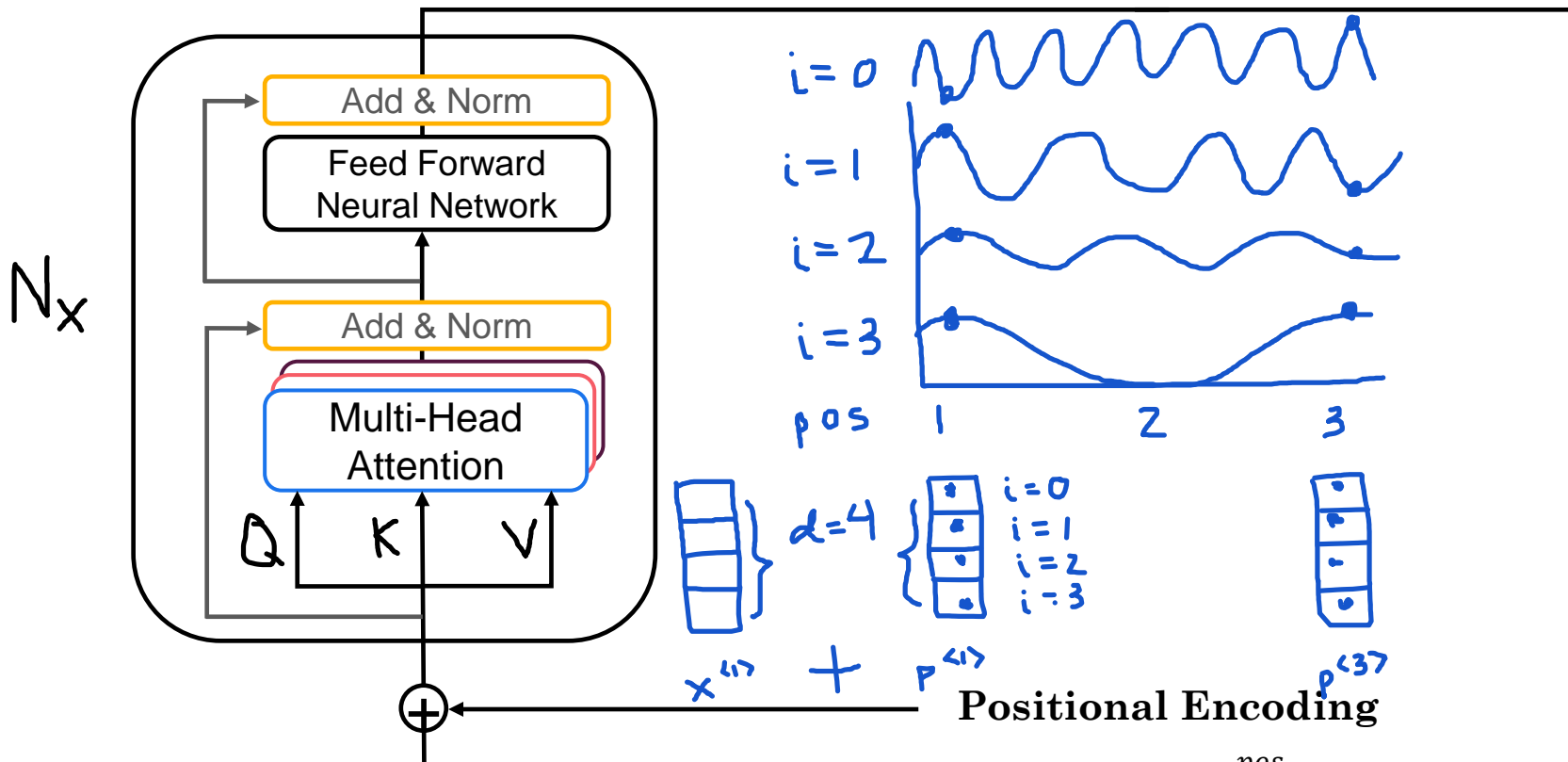
Sequence to sequence models

Transformers

Transformer Details

<SOS> Jane visits Africa in September <EOS>

Encoder



<SOS> $x^{(1)}$ $x^{(2)}$... $x^{(T_x-1)}$ $x^{(T_x)}$ <EOS>
Jane visite l'Afrique en septembre

Decoder

