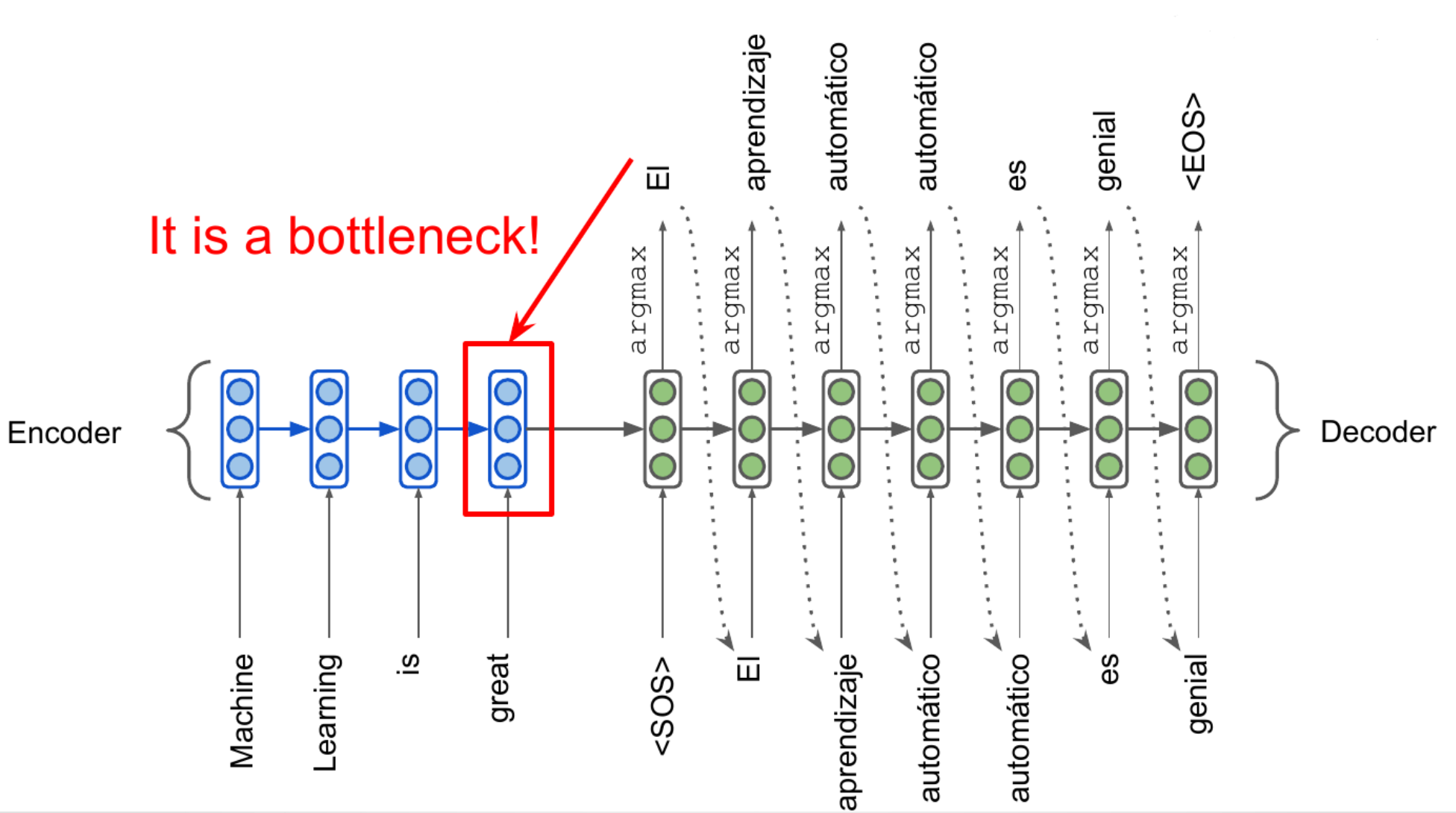
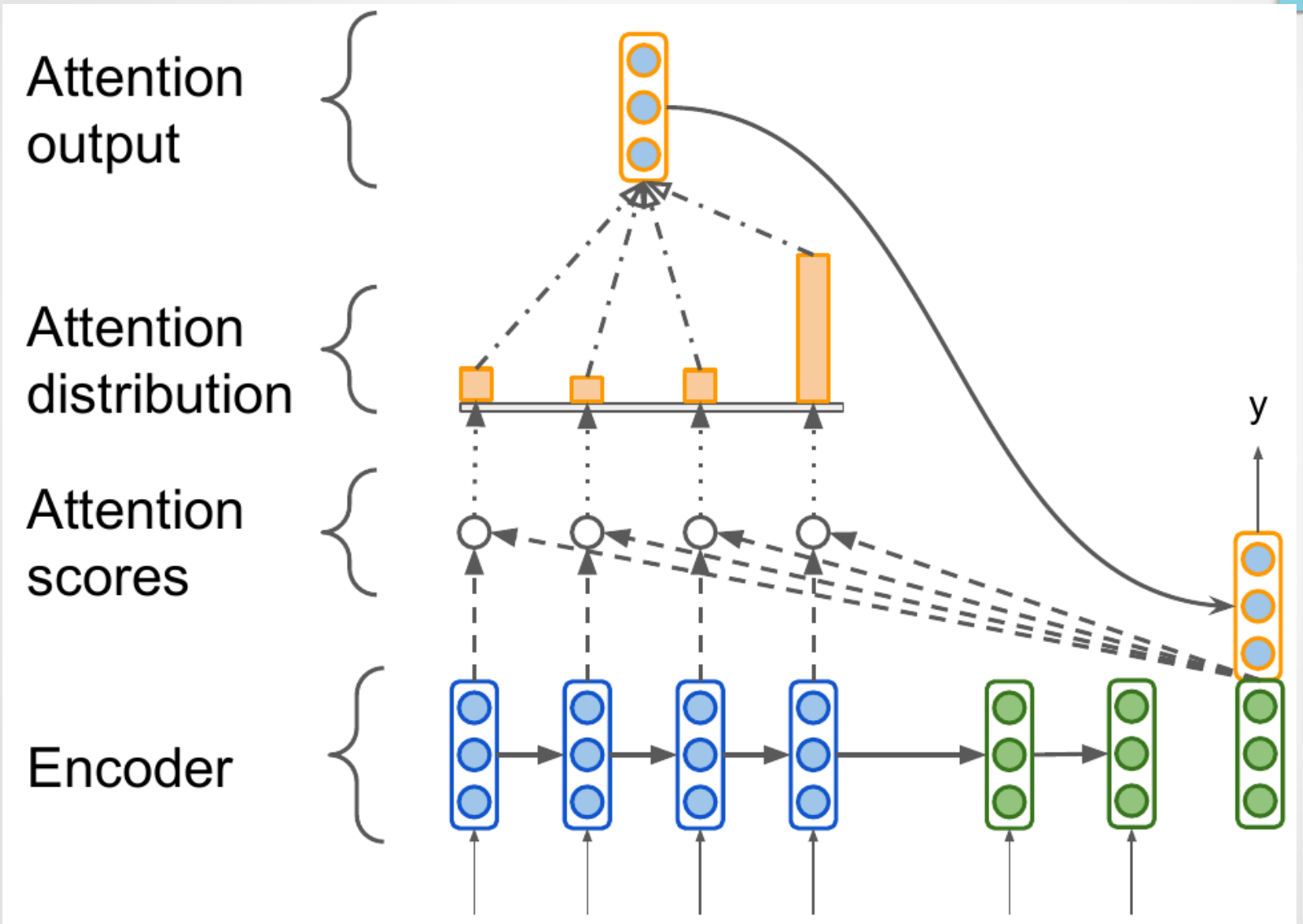
Модель Transformer и механизм внимания Self-Attention

Евгений Борисов

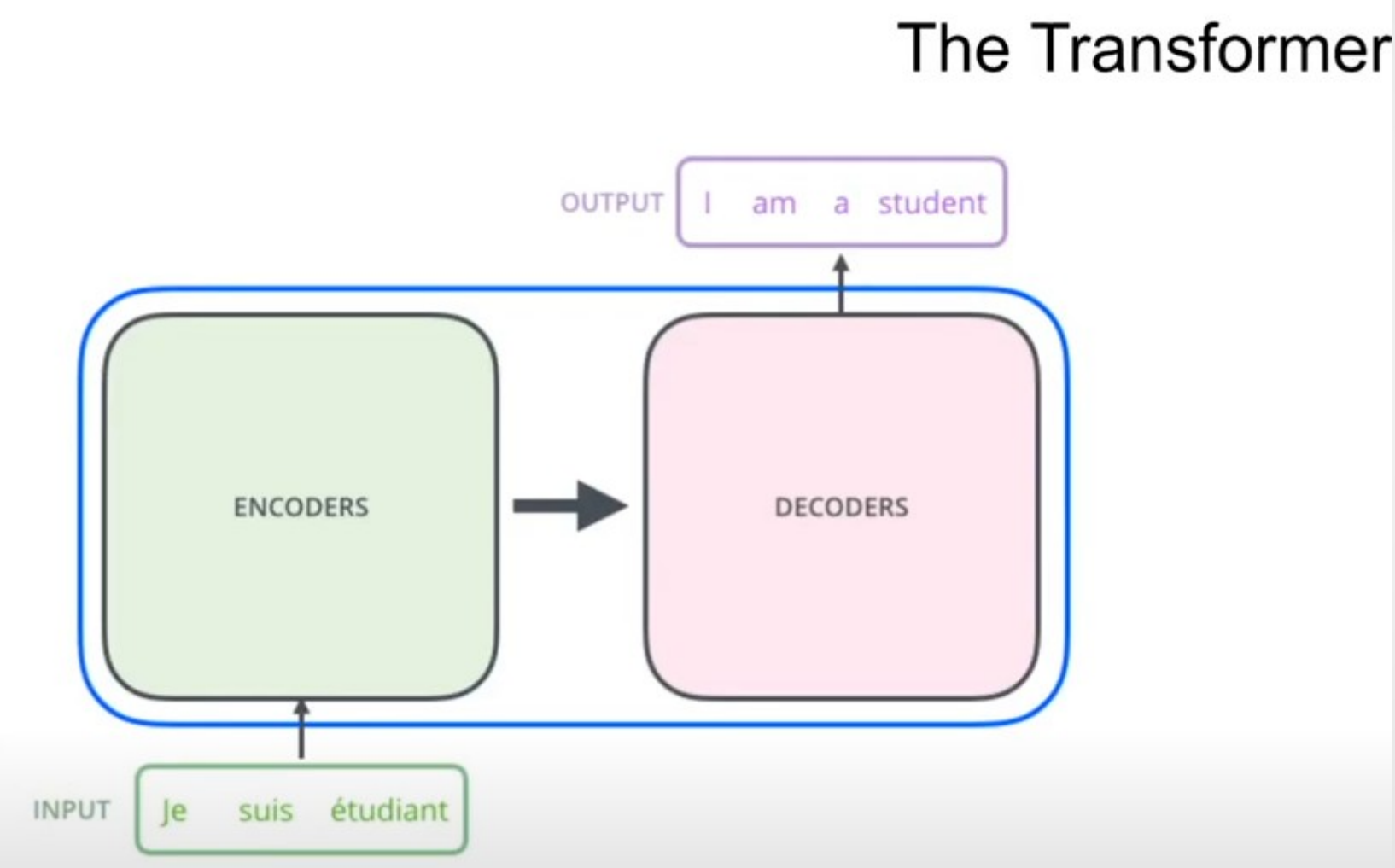
SEQ2SEQ NMT


It is a bottleneck!
Encoder
Decoder
Machine
Learning
is
great
<SOS>
El
aprendizaje
automático
automático
es
genial
<EOS>
argmax
El
aprendizaje
automático
automático
es
genial

SEQ2SEQ NMT with ATTENTION



Transformer



Transformer

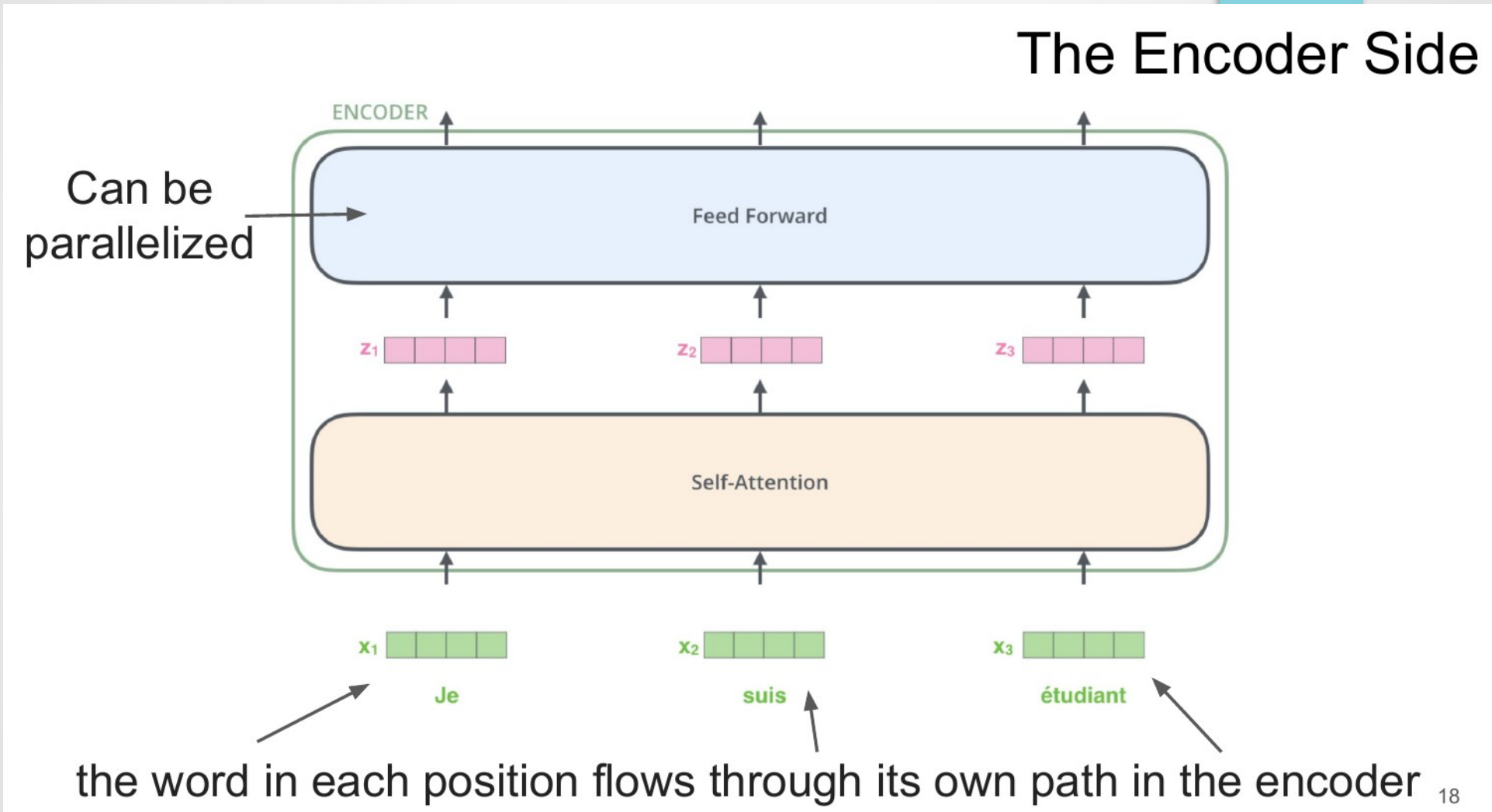
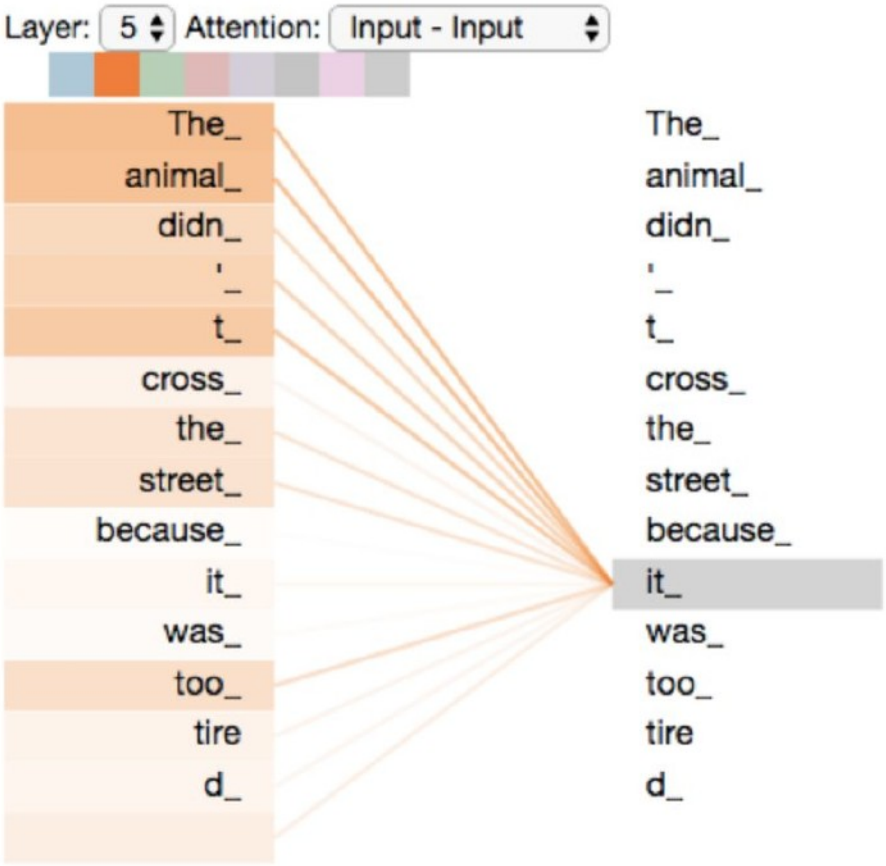


Image source: https://jalammar.github.io/illustrated-transformer/

Transformer

Self-Attention at a High Level



Transformer

Self-Attention

Transformer Self-Attention

query - откуда смотрим (из какого слова)

key - куда смотрим (на какое слово)

value - смысл (условно) слова

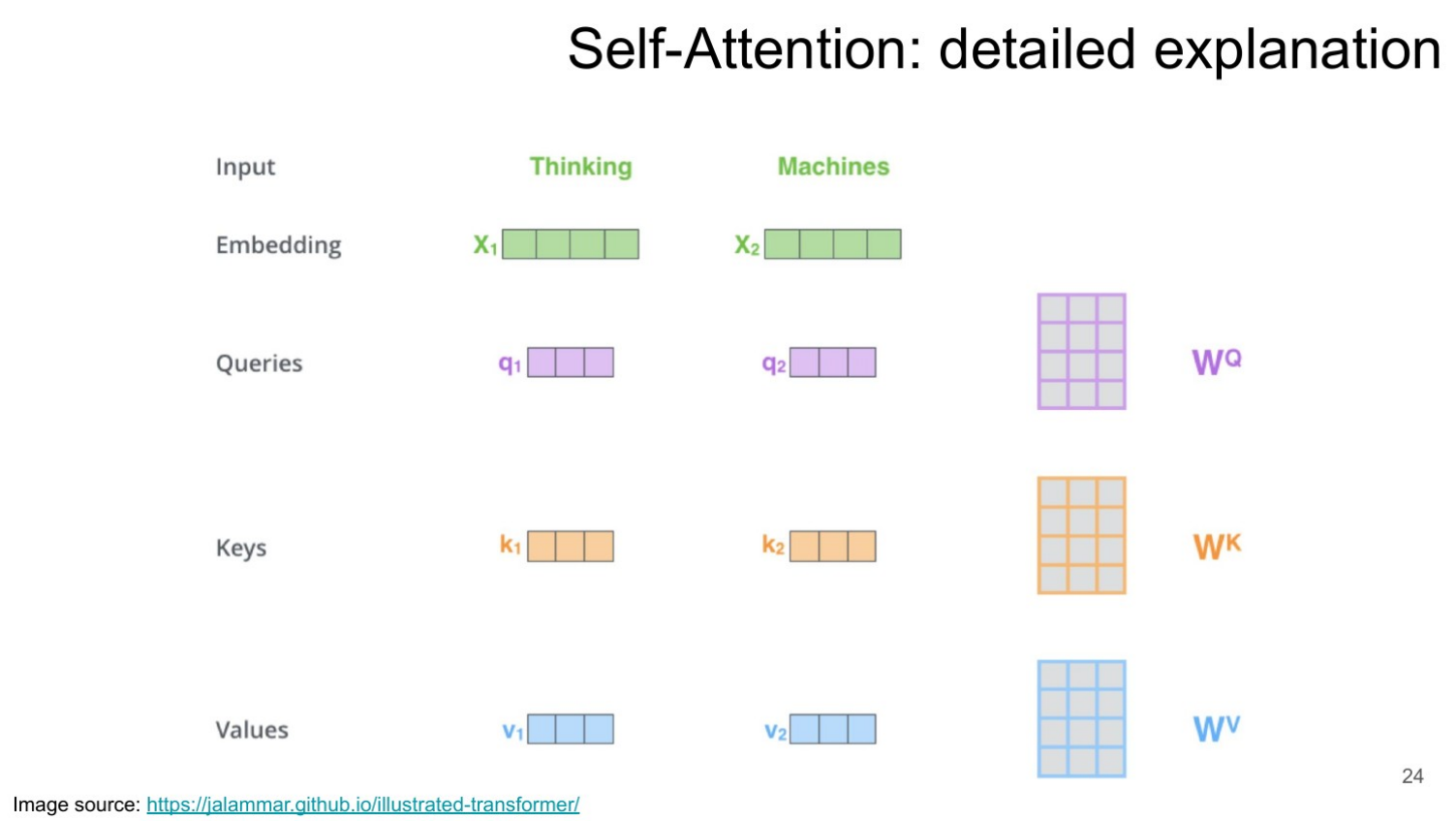


Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Self-Attention

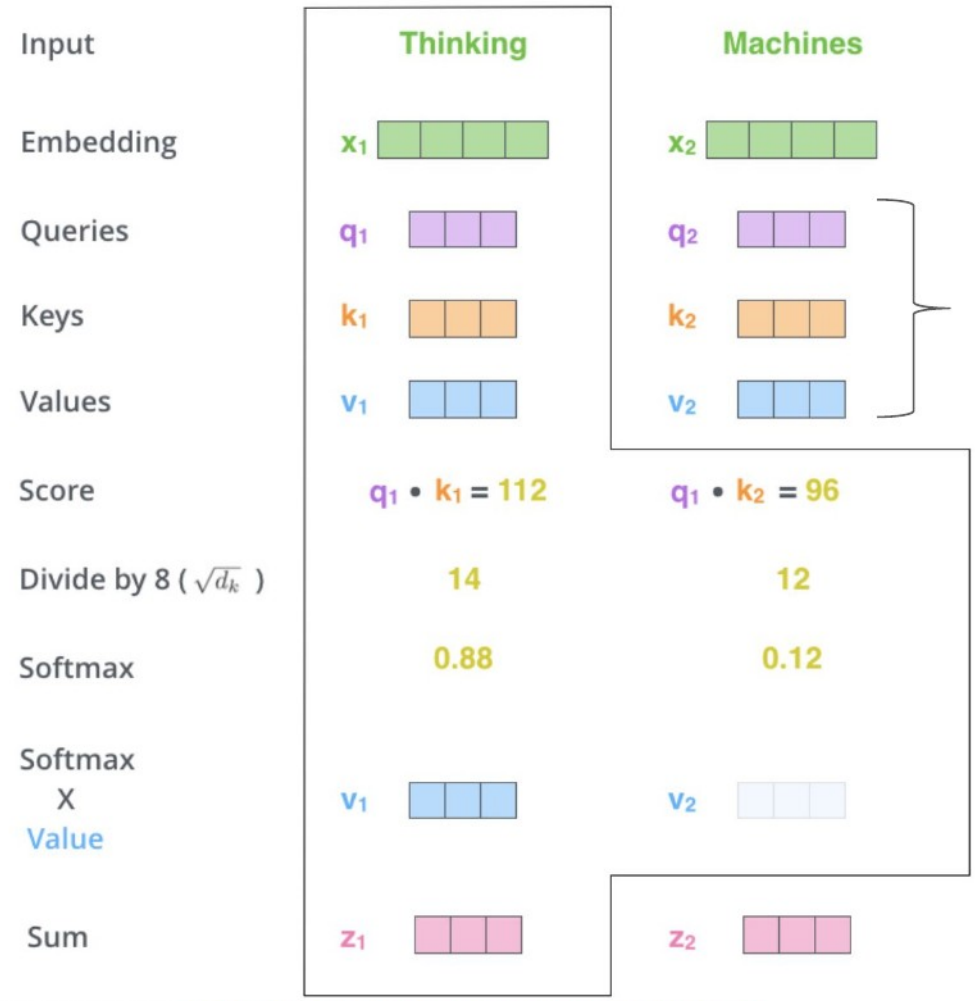
Self-Attention: Matrix Calculation

Pack embeddings into matrix **X**

Multiply **X** by weight matrices we've trained (**Wk, Wq, Wv**)

Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Self-Attention



Self-Attention

STEP 1: create Query, Key, Value

STEP 2: calculate scores

STEP 3: divide by $\sqrt{d_k}$

STEP 4: softmax

STEP 5: multiply each value vector by the softmax score

STEP 6: sum up the weighted value vectors

Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Self-Attention

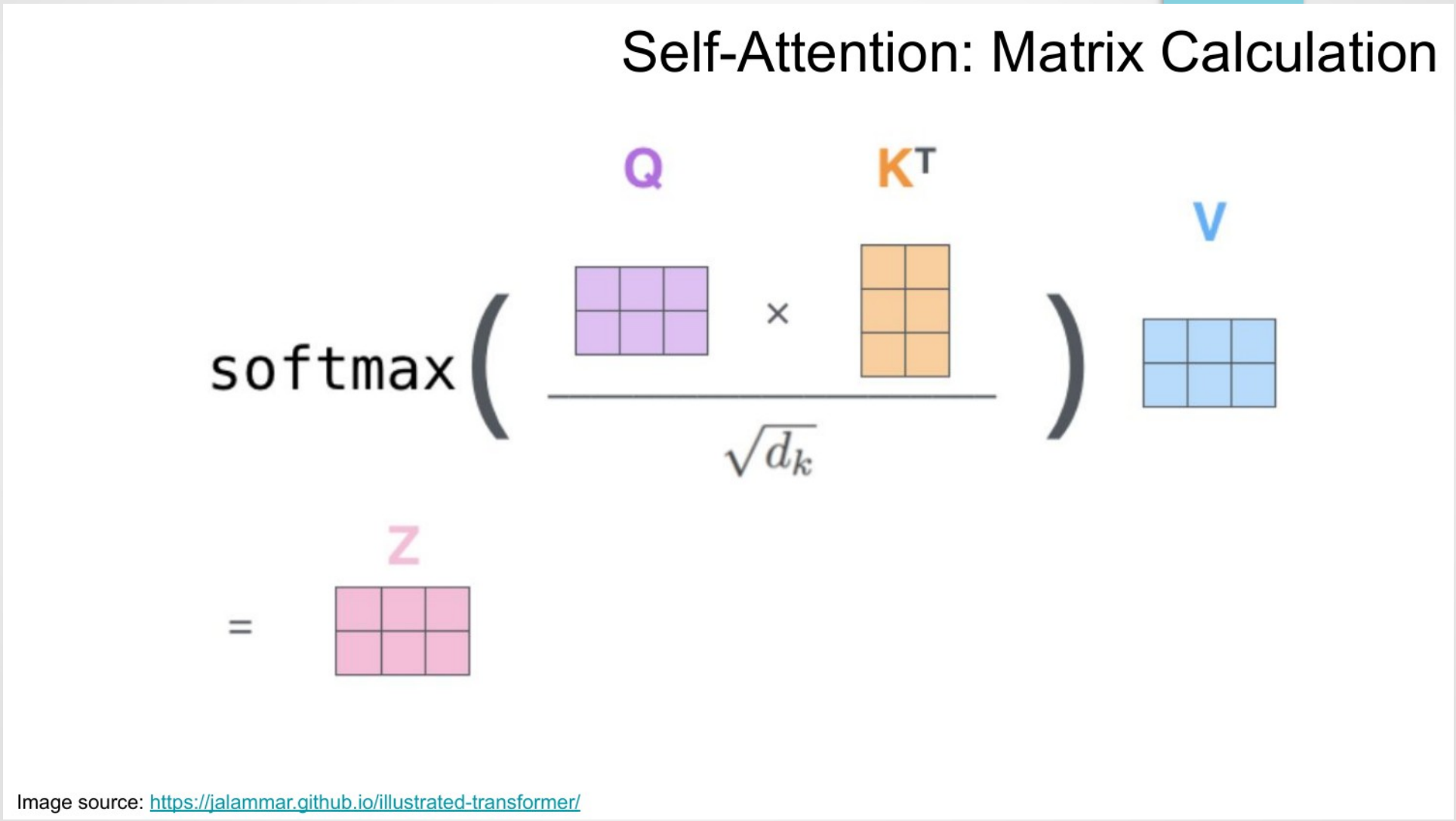


Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Self-Attention

каждая голова MHSA ищет свои связи между словами

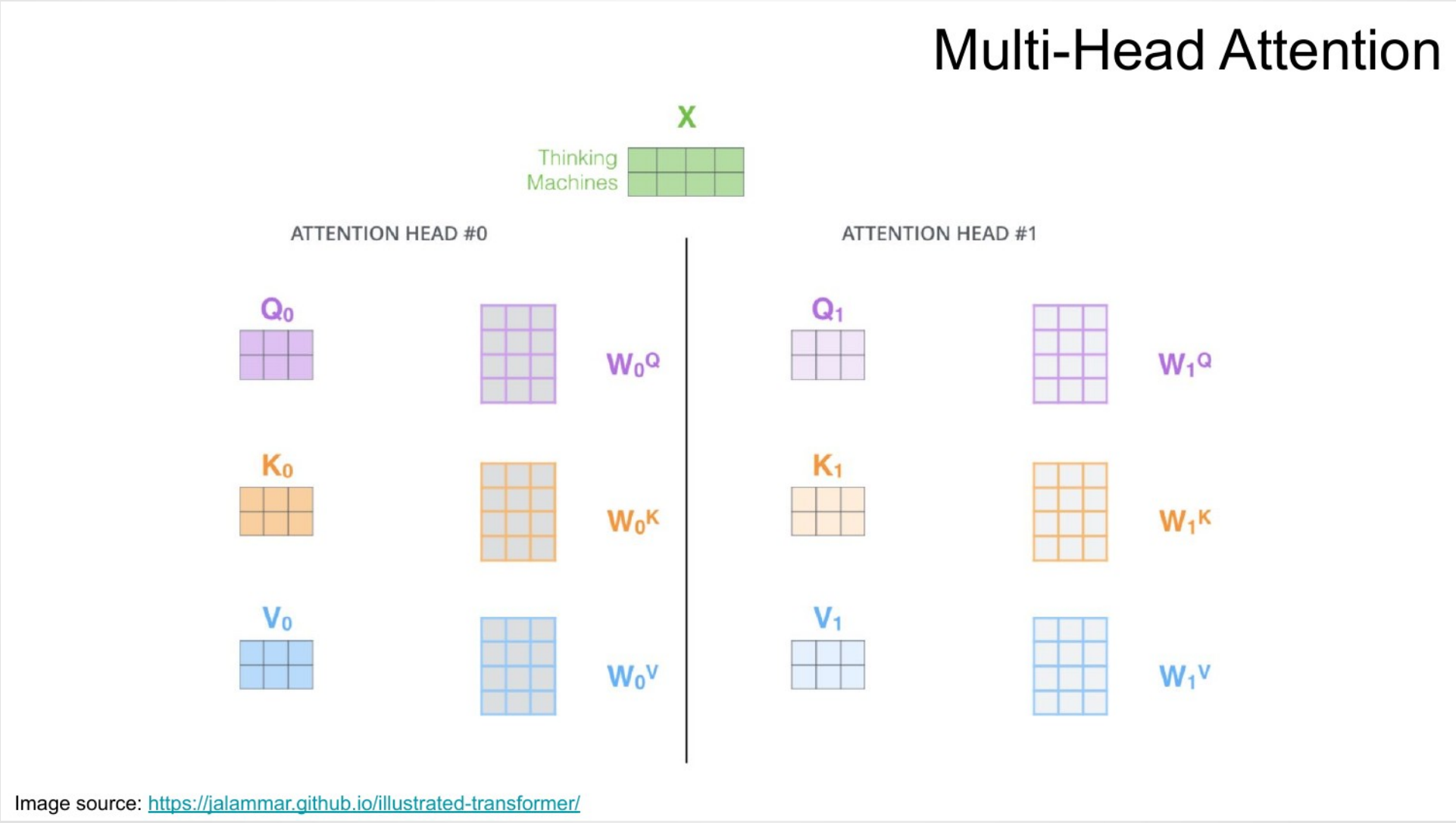


Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Self-Attention

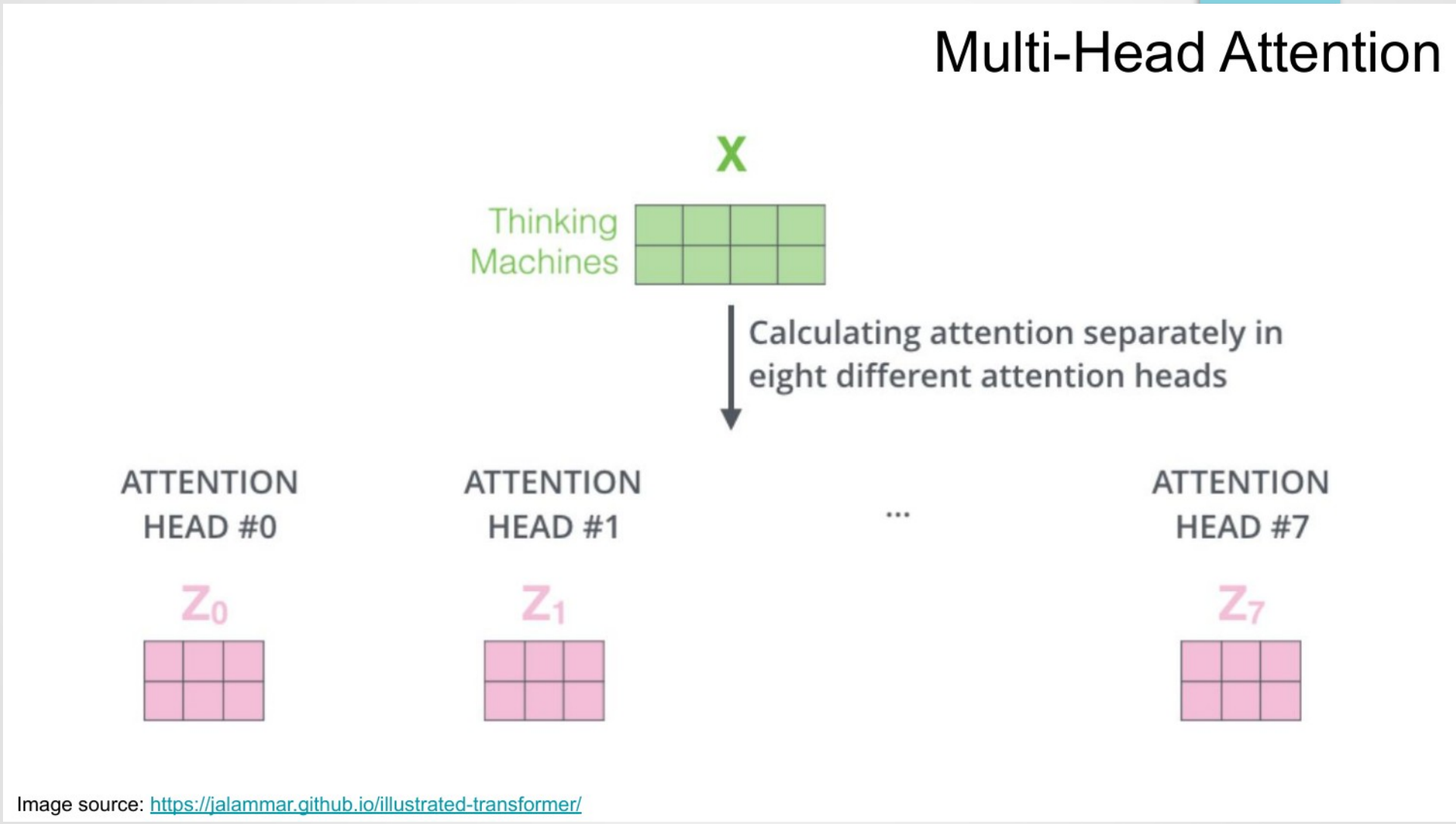


Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Self-Attention

можно сделать так, чтобы размерность входа и выхода MHSE была одинаковая

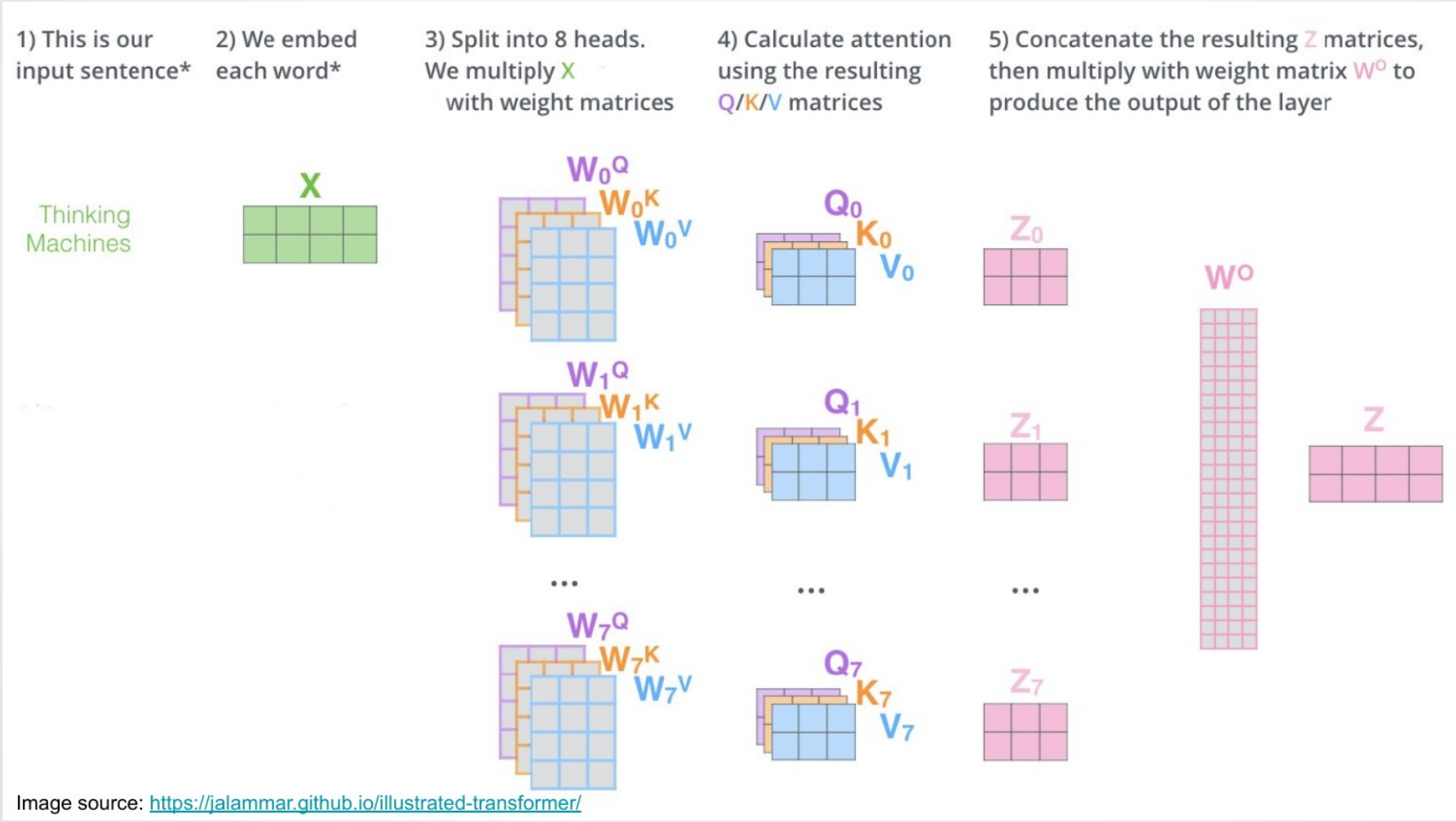


Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Positional Encoding

Positional Encoding

Transformer Positional Encoding

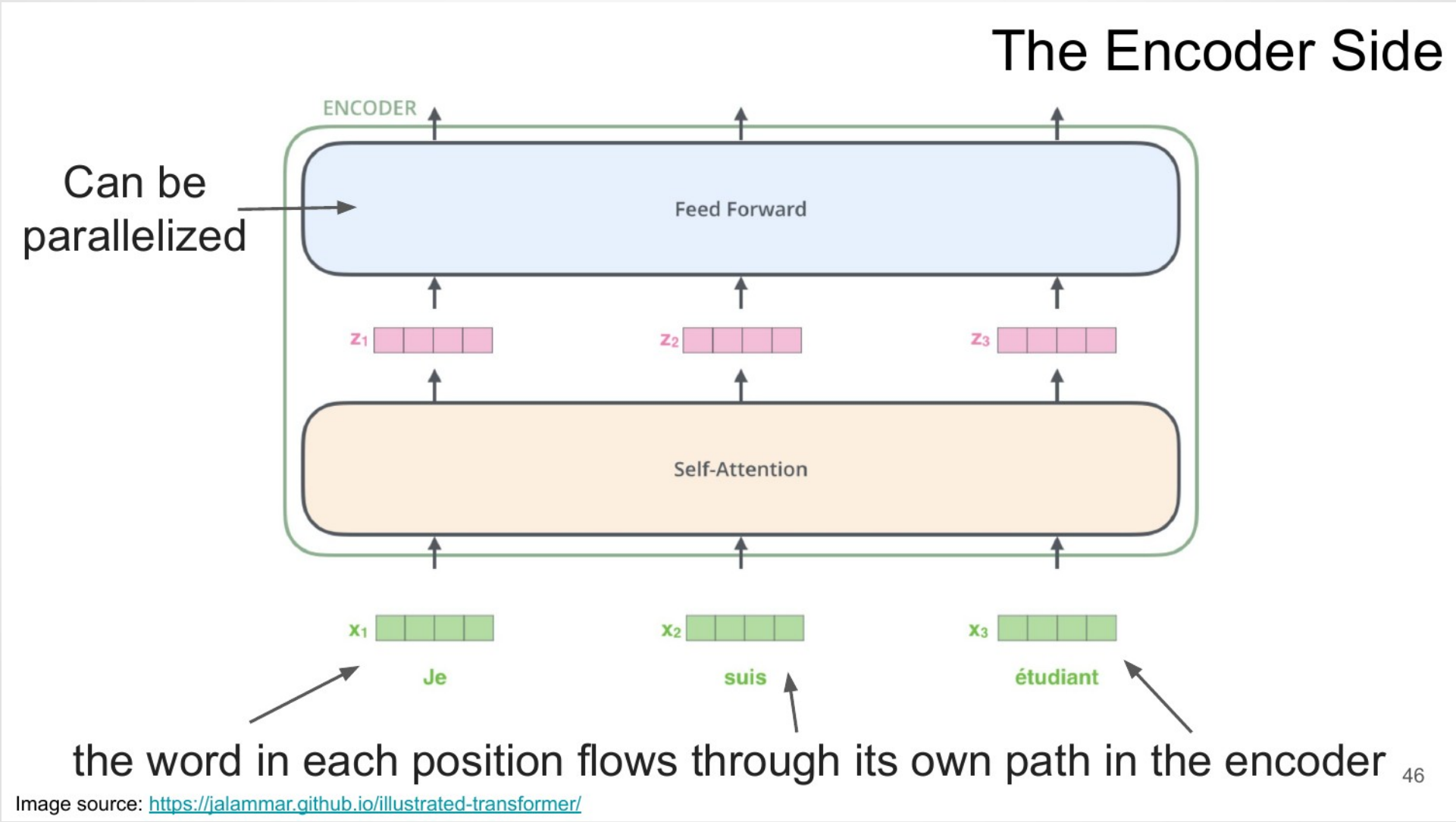


Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Positional Encoding

Необходимо обозначить позицию слова выполняя условия

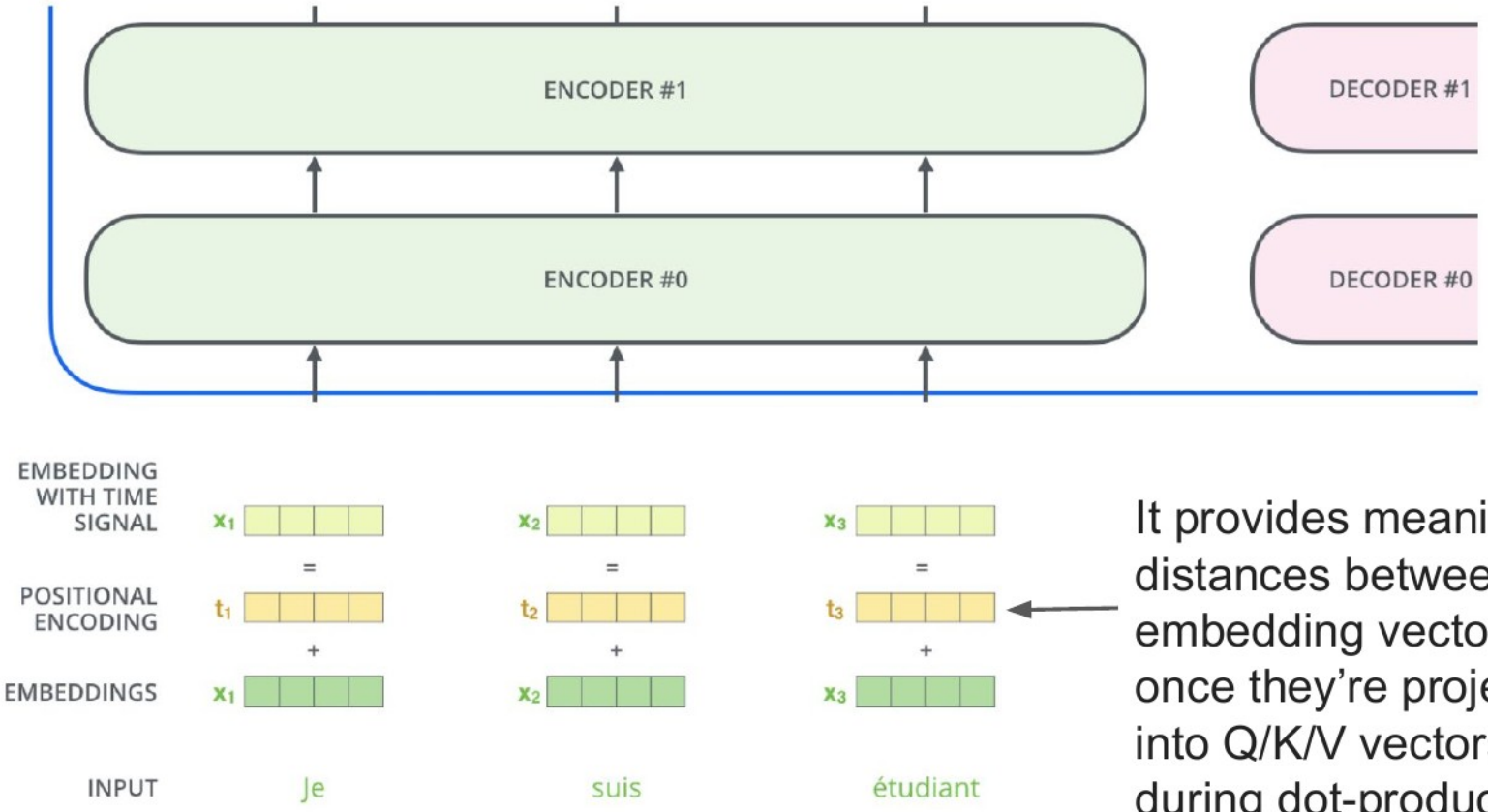
- уникальность для каждого слова

- не зависит от длинны предложения

- детерминирован (не стохастический)

Transformer Positional Encoding

Positional Encoding



It provides meaningful distances between the embedding vectors once they're projected into Q/K/V vectors and during dot-product attention

Image source: https://jalammar.github.io/illustrated-transformer/

Transformer Positional Encoding

Positional Encoding: why sin and cos?

$$\vec{p_t}^{(i)} = f(t)^{(i)} = \begin{cases} \sin(\omega_k t), & \text{if } i = 2k \\ \cos(\omega_k t), & \text{if } i = 2k+1 \end{cases}$$

$$\omega_k = \frac{1}{10000^{2k/d}}$$

$$\vec{p_t} = \begin{bmatrix} \sin(\omega_1.t) \\ \cos(\omega_1.t) \\ \\ \sin(\omega_2.t) \\ \cos(\omega_2.t) \\ \\ \vdots \\ \\ \sin(\omega_{d/2}.t) \\ \cos(\omega_{d/2}.t) \end{bmatrix}_{d \times 1}$$

t stays for position in the original sequence
k is the index of the element in the positional vector

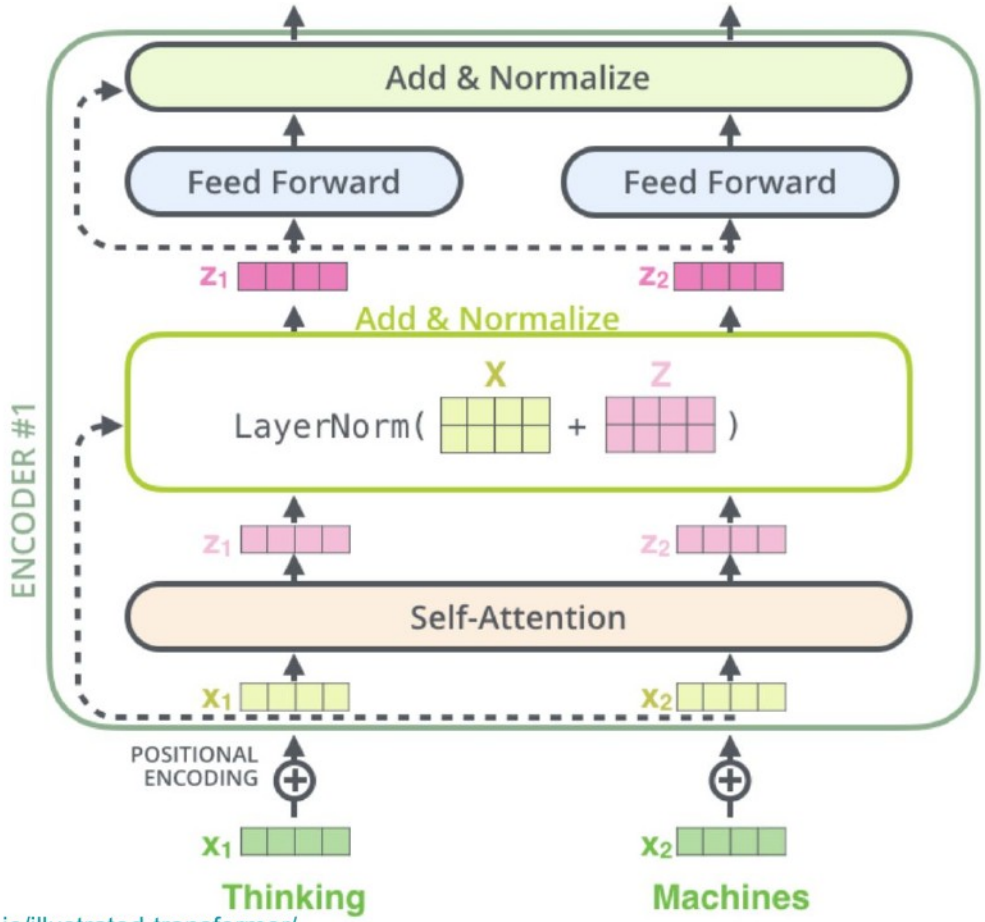
Image source: https://kazemnejad.com/blog/transformer_architecture_positional_encoding/

Transformer

Layer Normalization

Like BatchNorm

but normalize along all features representing latent vector



More info:
[Layer Normalization]

Image source: https://jalammar.github.io/illustrated-transformer/

Transformer

можно сделать так, чтобы размерность входа и выхода MHSE была одинаковая и состыковать несколько Encoder

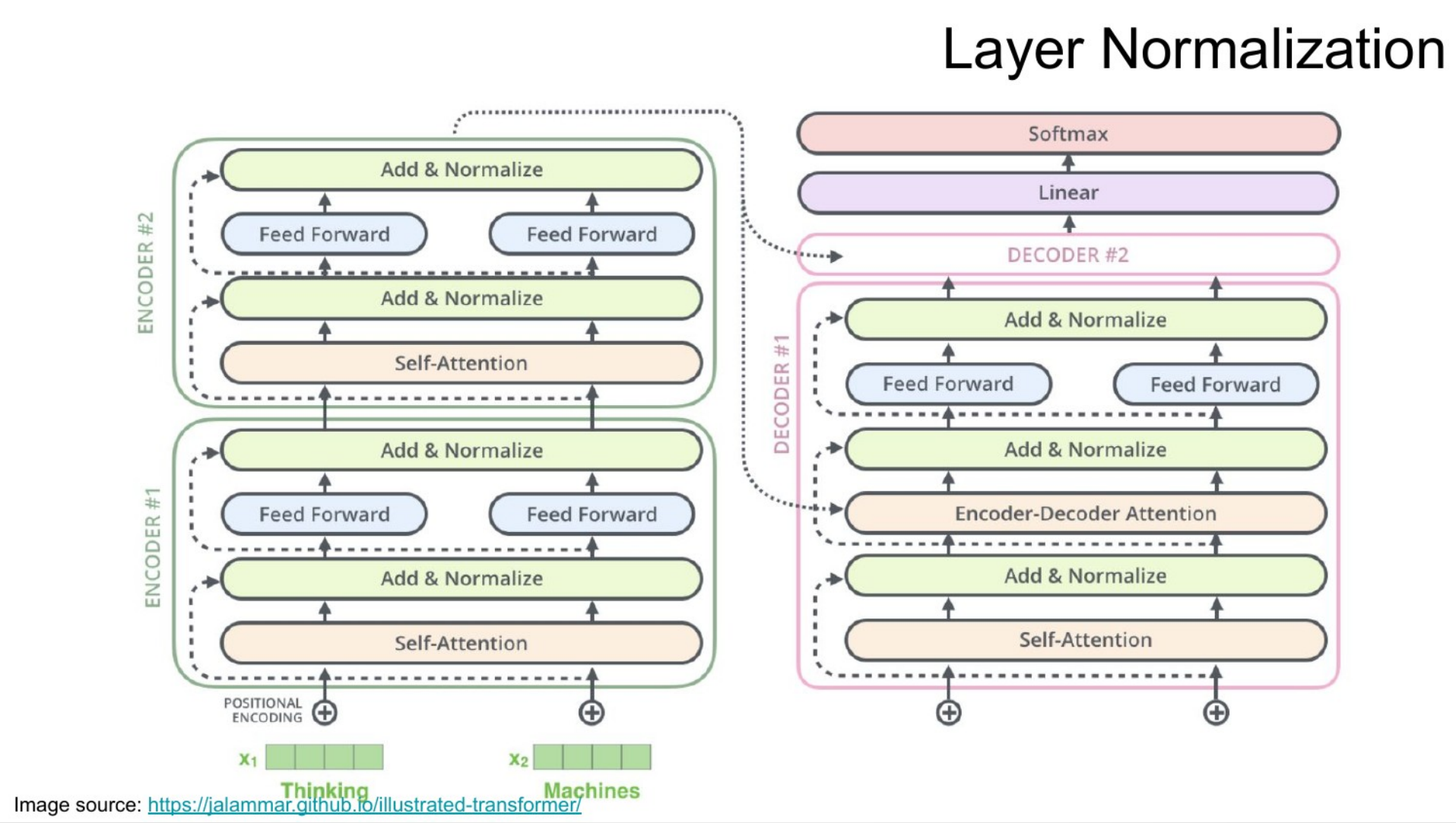


Image source: https://jalammar.github.io/illustrated-transformer/

SEQ2SEQ NMT: литература

git clone *https://github.com/mechanoid5/ml_nlp.git*

Евгений Борисов Неросетевой транслятор текстов. Использование рекуррентных нейронных сетей для создания систем машинного перевода и чатботов.
http://mechanoid.su/ml-chatbot.html

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin
Attention Is All You Need
https://arxiv.org/abs/1706.03762

Радослав Нейчев Прикладное машинное обучение. 4. Self-Attention. Transformer overview.
https://www.youtube.com/watch?v=UETKUIlYE6g&list=PL4_hYwCyhAvY7k32D65q3xJVo8X8dc3Ye&index=5