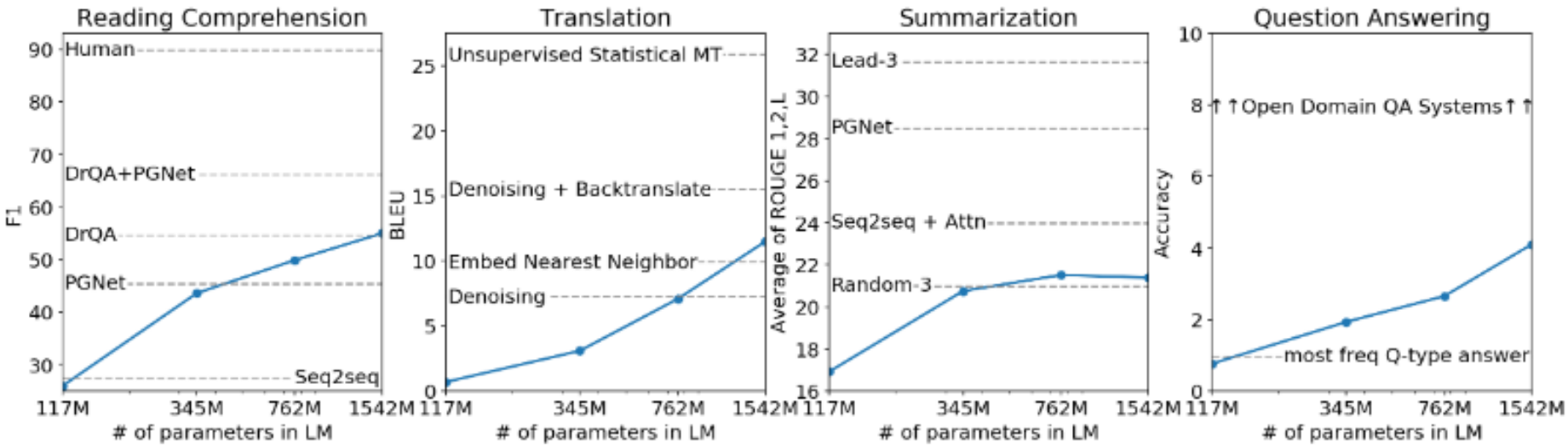
Language Models are Unsupervised Multitask Learners

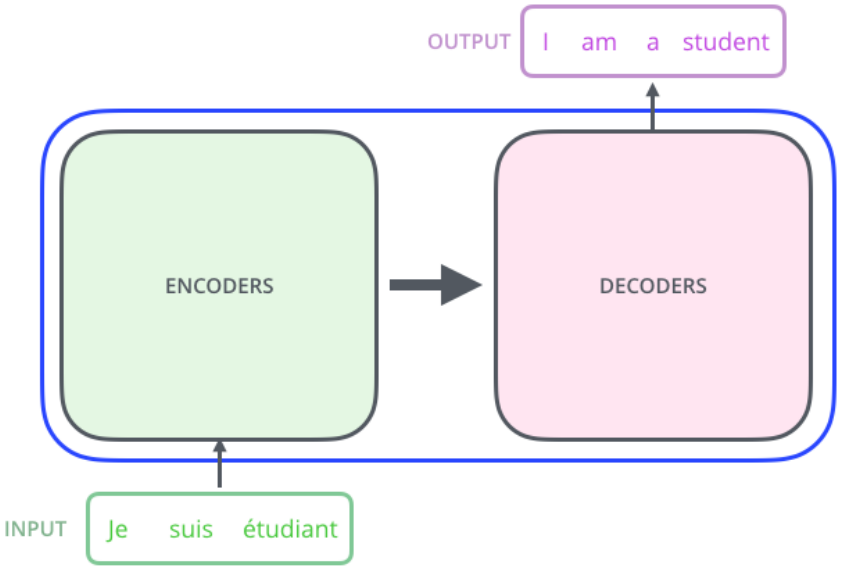
Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskeve

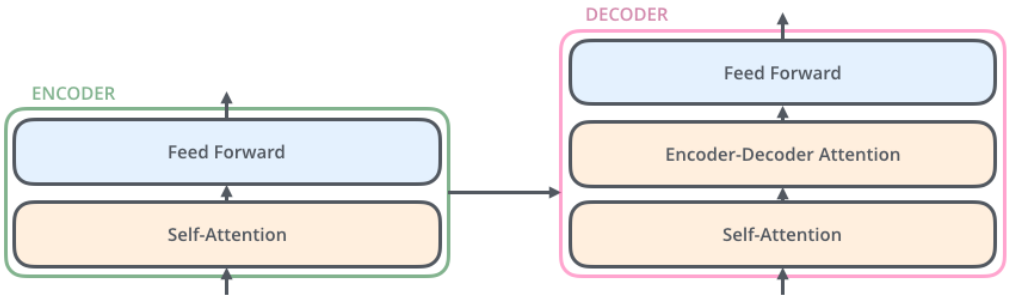
Zero-shot task performance



What GPT-2 Actually Is?

Generative Pretrained Transformer 2





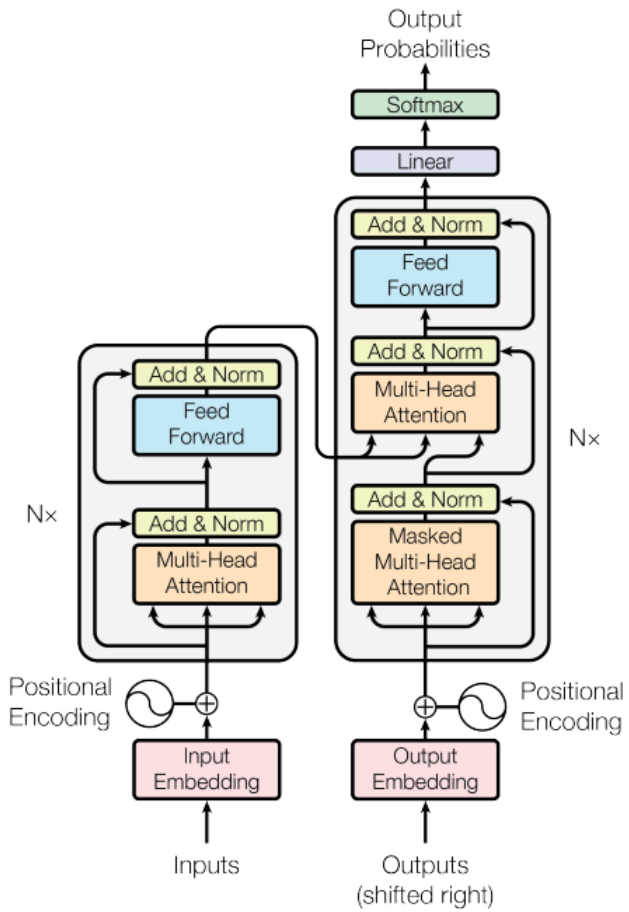
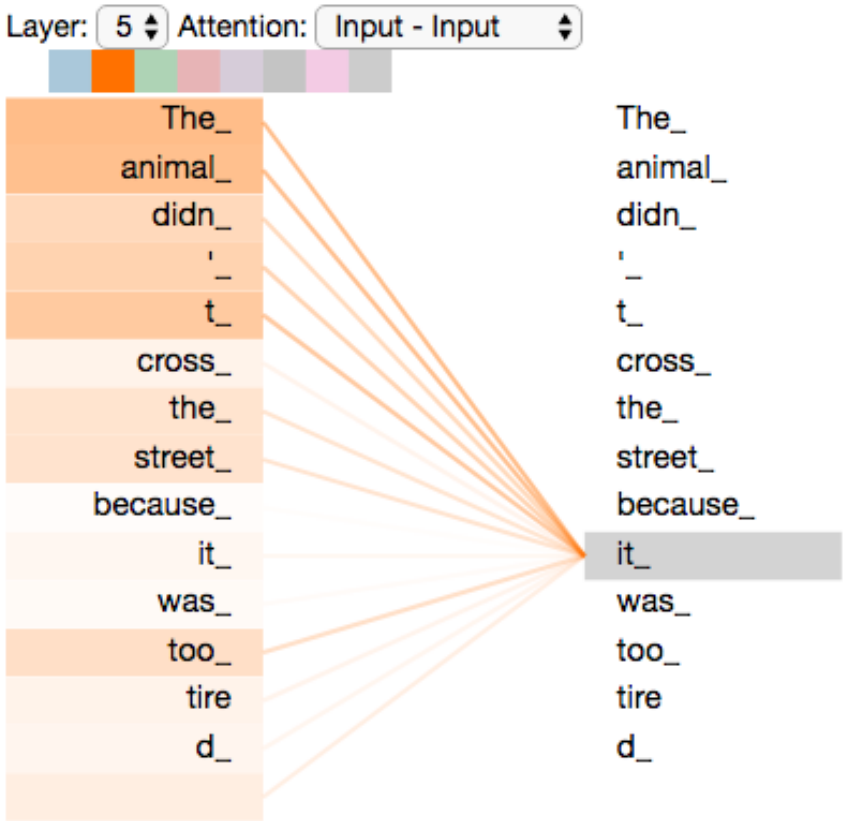


Figure 1: The Transformer - model architecture.

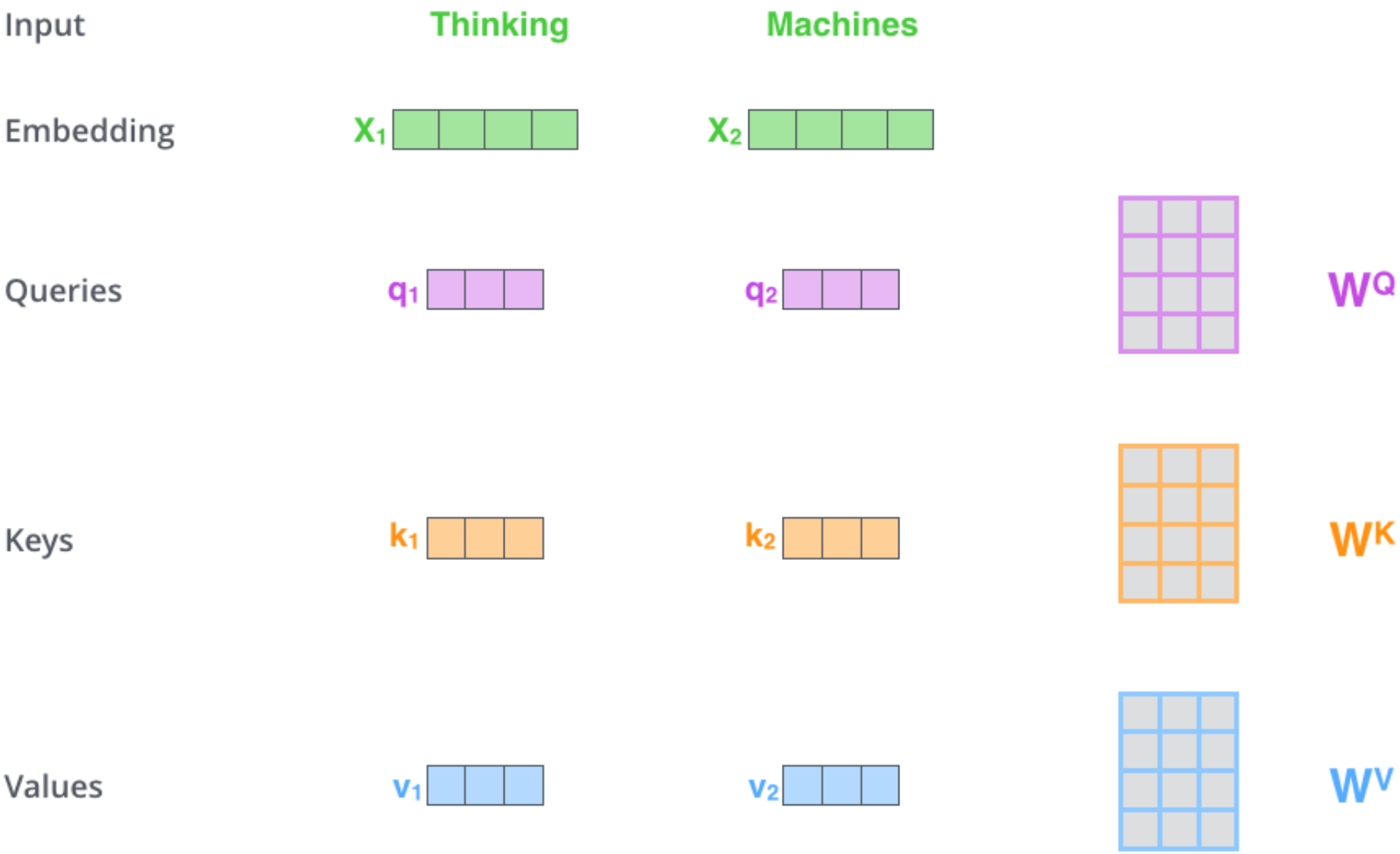
Self-attention layer

The animal didn't cross the street because it was too tired.

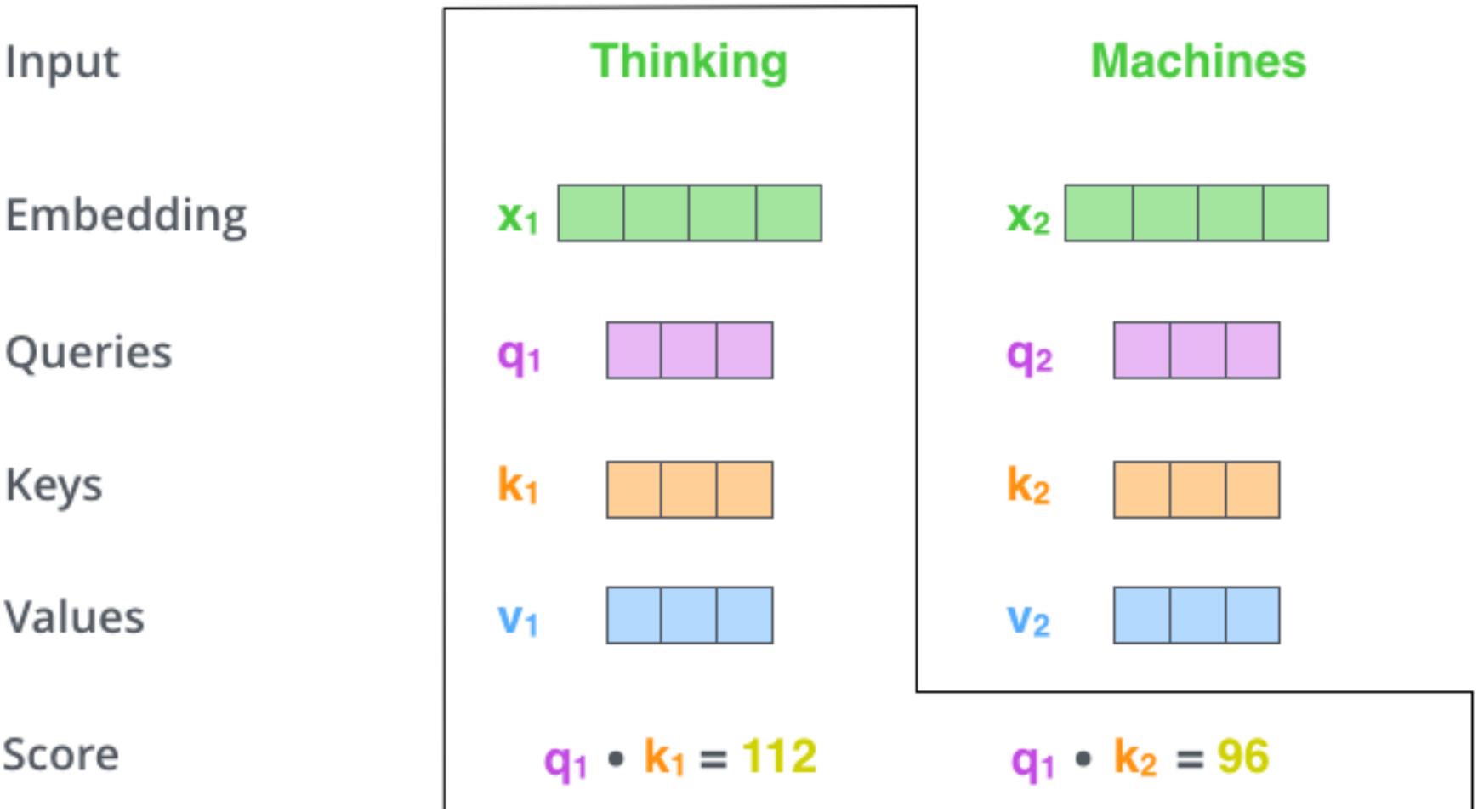
What does "it" refer to in this sentence?



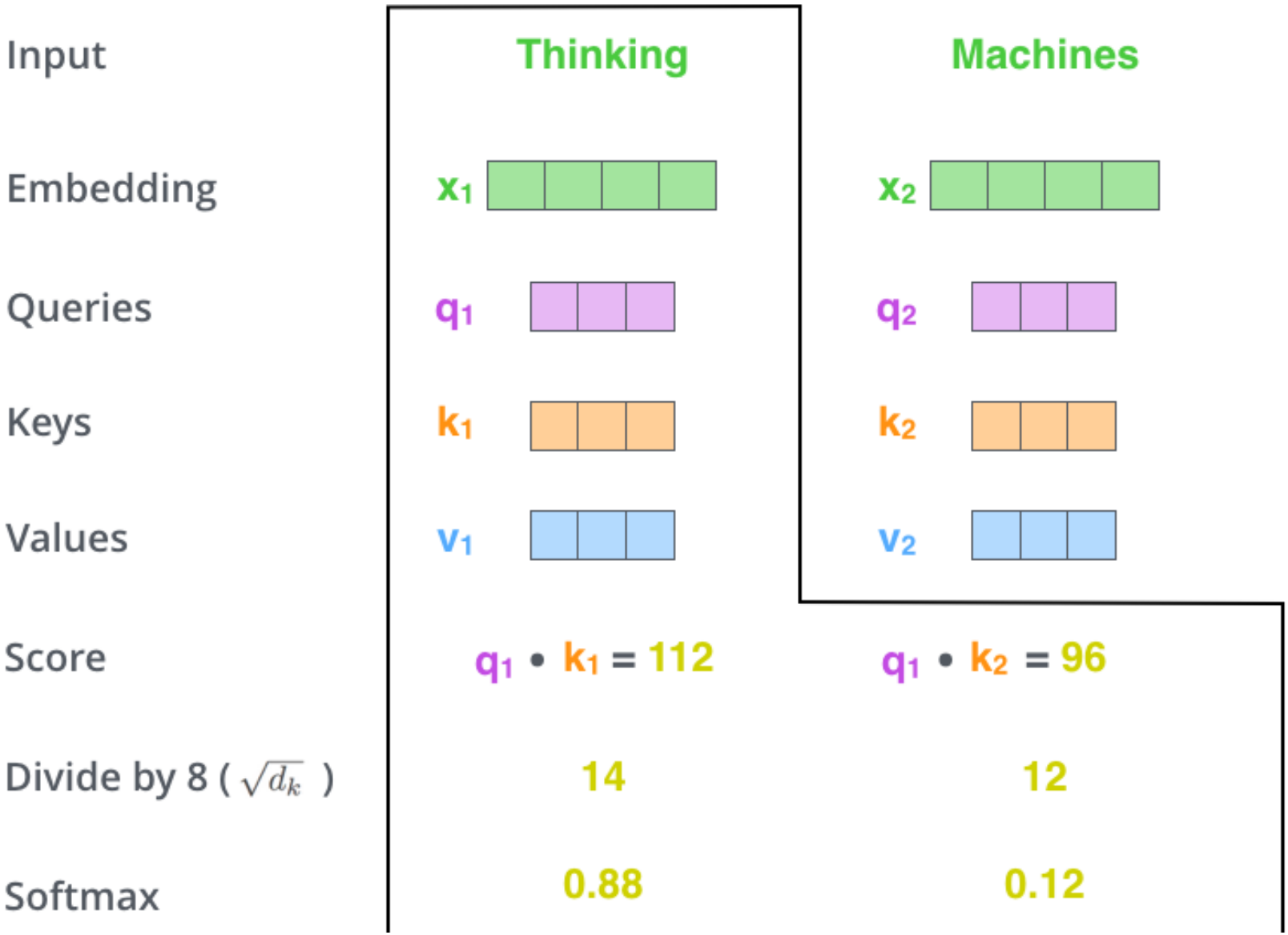
Step 1


Input
Thinking
Machines
Embedding
X1
X2
Queries
q1
q2
WQ
Keys
k1
k2
WK
Values
v1
v2
WV

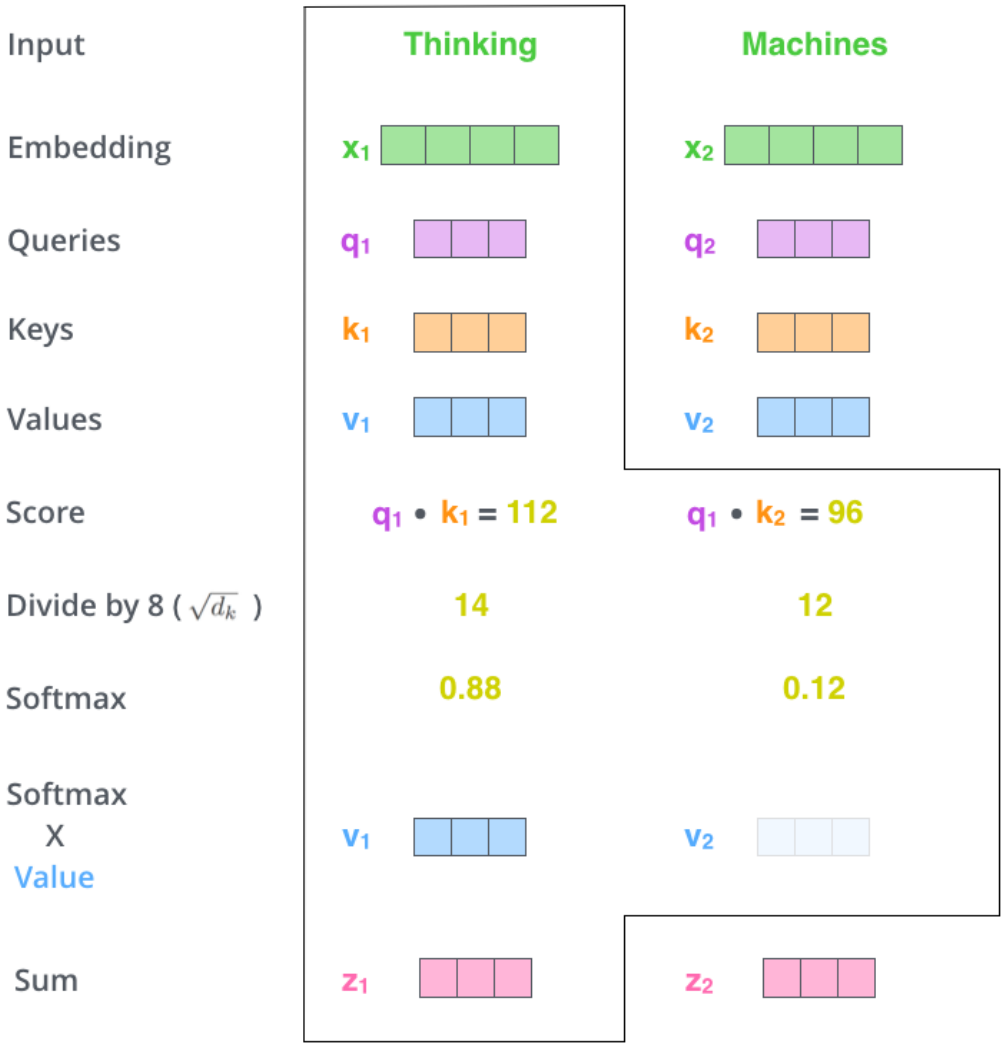
Step 2


Input
Thinking
Machines
Embedding
x1
x2
Queries
q1
q2
Keys
k1
k2
Values
v1
v2
Score
q1 • k1 = 112
q1 • k2 = 96

Step 3-4


Input
Thinking
Machines
Embedding
x1
x2
Queries
q1
q2
Keys
k1
k2
Values
v1
v2
Score
q1 • k1 = 112
q1 • k2 = 96
Divide by 8 (√dk)
14
12
Softmax
0.88
0.12

Step 5-6

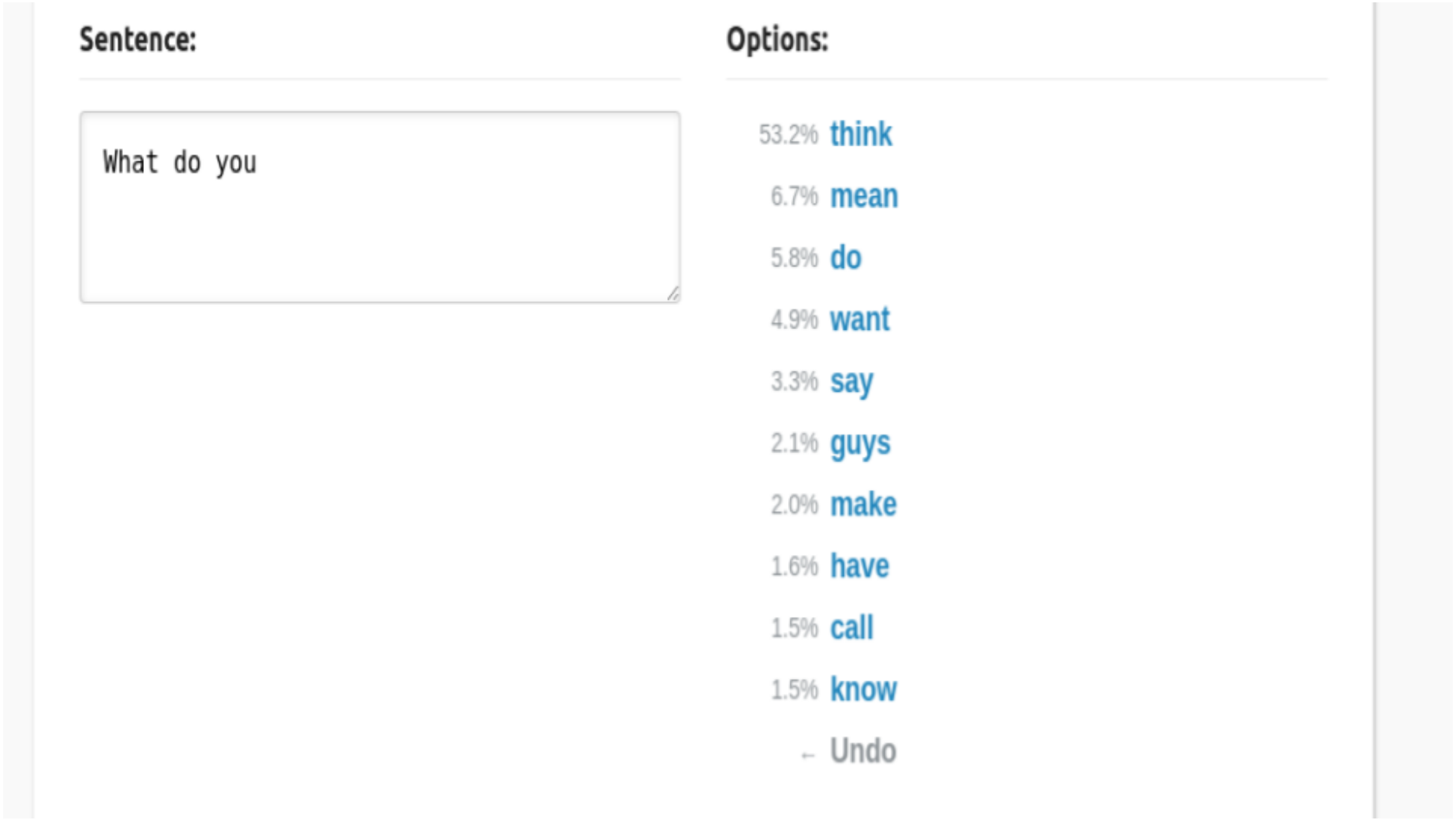

Input
Thinking
Machines
Embedding
x1
x2
Queries
q1
q2
Keys
k1
k2
Values
v1
v2
Score
q1 • k1 = 112
q1 • k2 = 96
Divide by 8 (√dk)
14
12
Softmax
0.88
0.12
Softmax X Value
v1
v2
Sum
z1
z2

GPT-2 can do fake news

Today, the United Nations has called for the immediate withdrawal of all nuclear weapons from the world

Generative Pretrained Transformer 2

Unidirectional Language model



Additional materials

https://transformer.huggingface.co
https://colab.research.google.com/drive/1esbpDOorf7DQJV8GXWON24
EQrSKOit