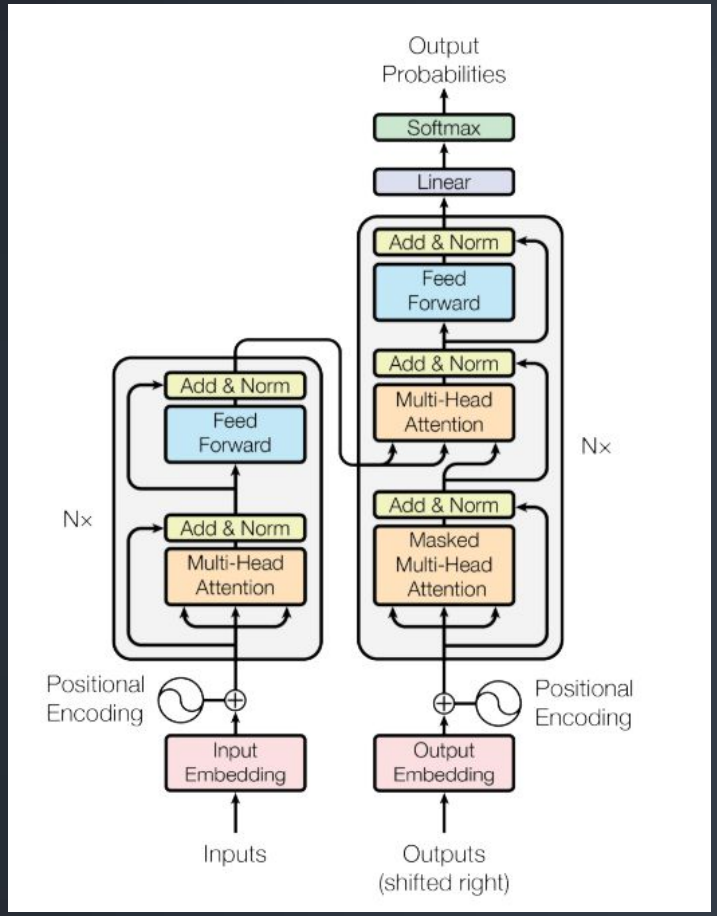
# Attention is all you need

*(Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin)*

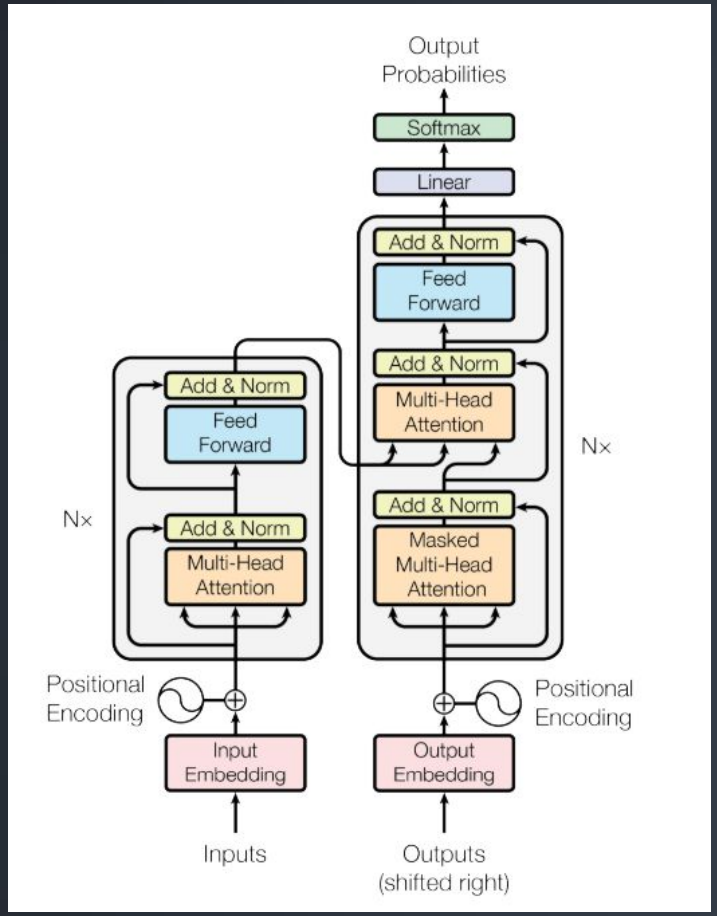
## Aprenentatge Automàtic

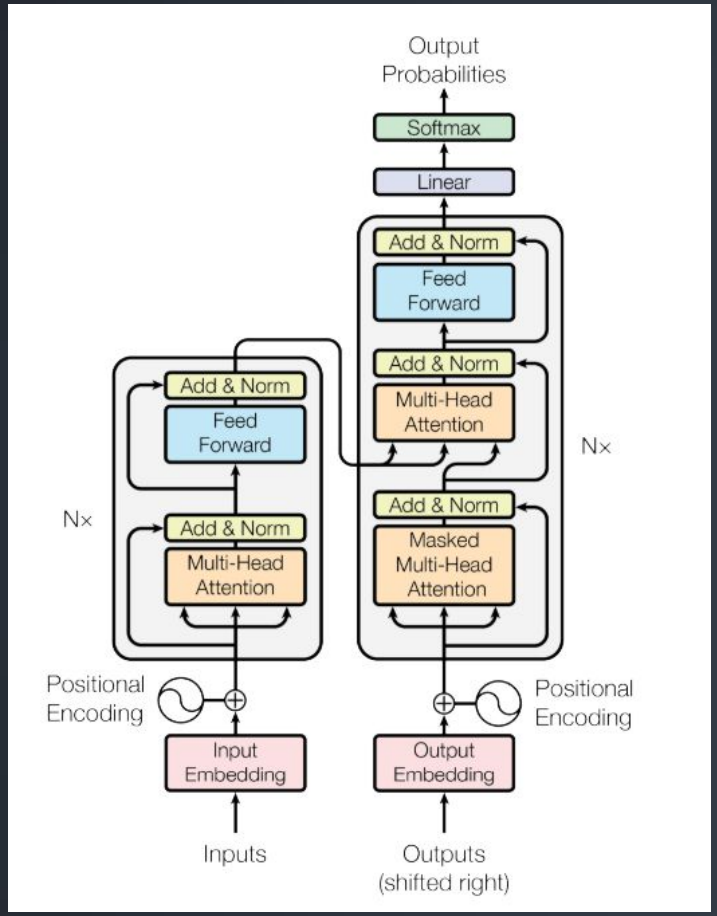
Marc Bracons

Juny de 2024


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)

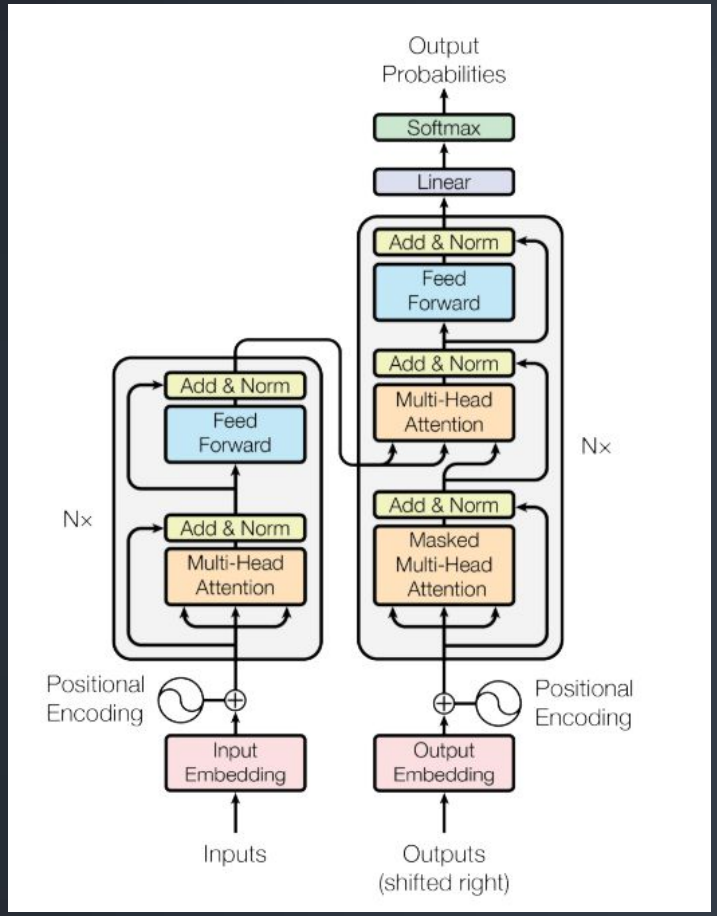
# Xarxa neuronal






## Xarxa neuronal

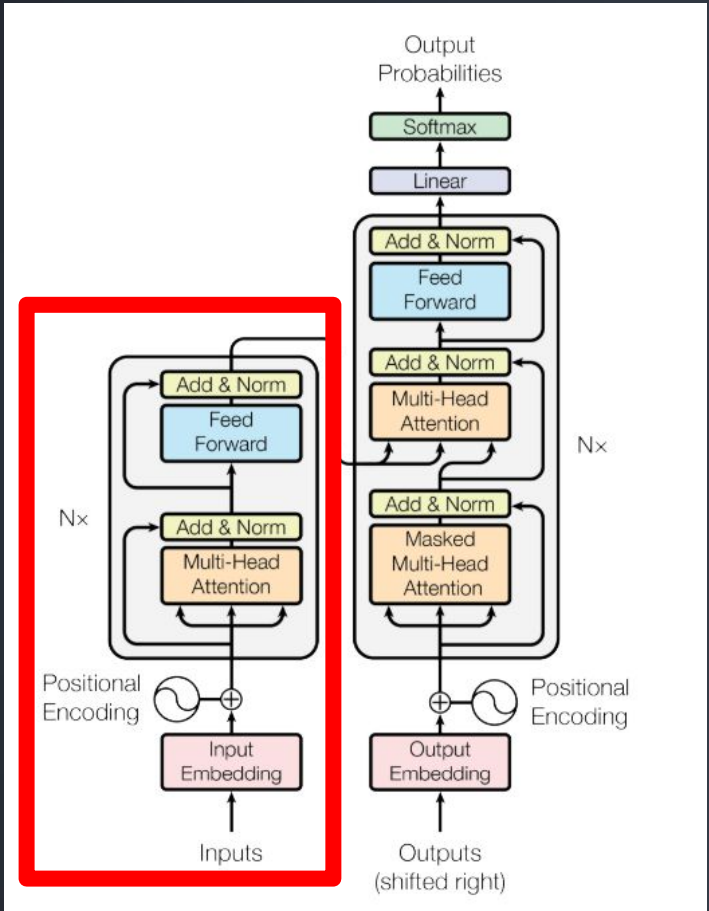
## Context i relacions




**Xarxa neuronal**

**Context i relacions**

**Aprenentatge supervisat**

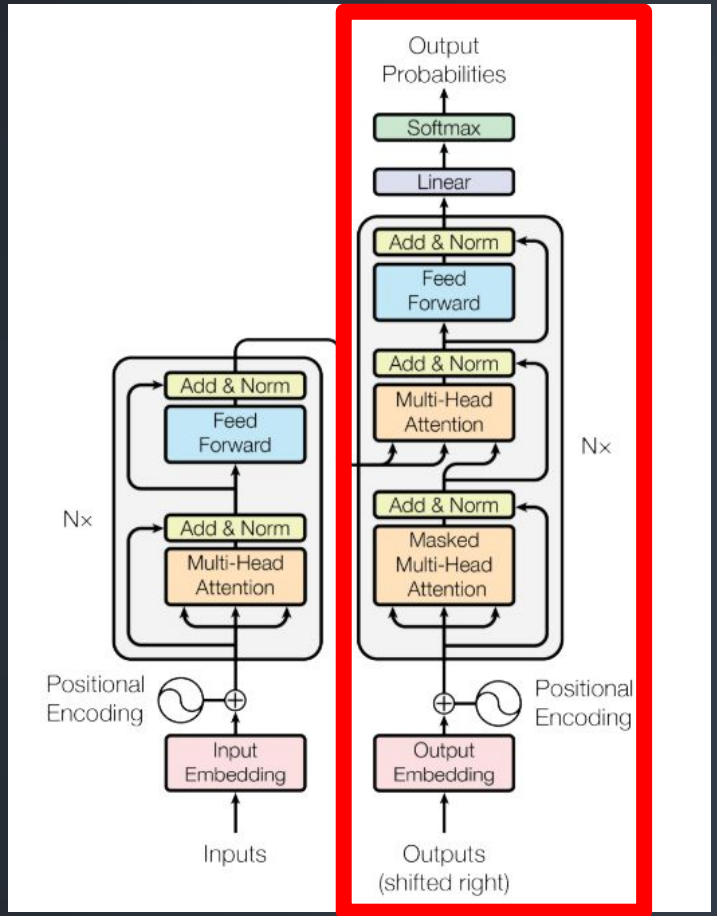



**Xarxa neuronal**

**Context i relacions**

**Aprenentatge supervisat**

**Encoder-Decoder**

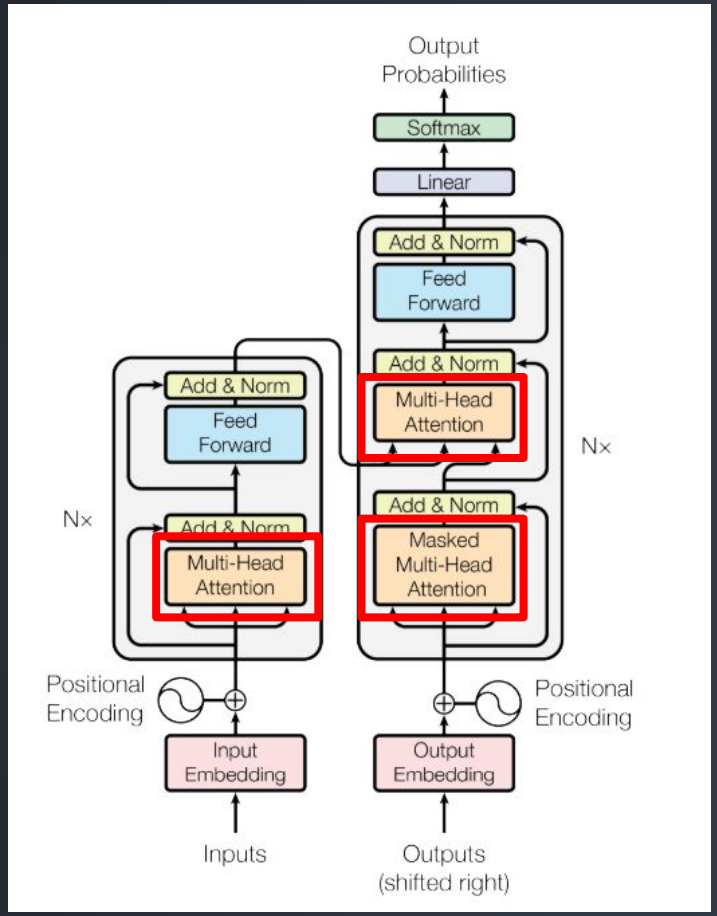



**Xarxa neuronal**

**Context i relacions**

**Aprenentatge supervisat**

**Encoder-Decoder**




**Xarxa neuronal**

**Context i relacions**

**Aprenentatge supervisat**

**Encoder-Decoder**

**Mecanisme d’atenció**

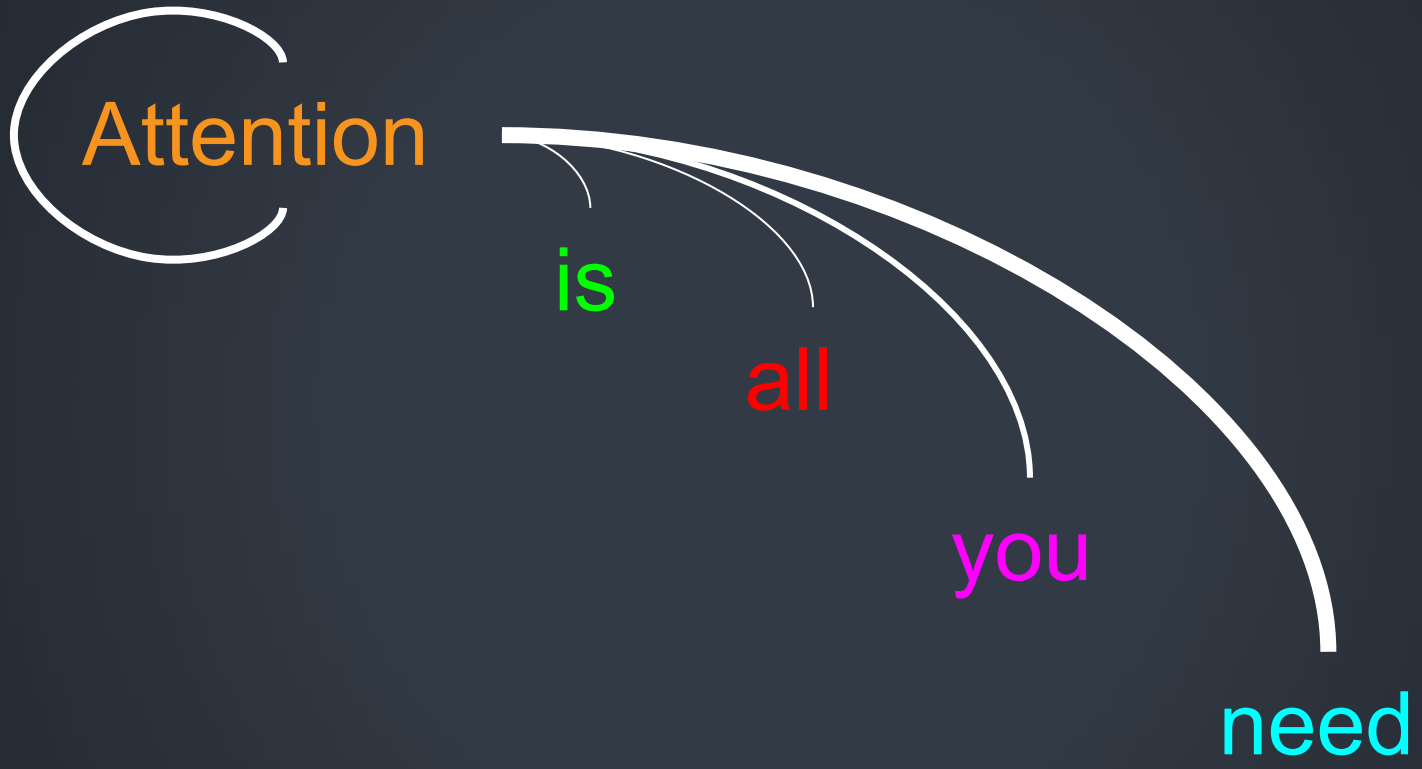
Attention

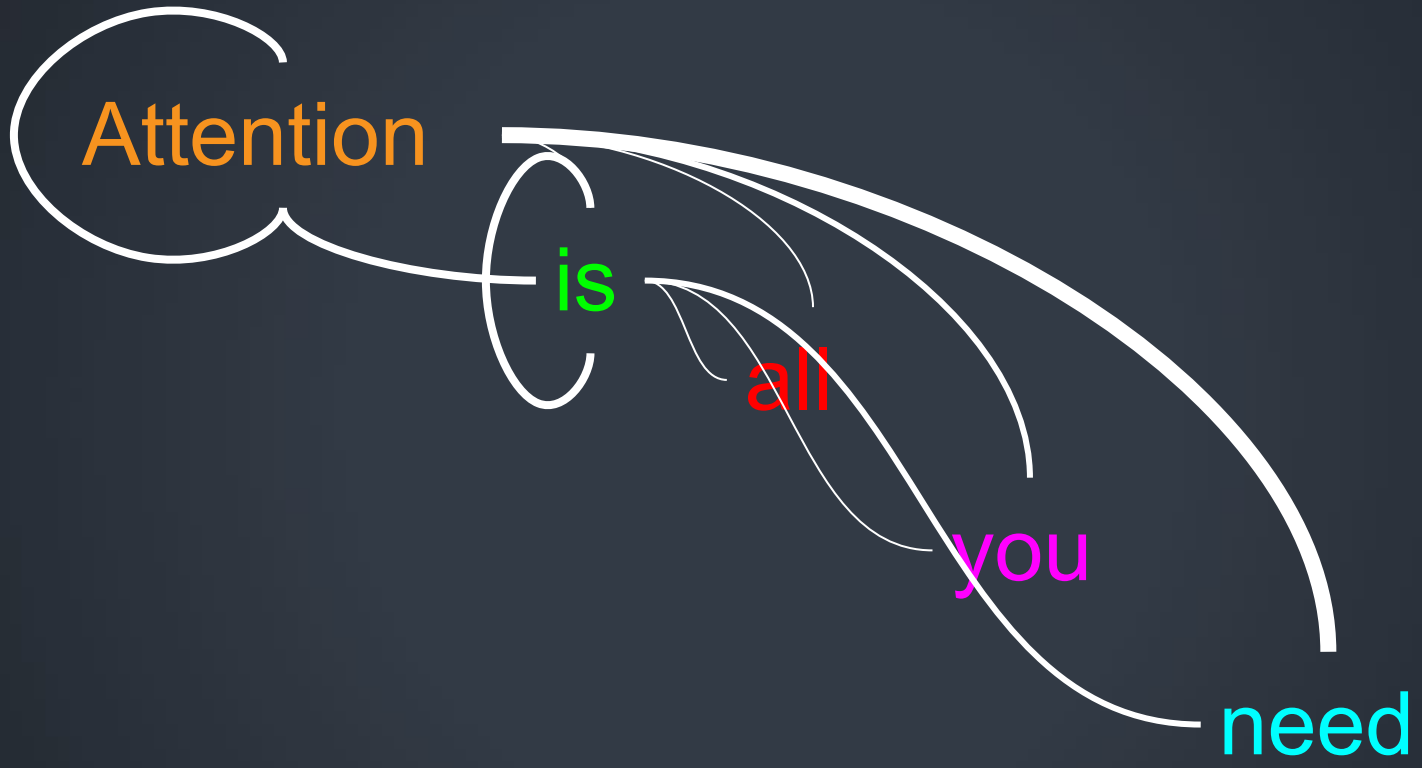
is

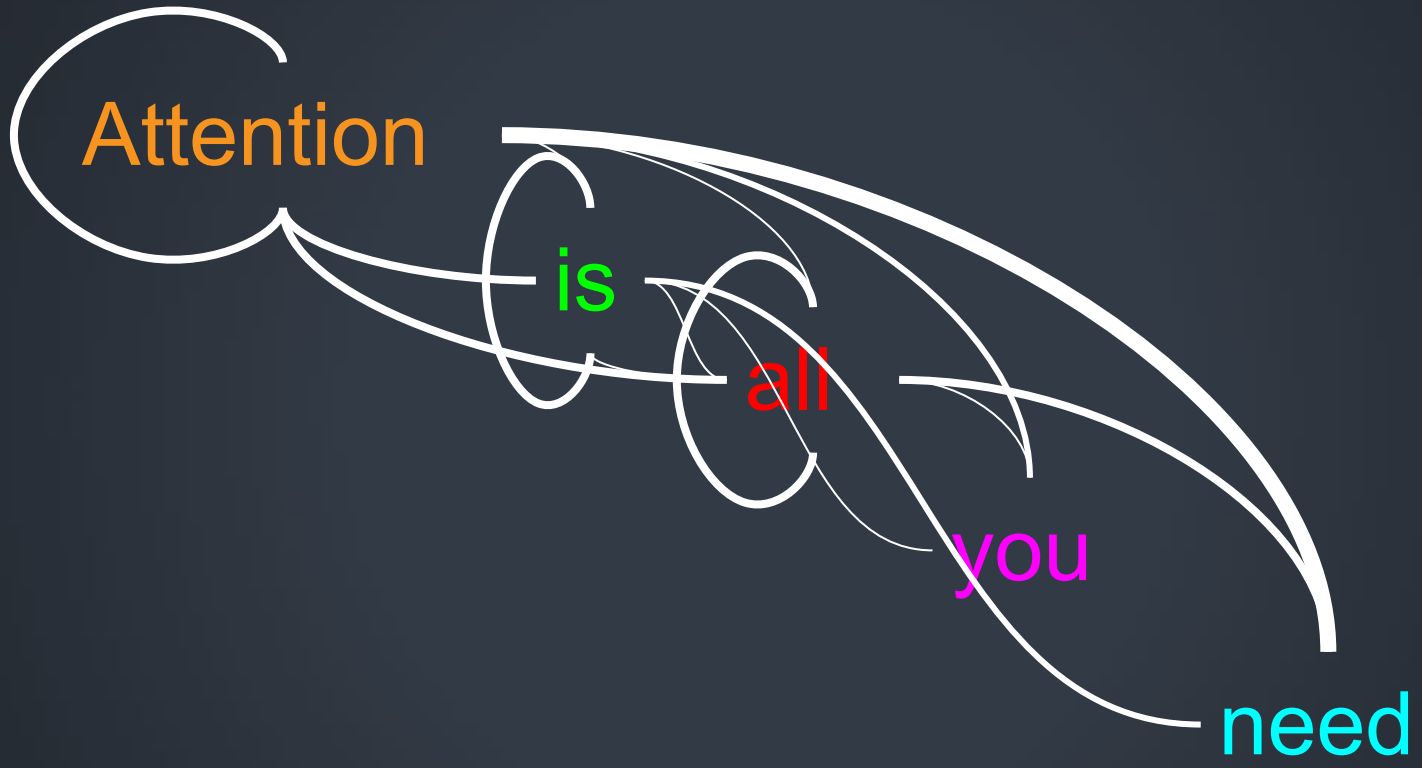
all

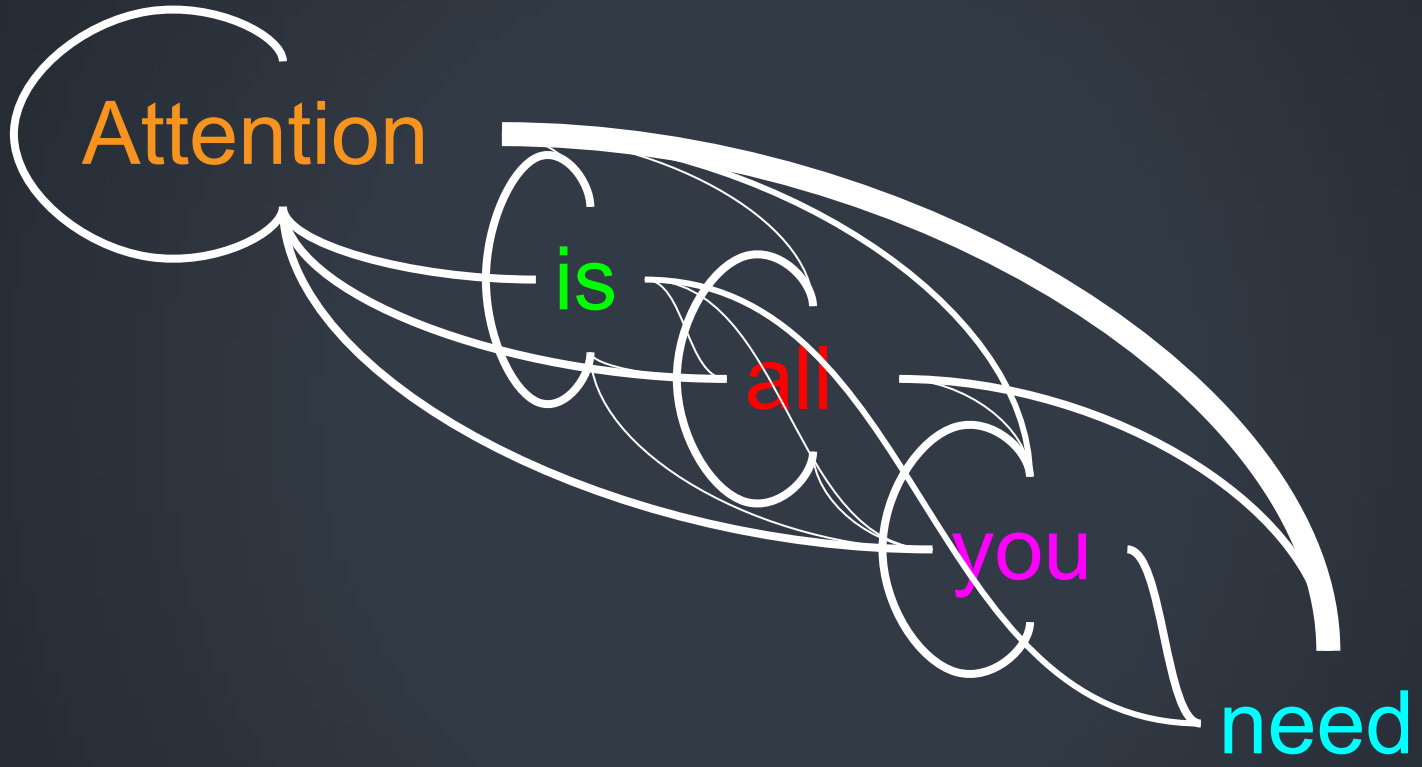
you

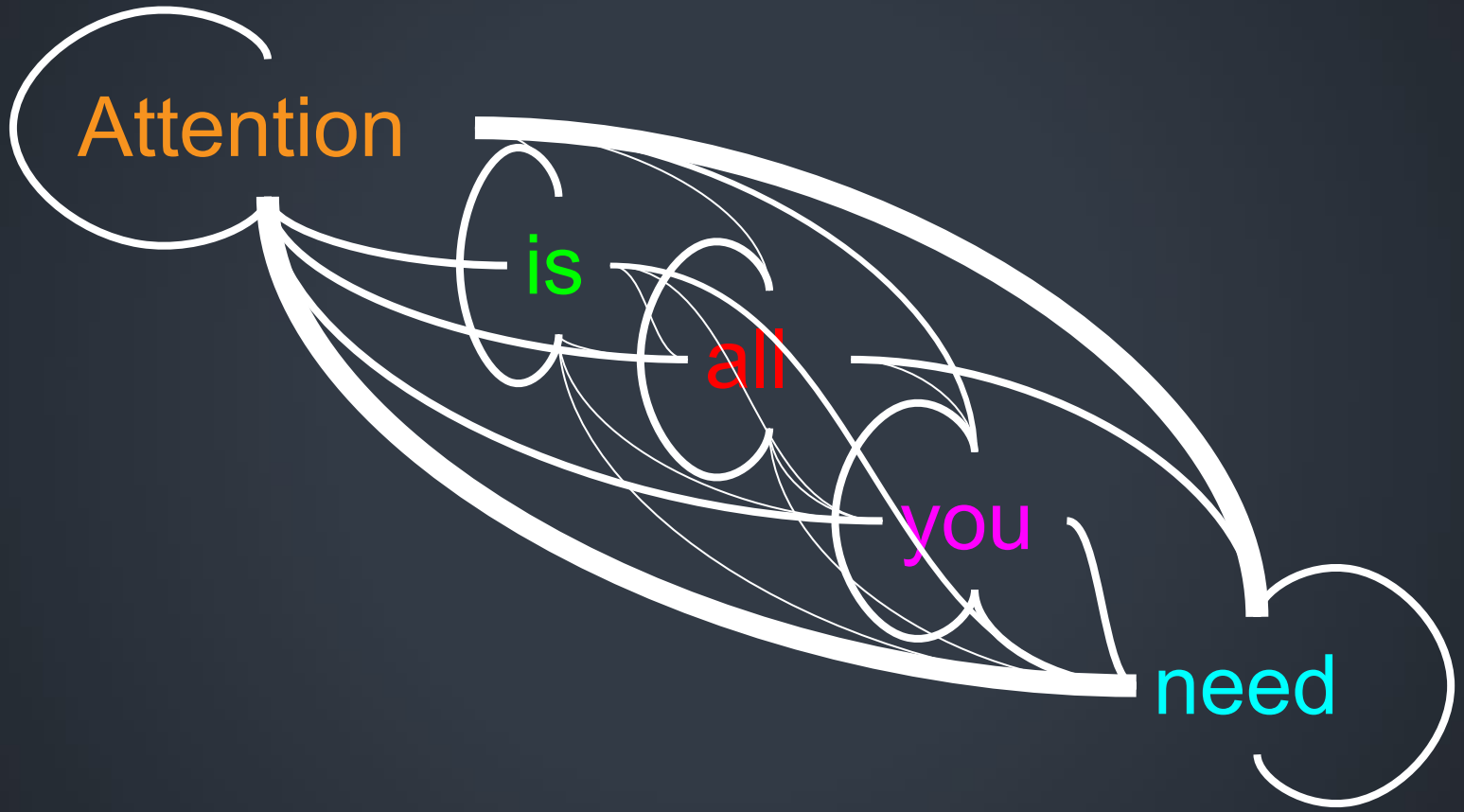
need


Attention
is
all
you
need


Attention
is
all
you
need


Attention
is
all
you
need


Attention
is
all
you
need


Attention
is
all
you
need

# Com es calculen aquestes relacions?

Attention

is

all

you

need

Query

Attention

is

all

you

need

Query

Attention

is

all

you

need

Query

Attention

is

all

you

need

Attention

is

all

you

need

Query

Key

Attention

is

all

you

need

Attention

is

all

you

need

Query

Key

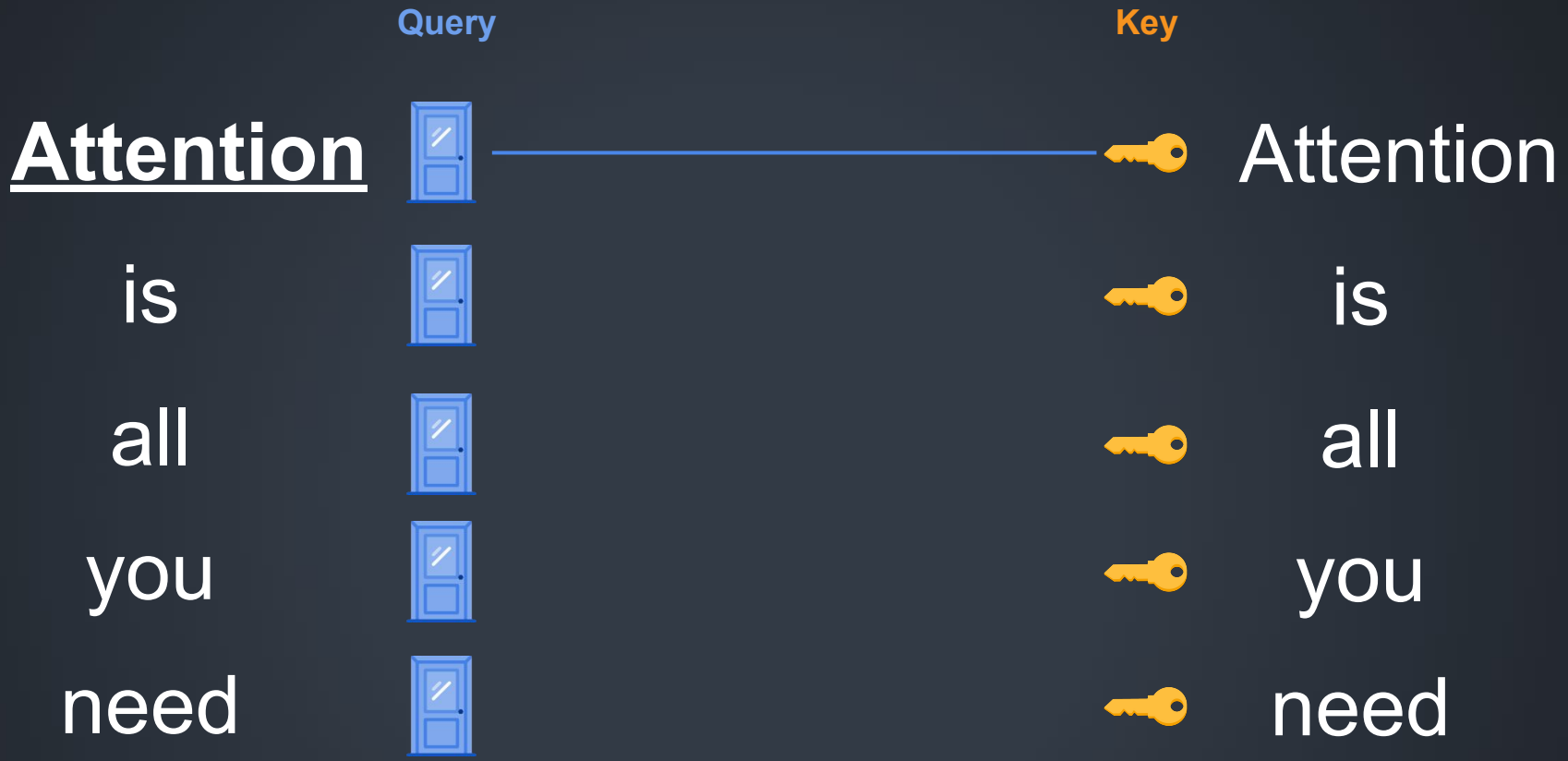
Attention Attention

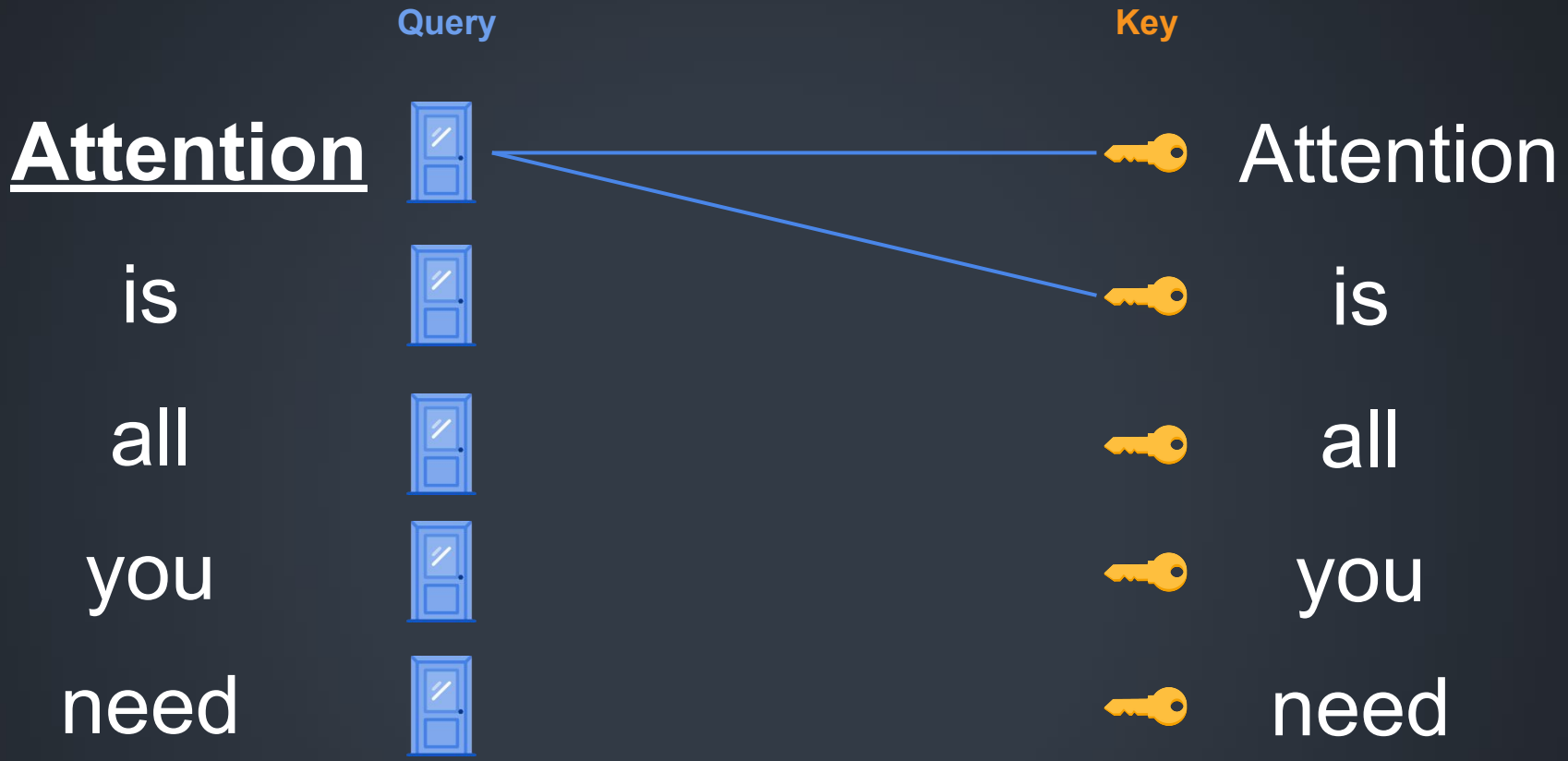
is is

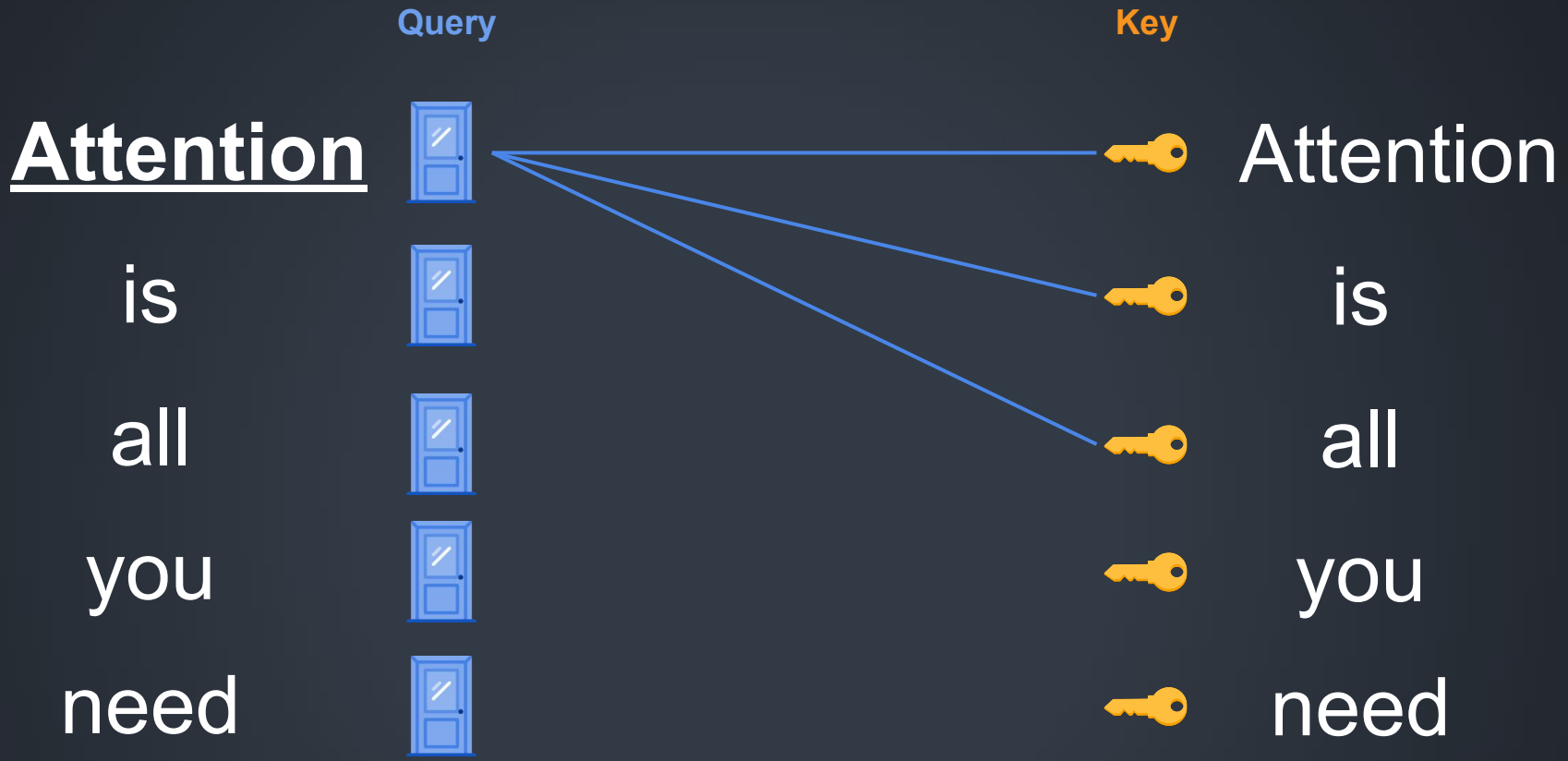
all all

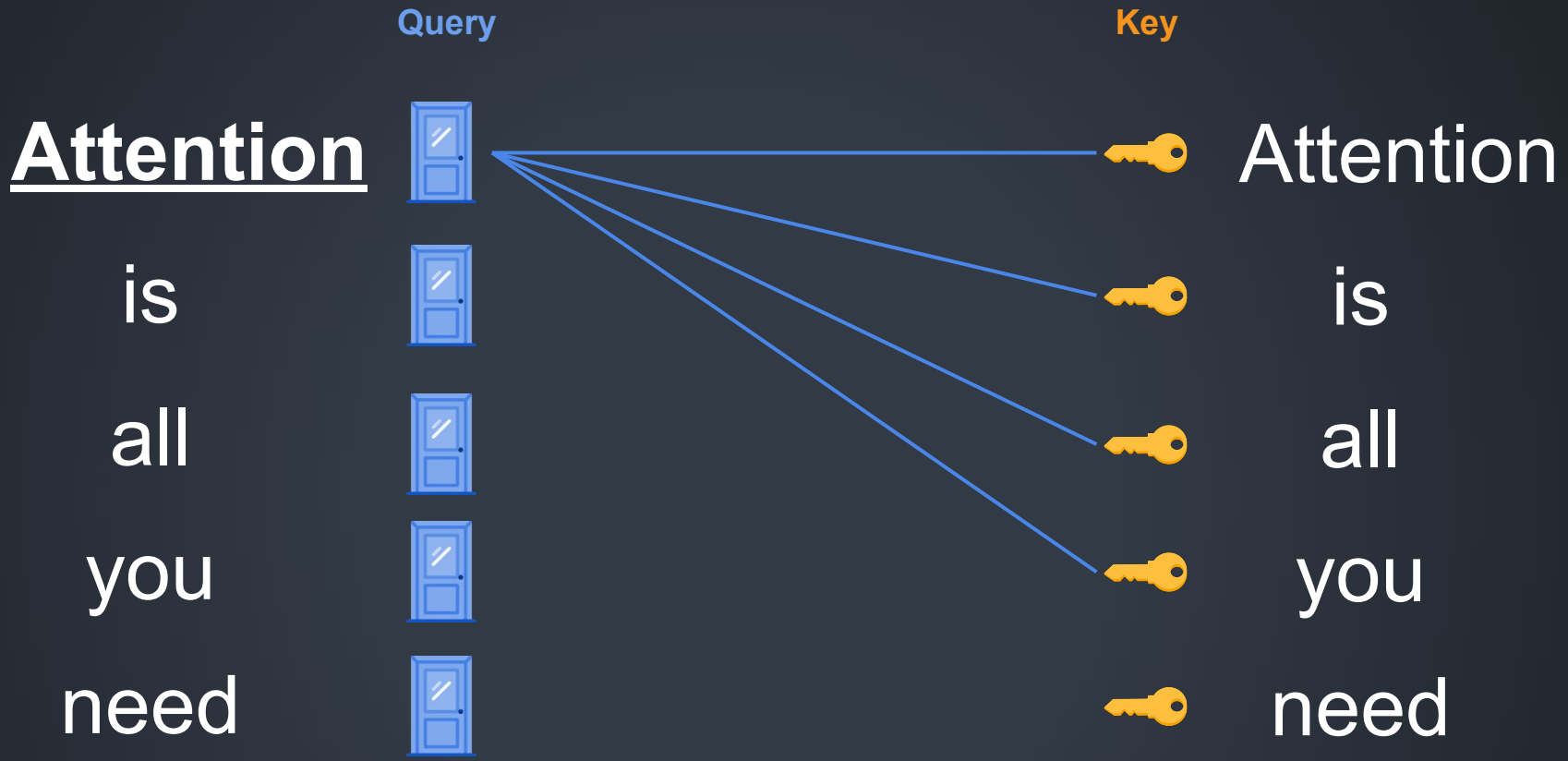
you you

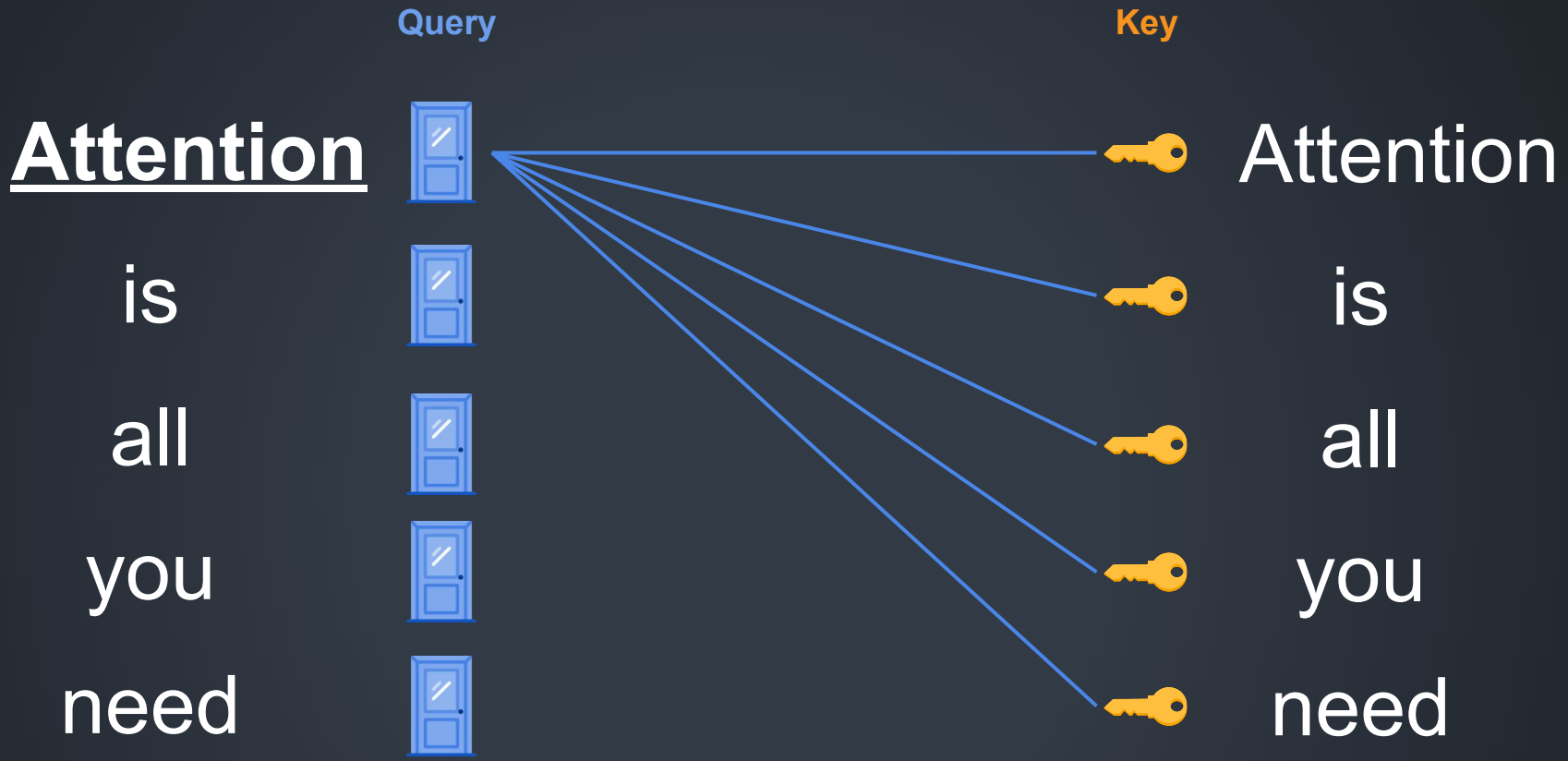
need need


Query
Key
Attention
is
all
you
need
Attention
is
all
you
need


Query
Key
Attention
is
all
you
need
Attention
is
all
you
need


Query
Key
Attention
is
all
you
need
Attention
is
all
you
need


Query
Key
Attention
is
all
you
need
Attention
is
all
you
need


Query
Key
Attention
is
all
you
need
Attention
is
all
you
need

Scores
Attention
is
all
you
need
0.25
0.07
0.09
0.32
0.27
Attention
is
all
you
need

Values

Attention

is

all

you

need

Values

Values

Scores

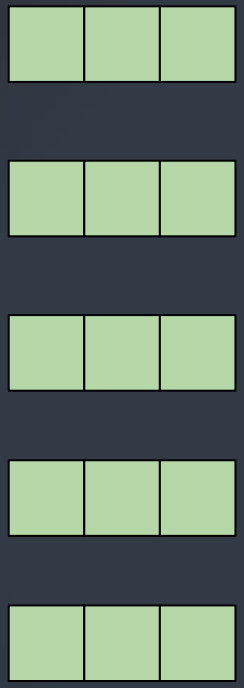
Attention

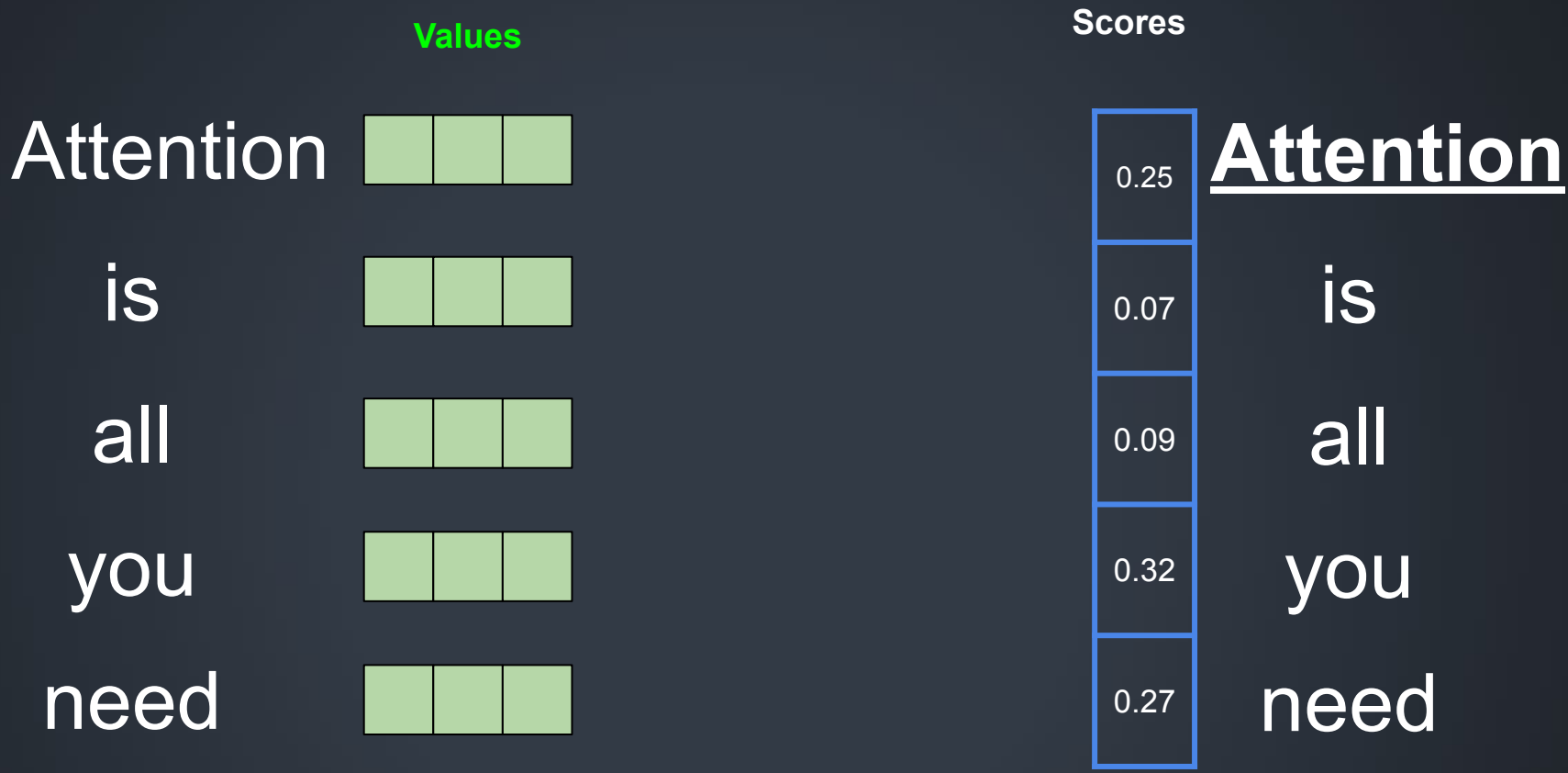
is

all

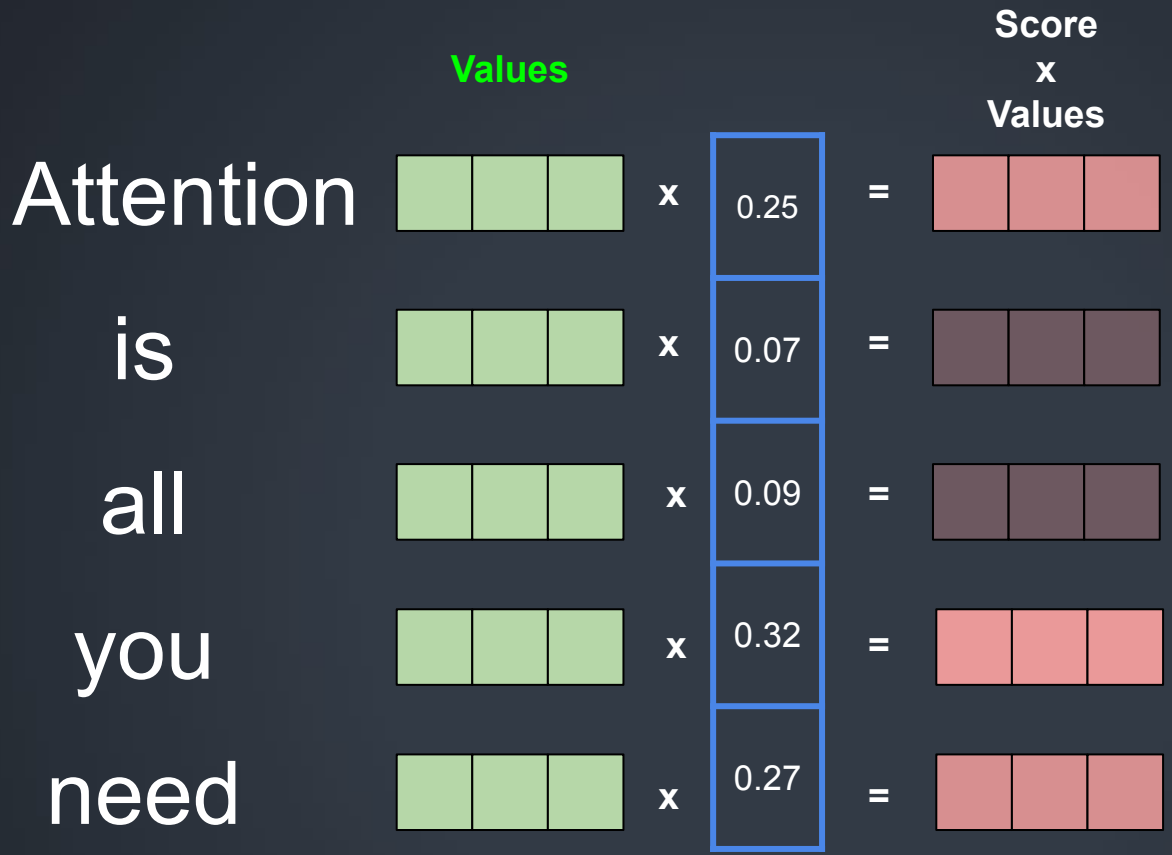
you

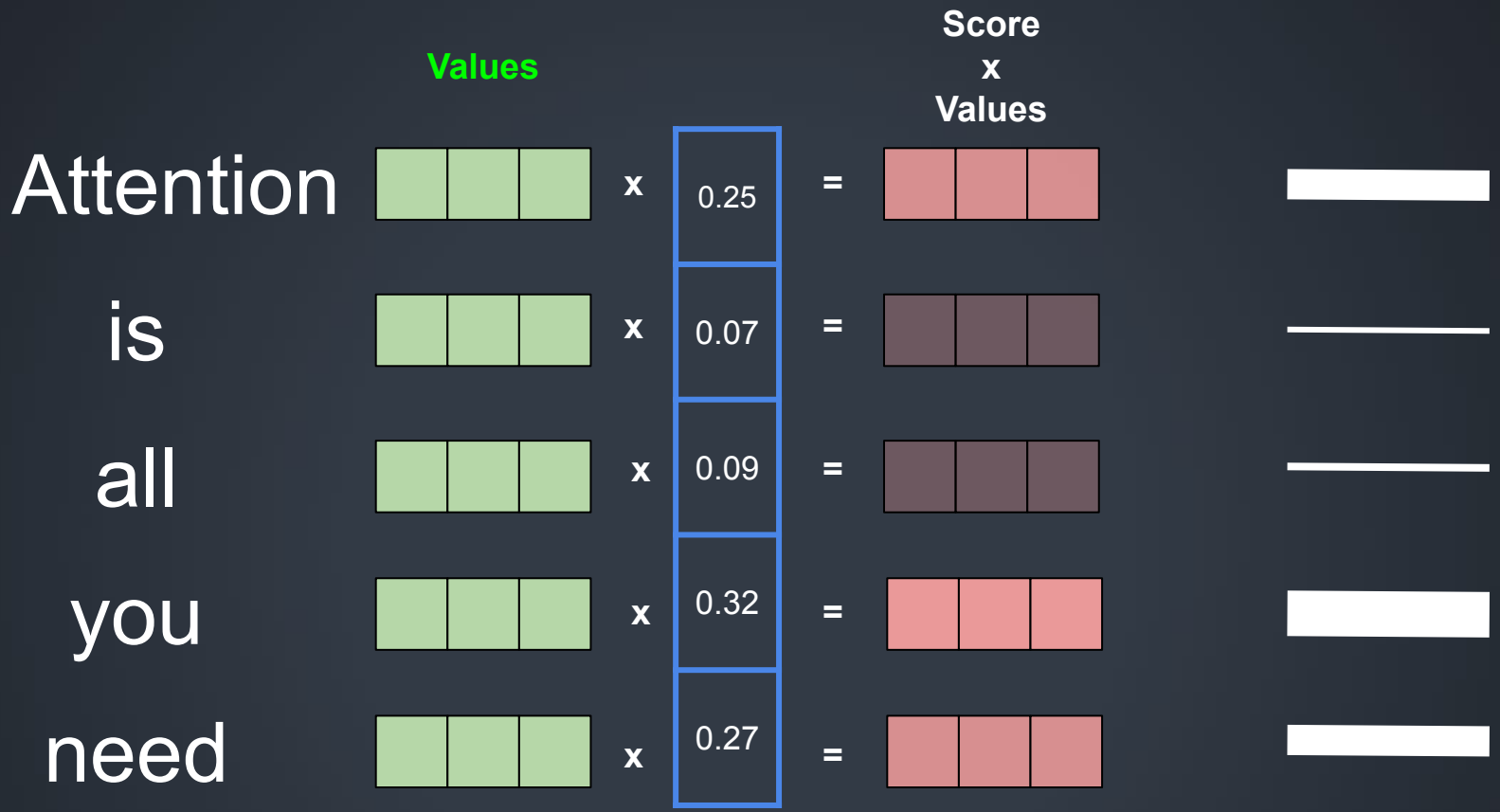
need




Values
Scores
Attention
is
all
you
need
0.25
0.07
0.09
0.32
0.27
Attention
is
all
you
need

Values
Attention
x
0.25
is
x
0.07
all
x
0.09
you
x
0.32
need
x
0.27


Values
Score
x
Values
Attention
is
all
you
need
x
0.25
0.07
0.09
0.32
0.27
=


Values
Score
x
Values
Attention
is
all
you
need
x
0.25
0.07
0.09
0.32
0.27
=

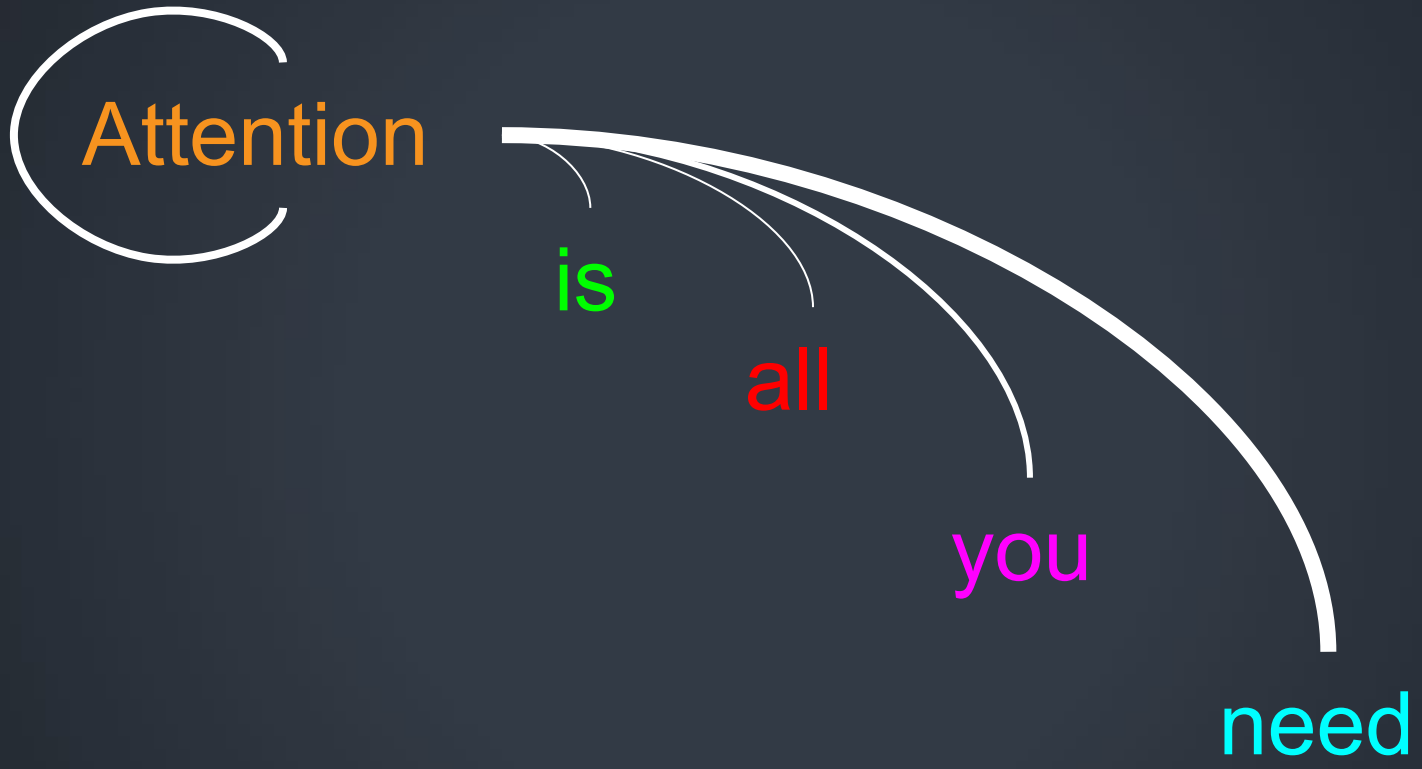
Attention

is

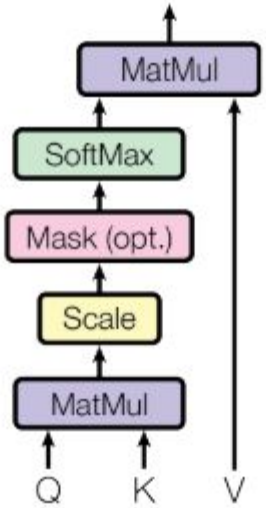
all

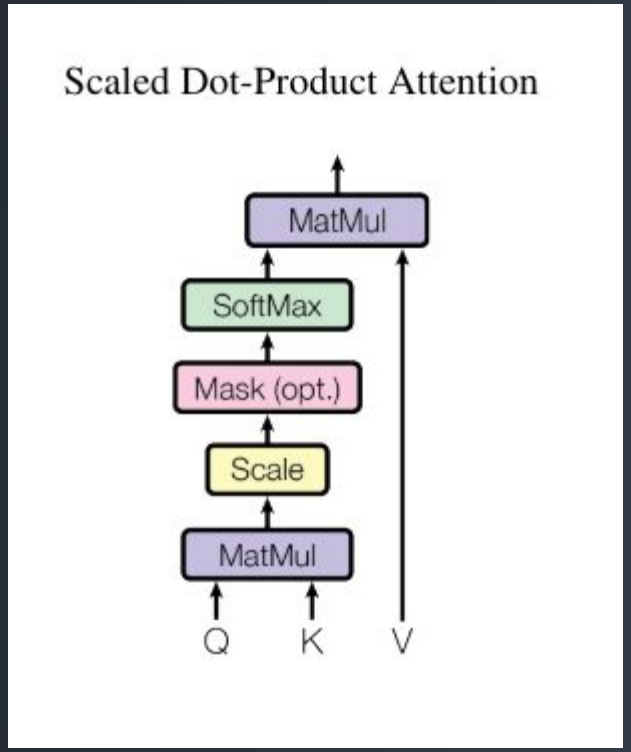
you

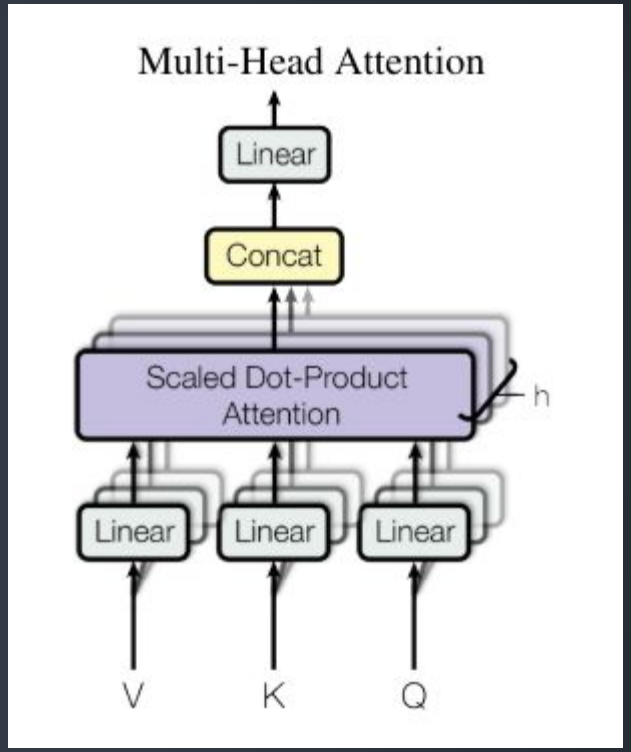
need


Attention
is
all
you
need

Scaled Dot-Product Attention

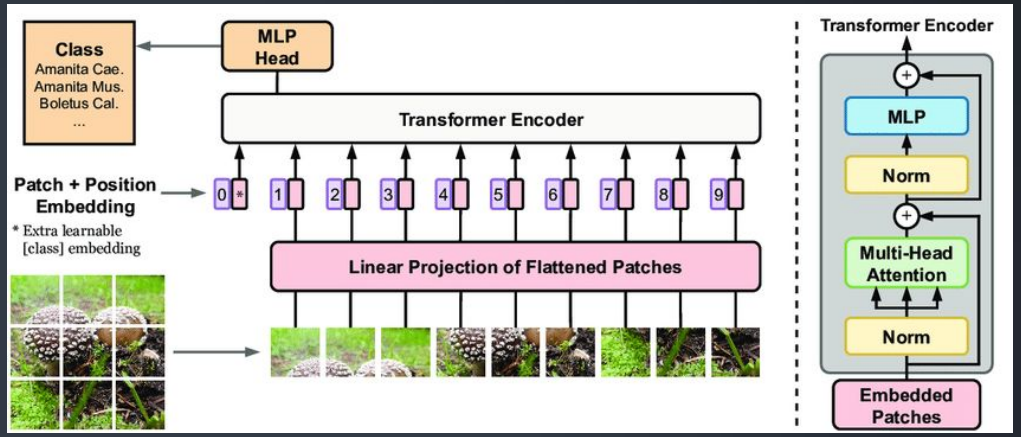



Scaled Dot-Product Attention
MatMul
SoftMax
Mask (opt.)
Scale
MatMul
Q
K
V

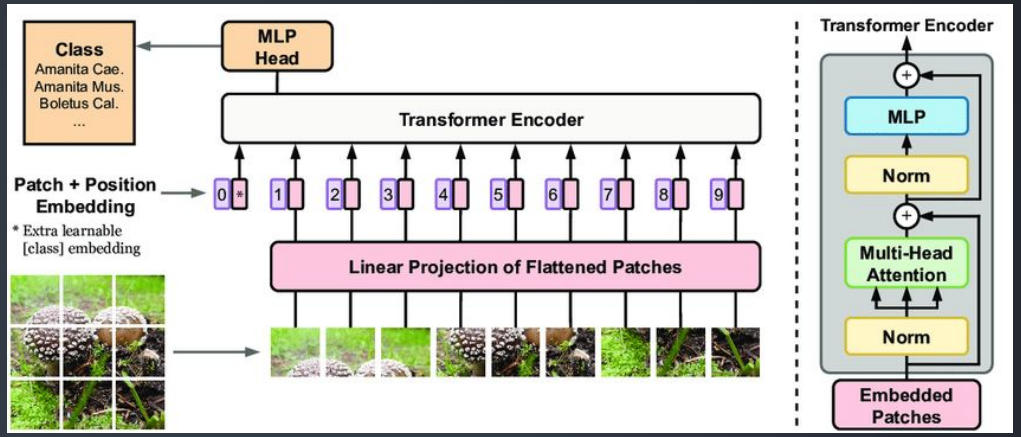

Multi-Head Attention
Linear
Concat
Scaled Dot-Product Attention
h
Linear
Linear
Linear
V
K
Q

NLP

NLP


Class
Amanita Cae.
Amanita Mus.
Boletus Cal.
...
MLP Head
Transformer Encoder
Patch + Position Embedding
* Extra learnable [class] embedding
0
*
1
2
3
4
5
6
7
8
9
Linear Projection of Flattened Patches
Transformer Encoder
MLP
Norm
Multi-Head Attention
Norm
Embedded Patches

NLP



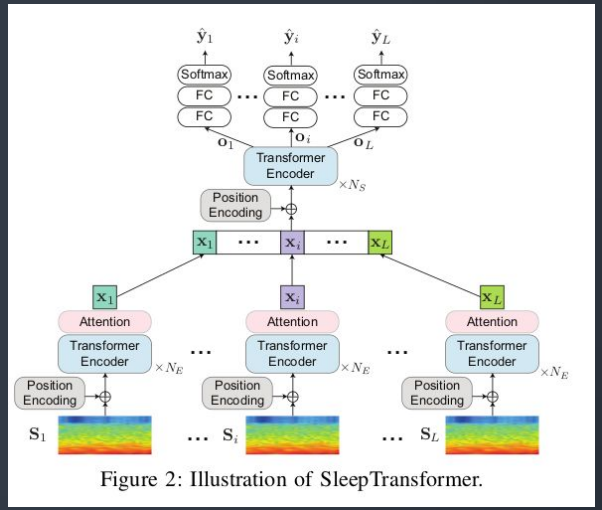




Figure 2: Illustration of SleepTransformer.