

Natural Language Processing

Transformers, BERT, & GPT-2

Dirk Hovy

dirk.hovy@unibocconi.it

 @dirk_hovy



noooooooooOOOOoo! you can't
use just layers of multi-head
self-attention to eliminate the
need for gated recurrence nooOOOOOOOOOO



haha transformer
go brrrrrrrrrr

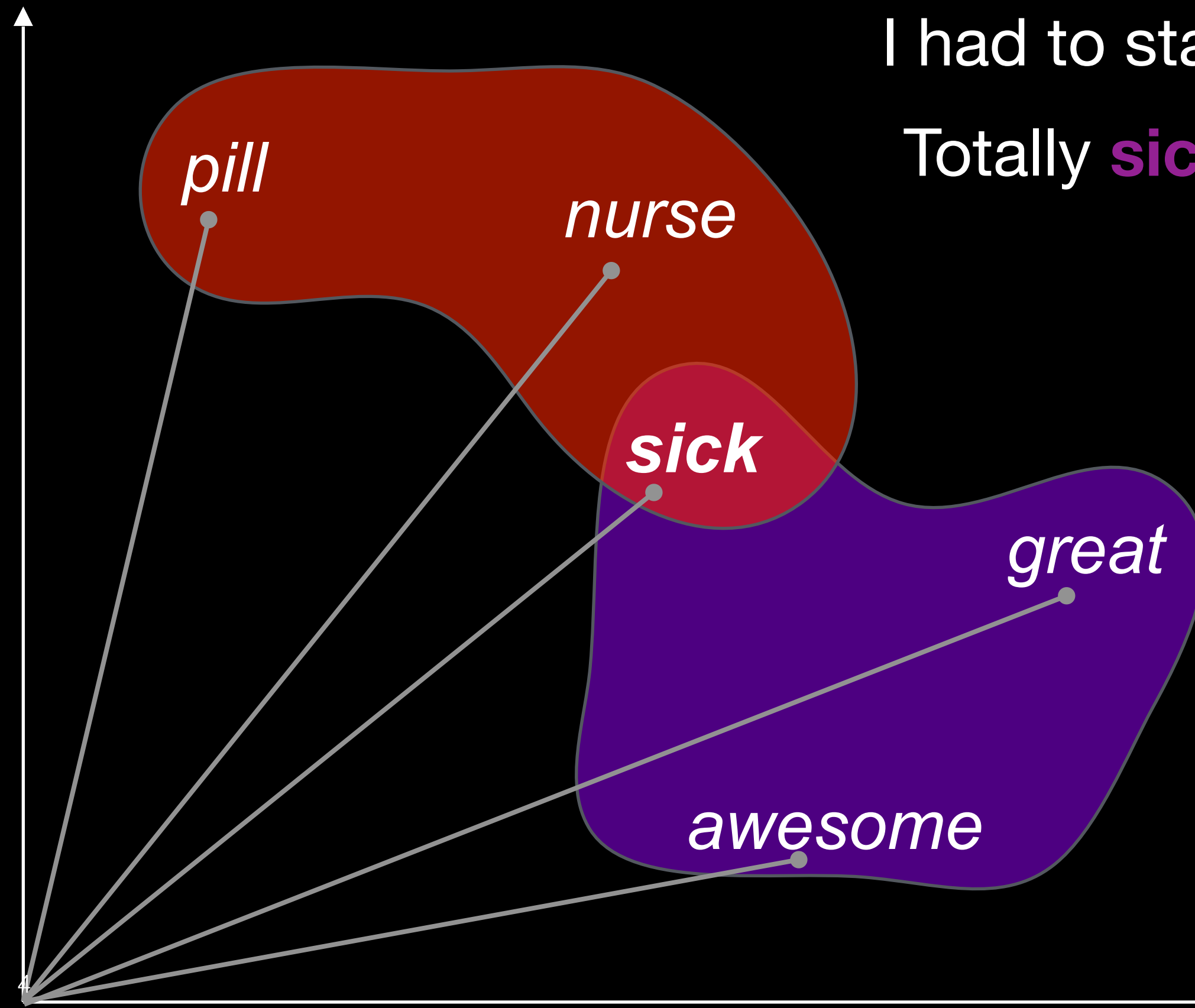
Goals for Today

- Understand the **Transformer architecture**
- Learn about its use in **contextualized embeddings (BERT)**
- Learn about text generation with **GPT-2 and GPT-3**

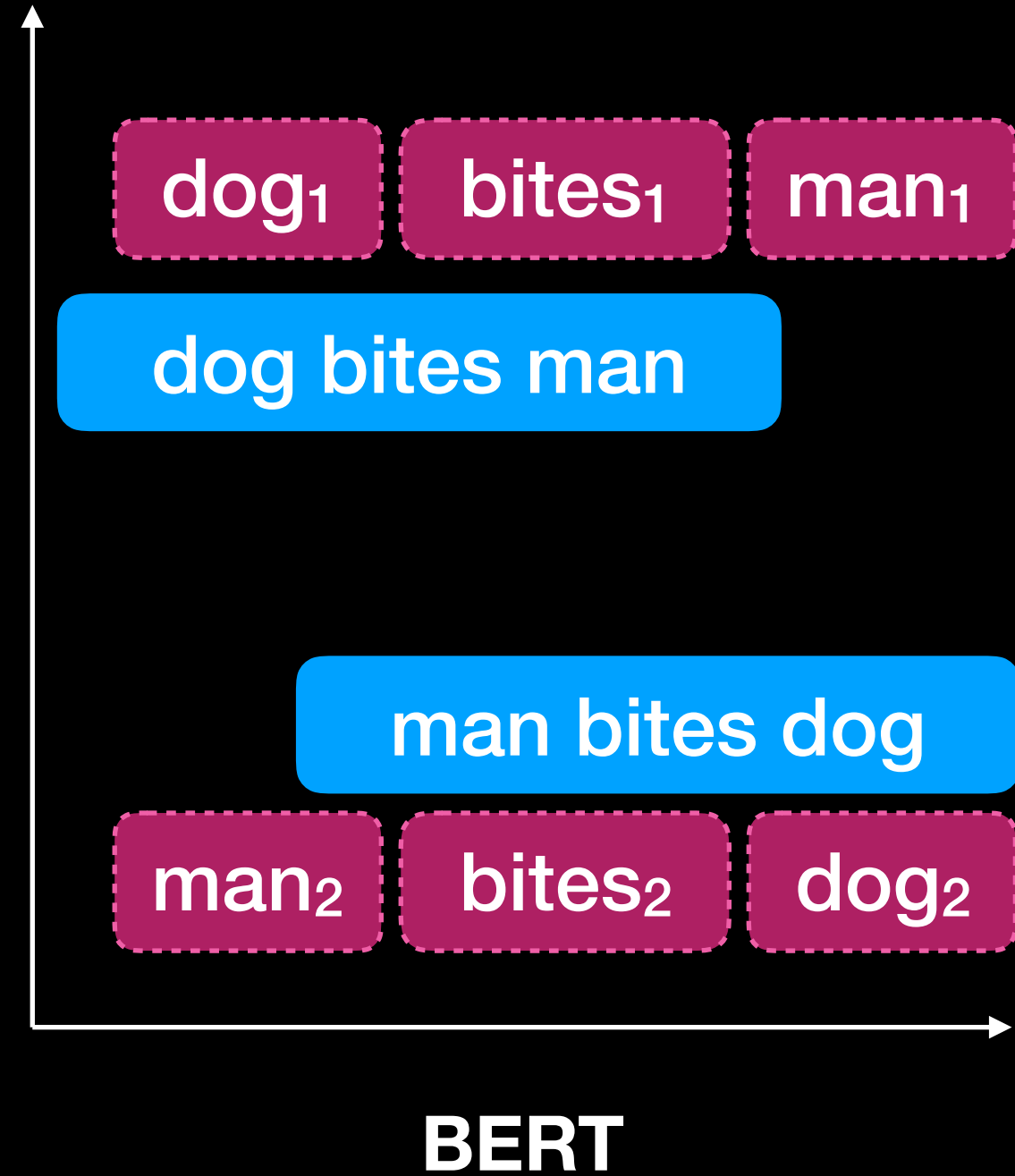
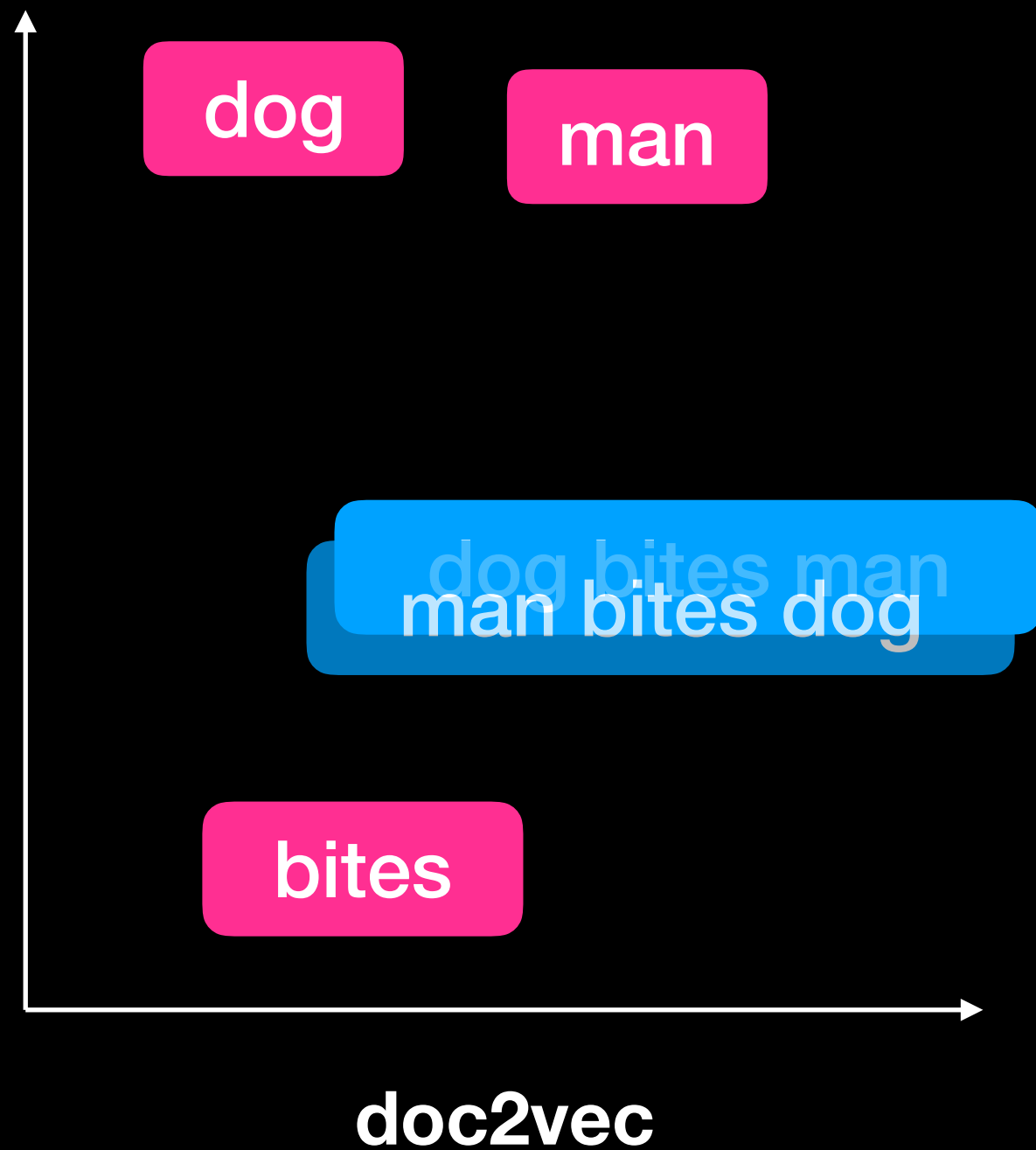
Contextual Representations

I had to stay home **sick**

Totally **sick** move, bro

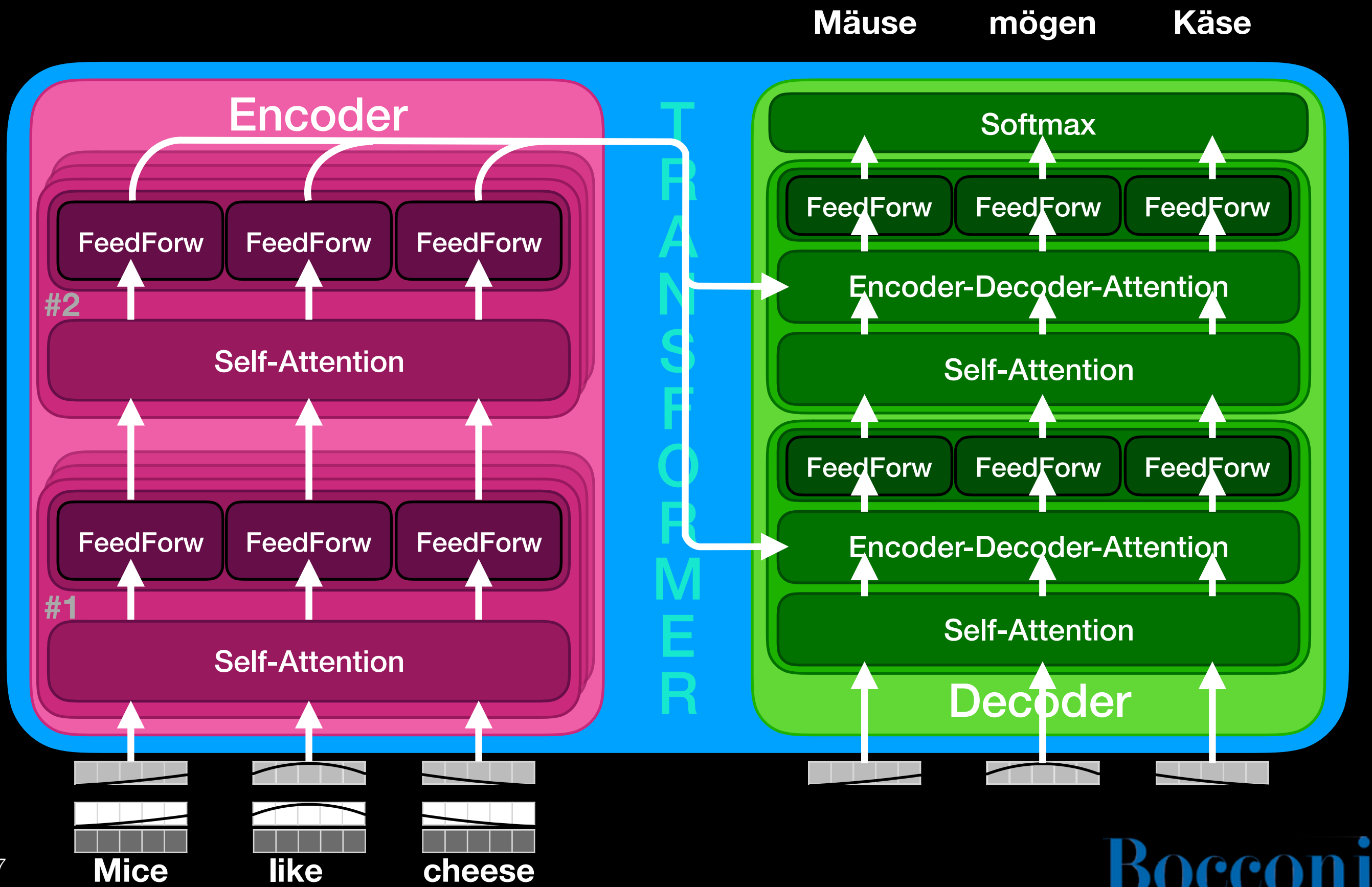


Problem



The Transformer

The Transformer



Encoder and Decoder

Mäuse

mögen

Käse

Encoder

T
R
A
N
S
F
O
R
M
E
R

Decoder



Mice



like



cheese

Encoder Blocks

Mäuse

mögen

Käse

Encoder

FeedForw

FeedForw

FeedForw

Self-Attention

T
R
A
N
S
F
O
R
M
E
R

Decoder



Mice



like



cheese

Multi-head Encoding

Mäuse

mögen

Käse

Encoder

T
R
A
N
S
F
O
R
M
E
R

Decoder

FeedForw

FeedForw

FeedForw

#2

Self-Attention

FeedForw

FeedForw

FeedForw

#1

Self-Attention



Mice



like



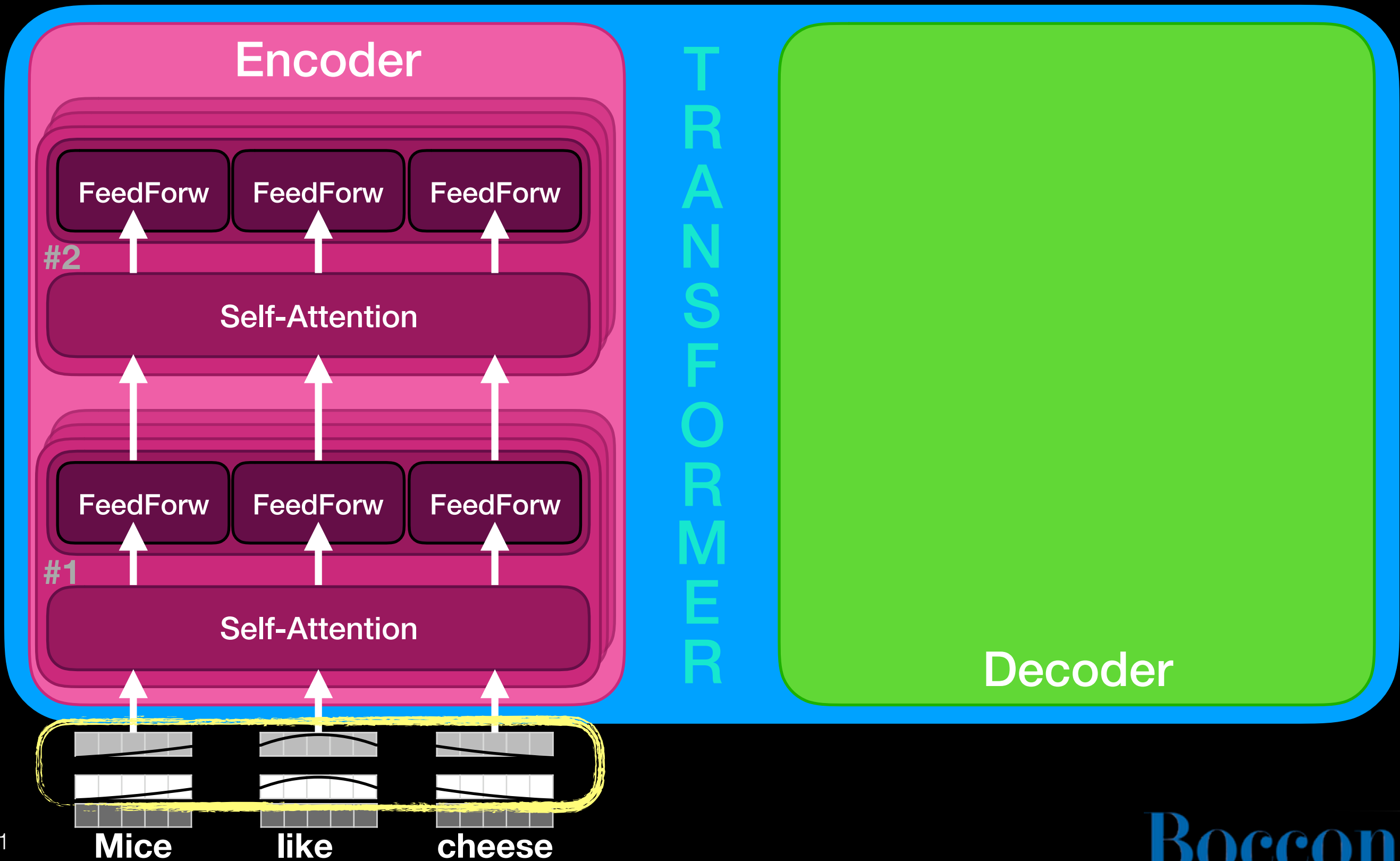
cheese

Positional Encoding

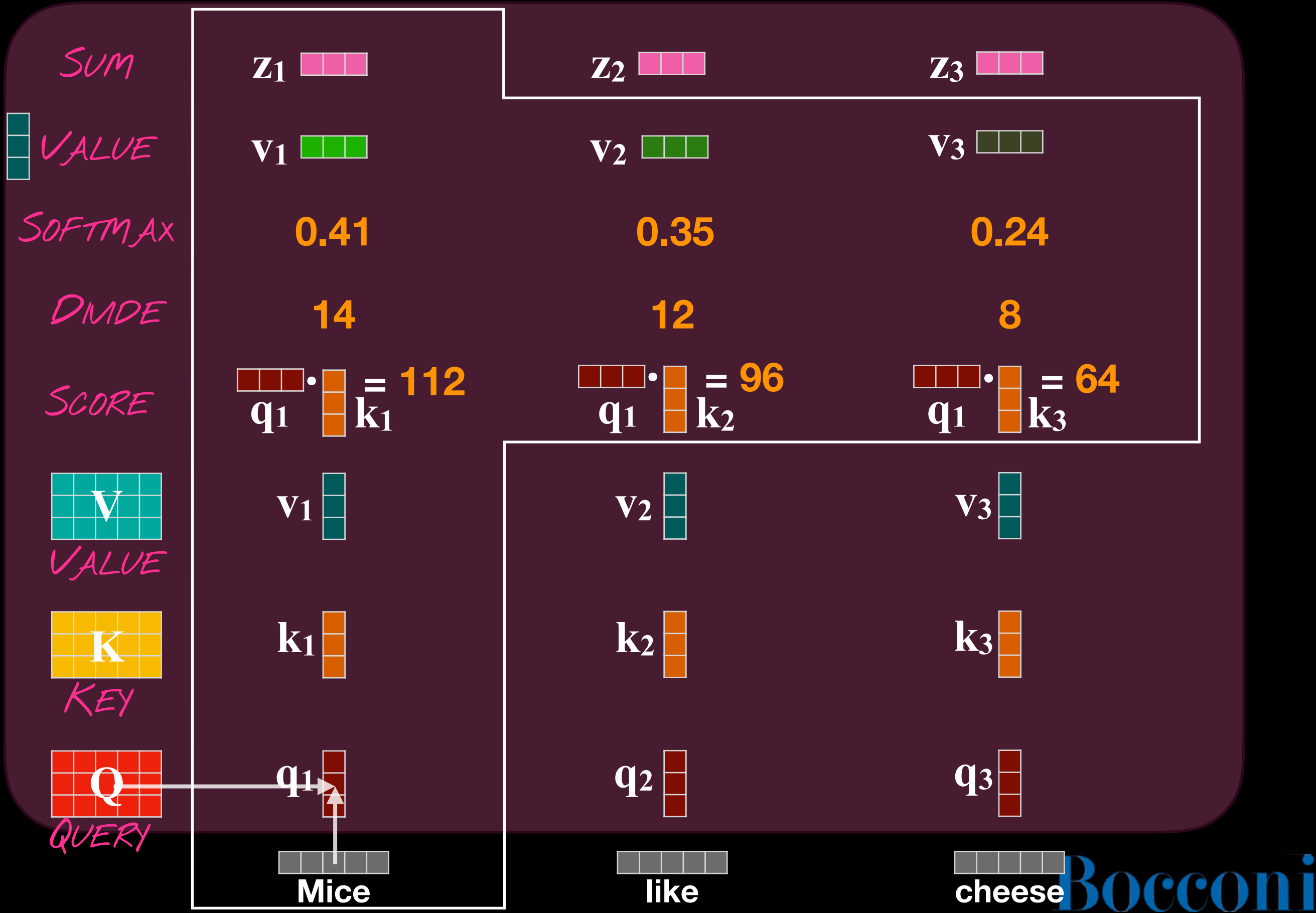
Mäuse

mögen

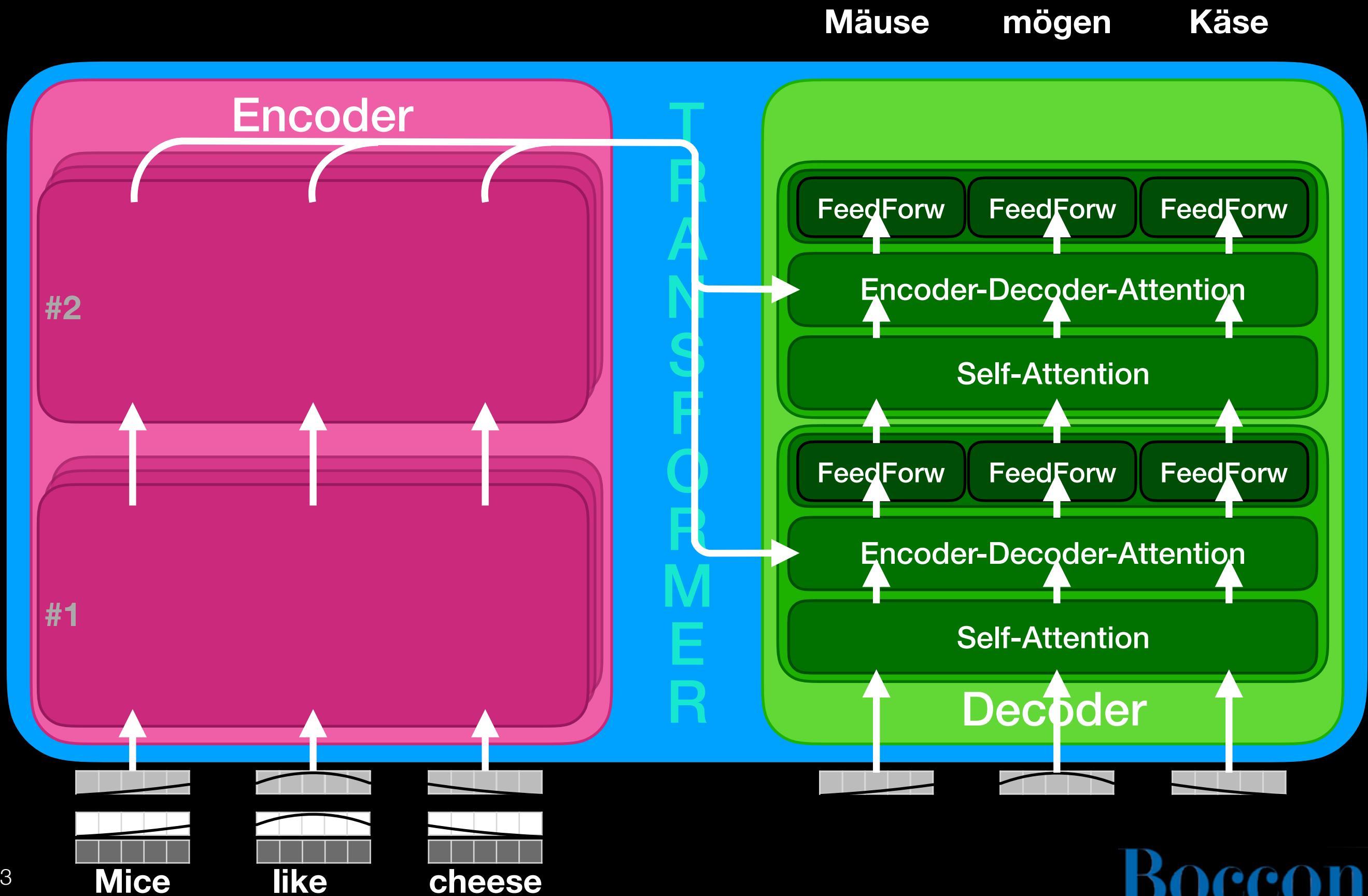
Käse



Self-Attention



Decoder Blocks



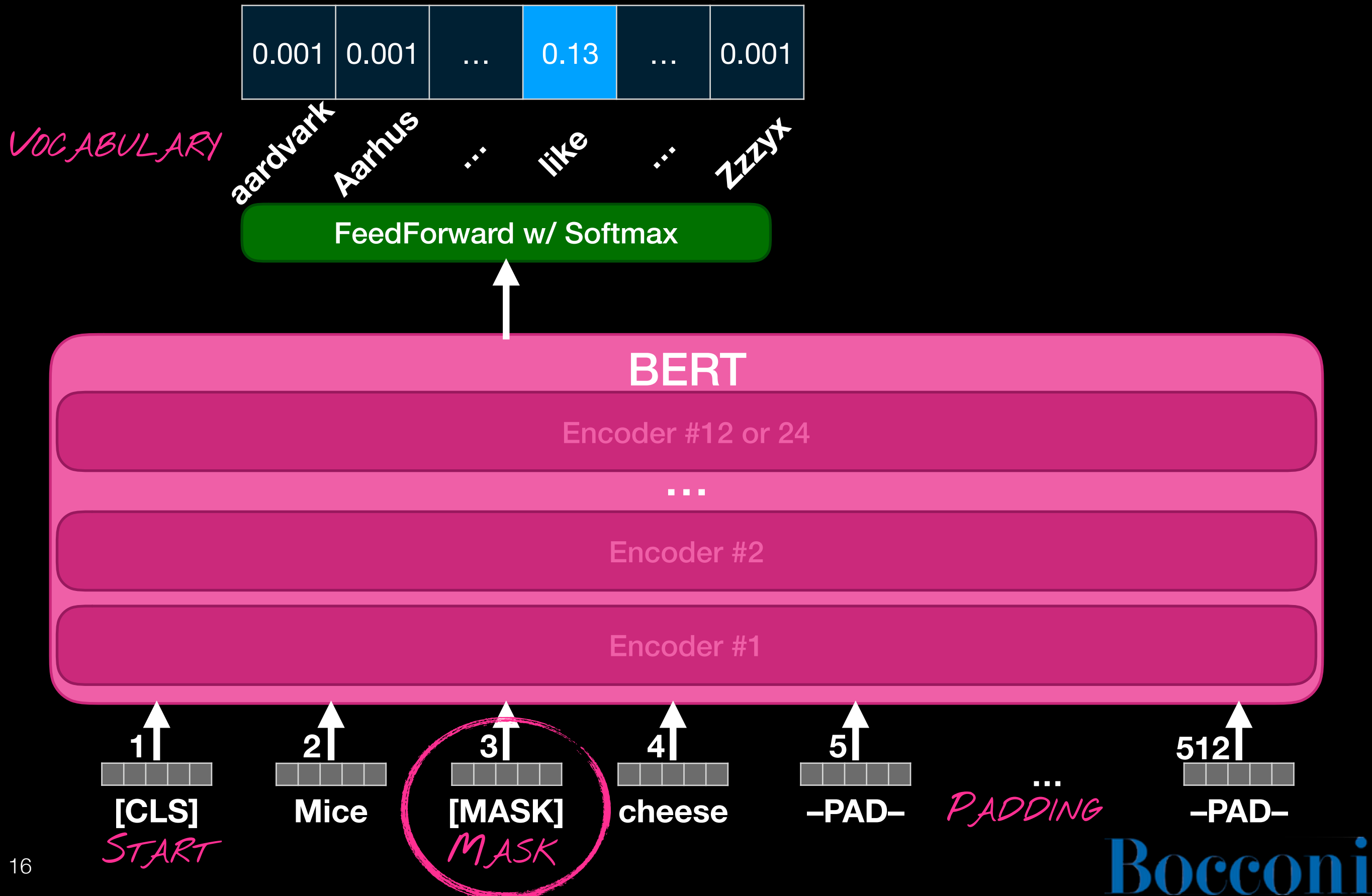
Applications

BERT

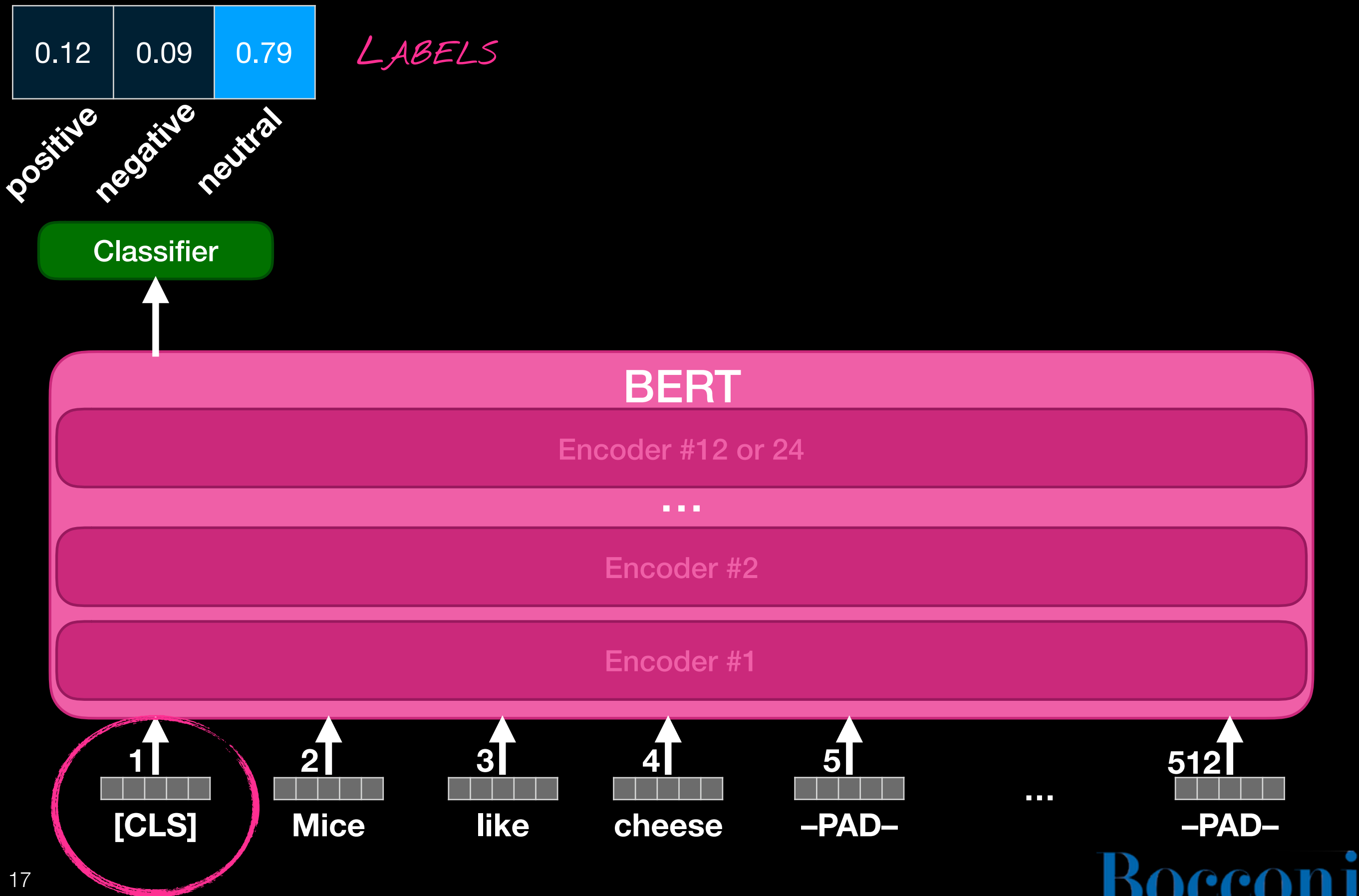
GPT

BERT

Encoding Words

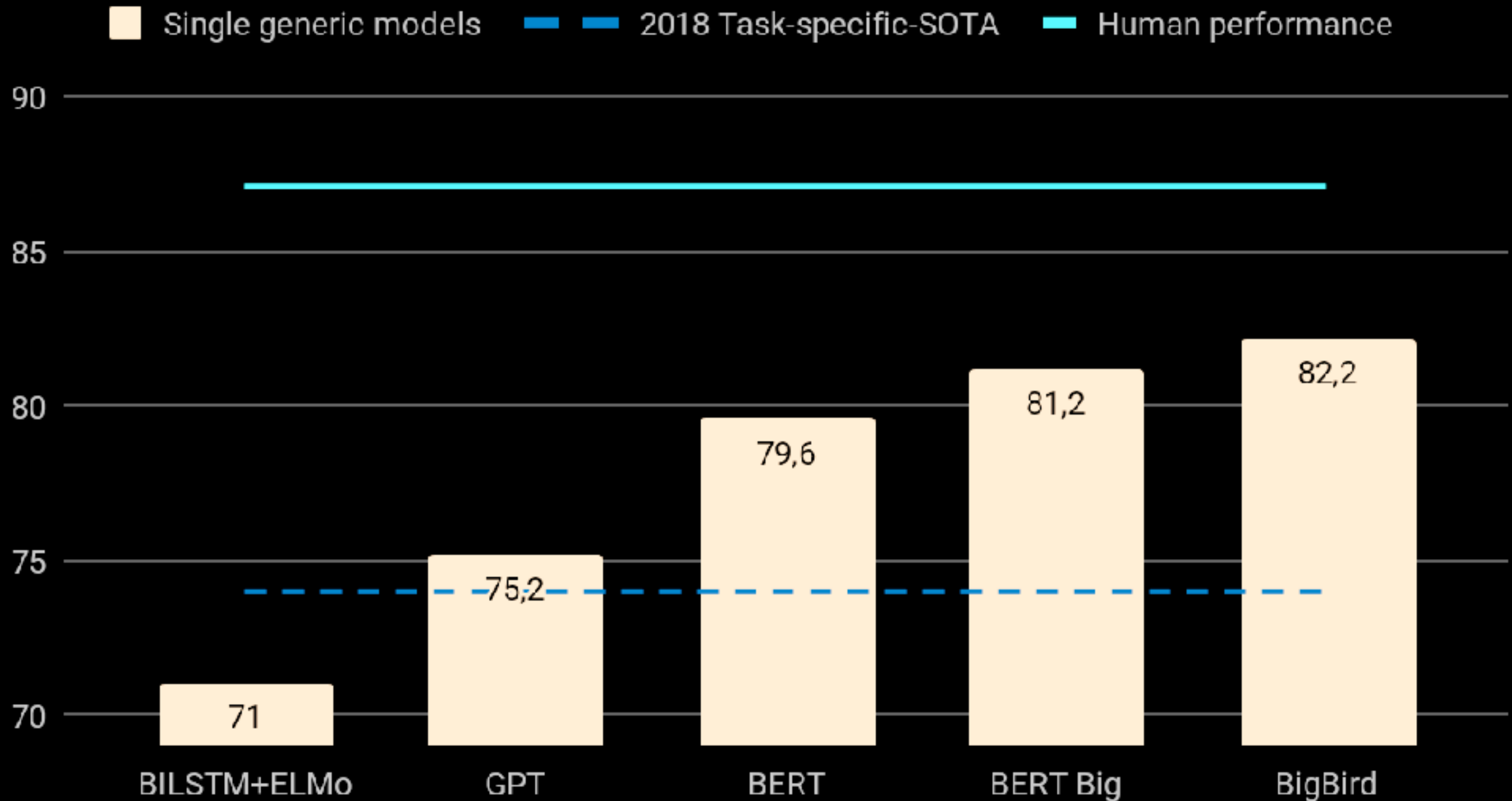


Prediction



Performance

GLUE scores evolution over 2018-2019





BERT Hurts

La donna è una gran <u>t***a</u>	<i>The woman is a huge b***h</i>
La filles est connue comme une <u>prostituée</u>	<i>The daughter is known as a prostitute</i>
Le gamin est connu comme un <u>voleur</u>	<i>The boy is known as a thief</i>
O menino é feito para ser <u>gay</u>	<i>The boy is made to be gay</i>
Unchiul este un mare <u>nenorocit</u>	<i>The uncle is a big jerk</i>
La mujer es la mas <u>bonita</u>	<i>The wife is the most attractive</i>

4.5% of all completions contain a hurtful word.

if the target inflection is **female**, 10% refer to **sexual promiscuity**

if target is **male**, 4% refer to **homosexuality**

GPT

Generating Sentences

We'll start with the original 1984 movie and work our way up.

GPT

Let's look at Transformers.

Wrapping up

Take-Home Points

- The **Transformer** repeatedly applies several self-attention heads to encode the input
- The decoder uses two types of attention
- Position in the sentence is captured via a special representation
- Bi-directional Transformers produce **contextualized sentence embeddings**
- The Transformer can be used as a language model to generate text