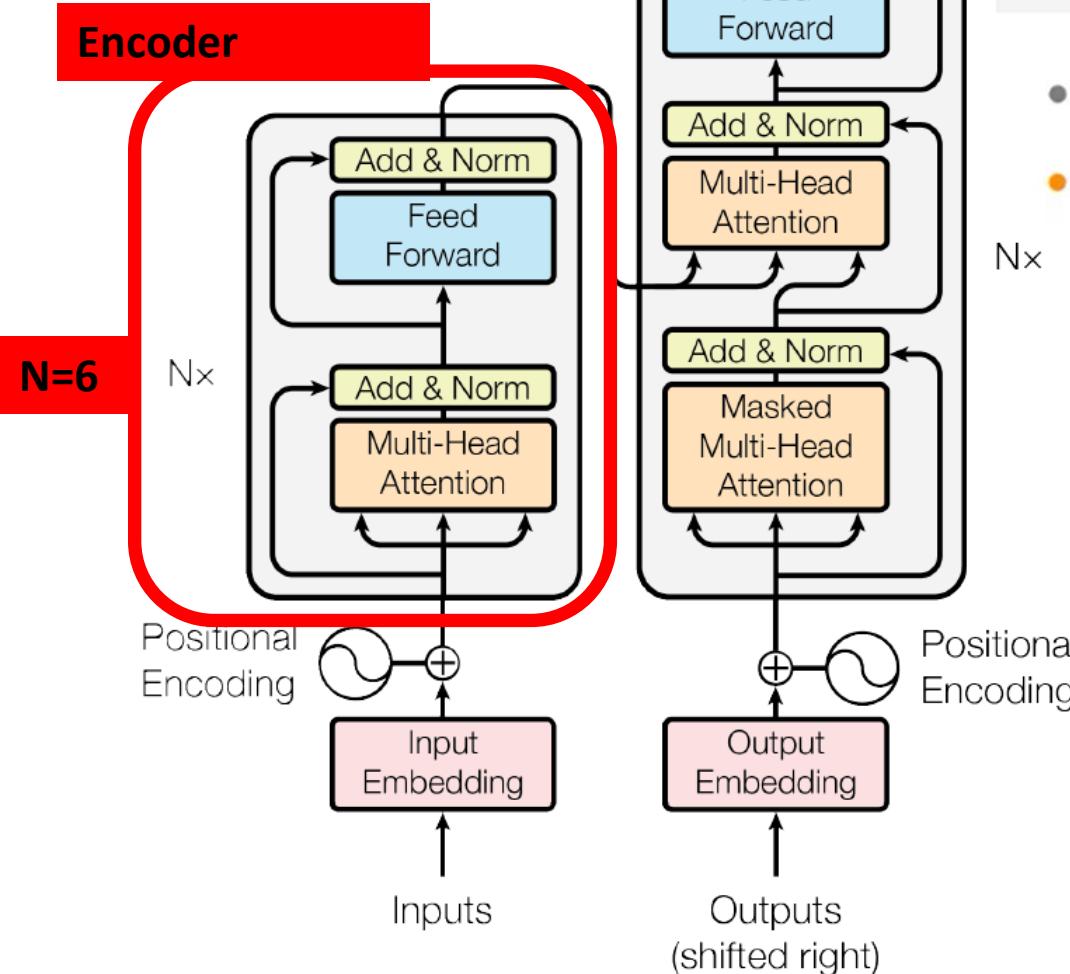


Transformers for NLP

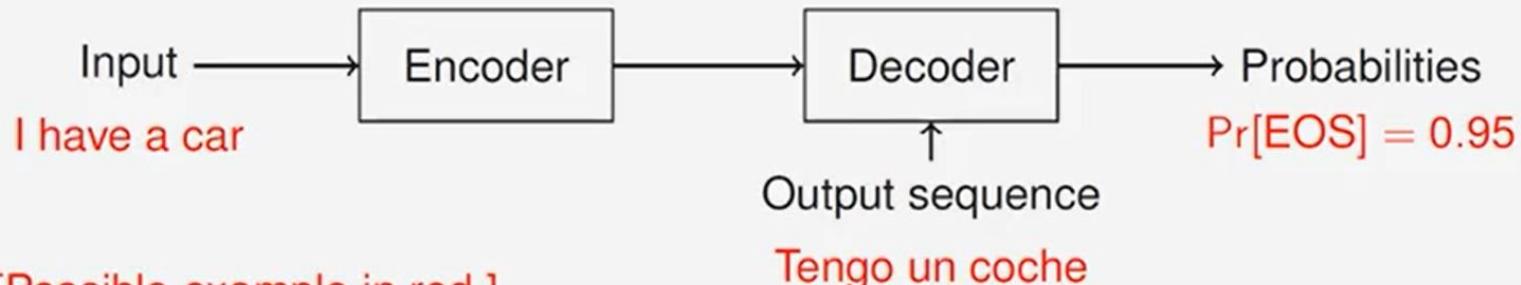
Dr. Kisor K. Sahu

IITBBS

Transformer architecture



High-level transformer architecture



- The encoder-decoder structure is standard in machine translations.
- What's new? Encoder and decoder use **deep self-attention networks**.

Training a Dialogue Completer

Dialogue PART 1

When you play the games of thrones

It is not our abilities that show who we truly are

Life happens where ever you are

All we have to do is decide what to do

There is some good in this world Mr. Frodo

Dialogue PART 2

<start> you win or you die <end>

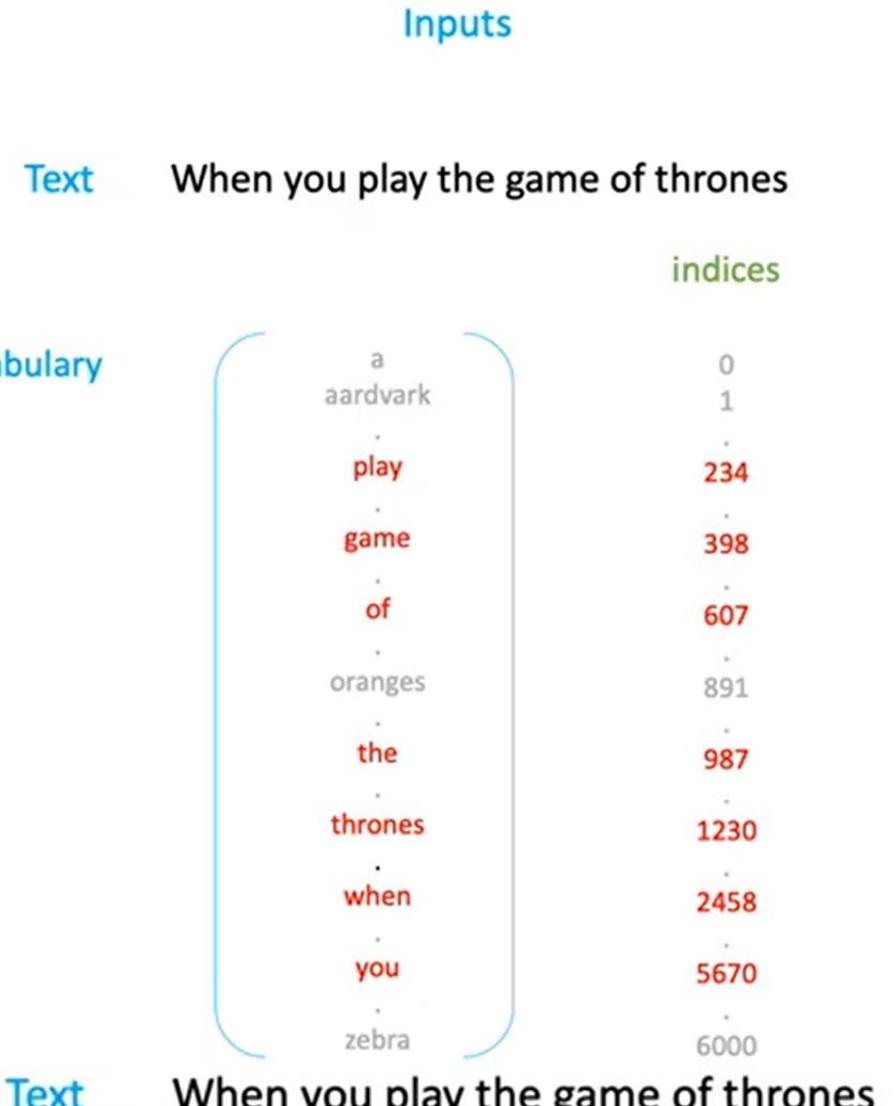
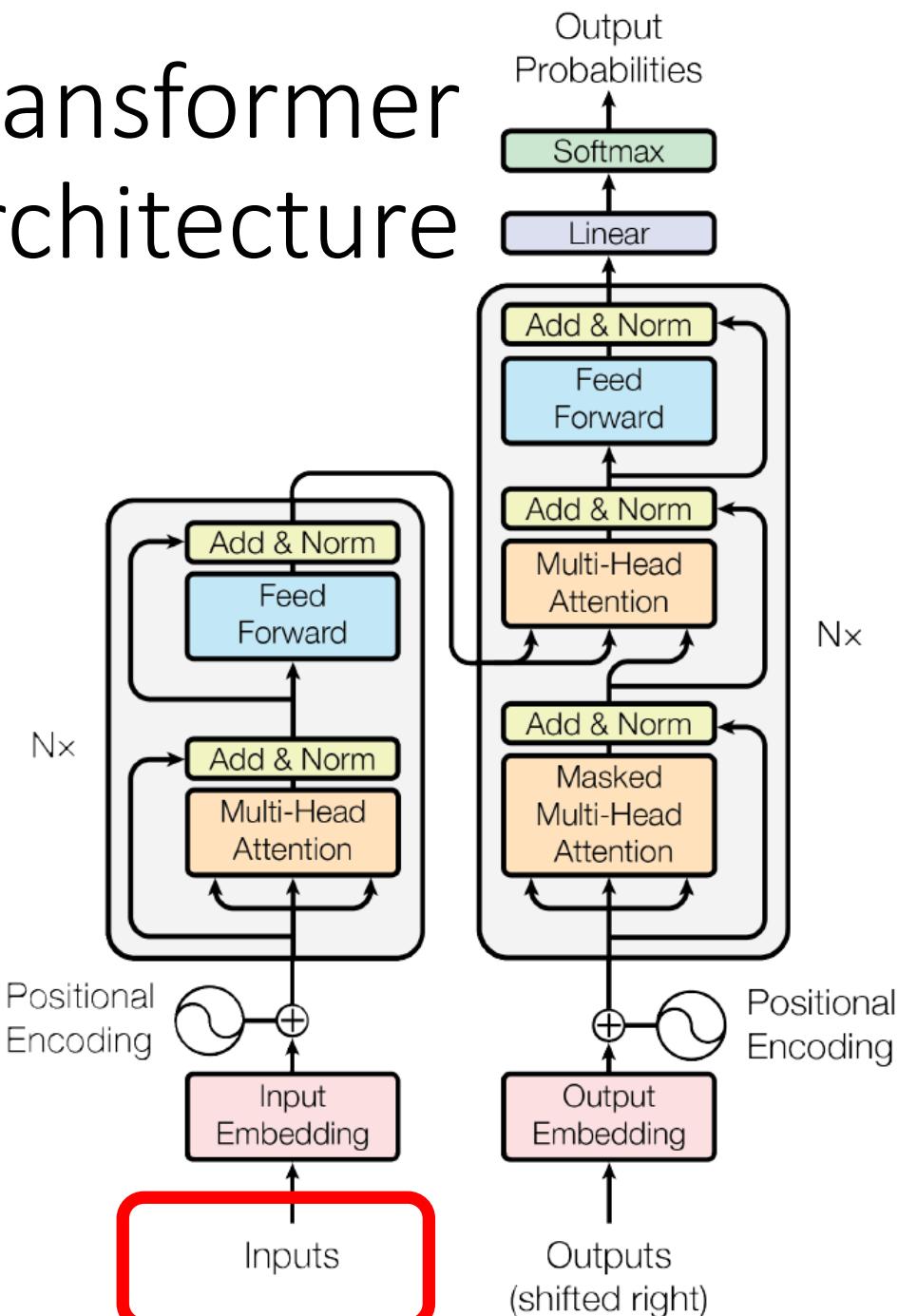
<start> it is our choices <end>

<start> whether you make it or not <end>

<start> with the time that has been given to us <end>

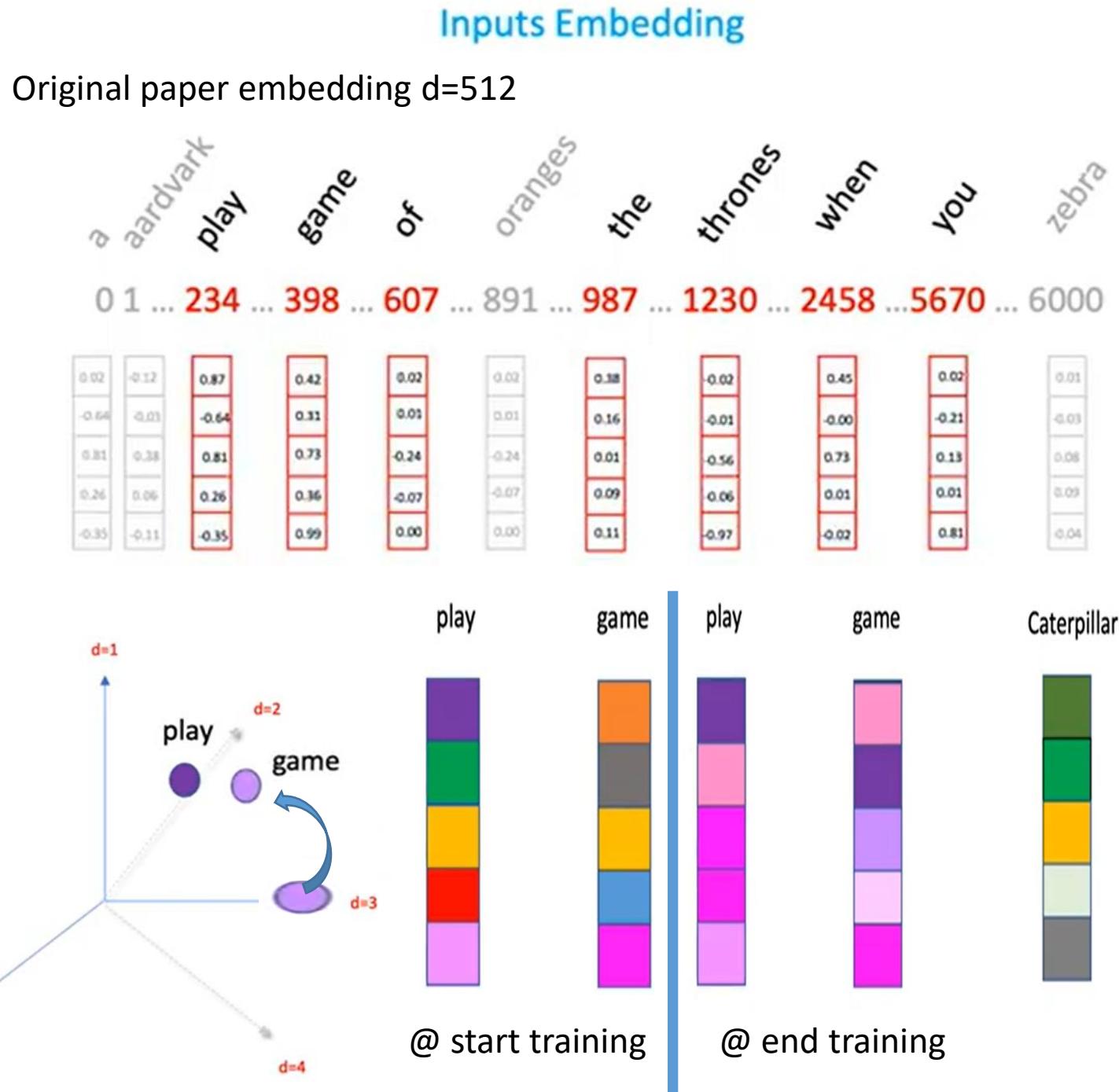
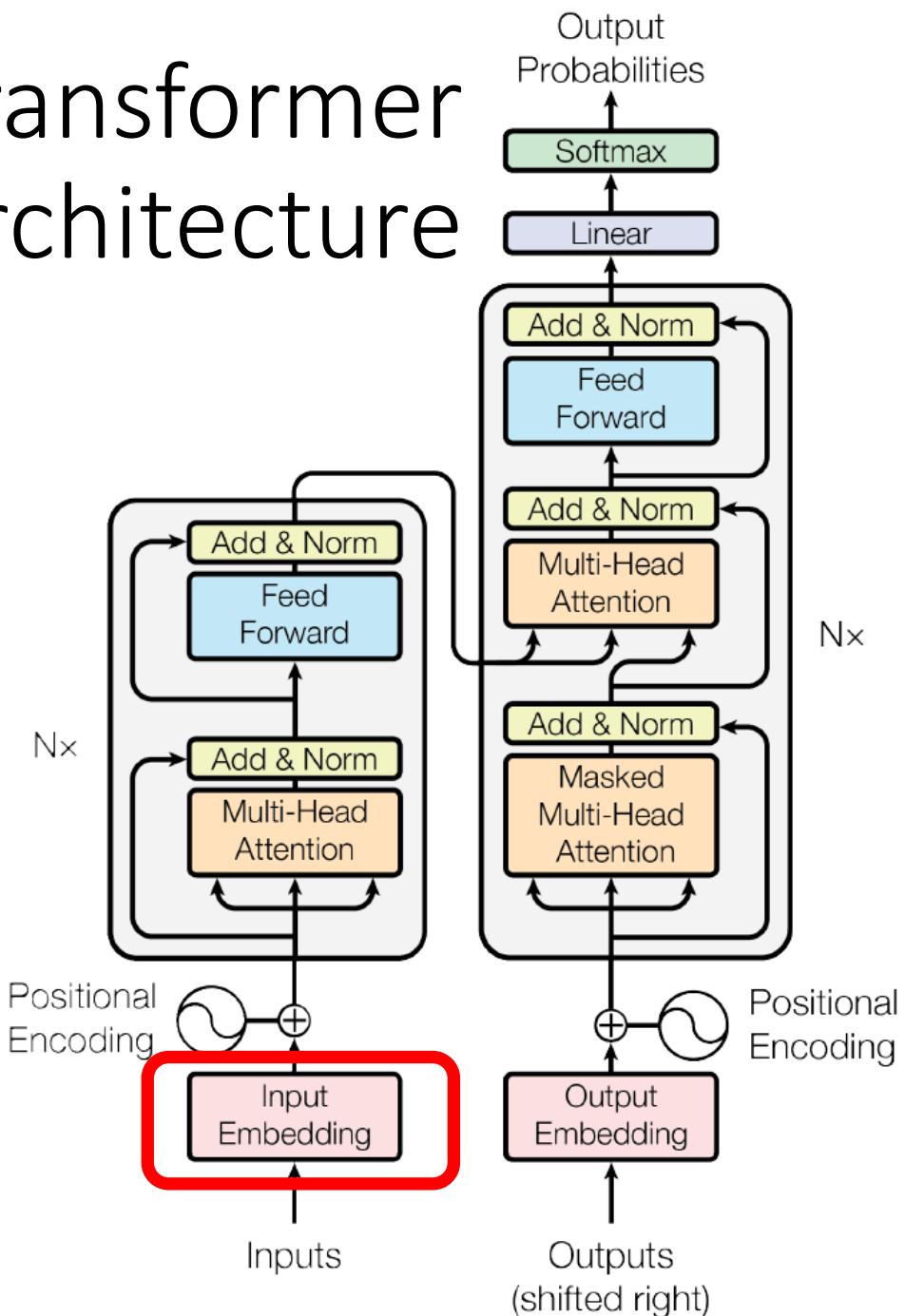
<start> and it is worth fighting for <end>

Transformer architecture

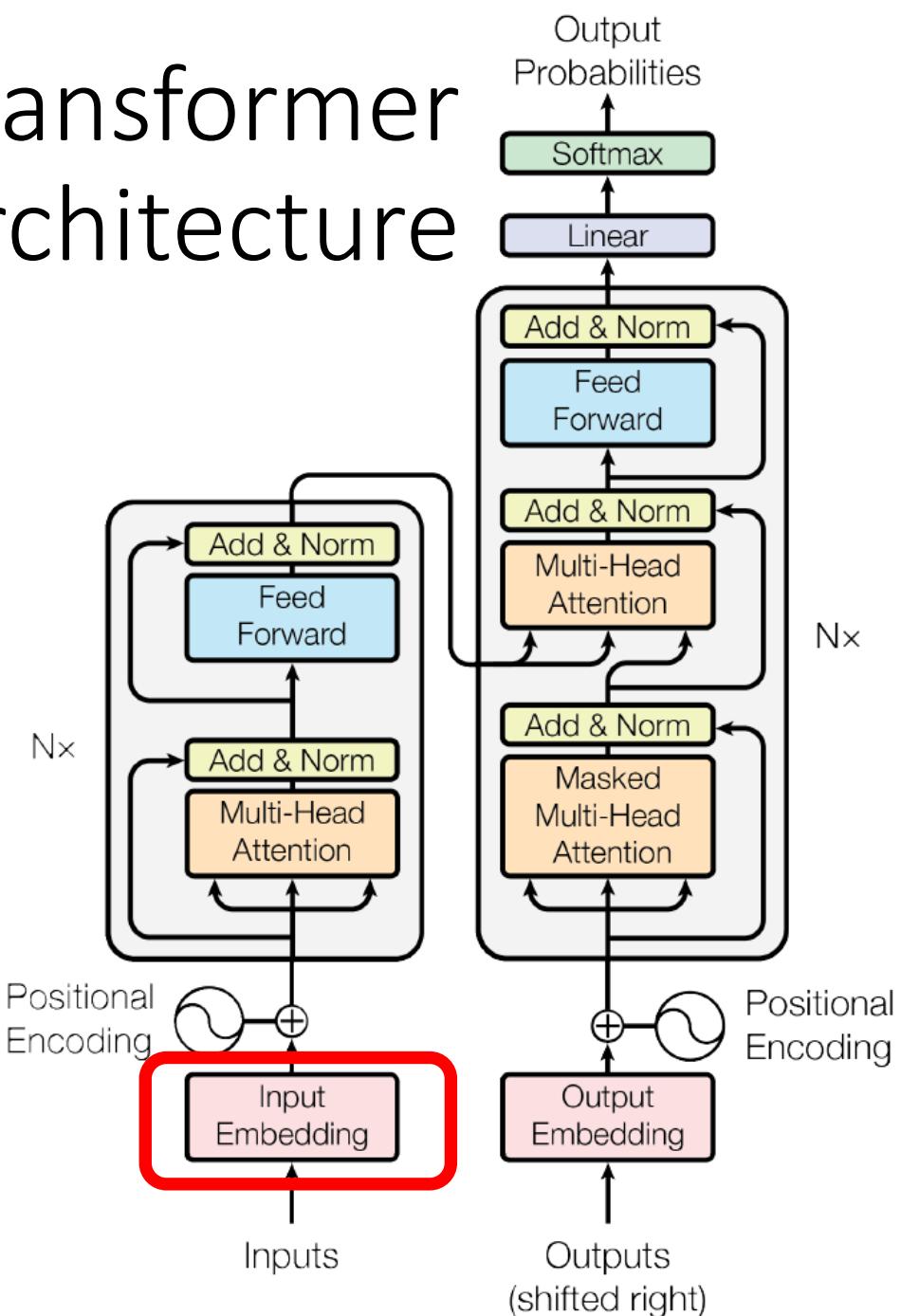


Vocabulary Indices
2458
5670
234
987
398
607
1230

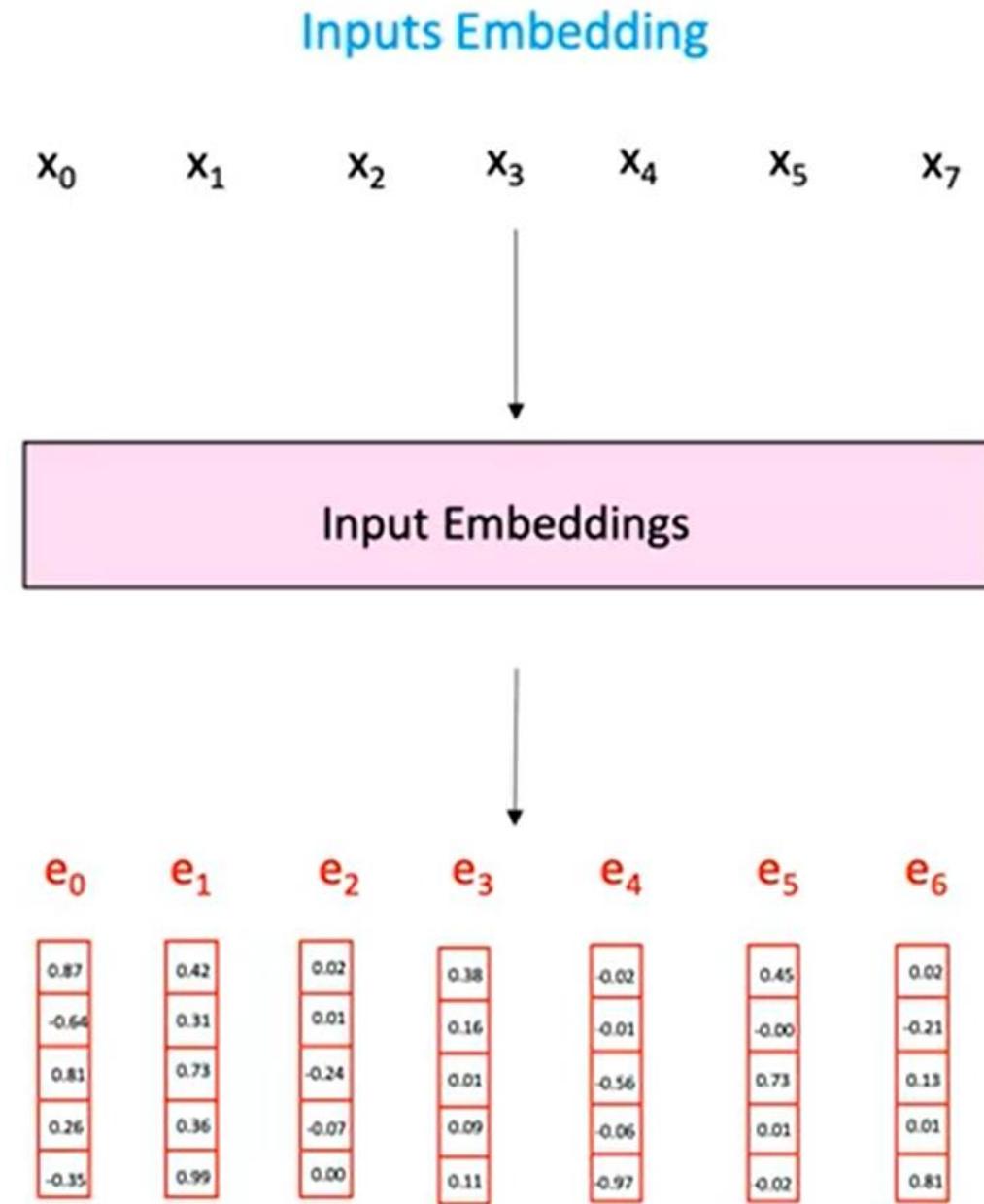
Transformer architecture



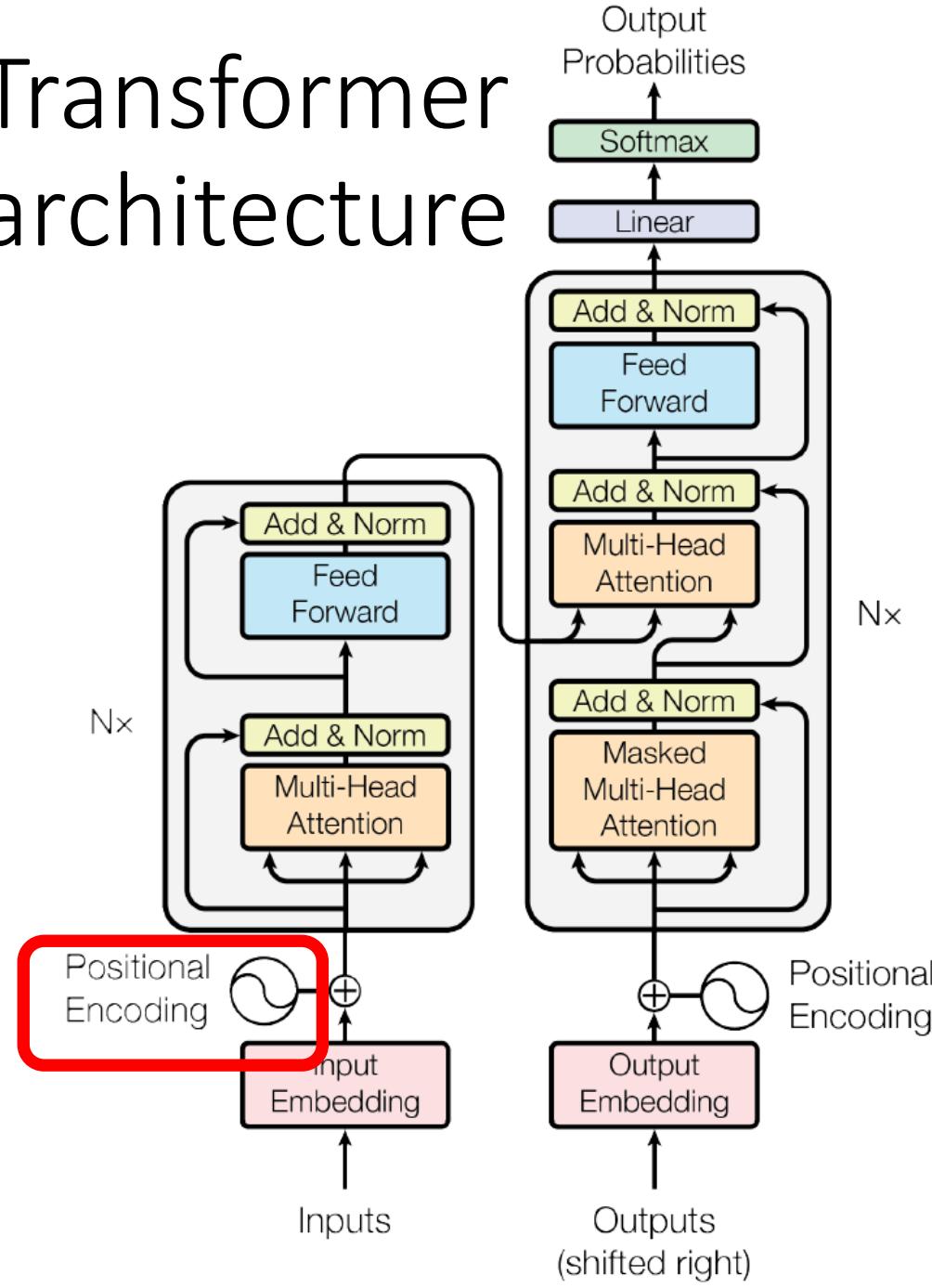
Transformer architecture



Original paper embedding d=512



Transformer architecture

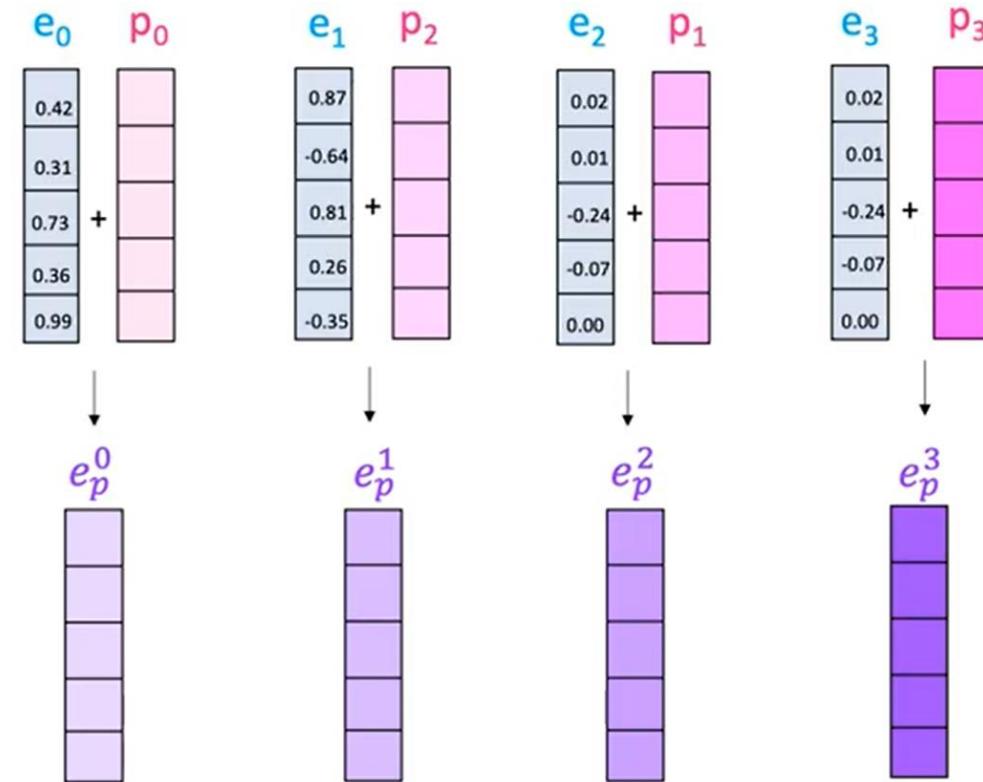


Position Embeddings (Intuition)

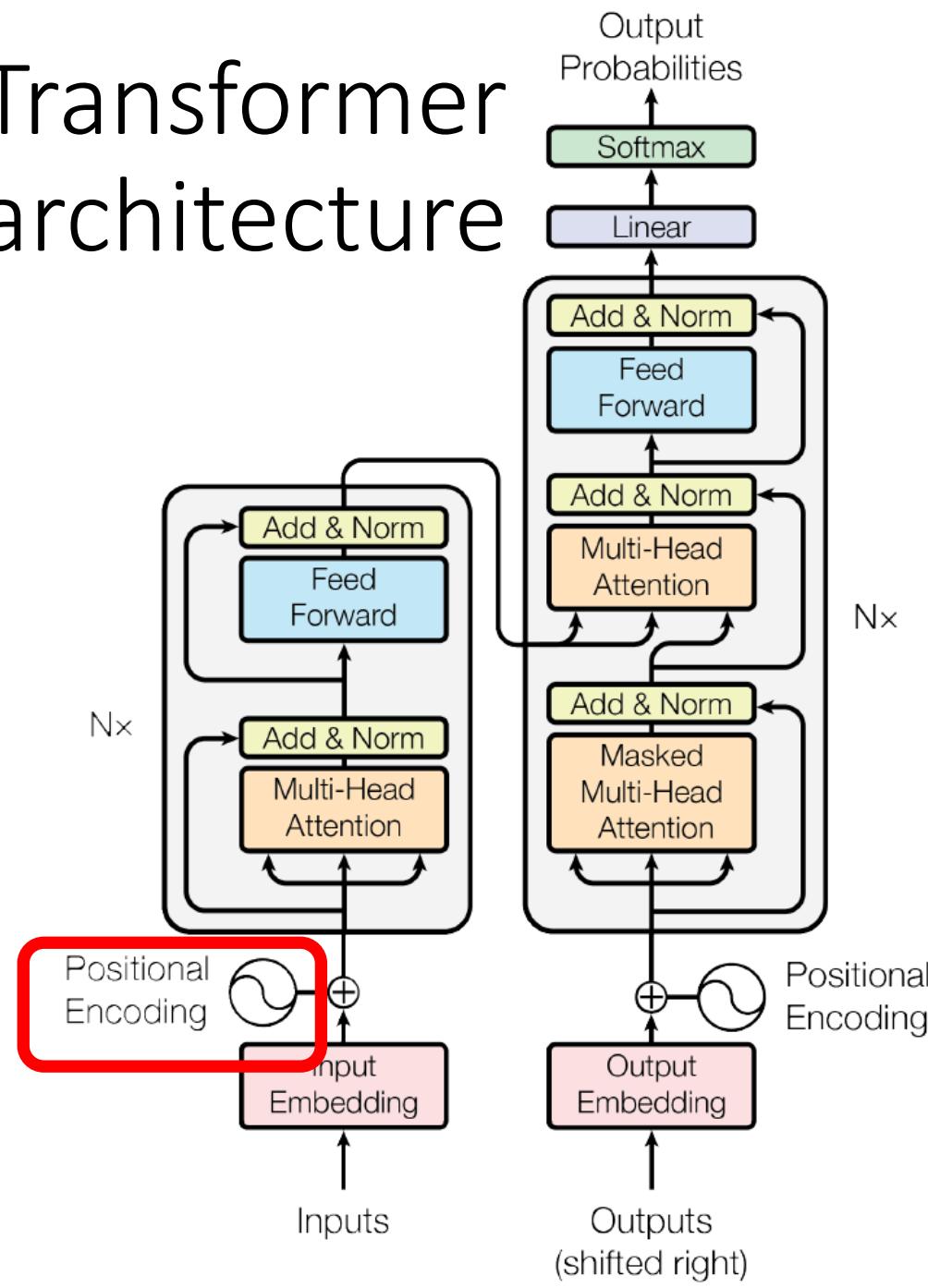
Here is why order matters

Even though she did **not** win the award, she was satisfied.

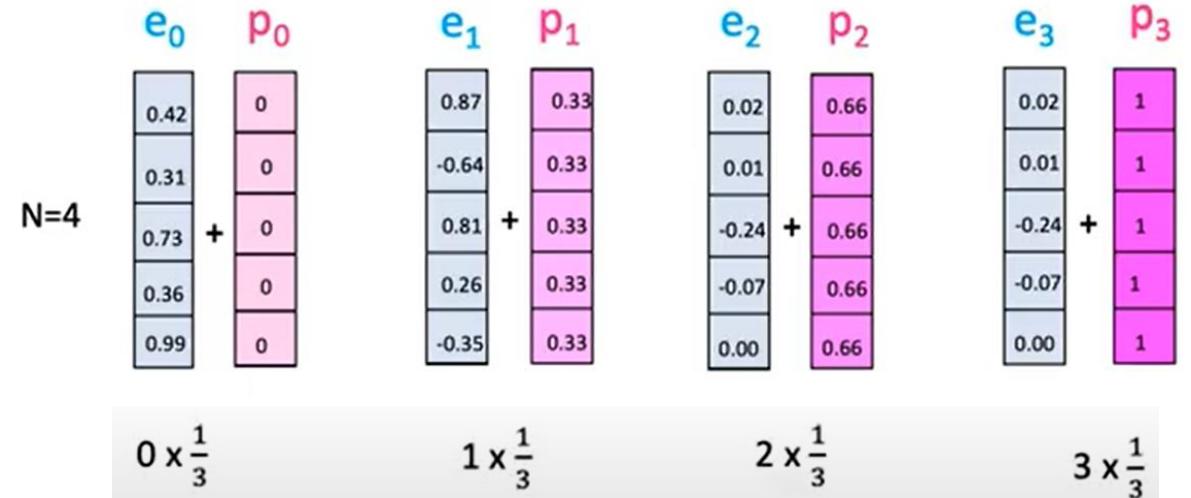
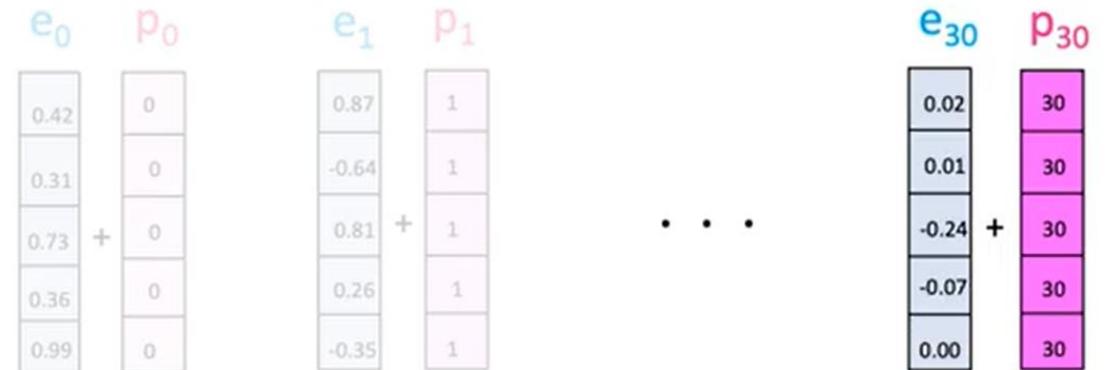
Even though she did win the award, she was **not** satisfied.



Transformer architecture

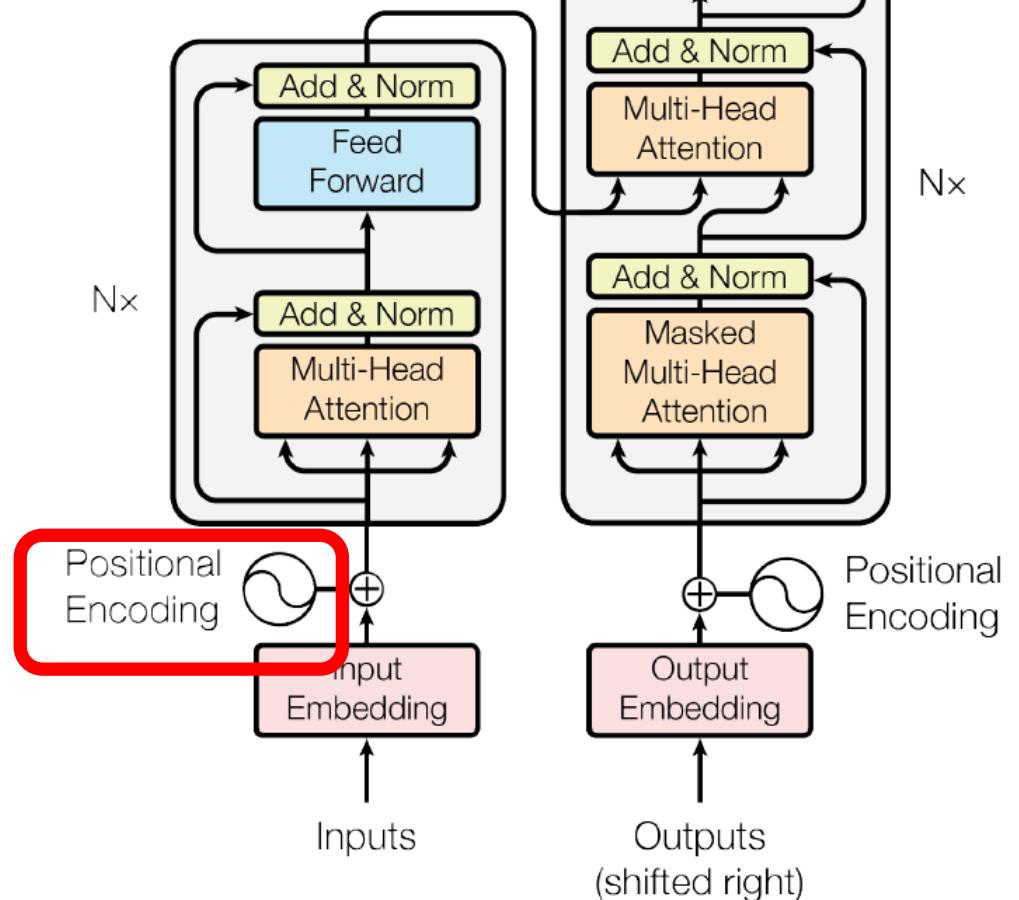


Position Embeddings



$$\frac{1}{N-1} = \frac{1}{3}$$

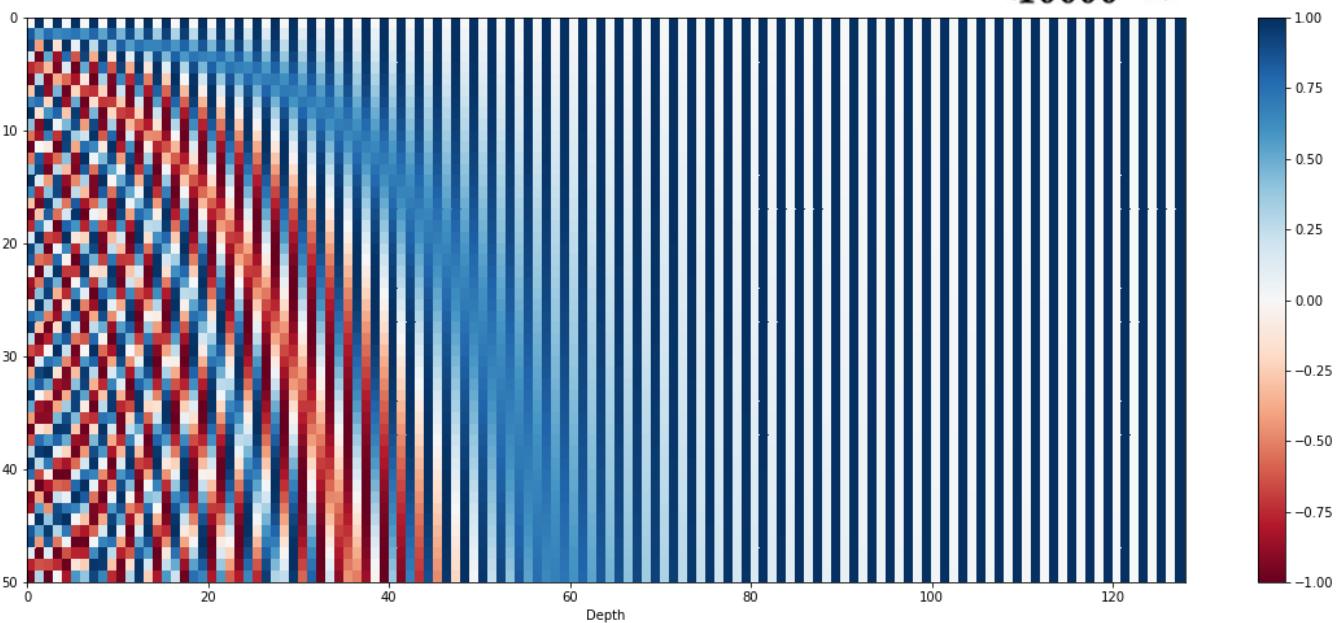
Transformer architecture



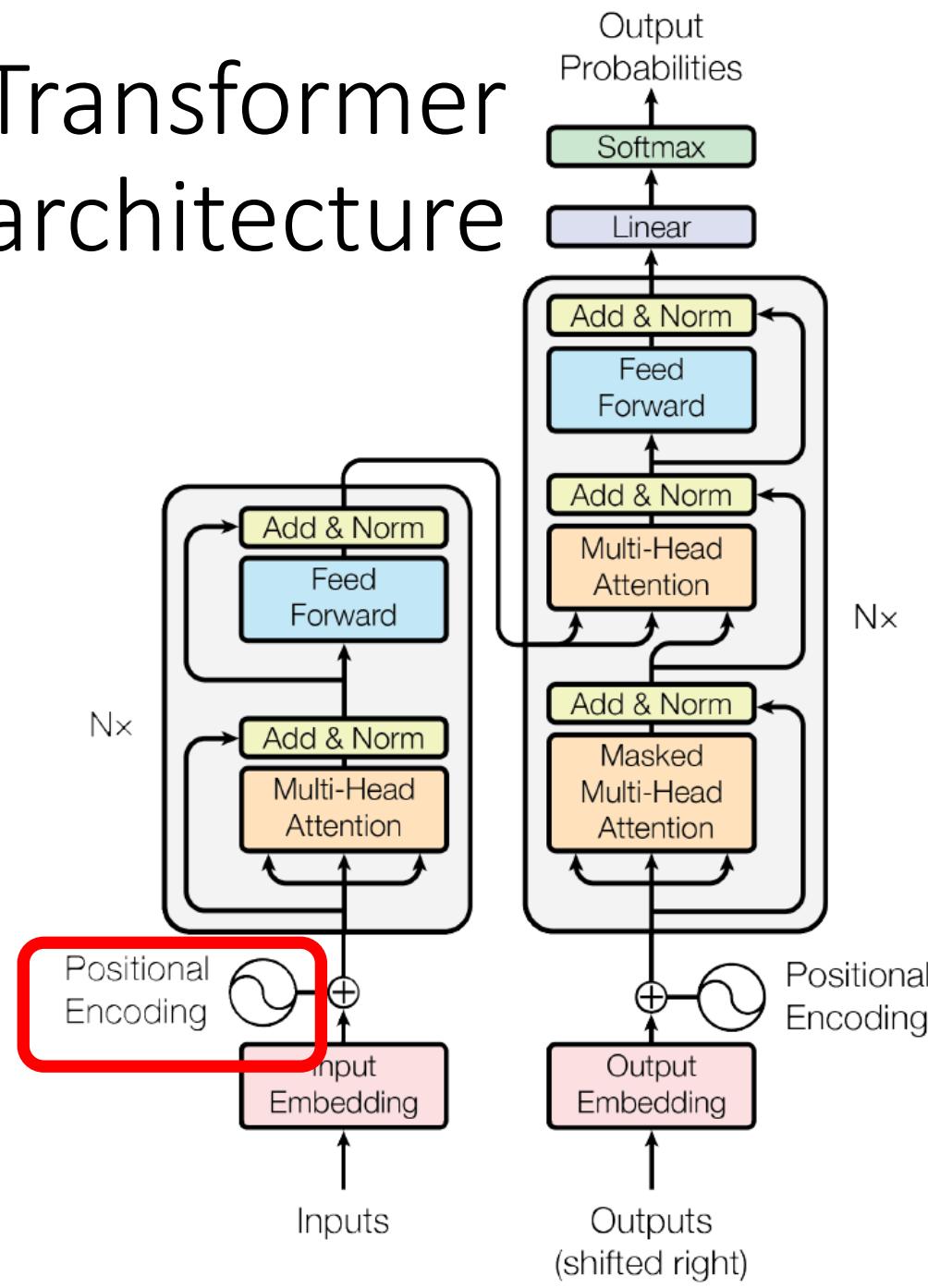
Position	Binary encoding				Modified binary encoding			
0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	0	0	0
2	0	0	1	0	0	1	0	0
3	0	0	1	1	1	1	0	0
4	0	1	0	0	0	0	1	0
5	0	1	0	1	1	0	0	1
6	0	1	1	0	0	1	1	0
7	0	1	1	1	1	1	1	0
8	1	0	0	0	0	0	0	1
9	1	0	0	1	1	0	0	1
10	1	0	1	0	0	1	0	1

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000} \frac{2i}{d}\right)$$

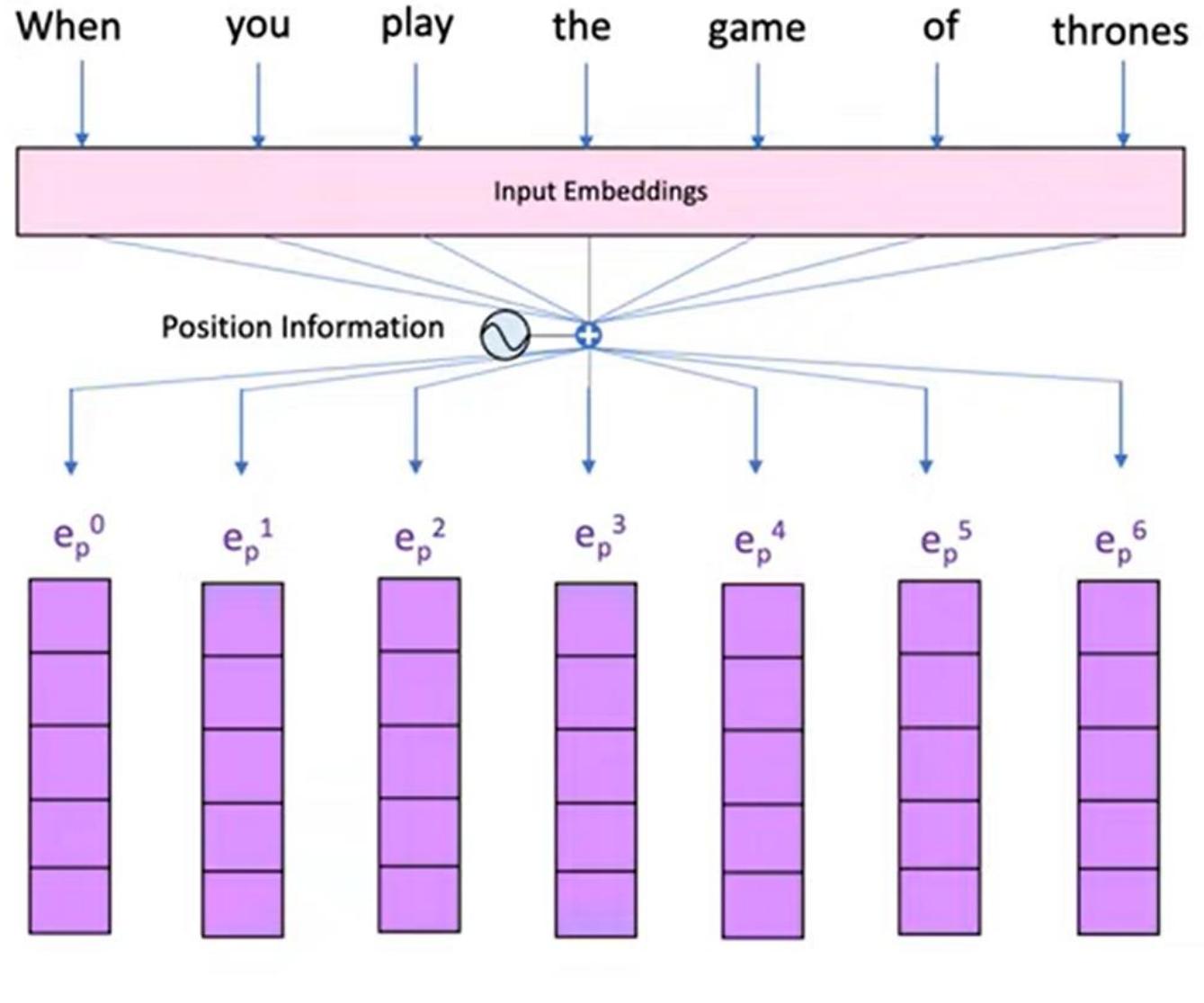
$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000} \frac{2i}{d}\right)$$



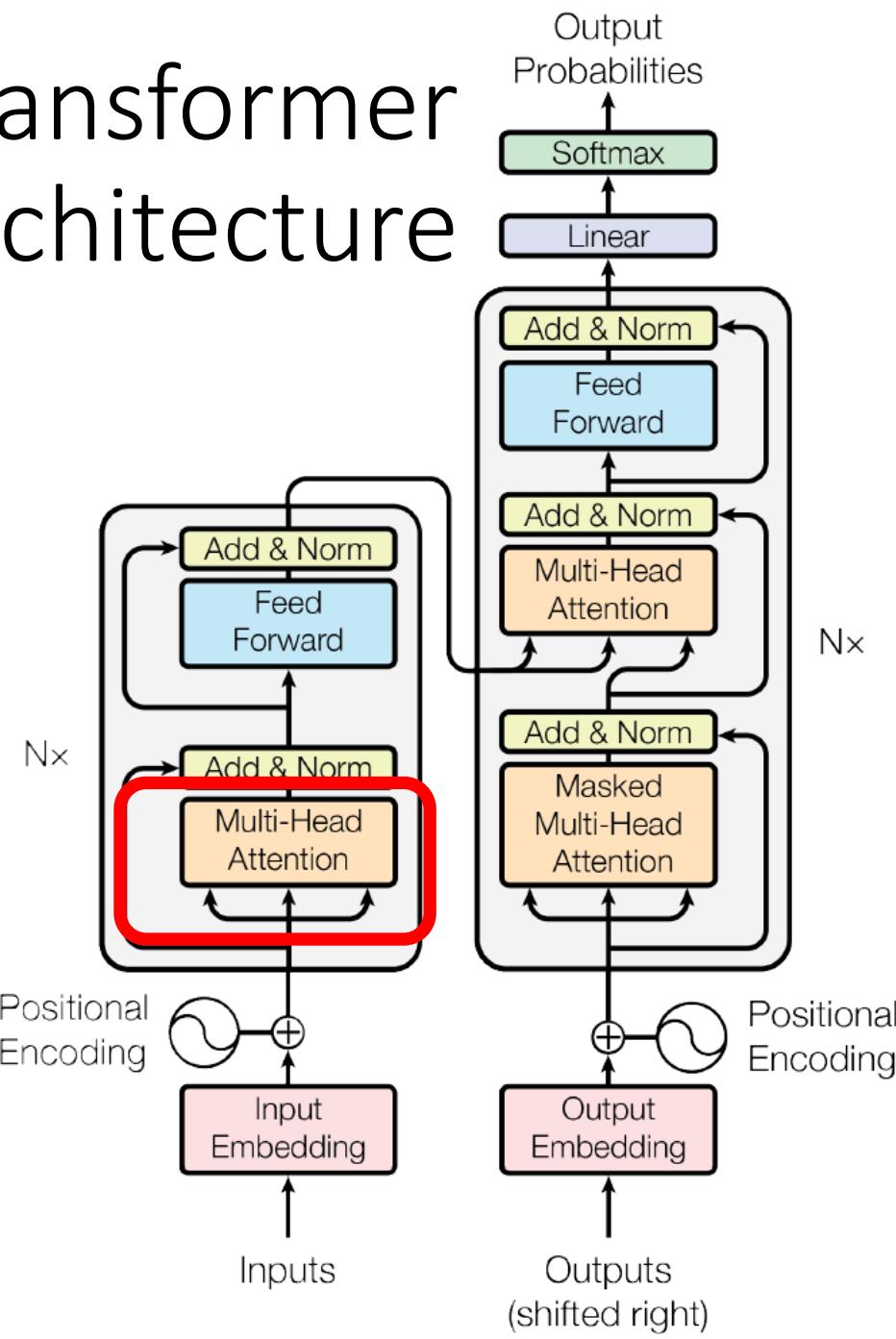
Transformer architecture



Embedding at a glance



Transformer architecture



Multi-Head Attention

Who is this Game of Thrones character?

She faced her enemies and whispered - DRACARYS

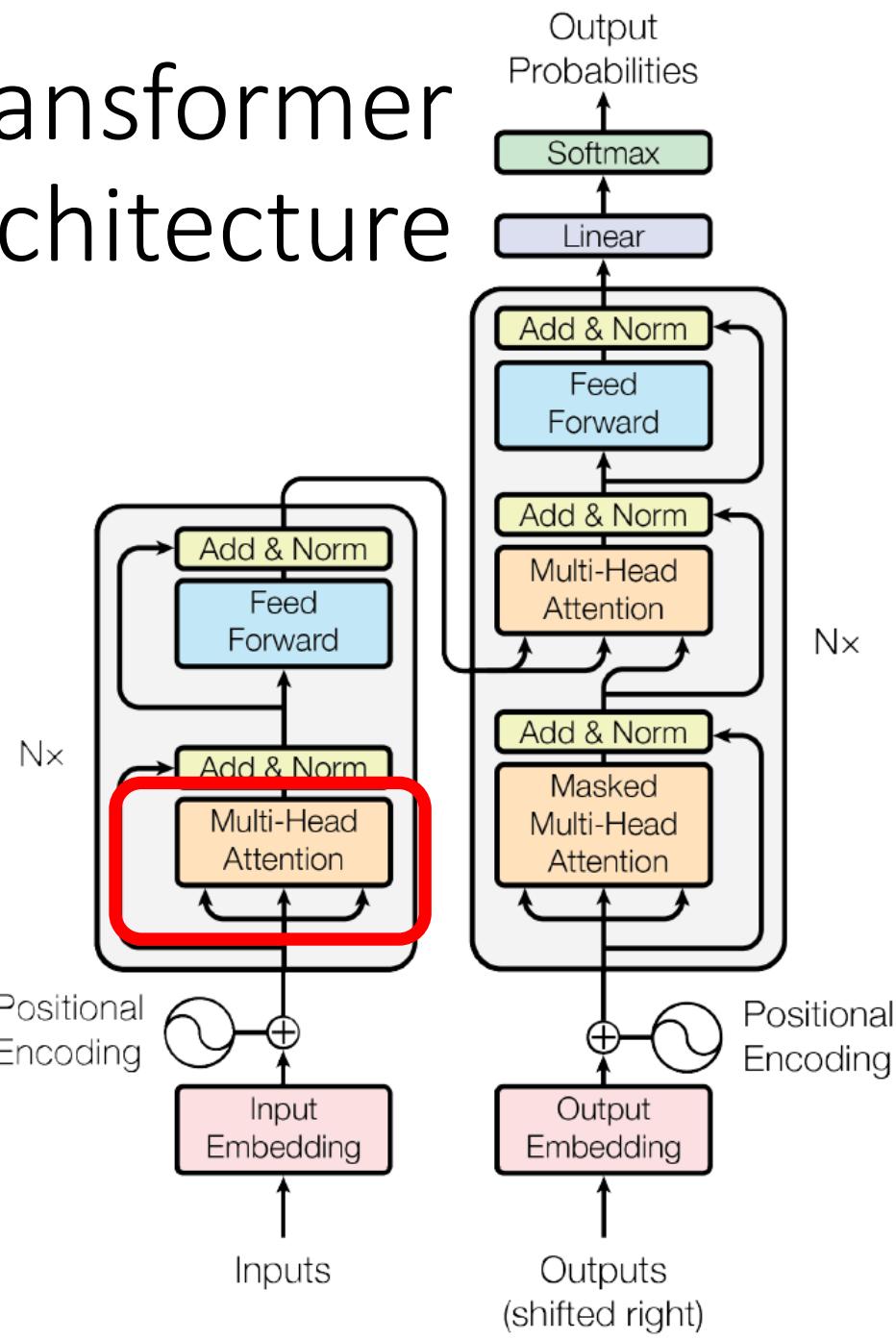
She faced her enemies and whispered - DRACARYS

Self-Attention

He went to the bank and learned of his empty account, after which he went to a river bank and cried.



Transformer architecture



Multi-Head Attention

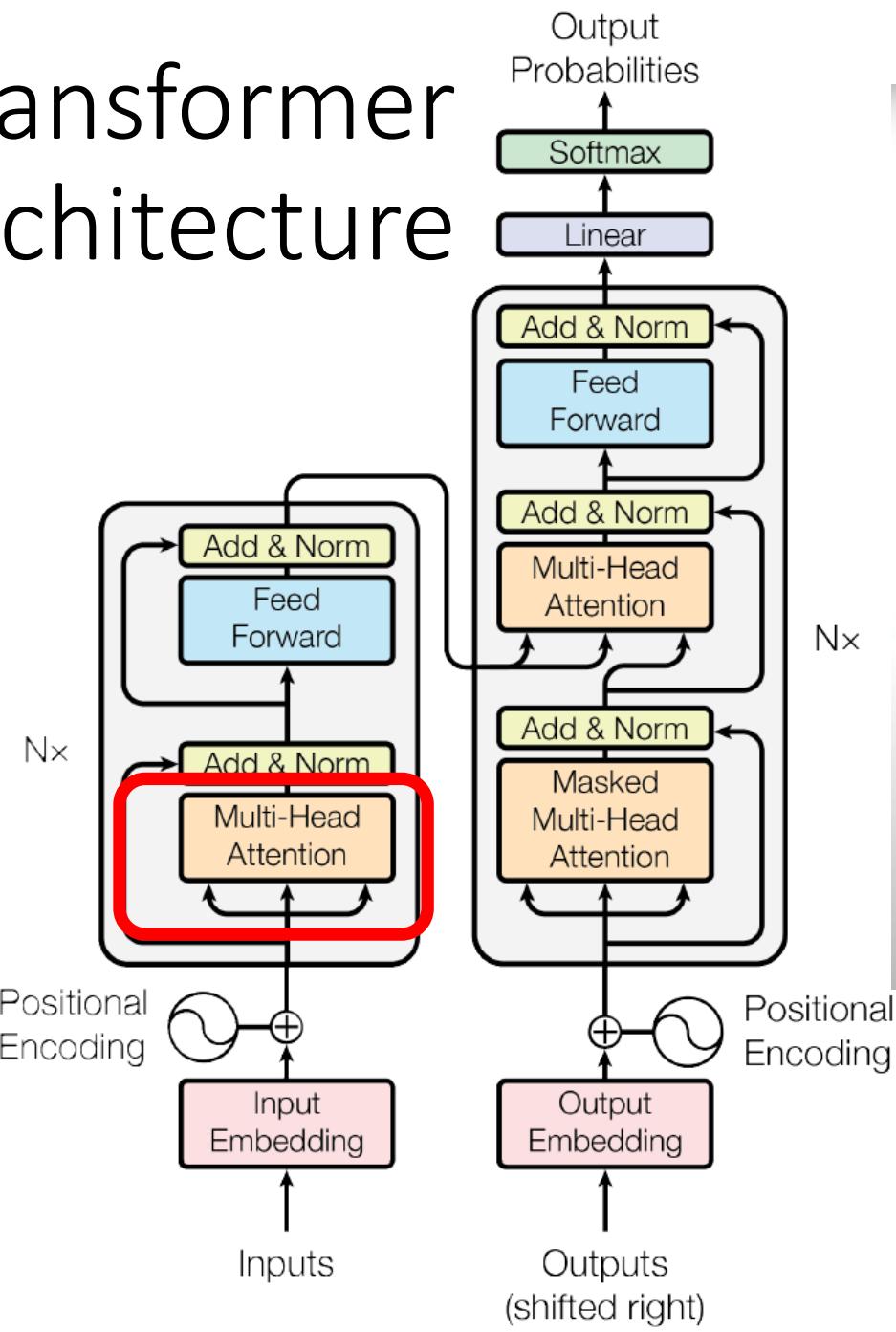
Self-Attention

He went to the bank and learned of his empty account, after which he went to a river bank and cried.



He went to the bank and learned of his empty account, after which he went to a river bank and cried.

Transformer architecture



Simple-Attention

Who is this Game of Thrones character?

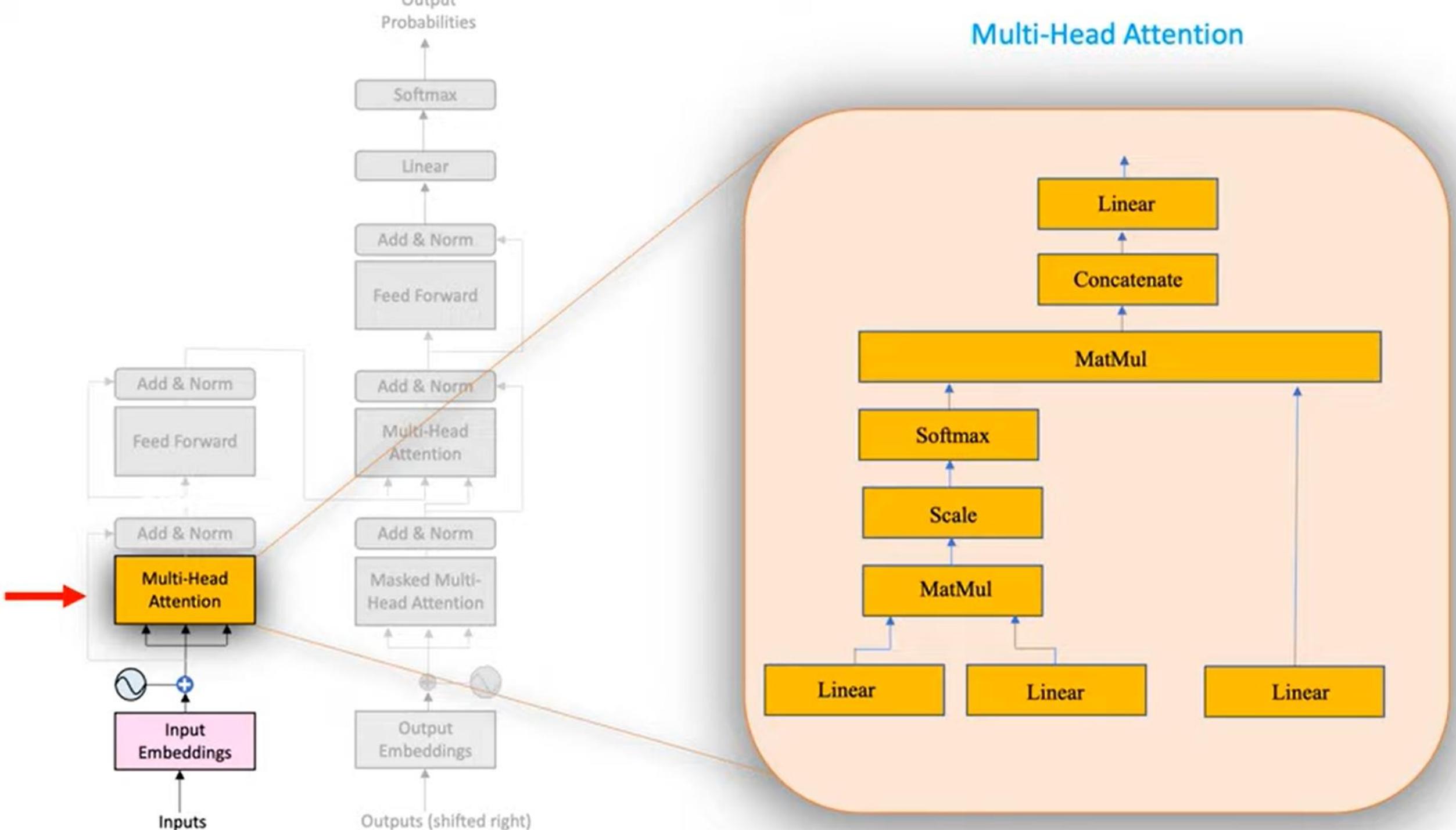
Query

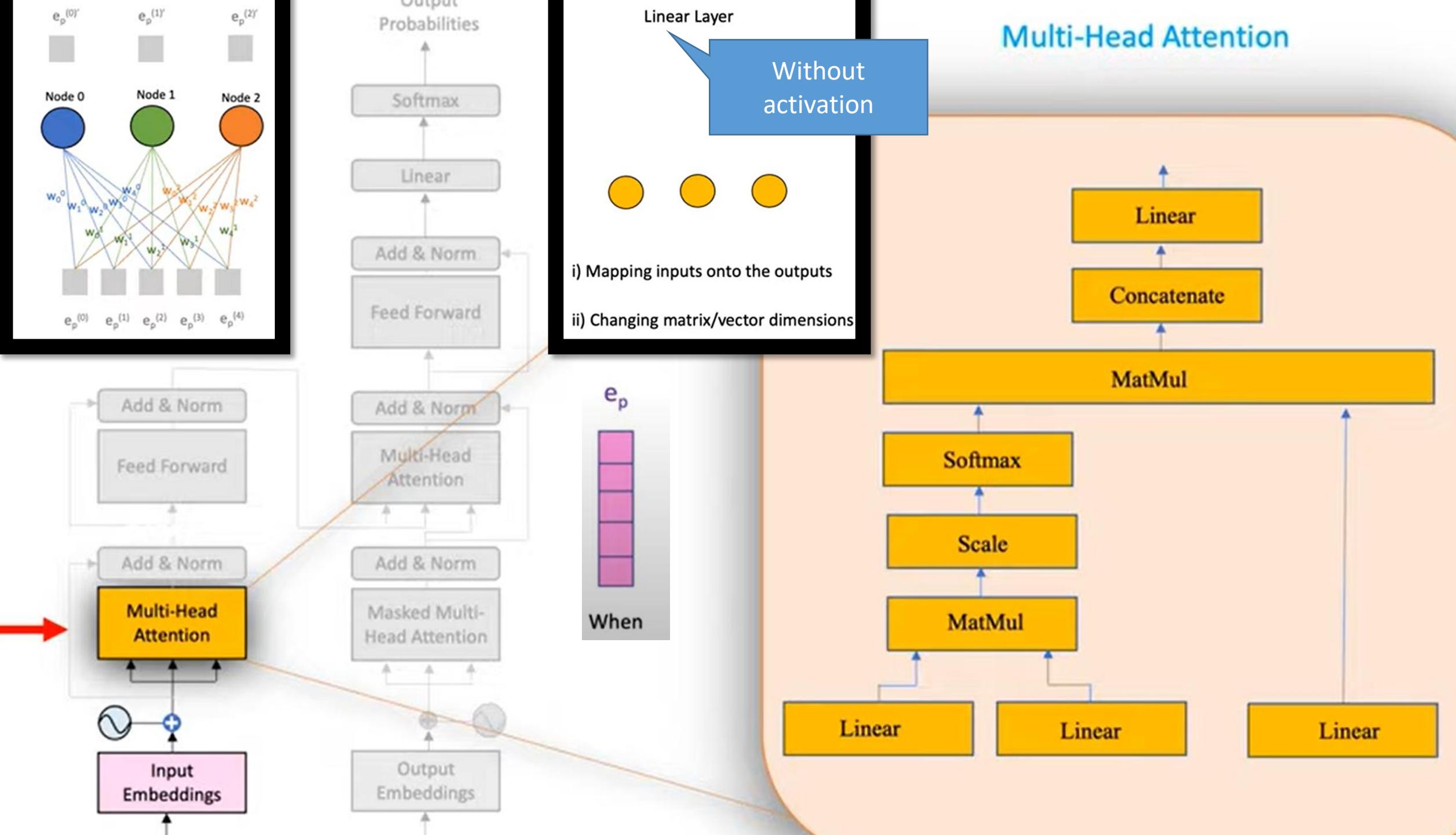
She faced her enemy and whispered - DRACARYS

Self-Attention

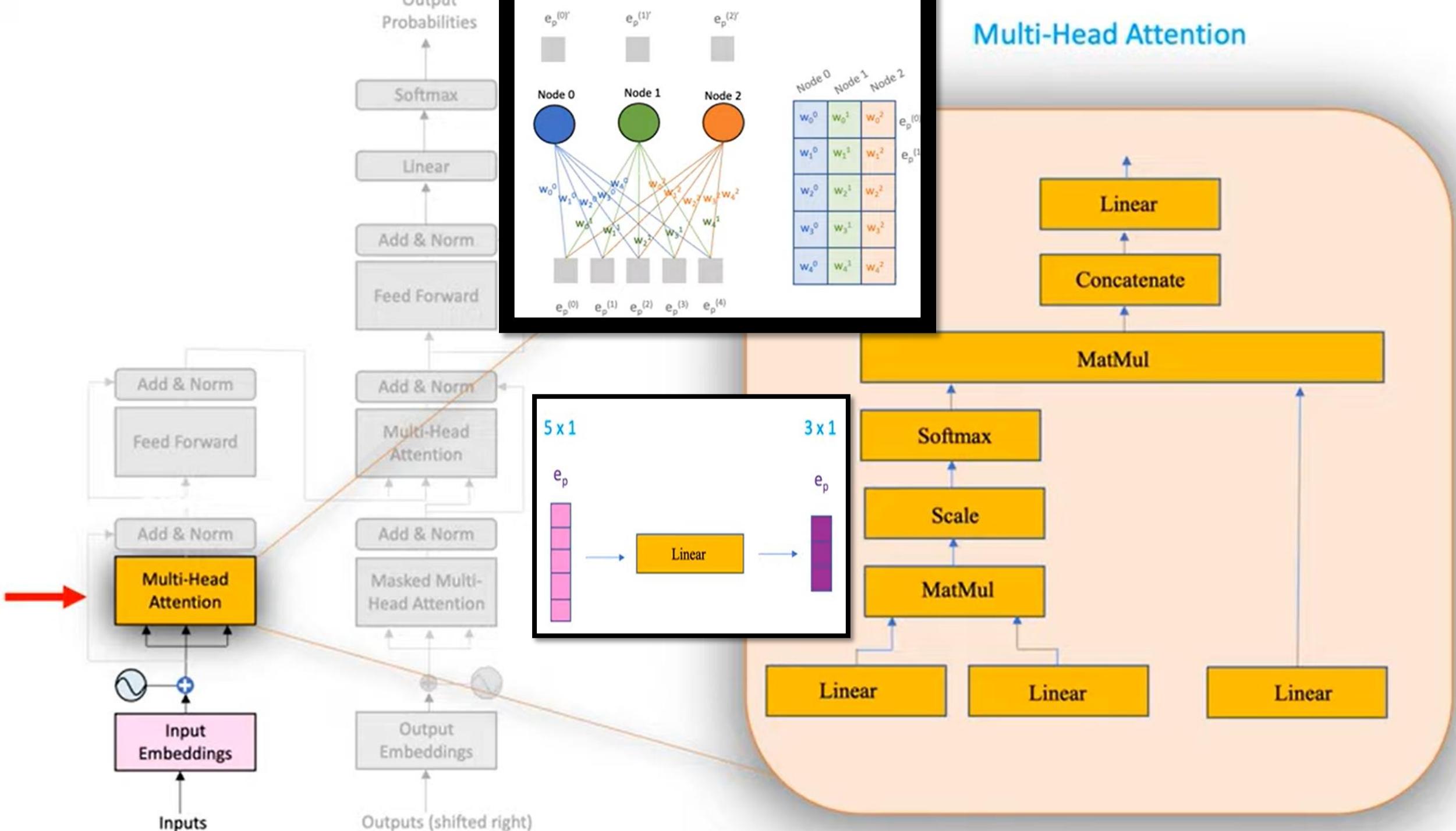
He went to the bank and learned of his empty account, after which he went to a river **bank** and cried.

Multi-Head Attention

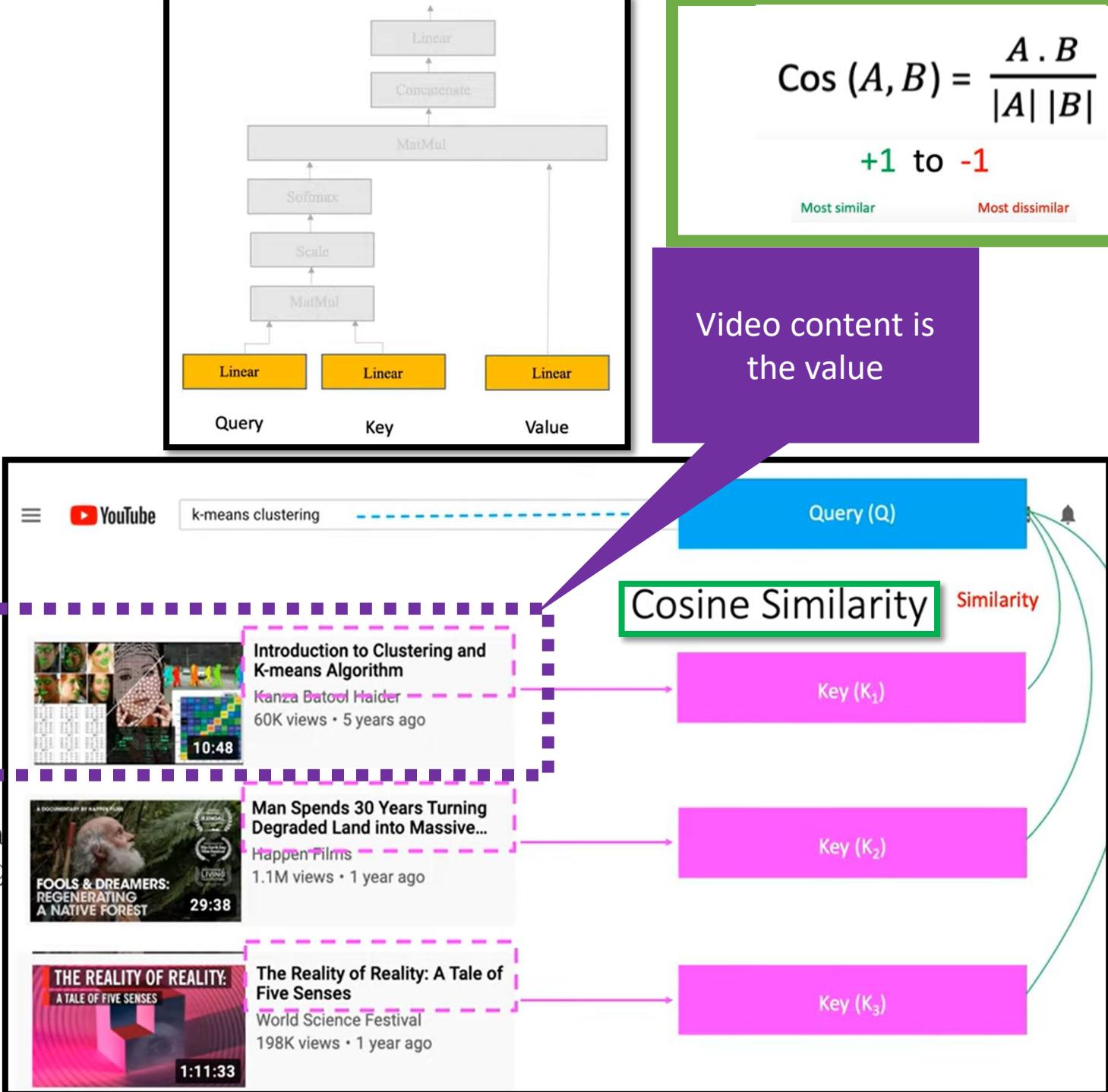
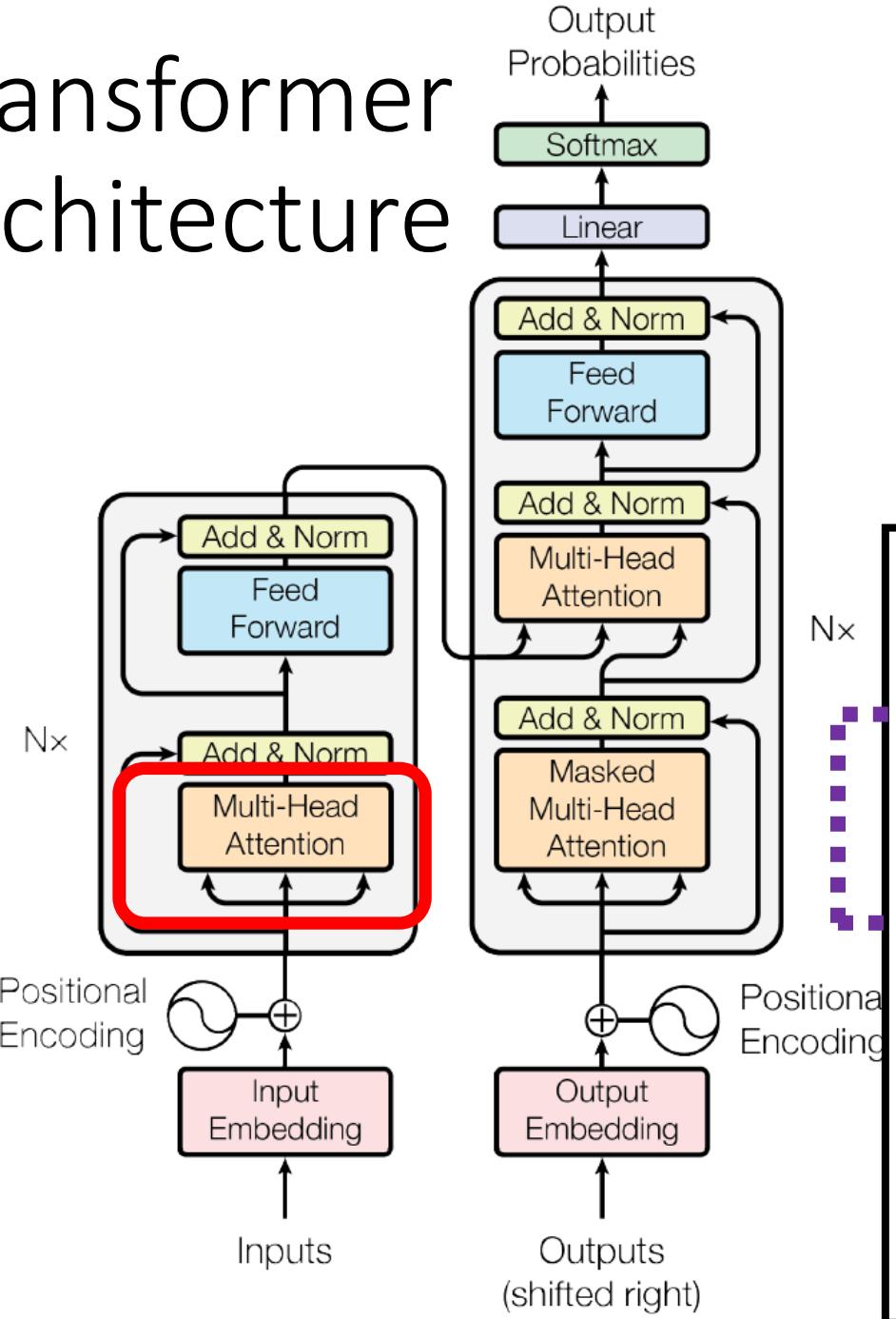




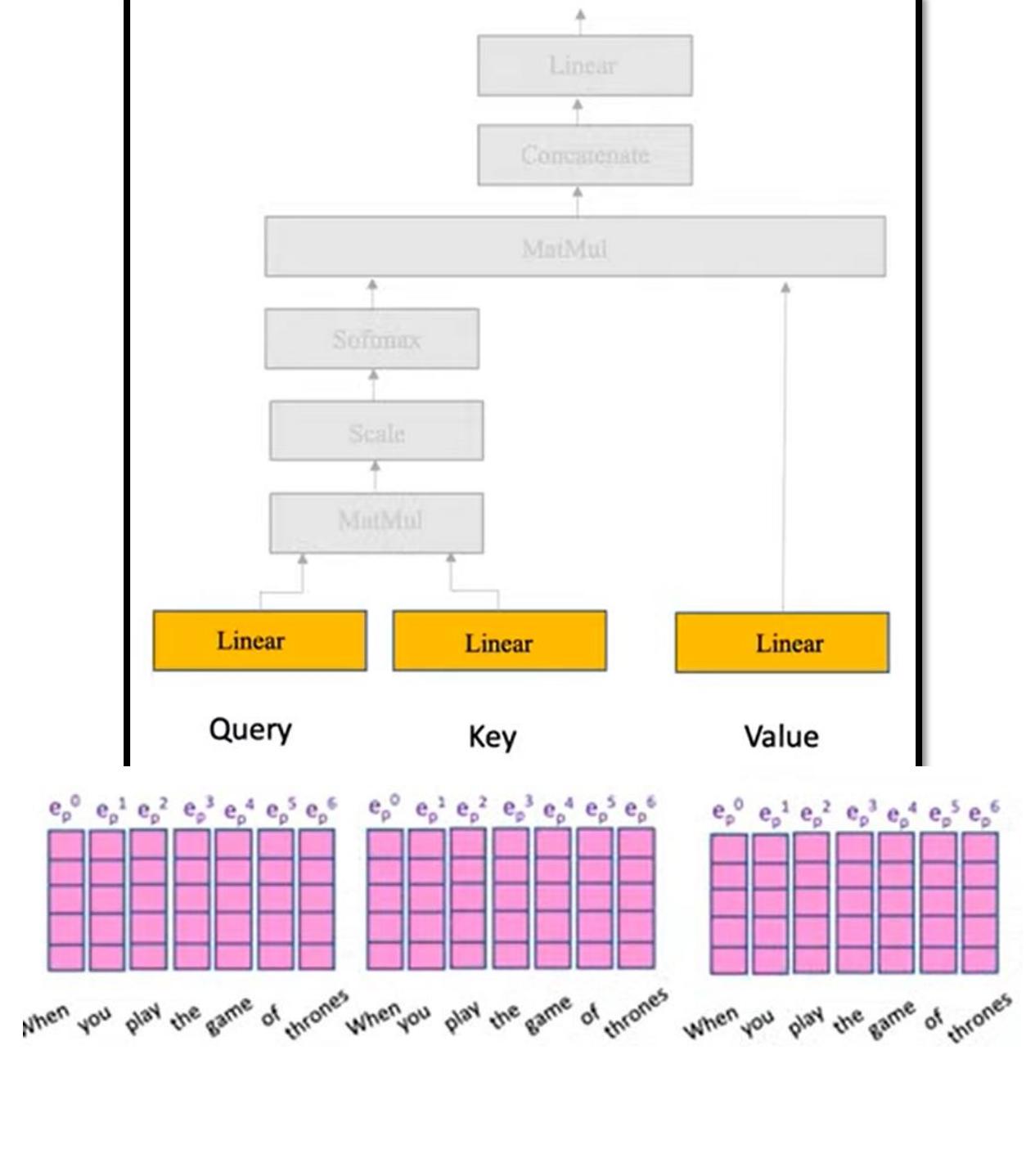
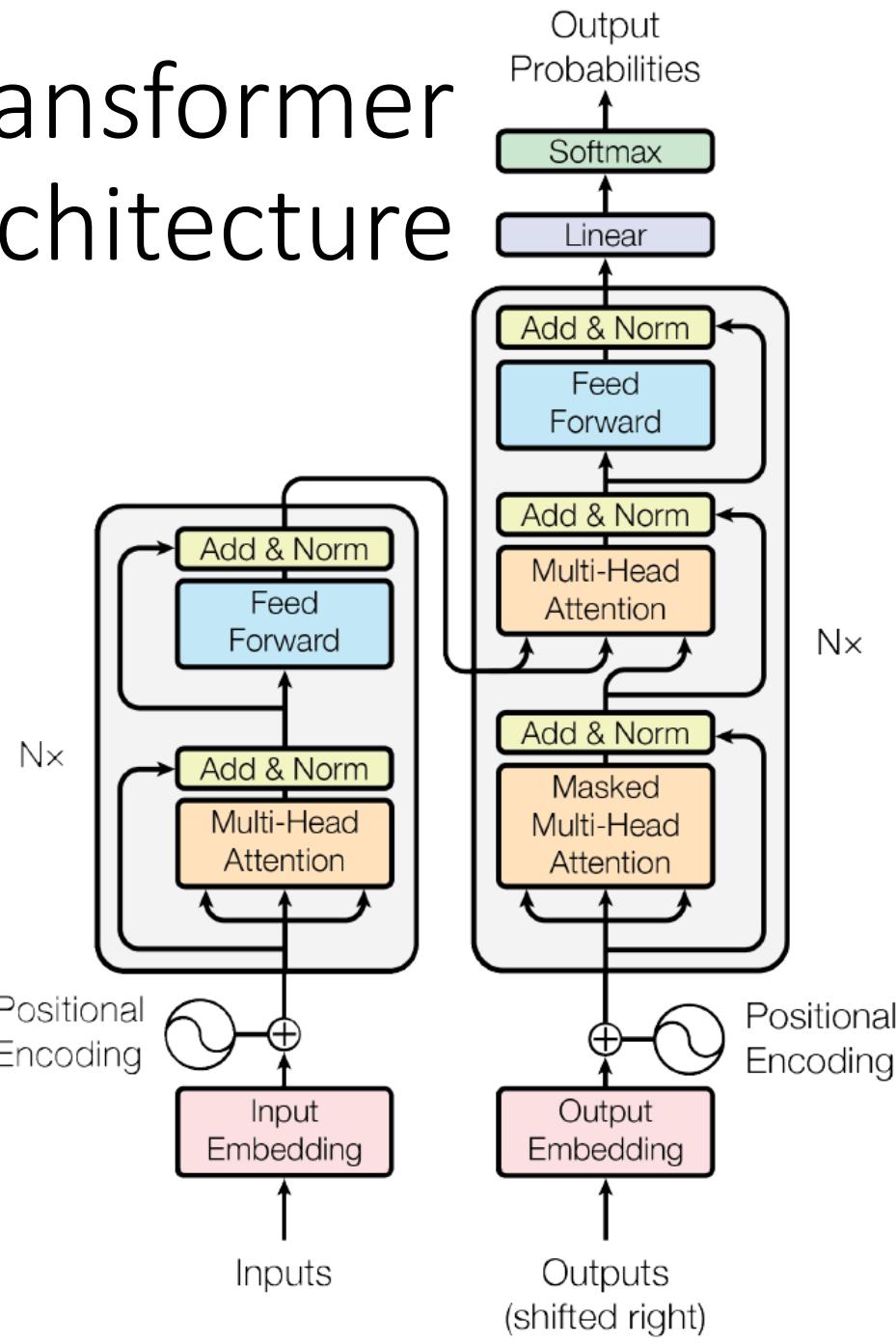
Multi-Head Attention



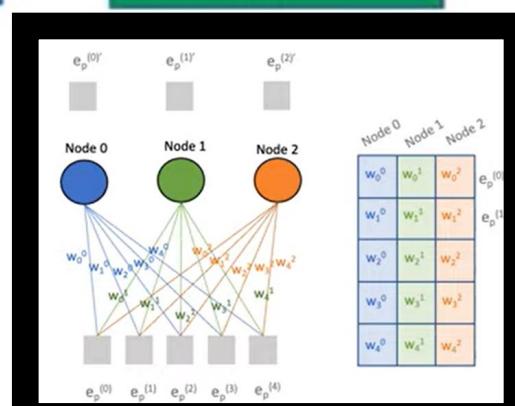
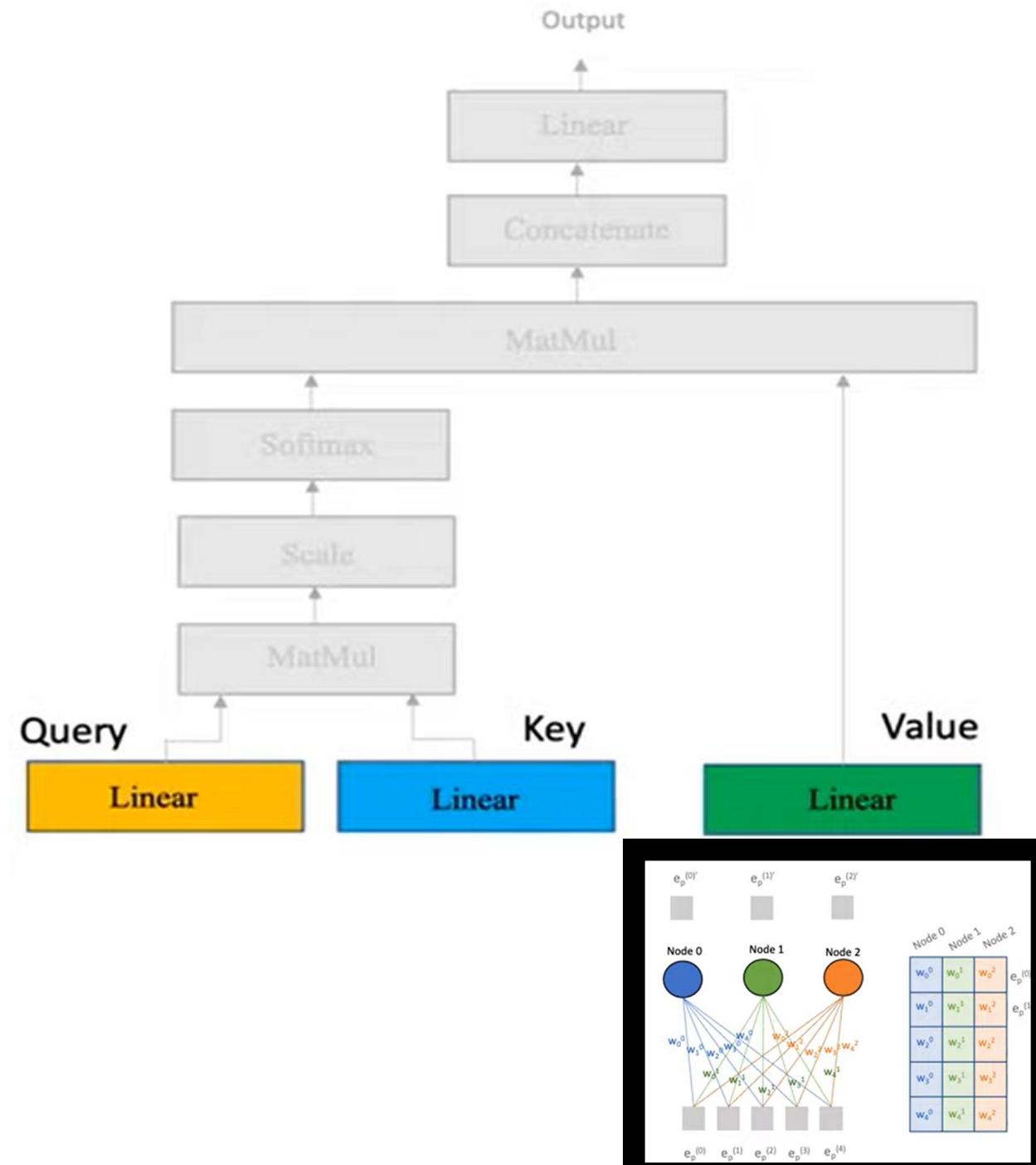
Transformer architecture



Transformer architecture



Multi-Head Attention



When
you
play
the
game
of
thrones



7 x 5

When
you
play
the
game
of
thrones



7 x 5

When
you
play
the
game
of
thrones



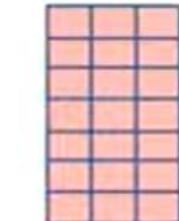
7 x 5

x



5 x 3

=

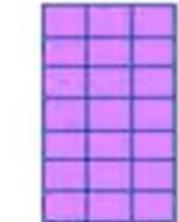


7 x 3

x



=

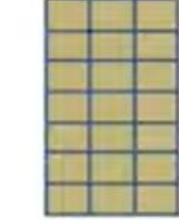


7 x 3

x

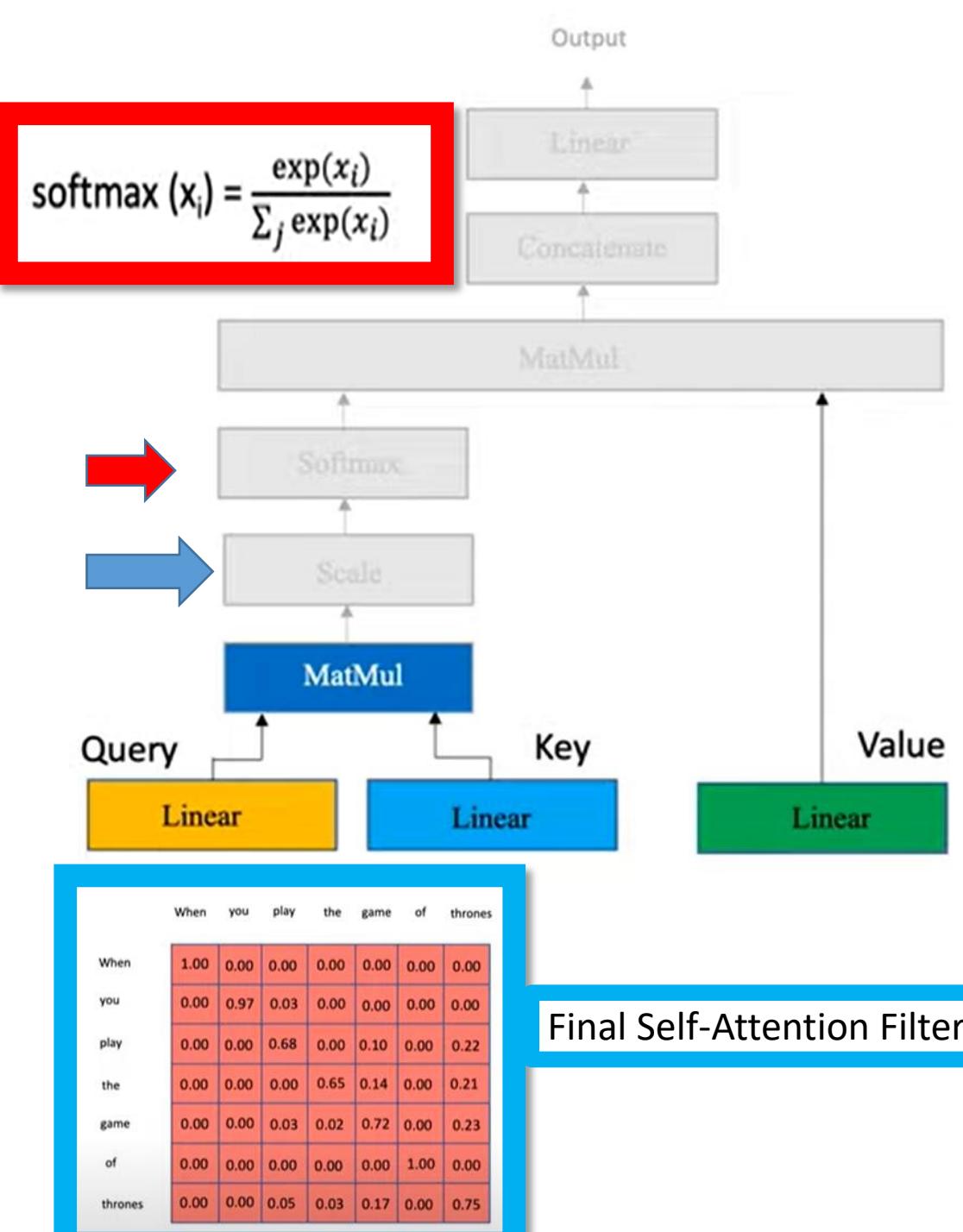


=



7 x 3

Multi-Head Attention



7 x 3

1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7

3 x 7

1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7

7 x 7

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

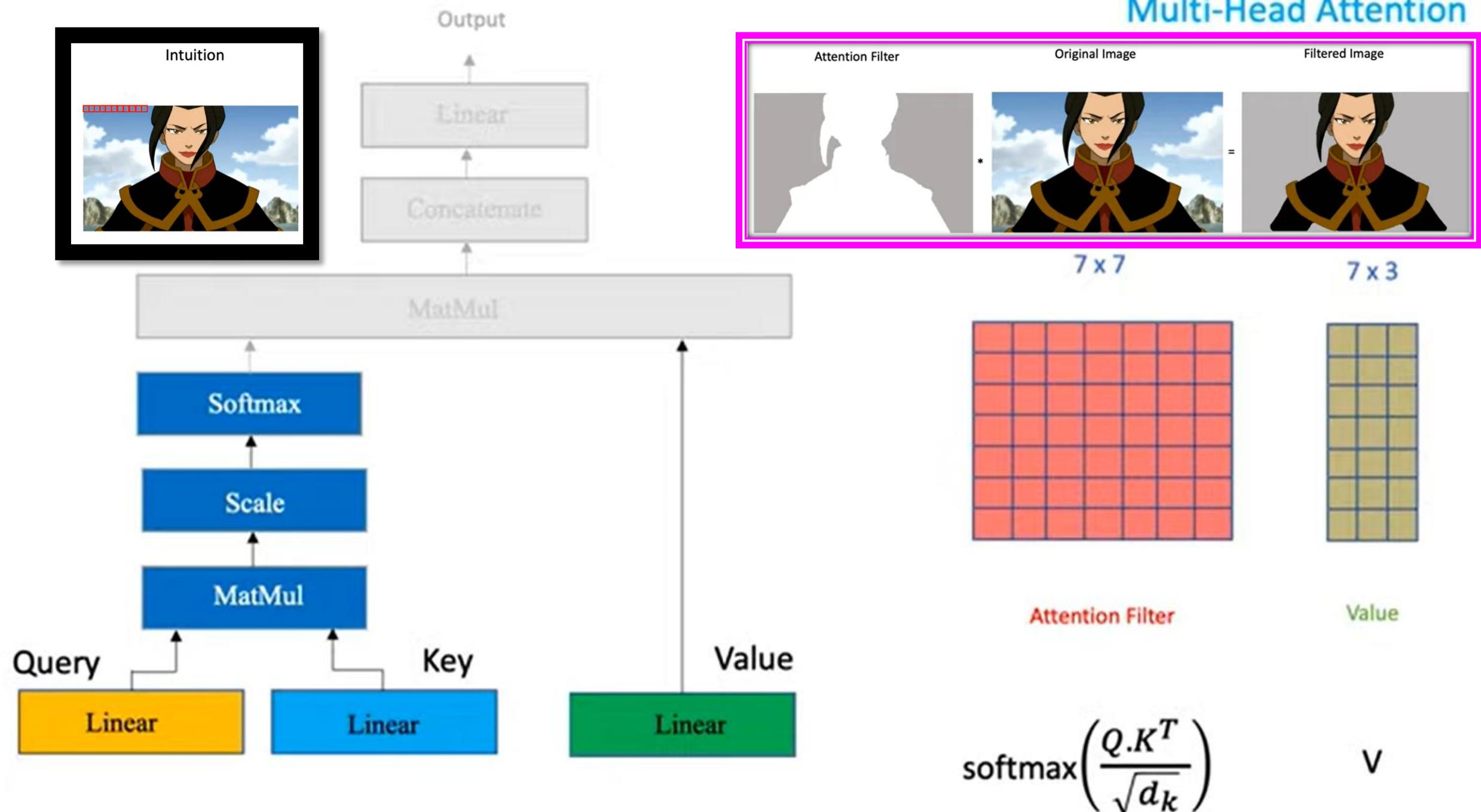
	When	you	play	the	game	of	thrones
When	89	48	41	36	35	40	19
you	67	91	11	92	17	99	11
play	91	10	11	11	12	41	98
the	11	96	28	12	98	11	00
game	76	11	91	24	12	12	12
of	11	29	77	78	22	93	13
thrones	11	87	12	12	13	98	19

Initially

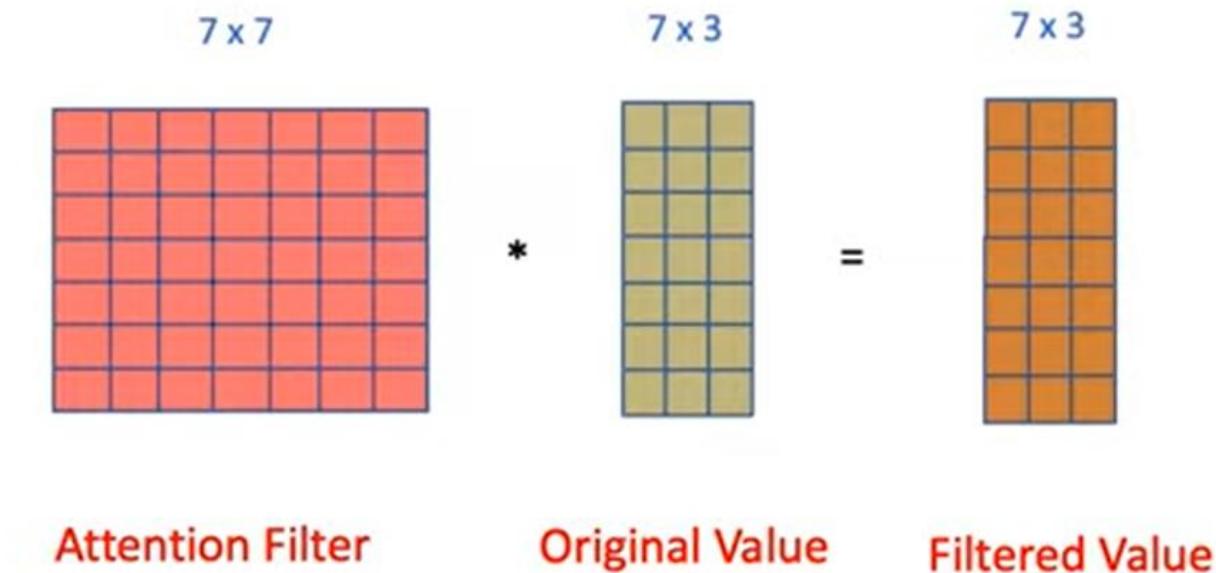
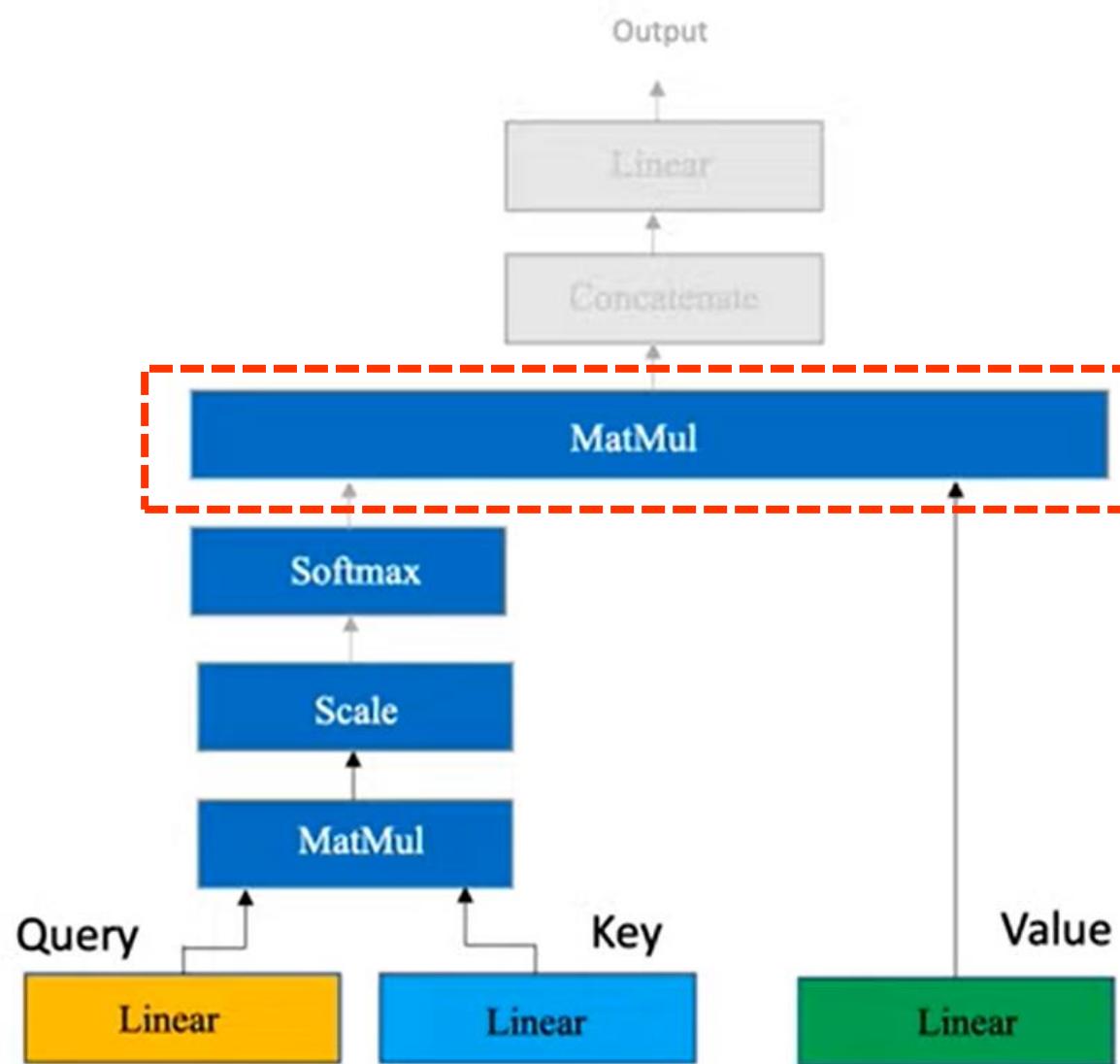
	When	you	play	the	game	of	thrones
When	89	20	41	10	55	10	59
you	20	90	81	22	70	15	72
play	41	81	95	10	90	30	92
the	10	22	10	92	88	40	89
game	55	70	90	88	98	44	87
of	10	15	30	40	44	85	59
thrones	59	72	92	90	95	59	99

$$\sqrt{d_k}$$

Multi-Head Attention

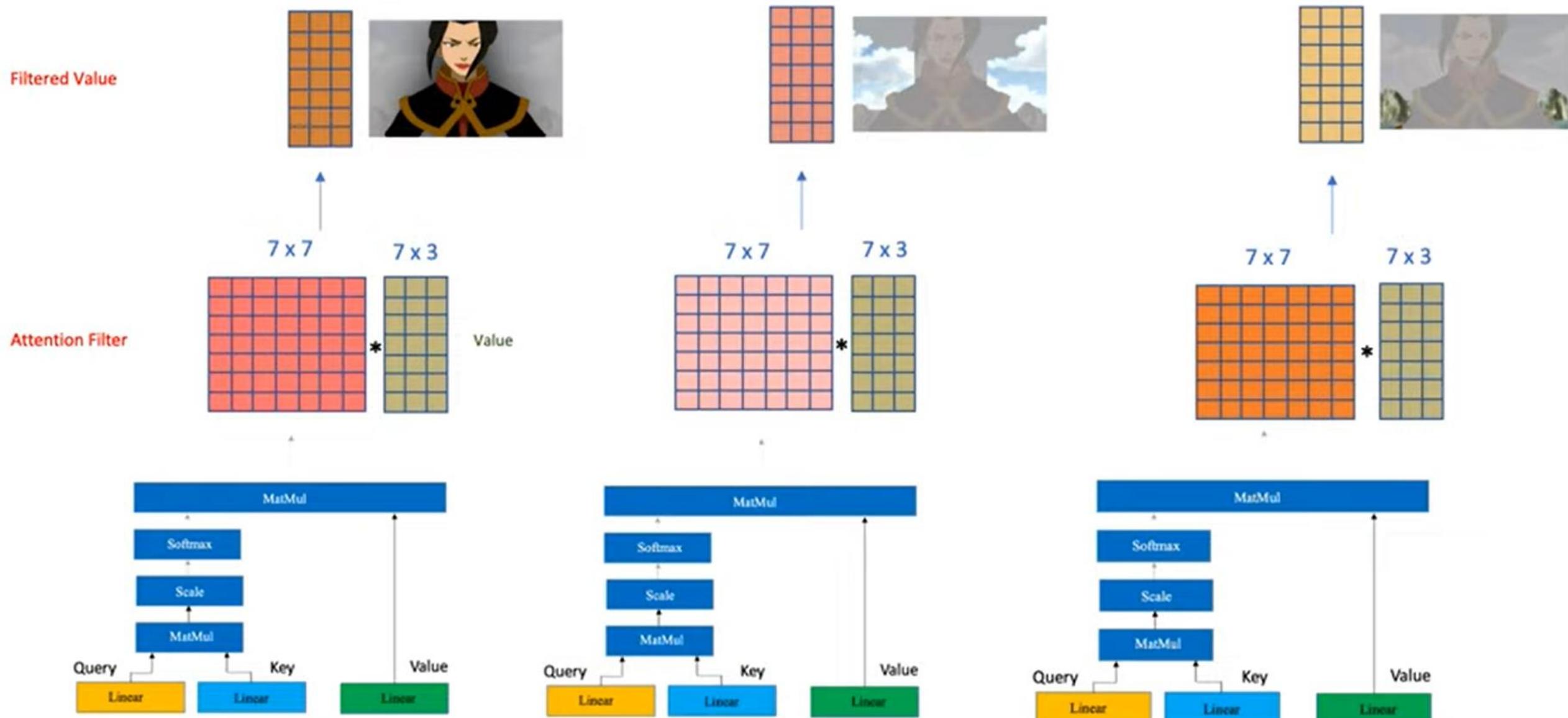


Multi-Head Attention



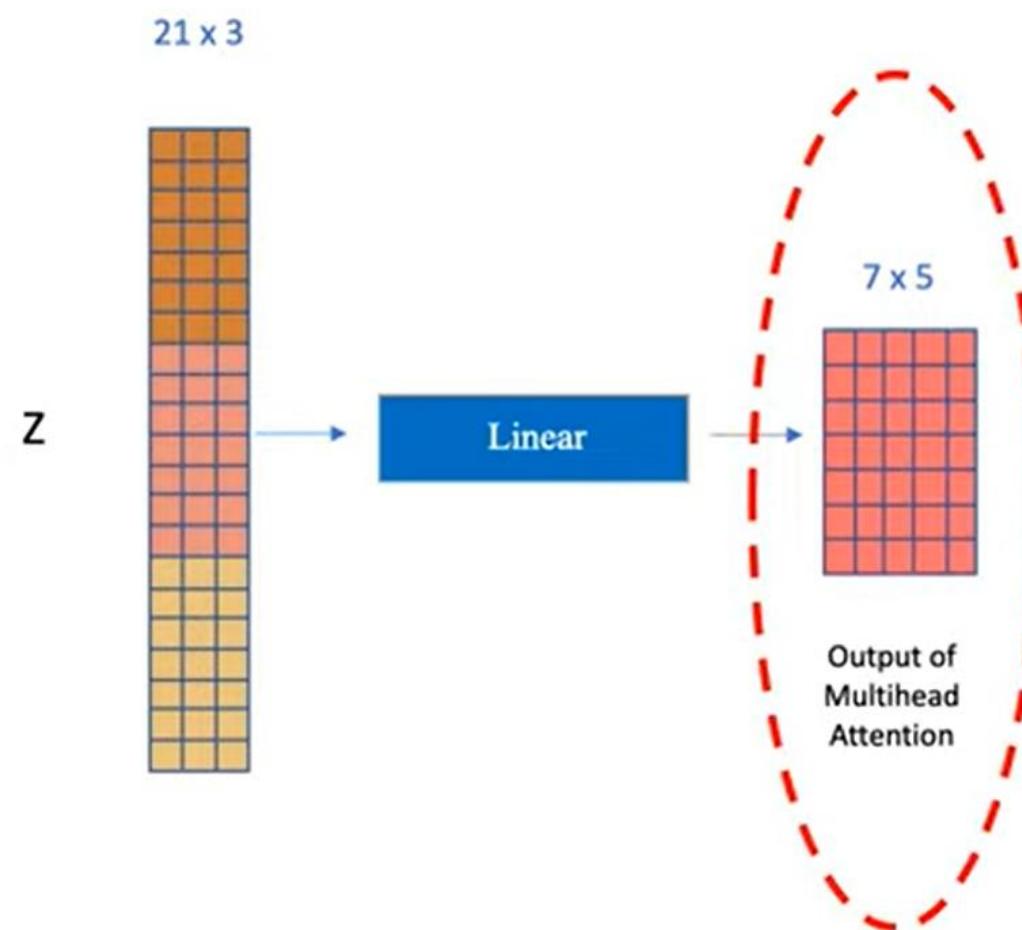
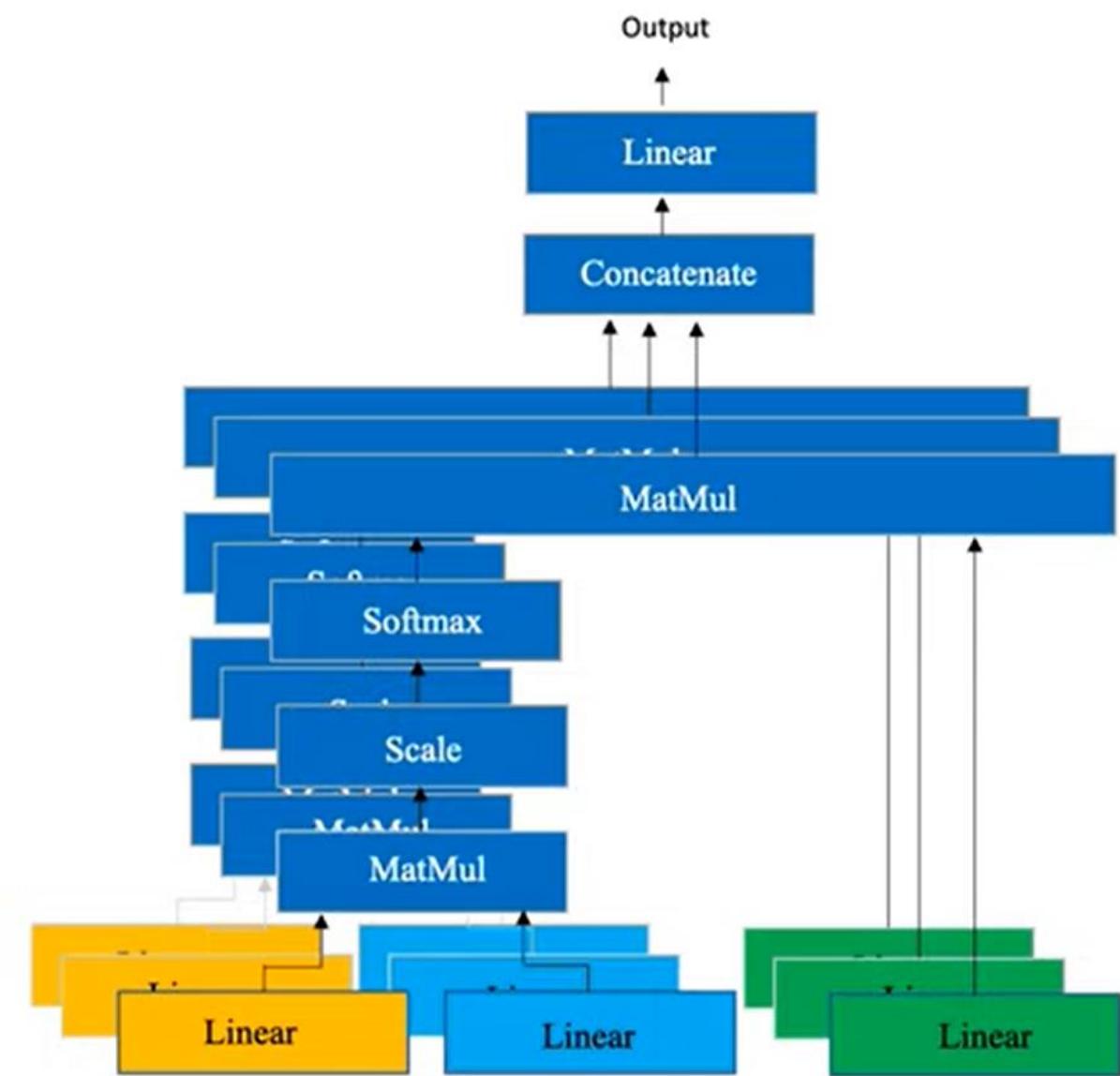
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V$$

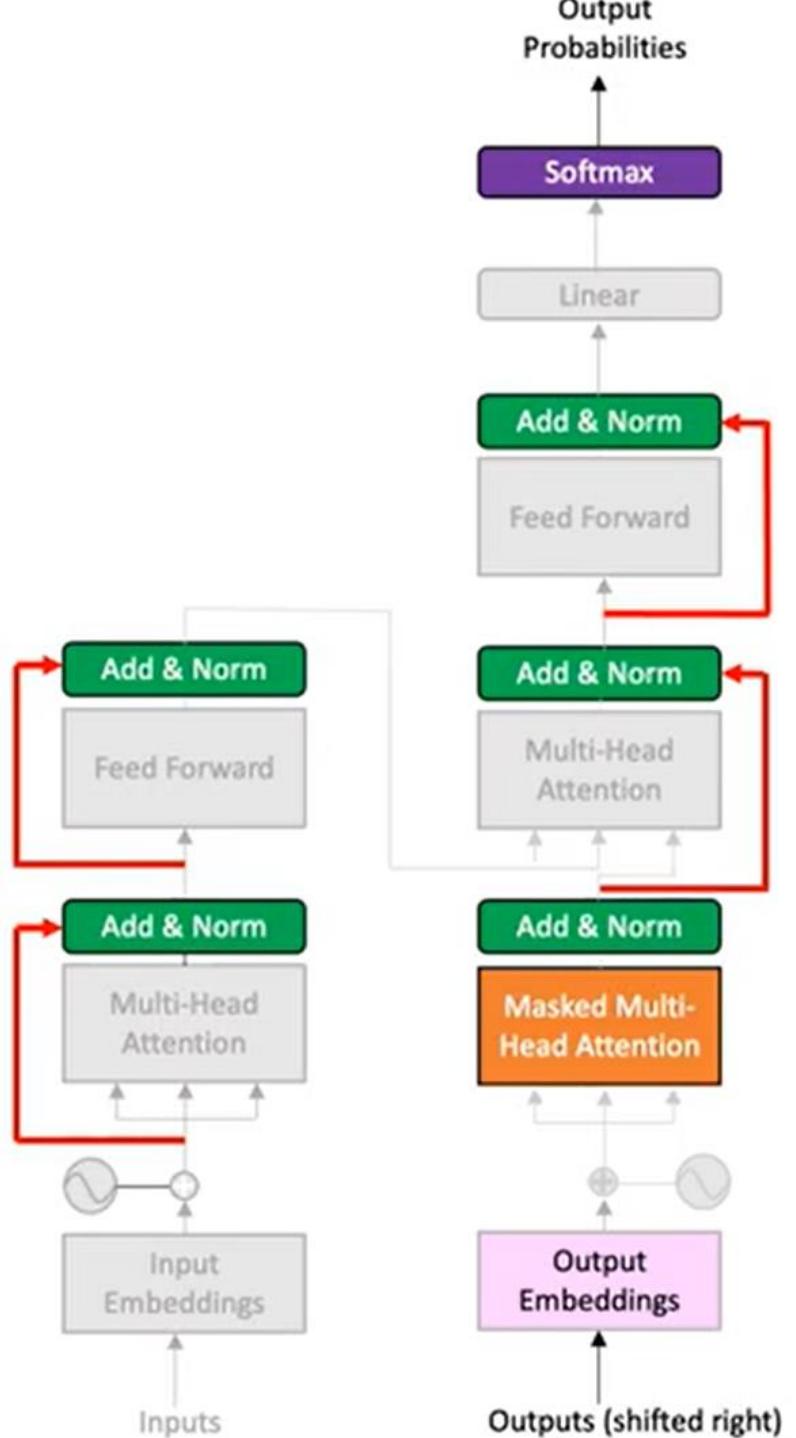
Multi-Head Attention



Original paper: 8 attention heads

Multi-Head Attention





- i) Knowledge Preservation
- ii) Vanishing Gradient Problem

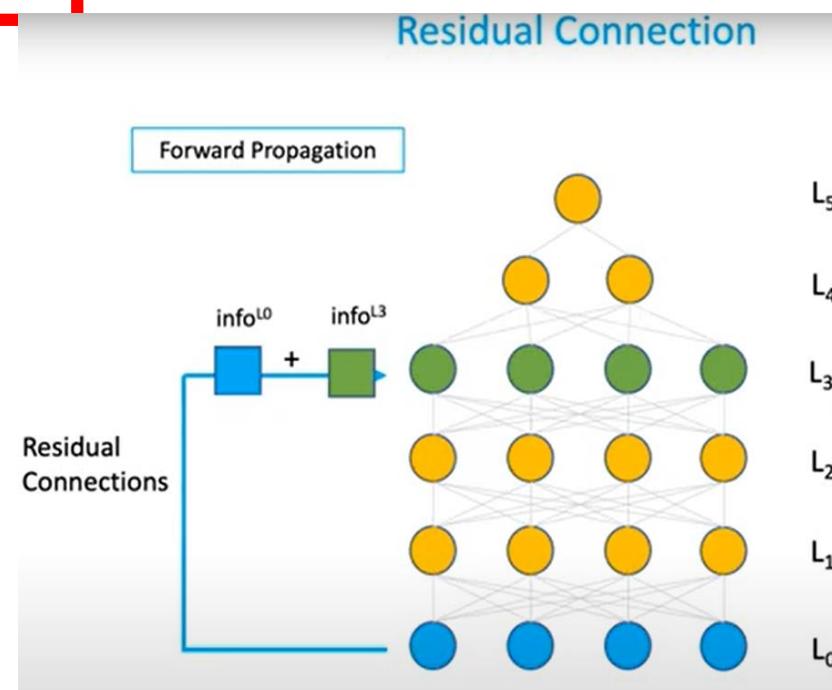
1. Residual Connections

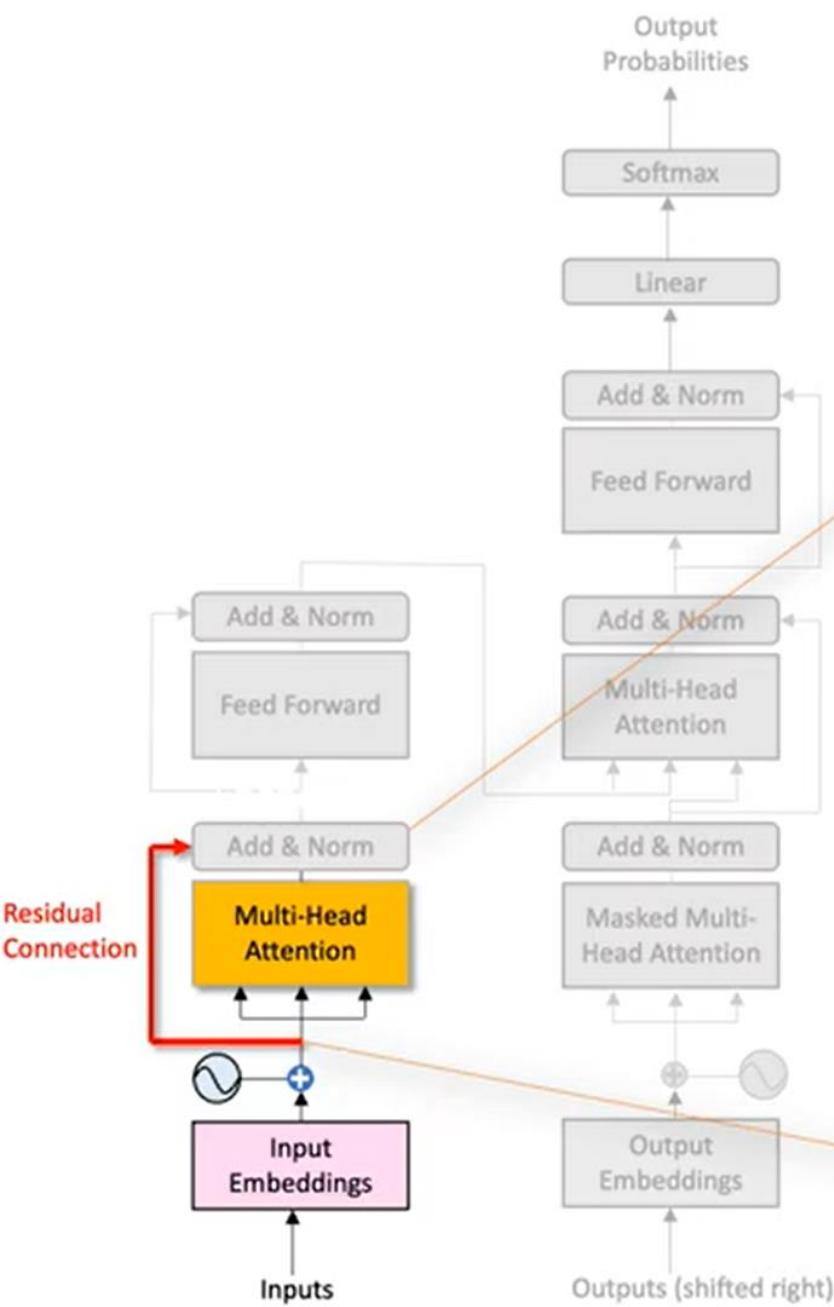
2. Layer Normalization

3. Decoder

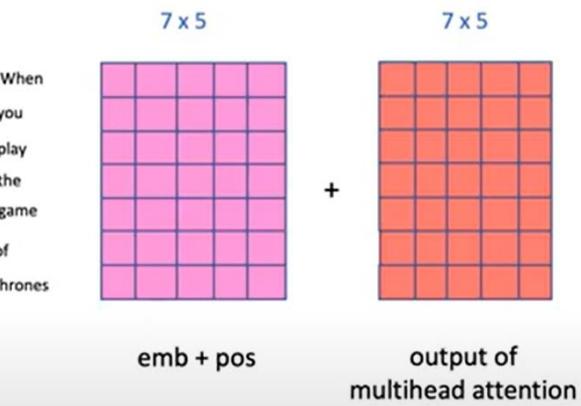
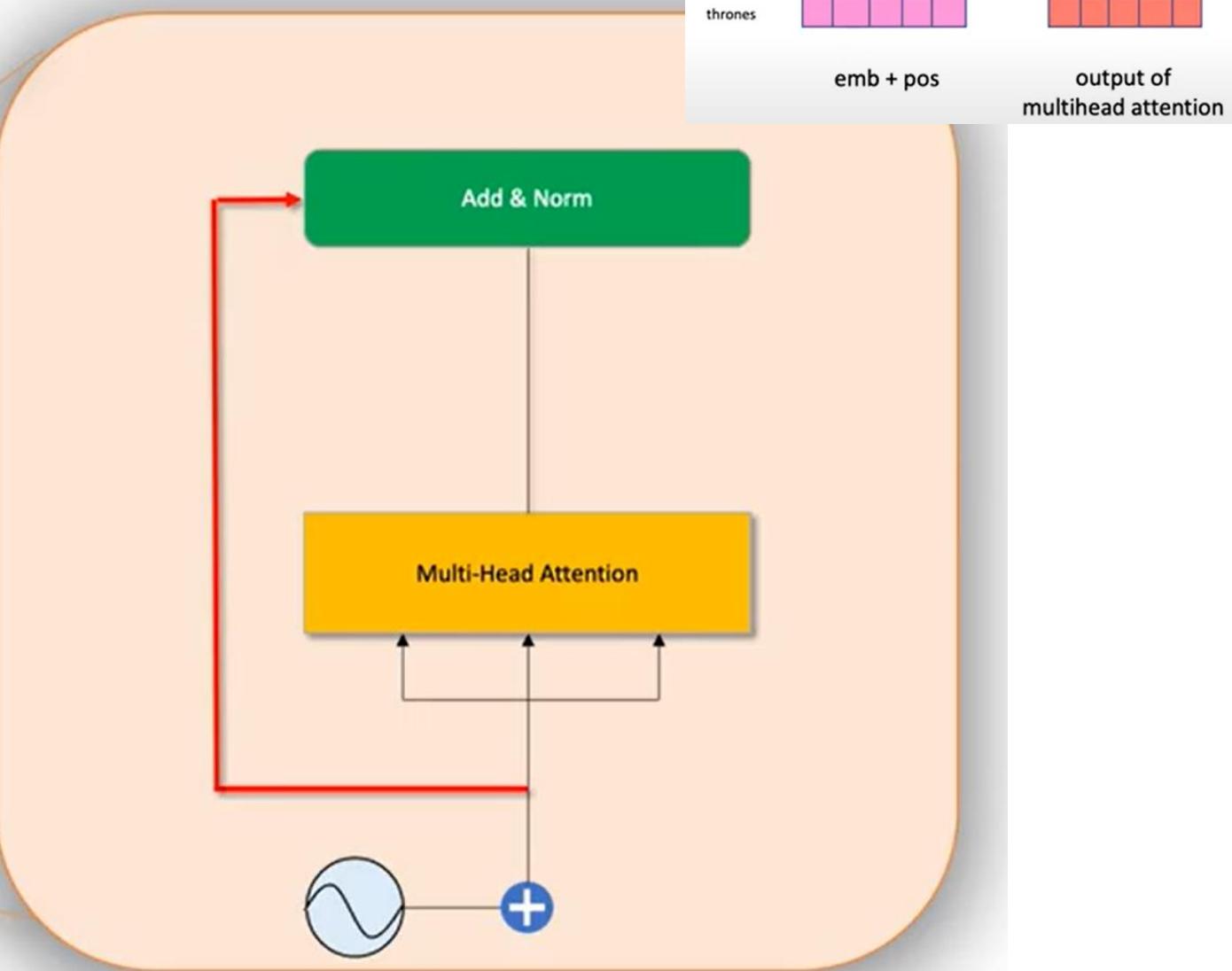
i. Important components

ii. Masked-Attention



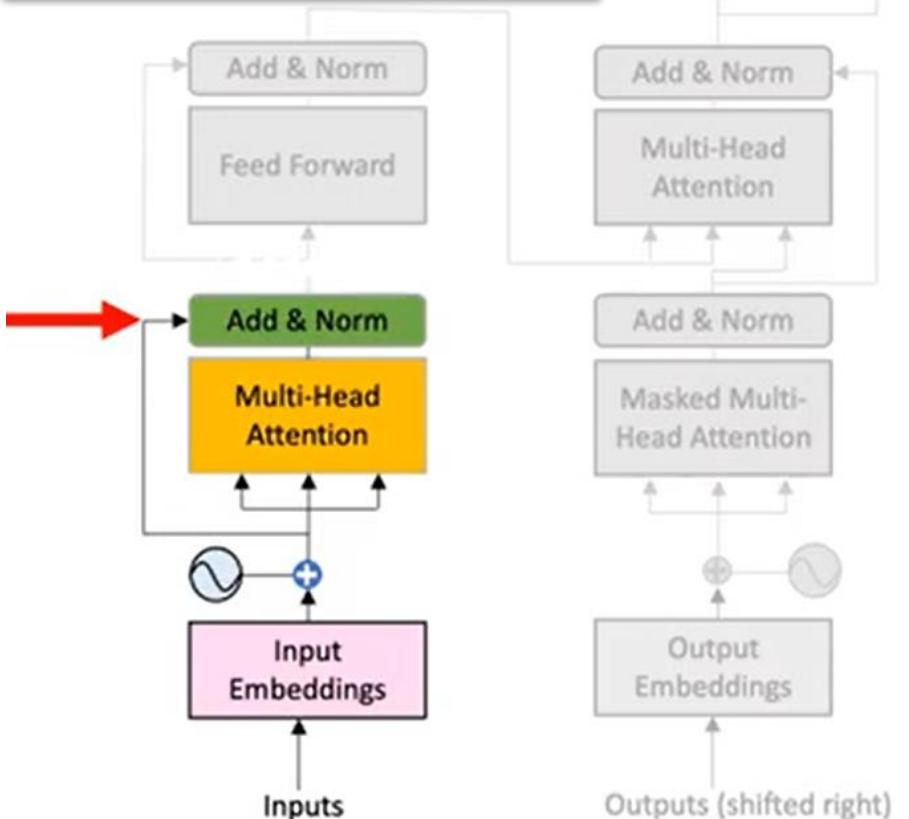


Residual Connection



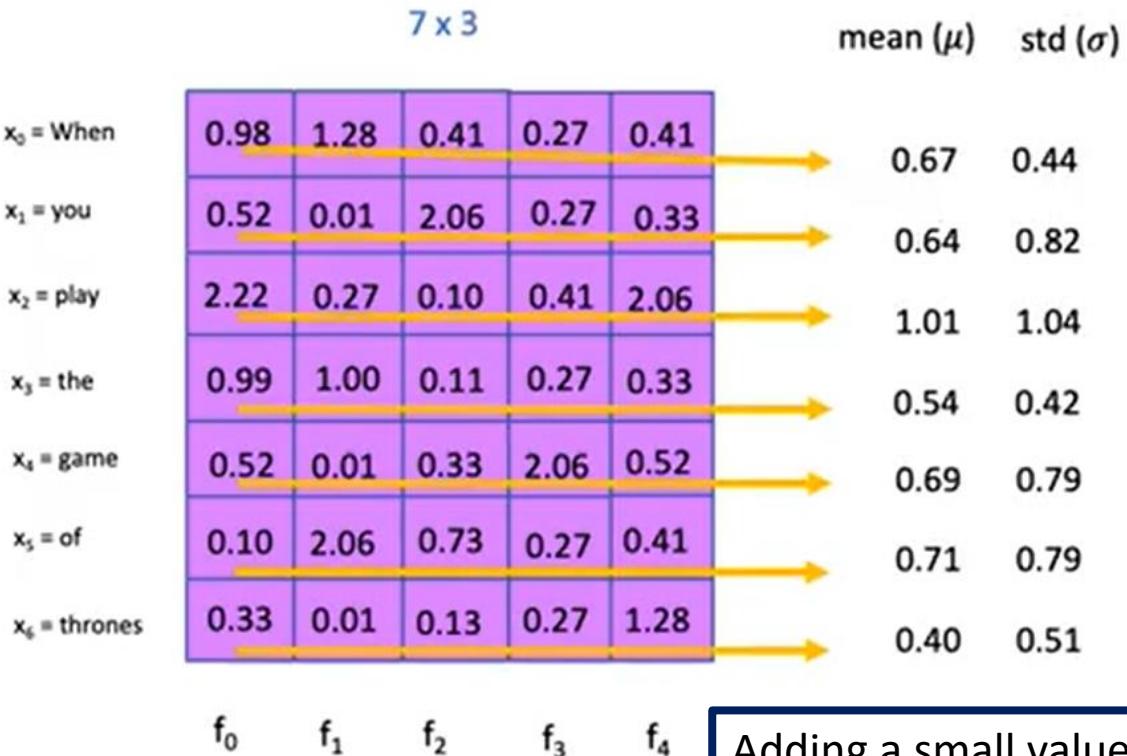
Post-normalization

$x_0 = \text{When}$	0.71	1.49	-0.59	-0.92	-0.59
$x_1 = \text{you}$	0.14	-0.77	1.74	-0.85	-0.38
$x_2 = \text{play}$	1.16	-0.72	-0.88	-0.58	1.01
$x_3 = \text{the}$	1.06	1.09	-1.02	-0.64	0.50
$x_4 = \text{game}$	-0.21	-0.85	-0.45	1.73	-0.21
$x_5 = \text{of}$	-0.78	1.71	0.02	-0.56	-0.39
$x_6 = \text{thrones}$	0.15	-0.78	-0.54	-0.27	1.73



Layer Normalization

Layer Normalization means standardization of neuron activation along the axis of feature

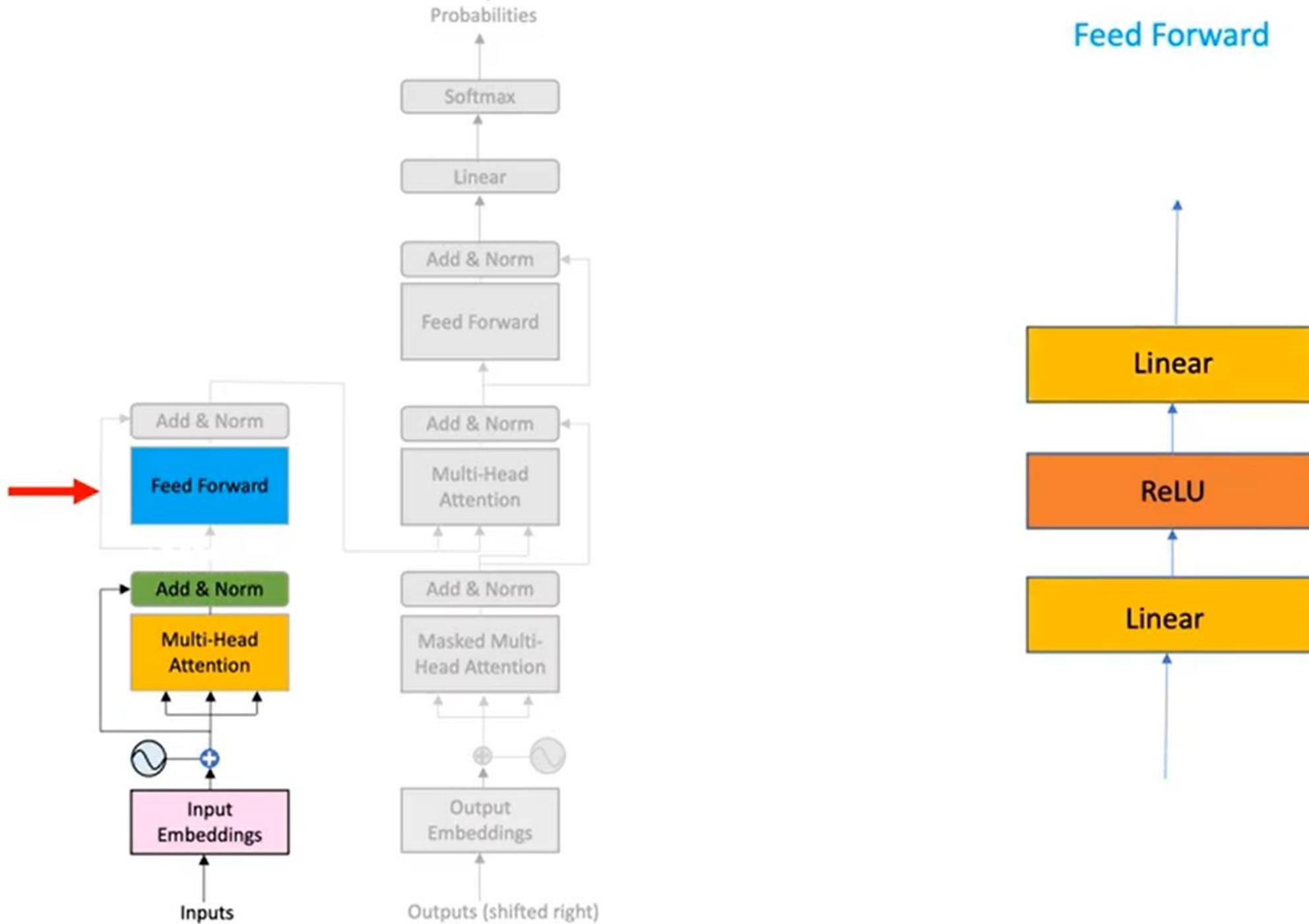


Adding a small value to avoid dividing by 0

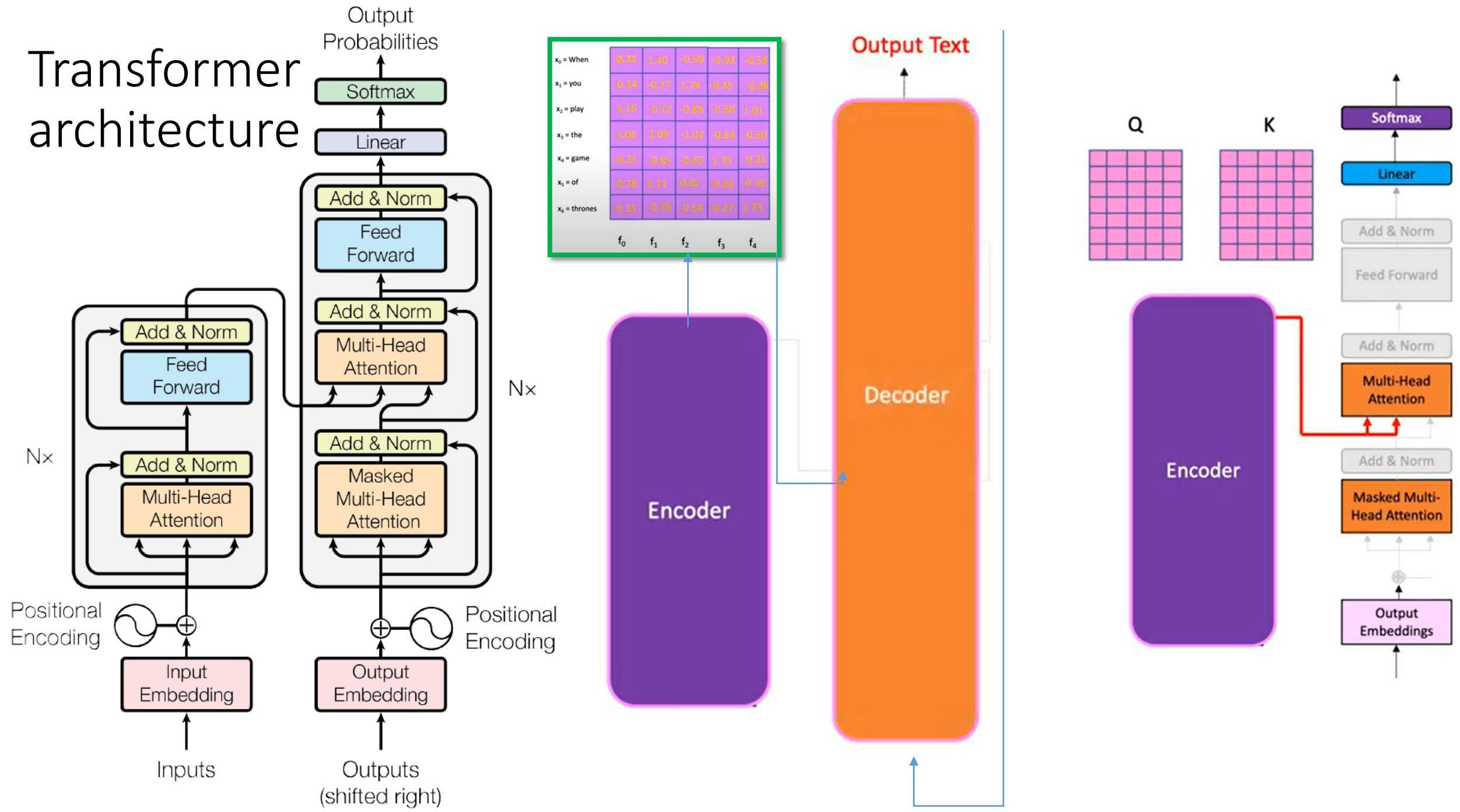
$$x_i = \frac{x_i^d - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

$$x_0 = \frac{0.98 - 0.67}{\sqrt{0.44^2 + 0.0001}}$$

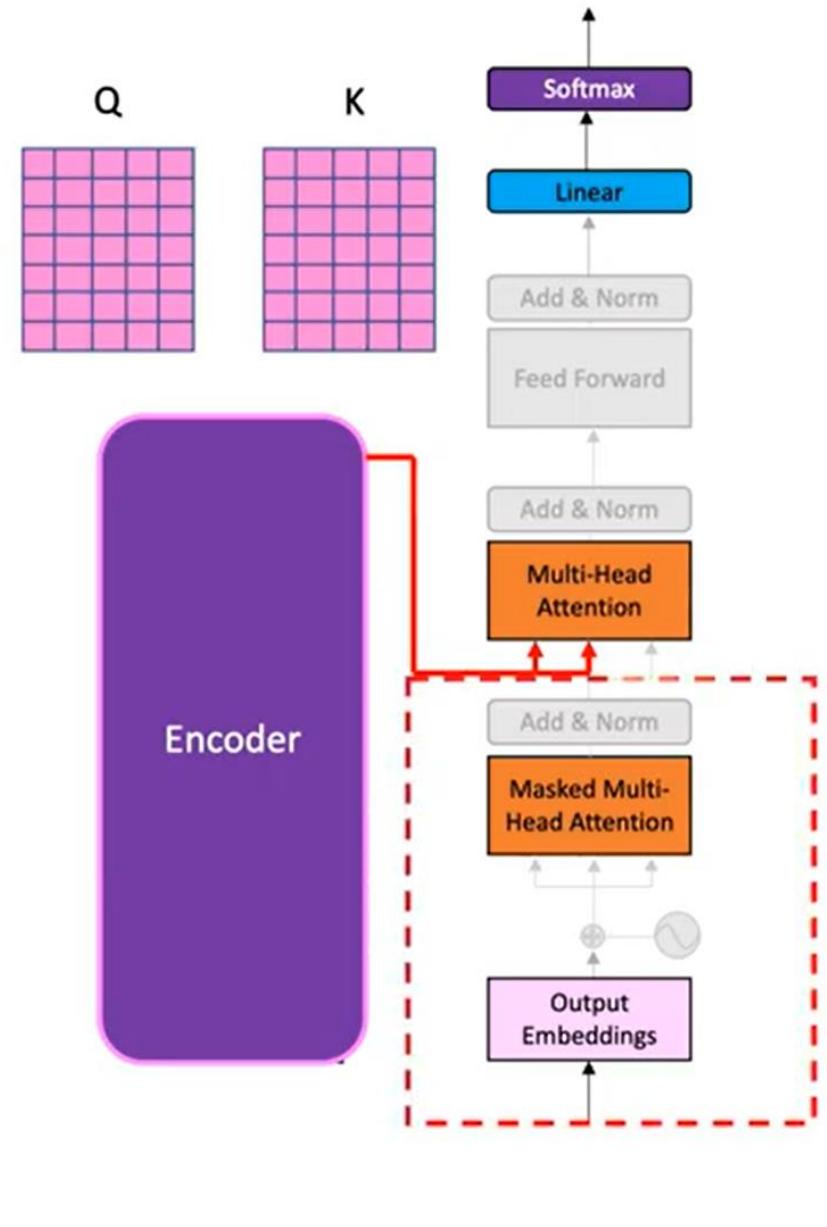
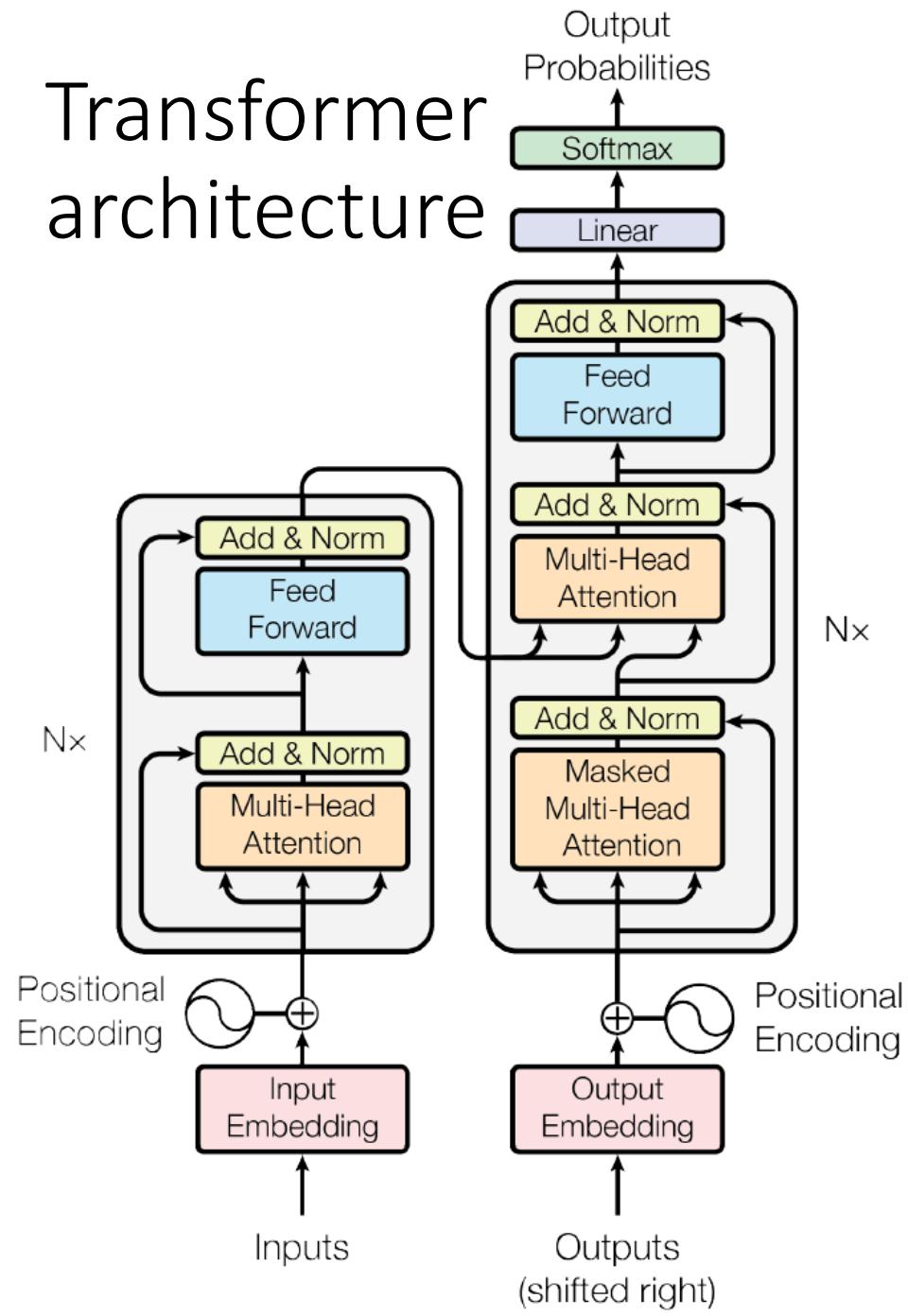
Feed Forward

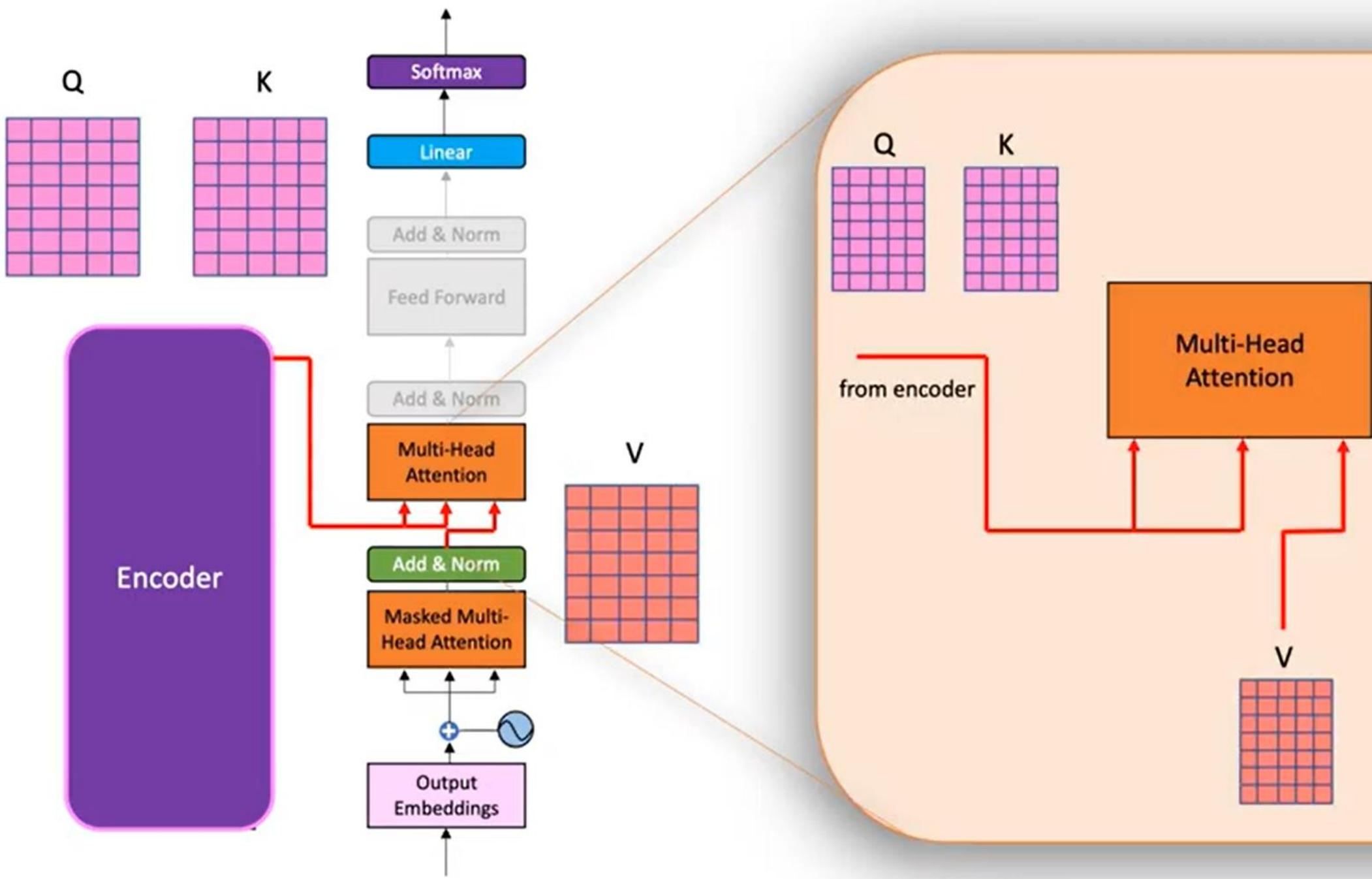


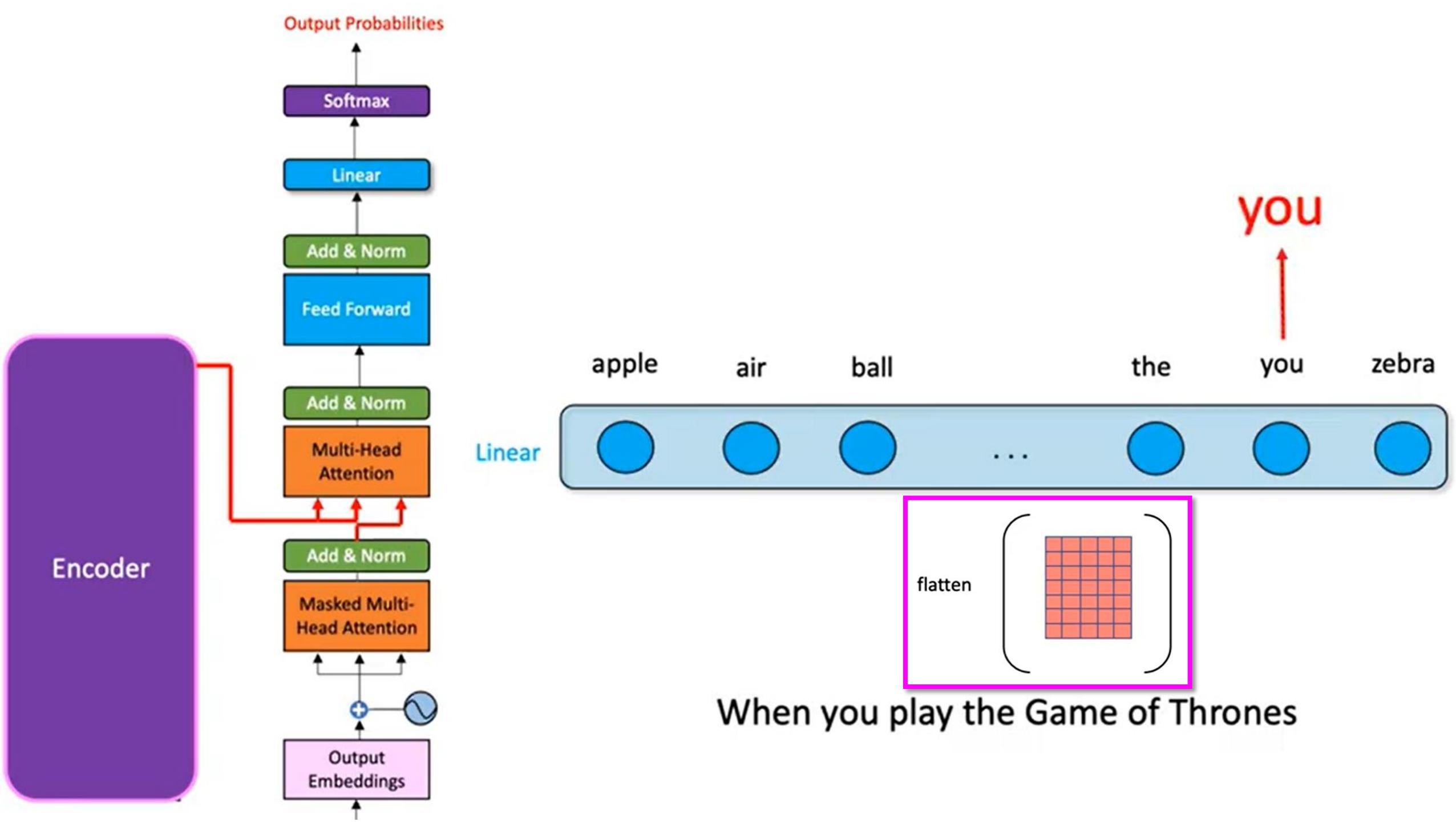
Transformer architecture

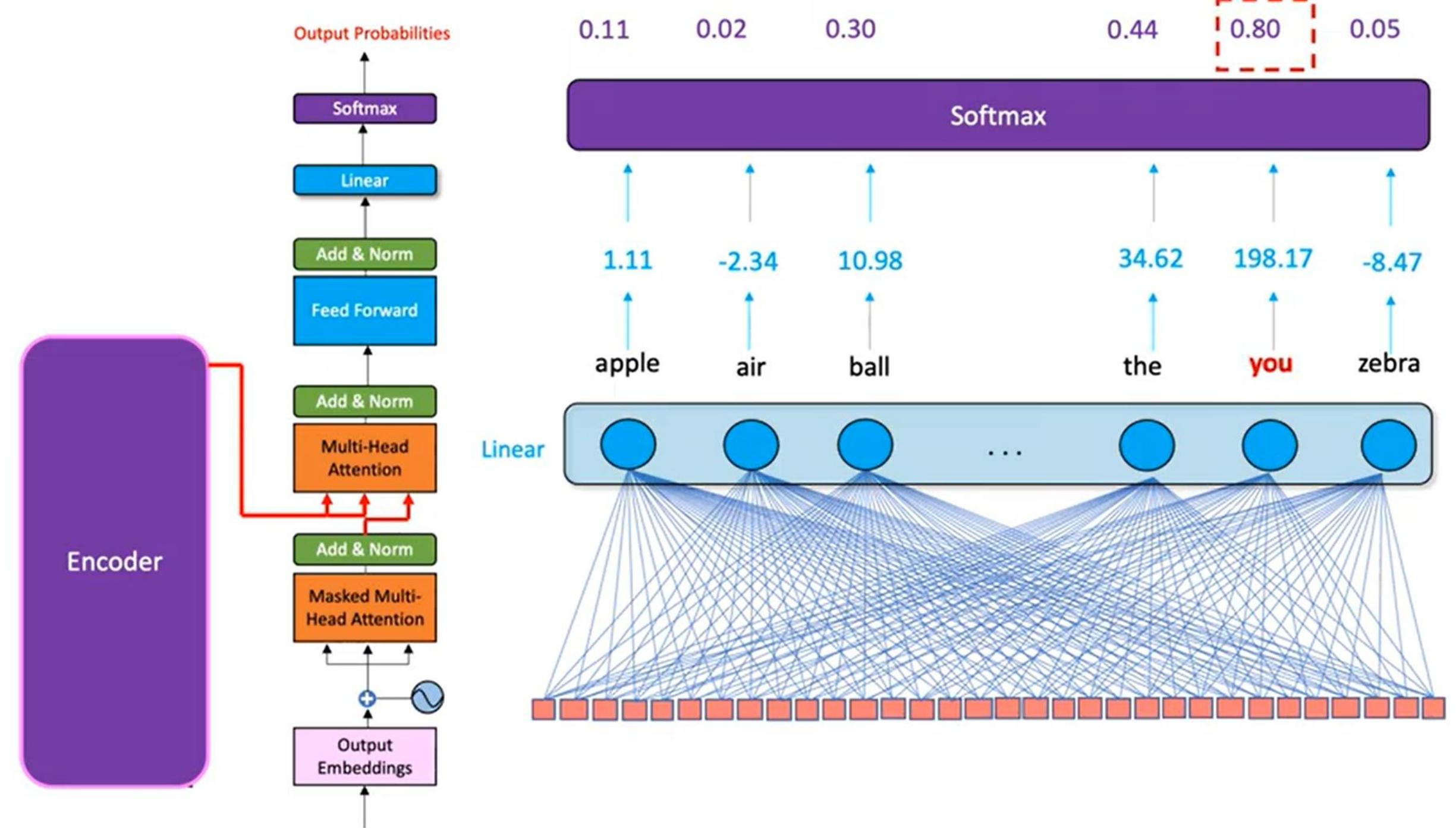


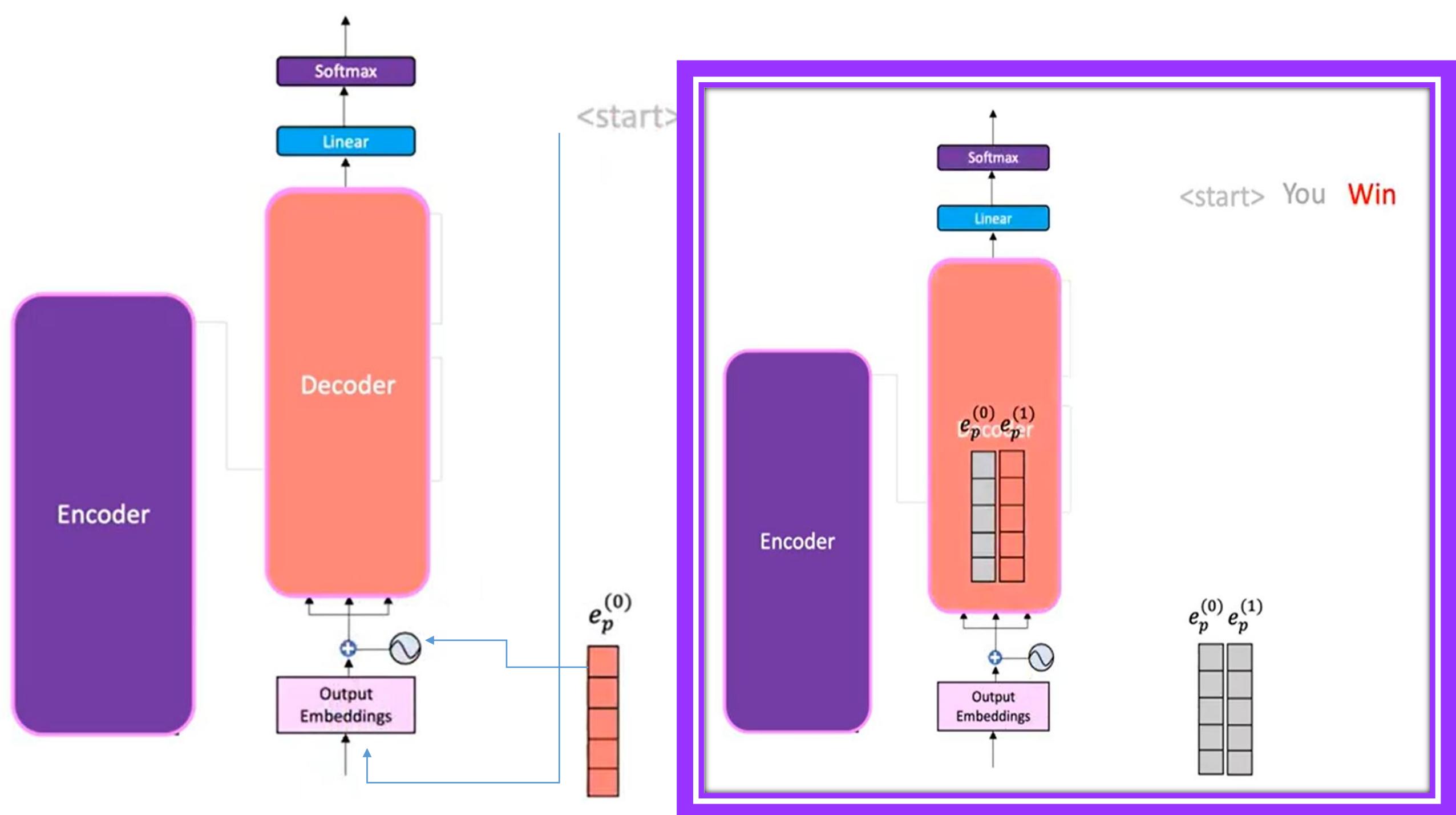
Transformer architecture

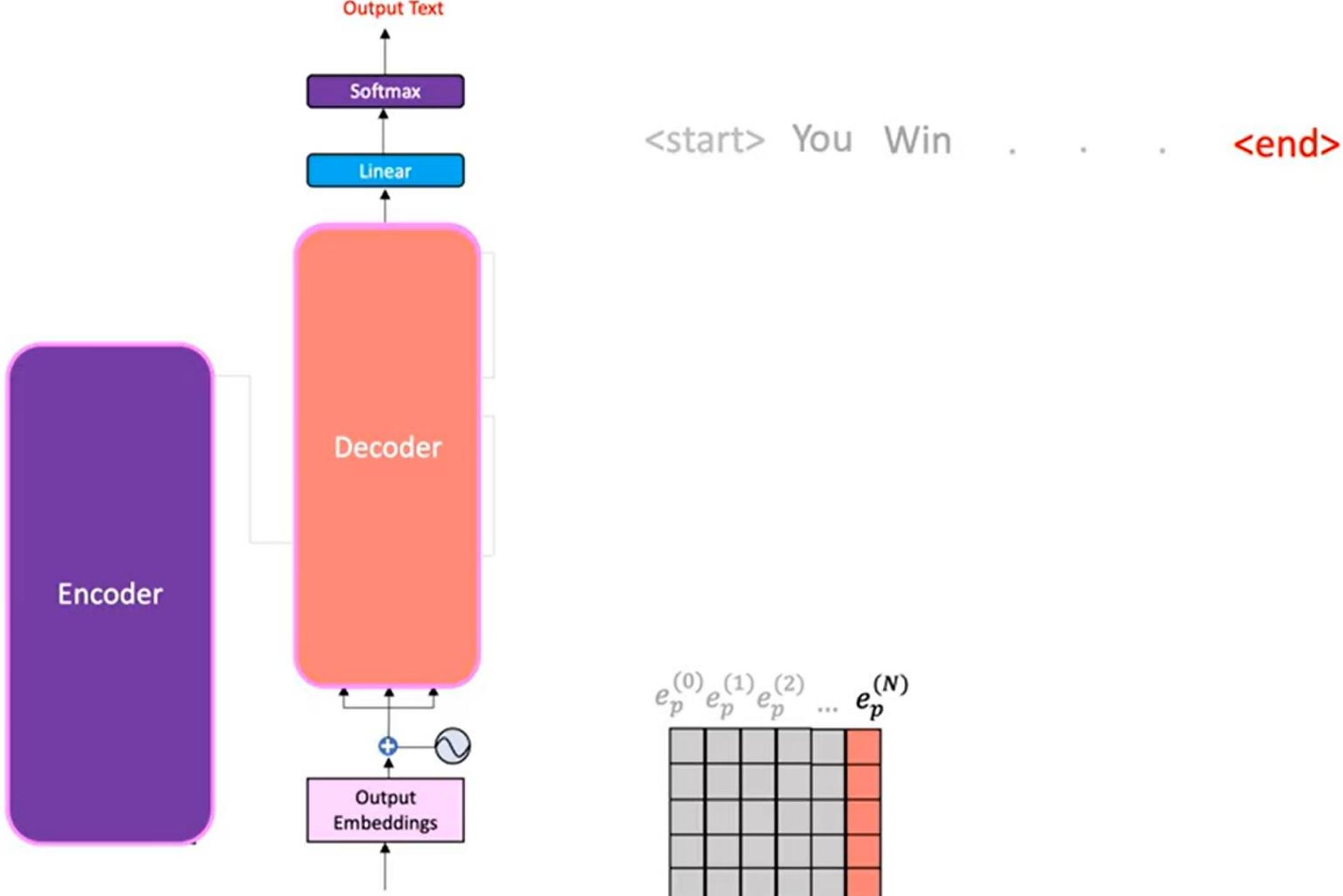


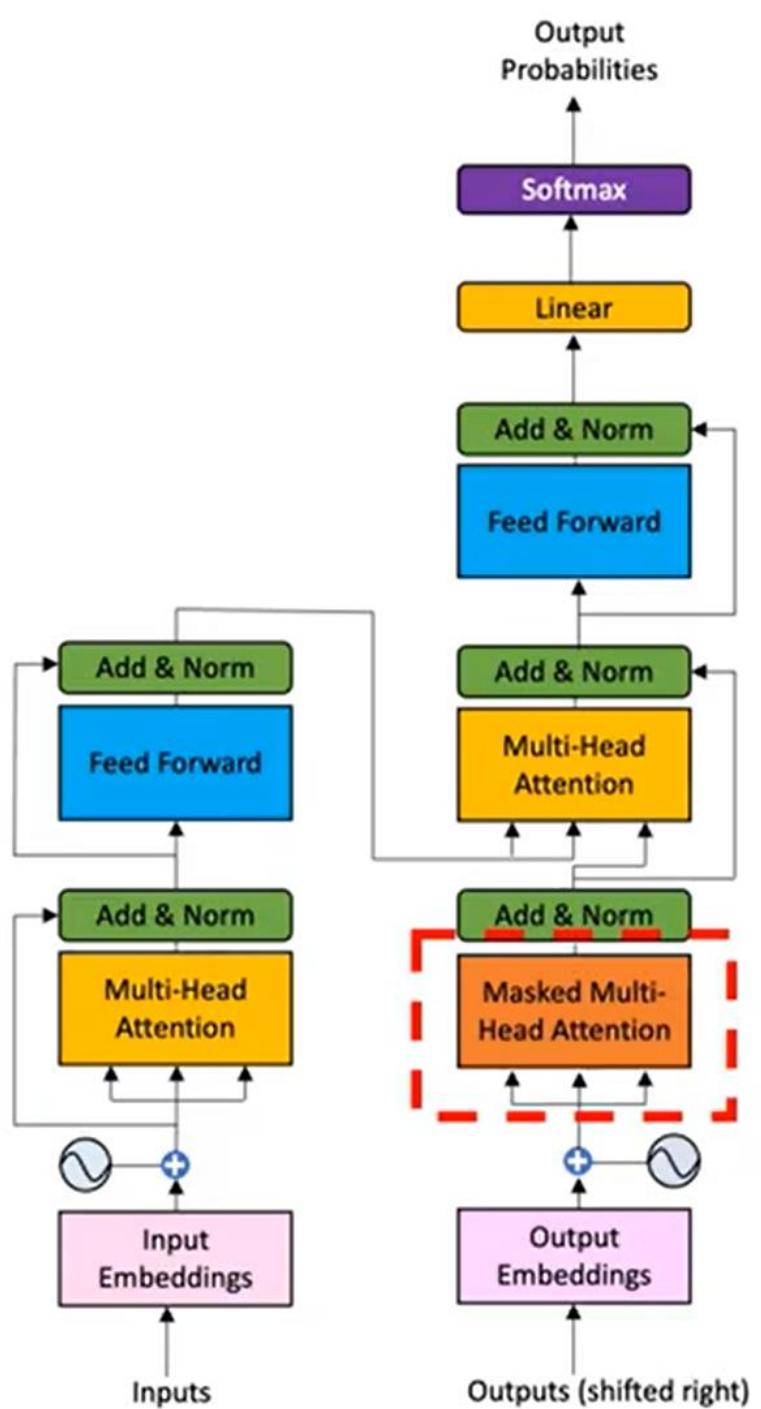












Training

Dialogue PART 1

What do you say to the God of Death
It is not our abilities that show who we truly are
Life happens where ever you are
All we have to do is decide what to do
There is some good in this world Mr. Frodo

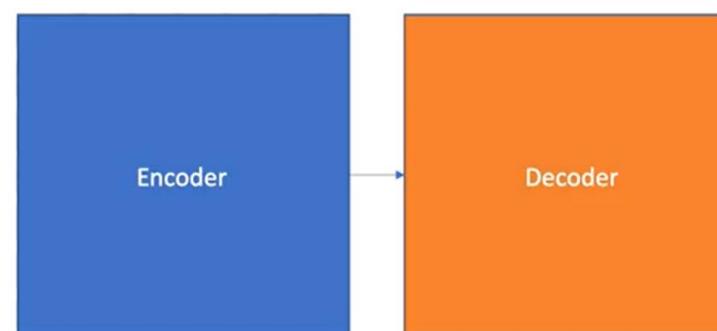
Dialogue PART 2

<start> Not today <end>
<start> it is our choices <end>
<start> whether you make it or not <end>
<start> with the time that has been given to us <end>
<start> and it is worth fighting for <end>

Transformer - Training

Dialogue PART 1

Fool
!
No
man
can
kill
me

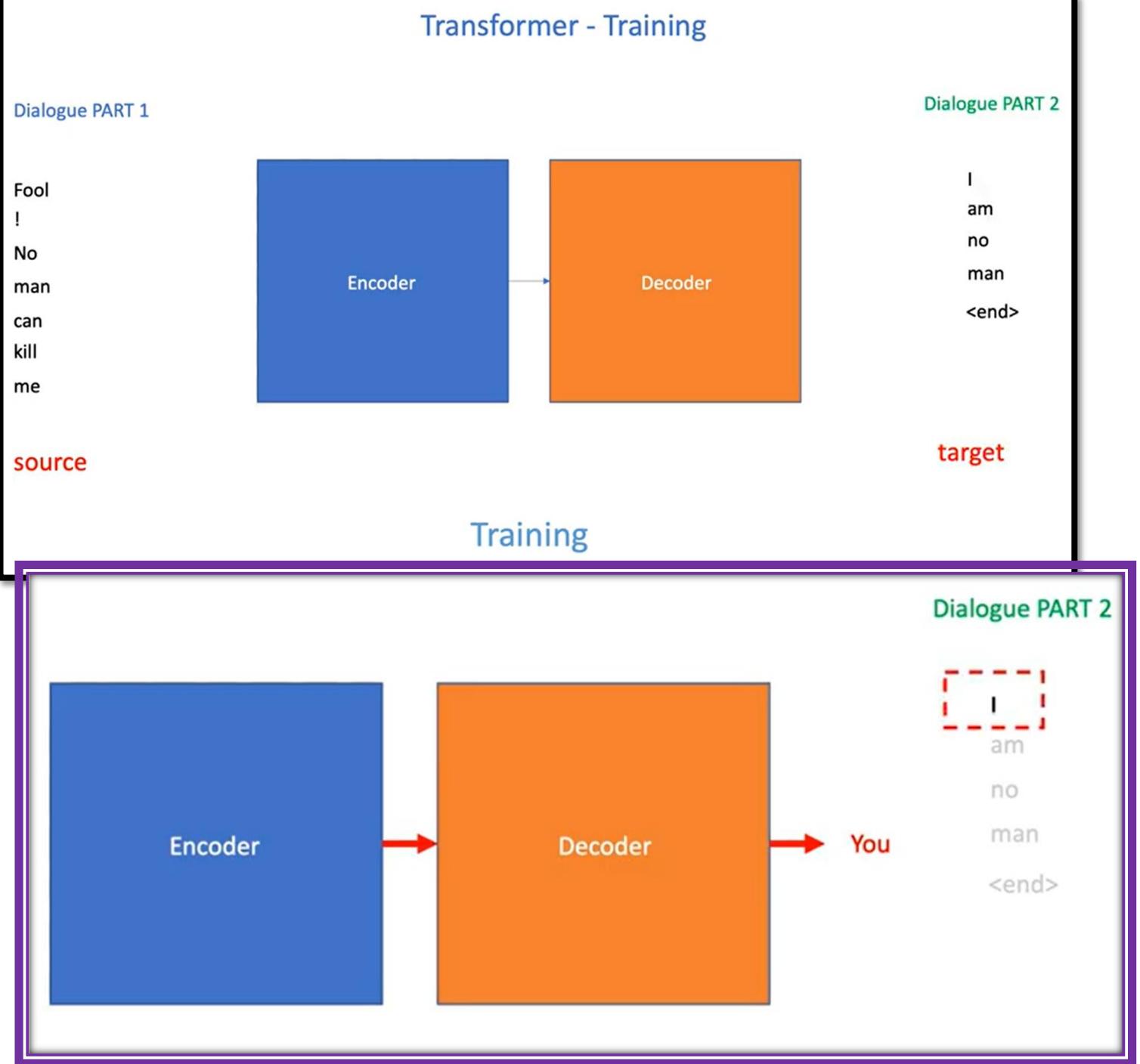
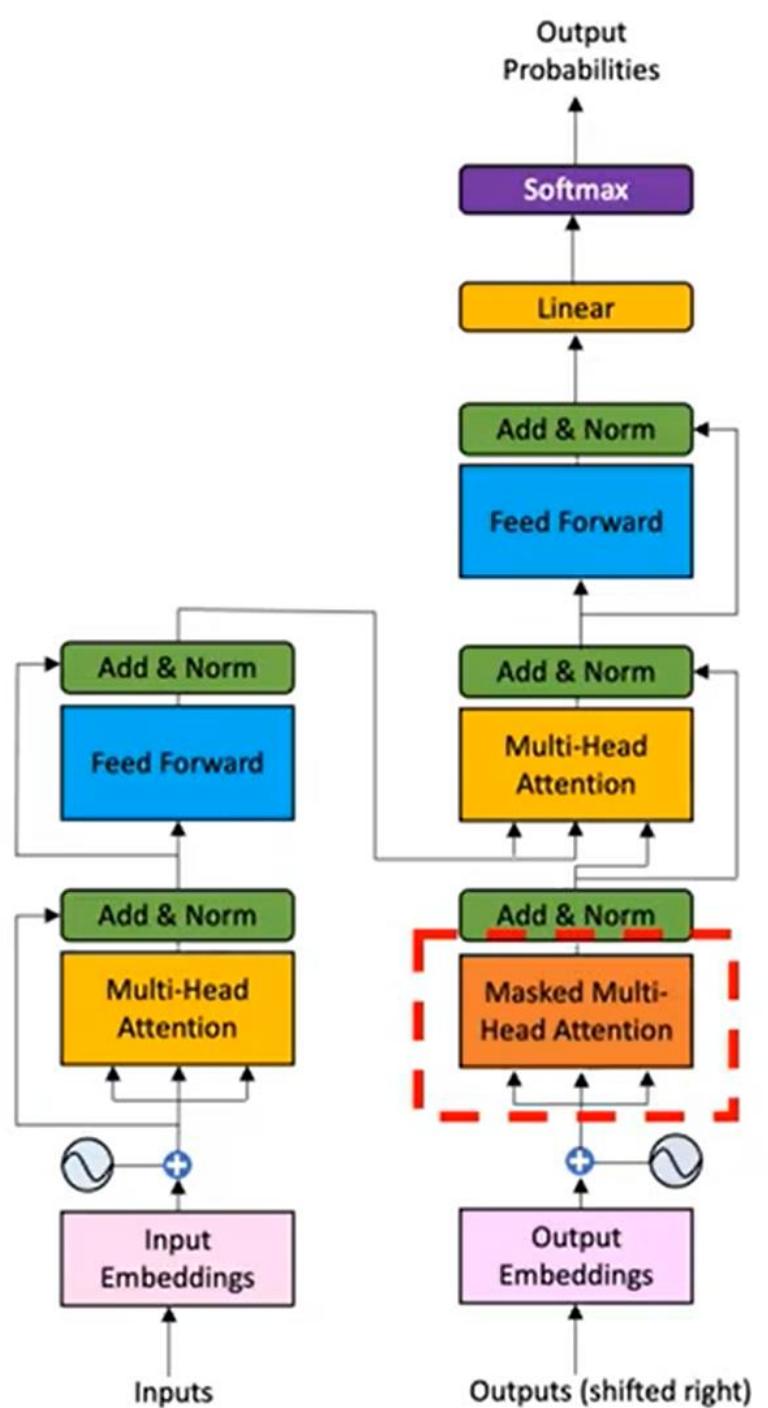


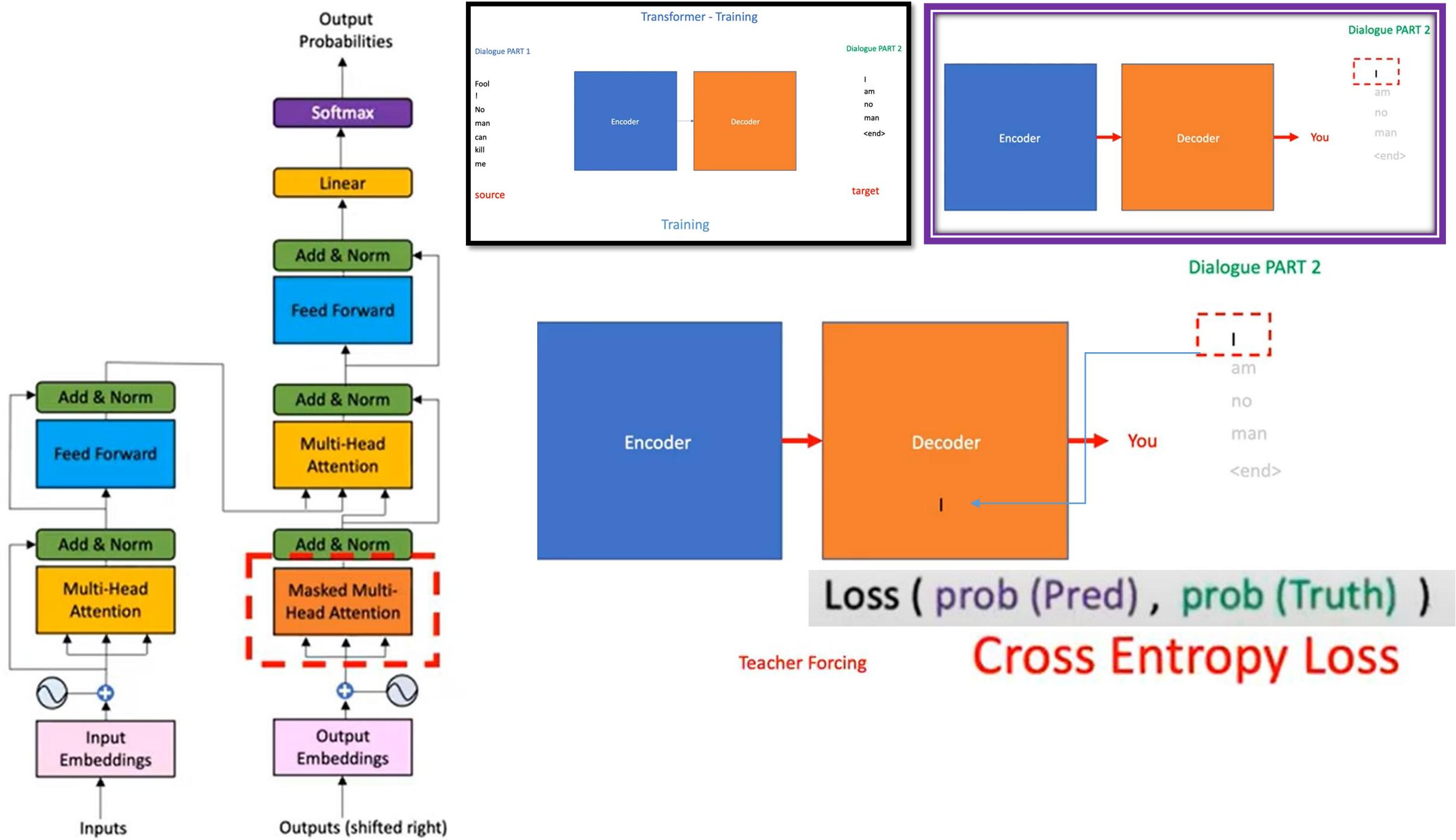
Dialogue PART 2

I
am
no
man
<end>

target

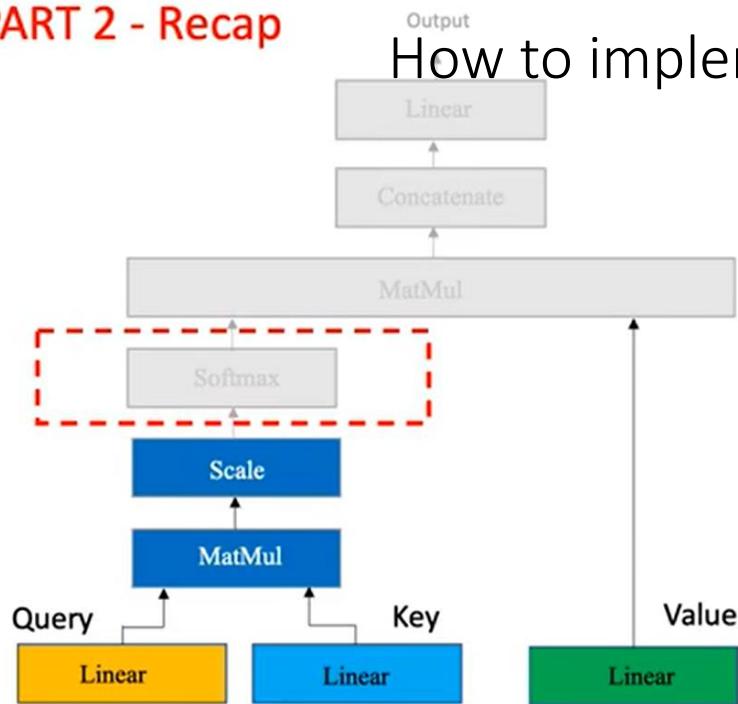
Training





PART 2 - Recap

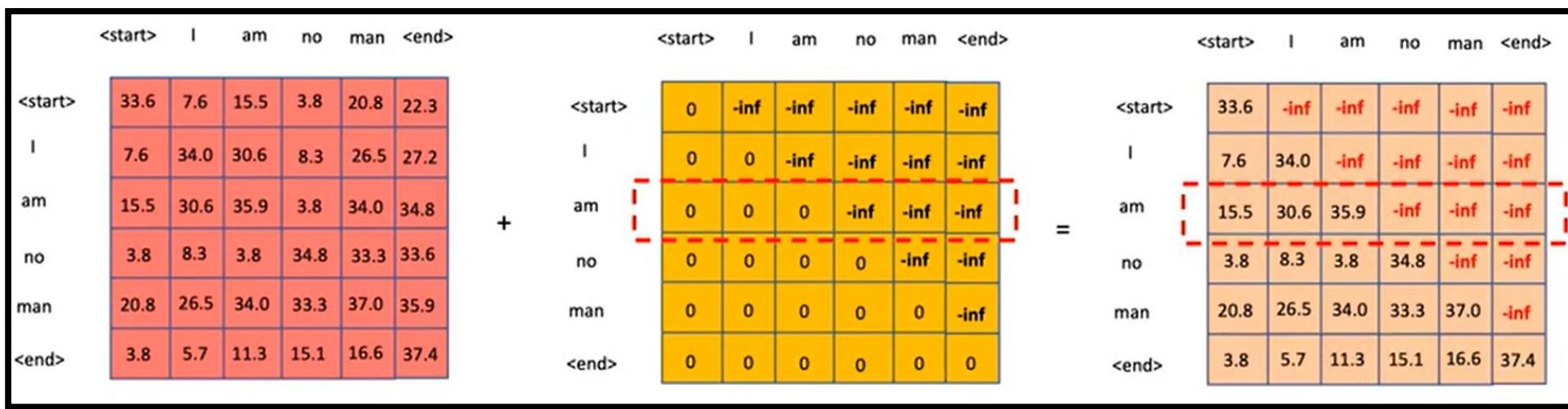
Part 2 - Multi-Head Attention



	When	you	play	the	game	of	thrones
When	33.6	7.6	15.5	3.8	20.8	3.8	22.3
you	7.6	34.0	30.6	8.3	26.5	5.7	27.2
play	15.5	30.6	35.9	3.8	34.0	11.3	34.8
the	3.8	8.3	3.8	34.8	33.3	15.1	33.6
game	20.8	26.5	34.0	33.3	37.0	16.6	35.9
of	3.8	5.7	11.3	15.1	16.6	32.1	22.3
thrones	22.3	27.2	34.8	34.0	35.9	22.3	37.4

Masking

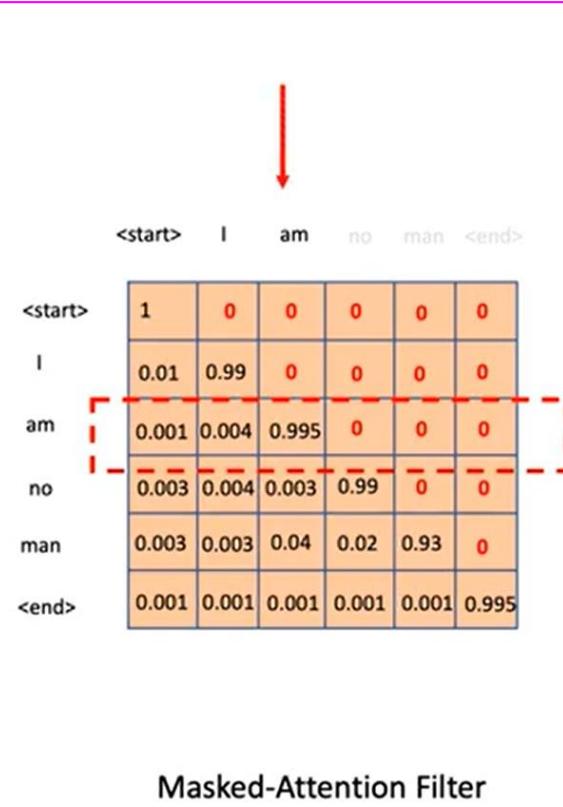
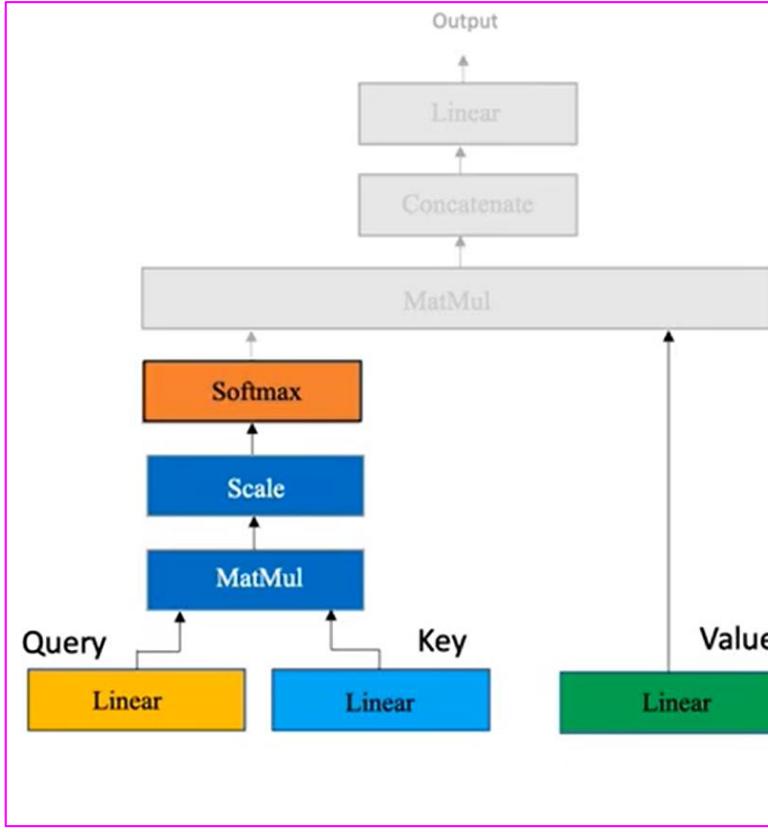
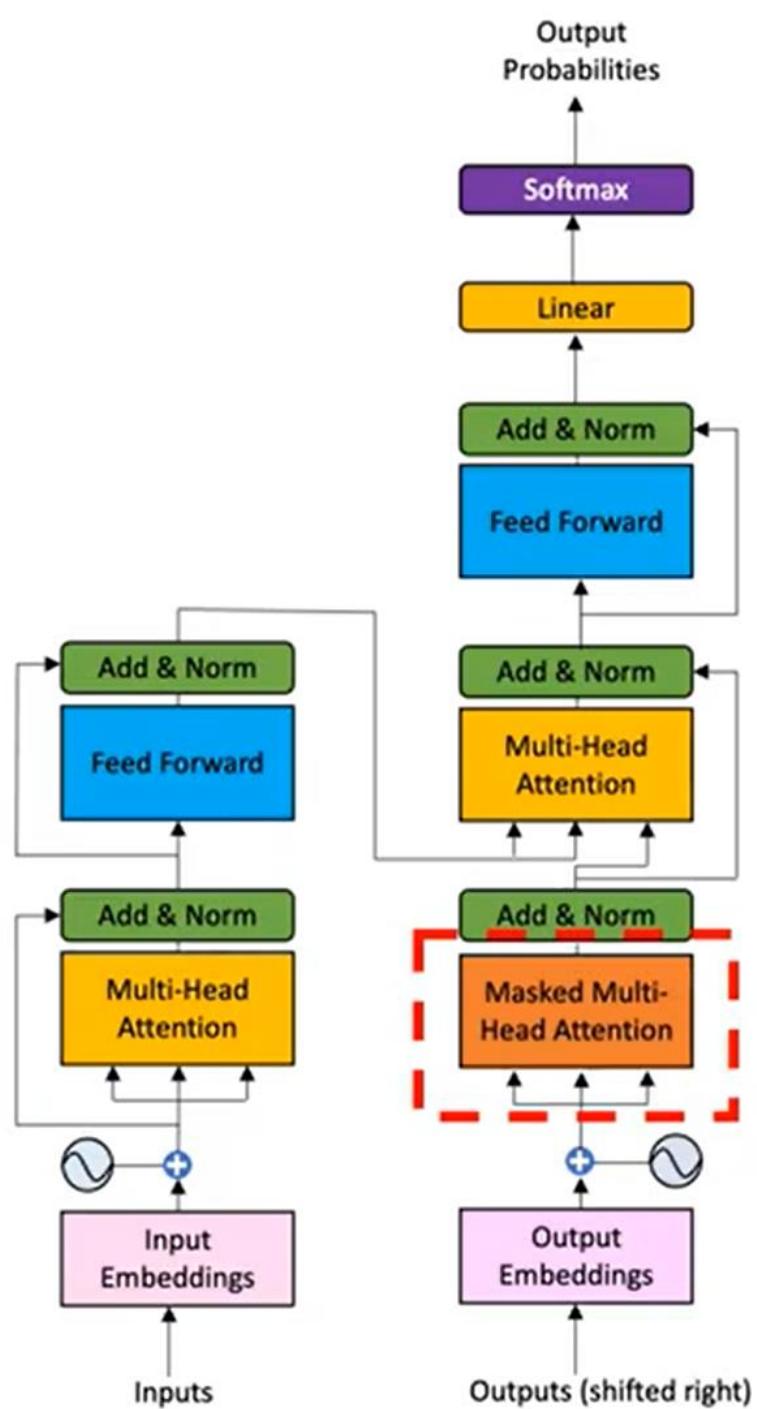
<start> I am no man <end>



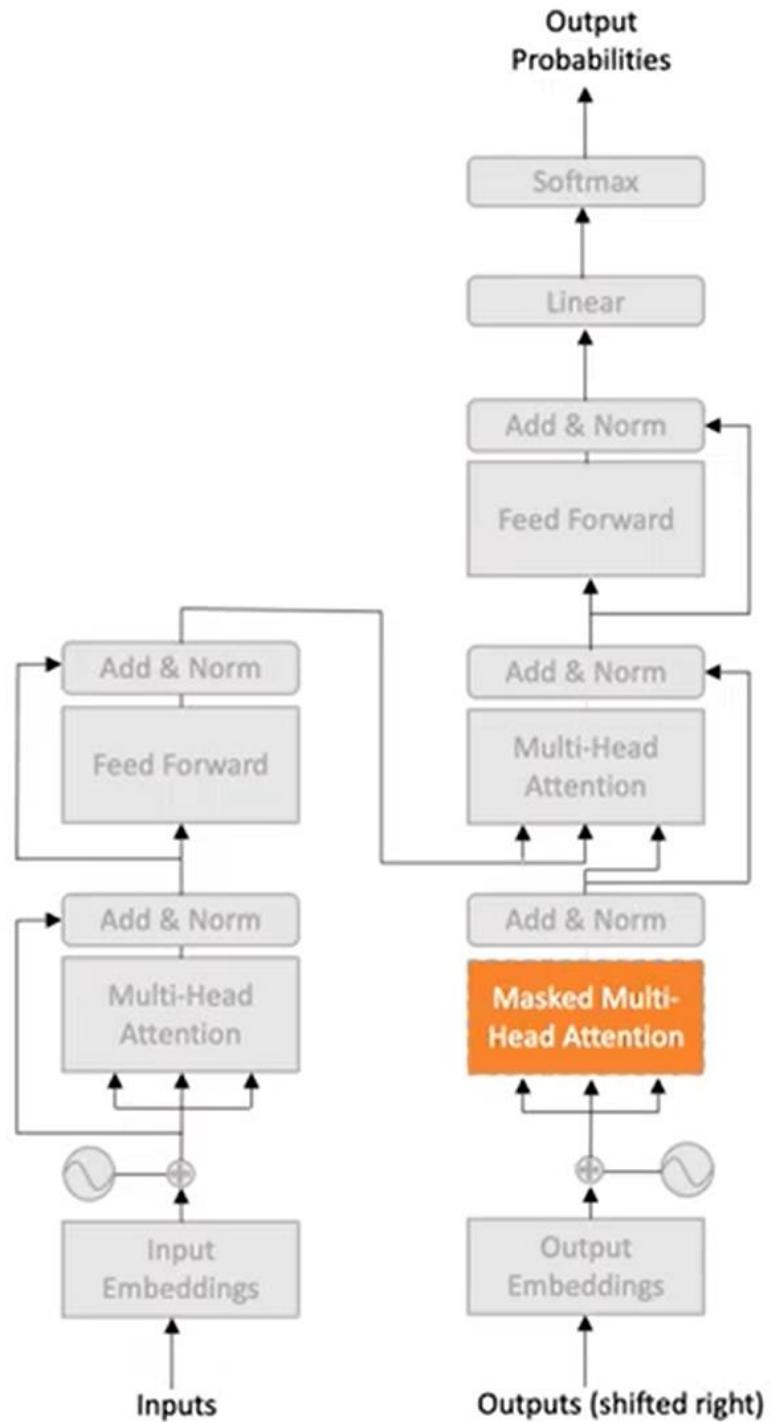
Attention Filter

Mask Filter

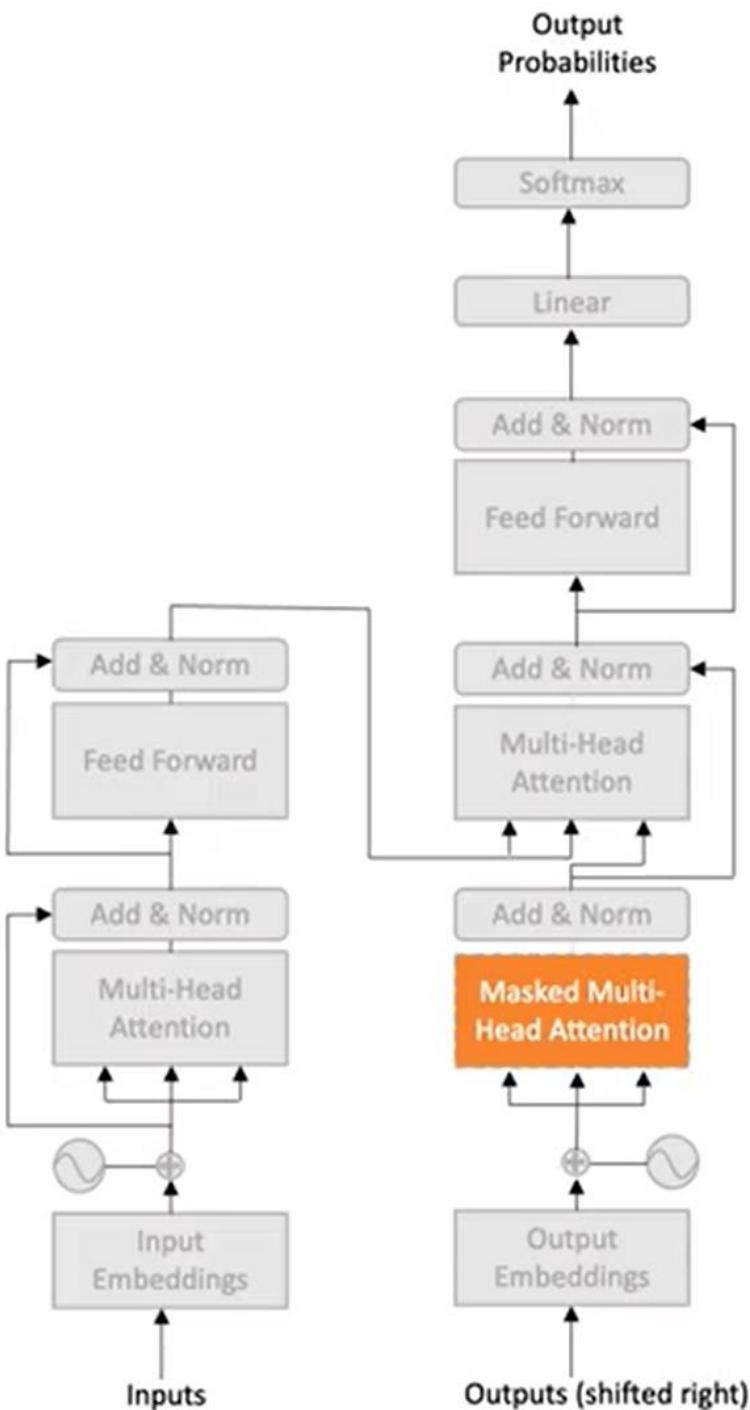
Masked-Attention Filter



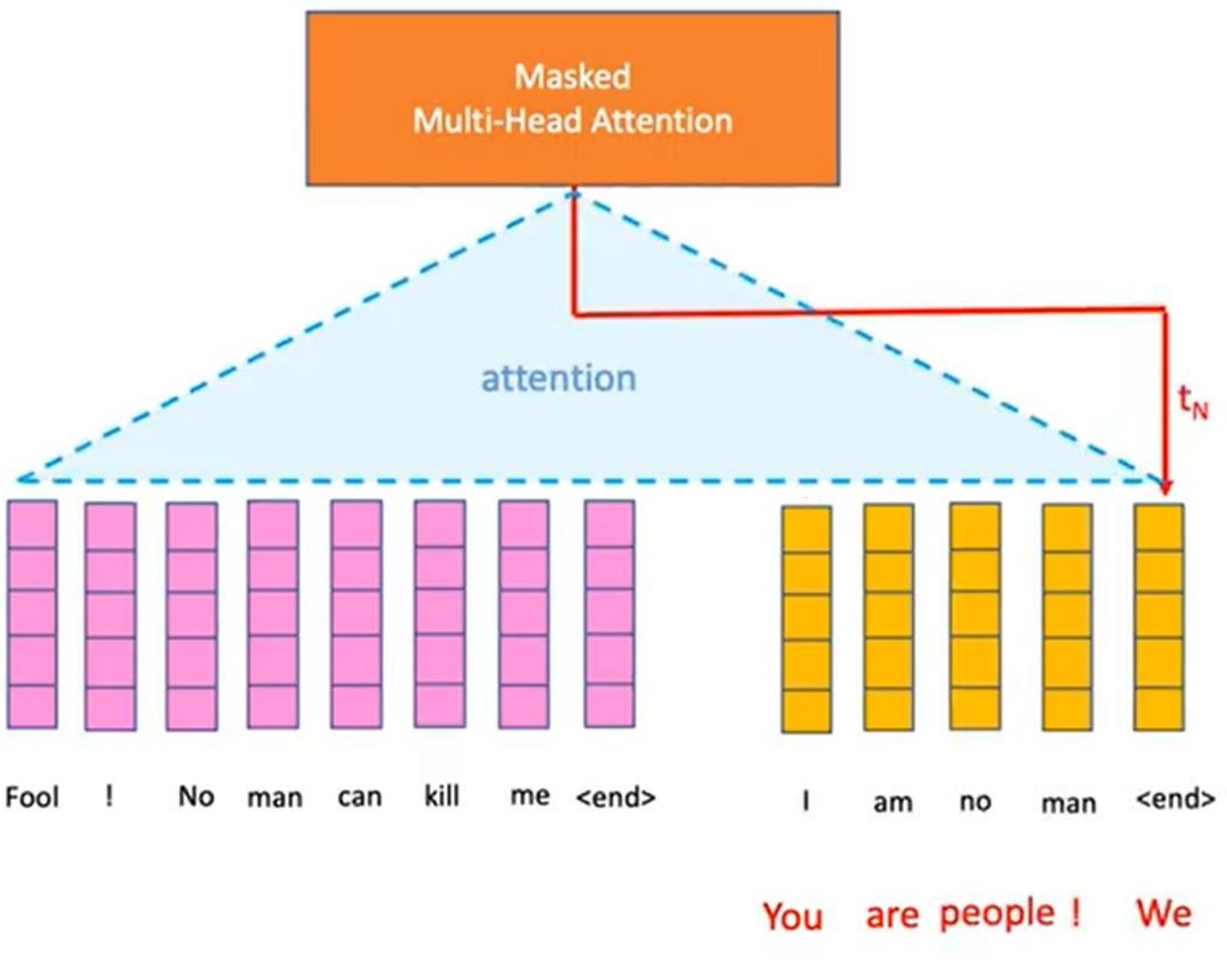
Masked Multi-Head Attention



You



Masked Multi-Head Attention



Dialogue Completer



Cersei

When you play the Game of Thrones, **you win or you die**

FUN WITH TRANSFORMERS

Link: <https://transformer.huggingface.co/doc/distil-gpt2>

The screenshot shows the "Write With Transformer" interface using the "distil-gpt2" model. The input text is "when you play the Game of Thones". A dropdown menu shows three suggestions:

- it's like a kid playing with a bag.
- you just love playing.

The end

Hope that you had a good time with transformer
Thank you!!!!!!