TRANSFORMER NEURAL NETWORK FOR EXTRACTIVE TEXT SUMMARIZATION

MERRIN SANCIA S
URK20AI1050
DEPT OF AI & DATA SCIENCE

TEXT SUMMARIZATION

Researchers have discovered a new species of frog in the Amazon rainforest. The frog, named 'Hyloscirtus hillisi,' was found during an expedition to a remote region of the forest. It has distinctive orange markings on its skin and emits a unique call that distinguishes it from other frog species in the area. This discovery highlights the importance of preserving the biodiversity of the Amazon.

A new frog species, 'Hyloscirtus hillisi,' was discovered in the Amazon rainforest. It has distinctive orange markings and a unique call. This finding underscores the need to protect the Amazon's biodiversity

POTENTIAL MODELS

1

RNN

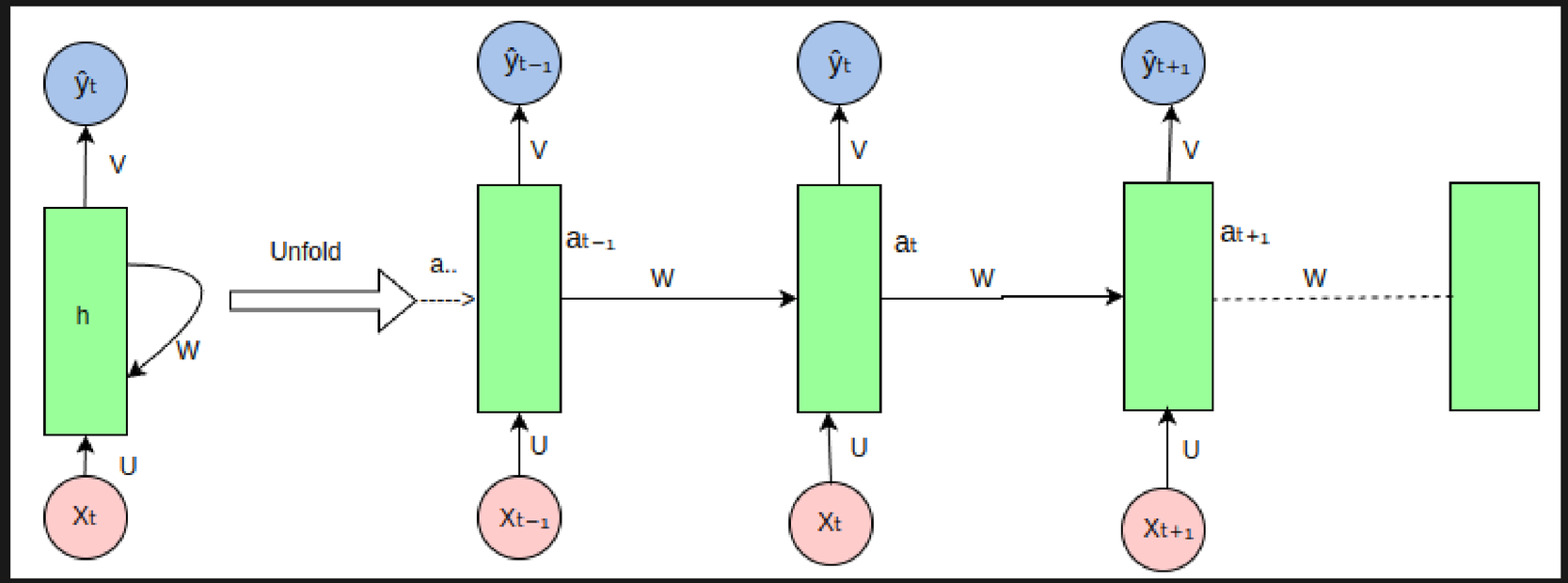
2

LSTM

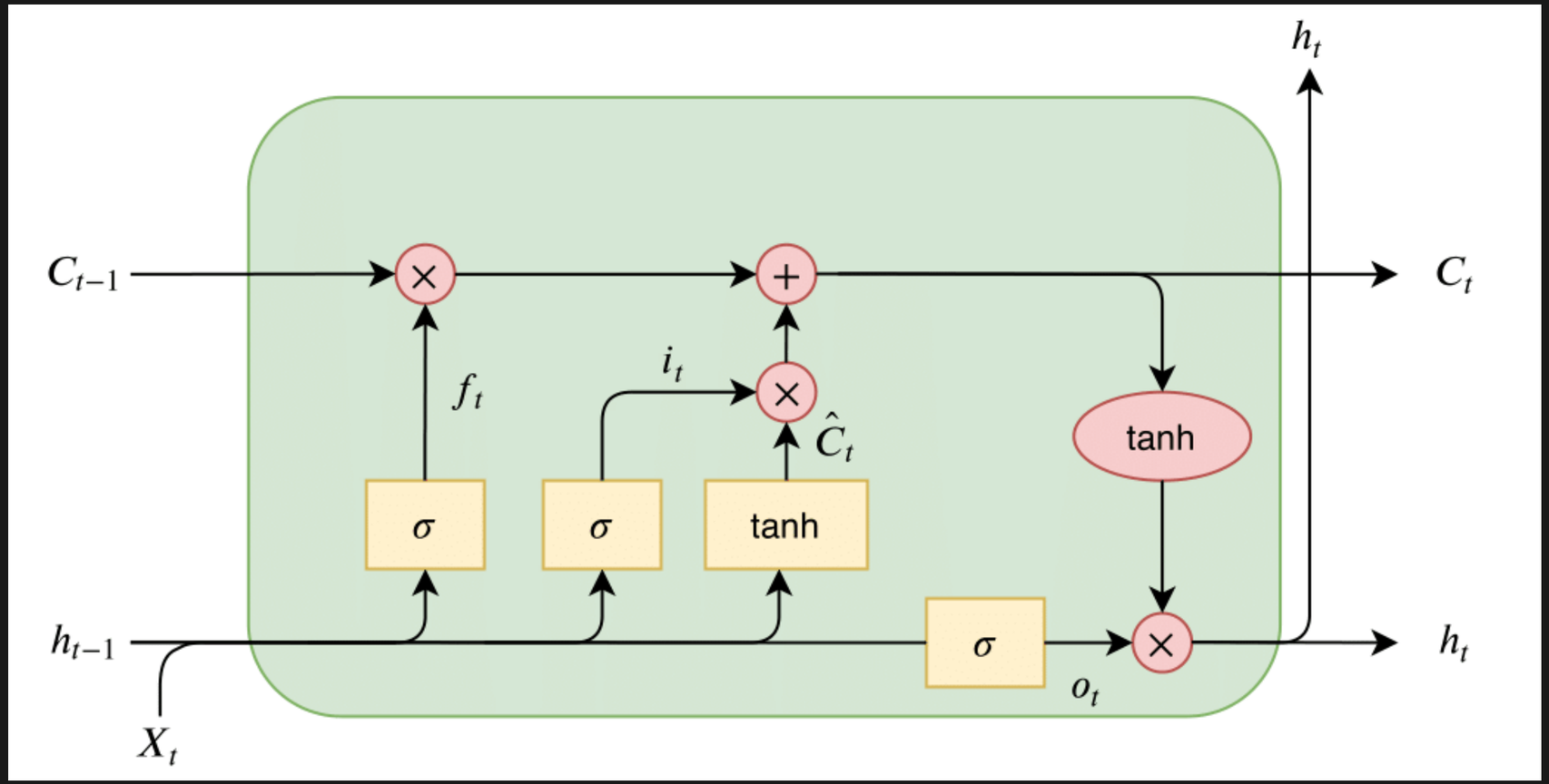
3

TRANSFORMER NEURAL NETWORK

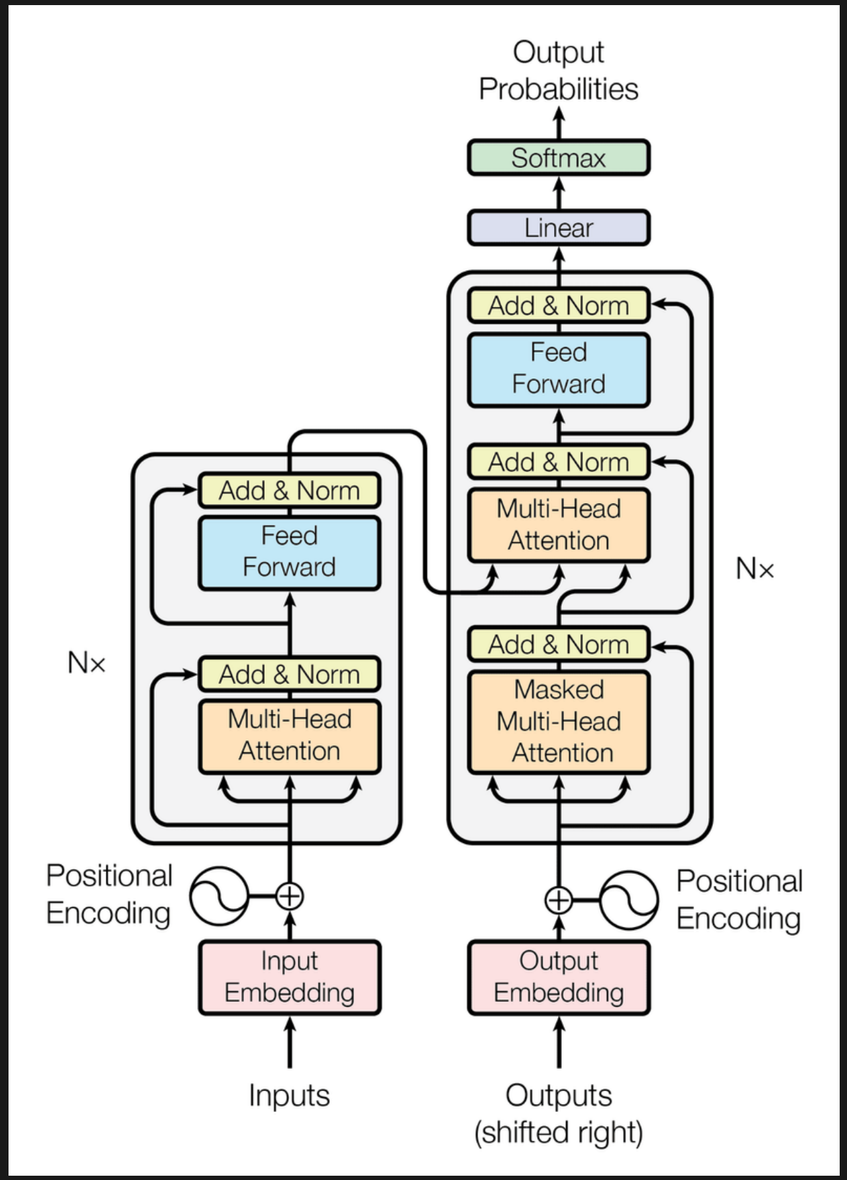
RNN



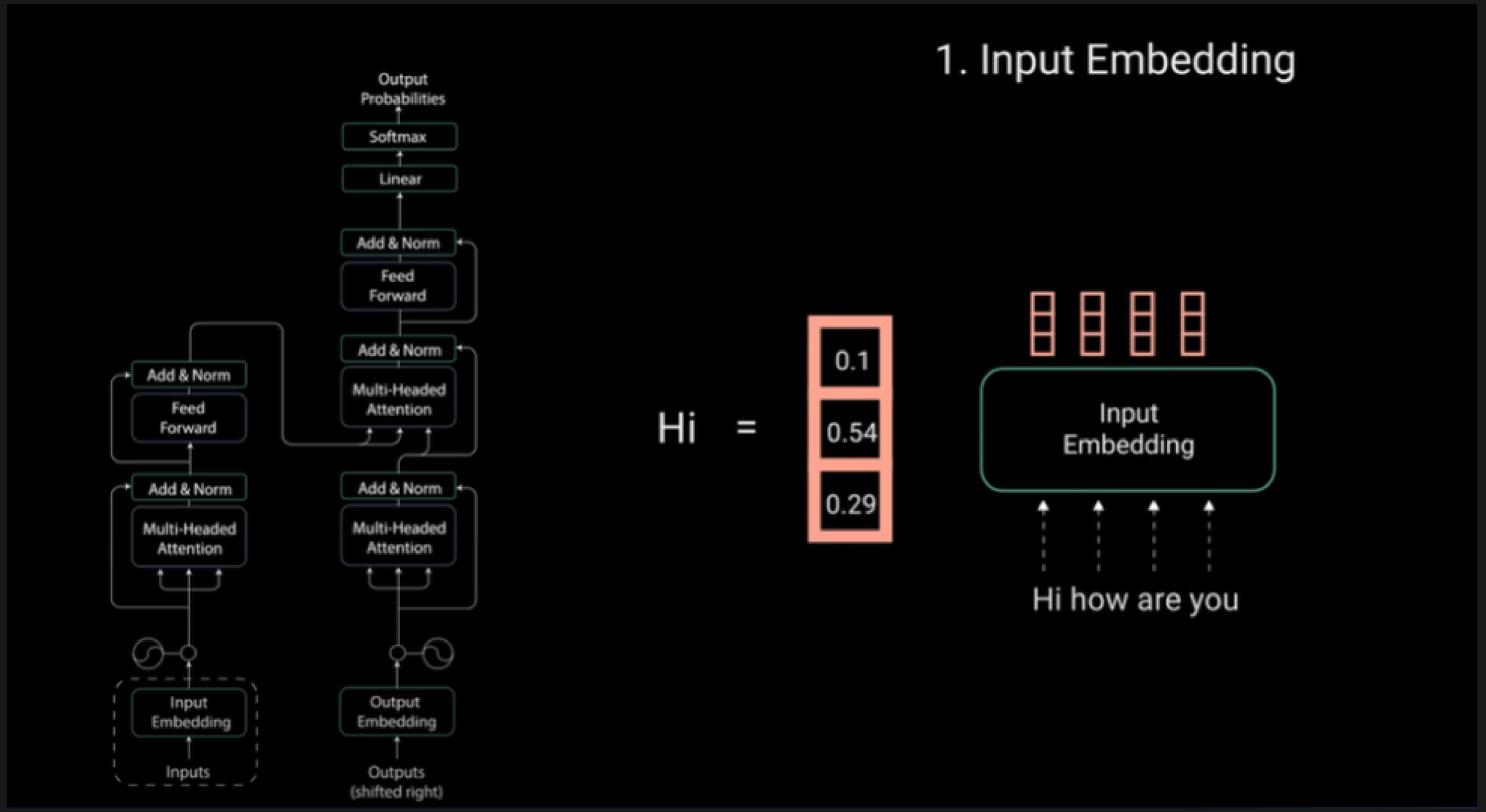
LSTM


h_t
C_{t-1}
C_t
f_t
i_t
$\hat{C}_t$
tanh
σ
σ
tanh
tanh
h_{t-1}
σ
h_t
o_t
X_t

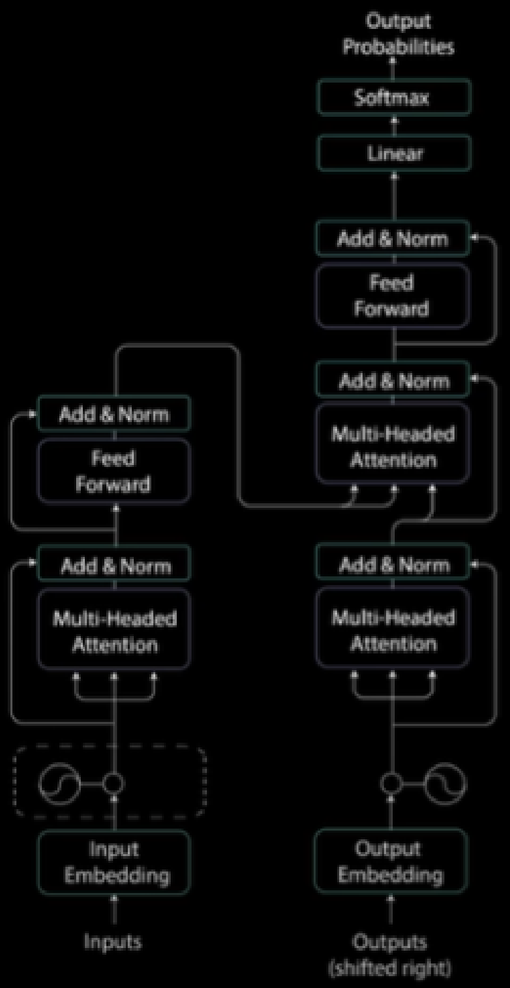
TRANSFORMER NEURAL NETWORK



STEP -1


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Add & Norm
Multi-Headed Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Input Embedding
Inputs
Output Embedding
Outputs (shifted right)
1. Input Embedding
Hi =
0.1
0.54
0.29
Input Embedding
Hi how are you

STEP -2



2. Positional Encoding

Positional Input Embeddings

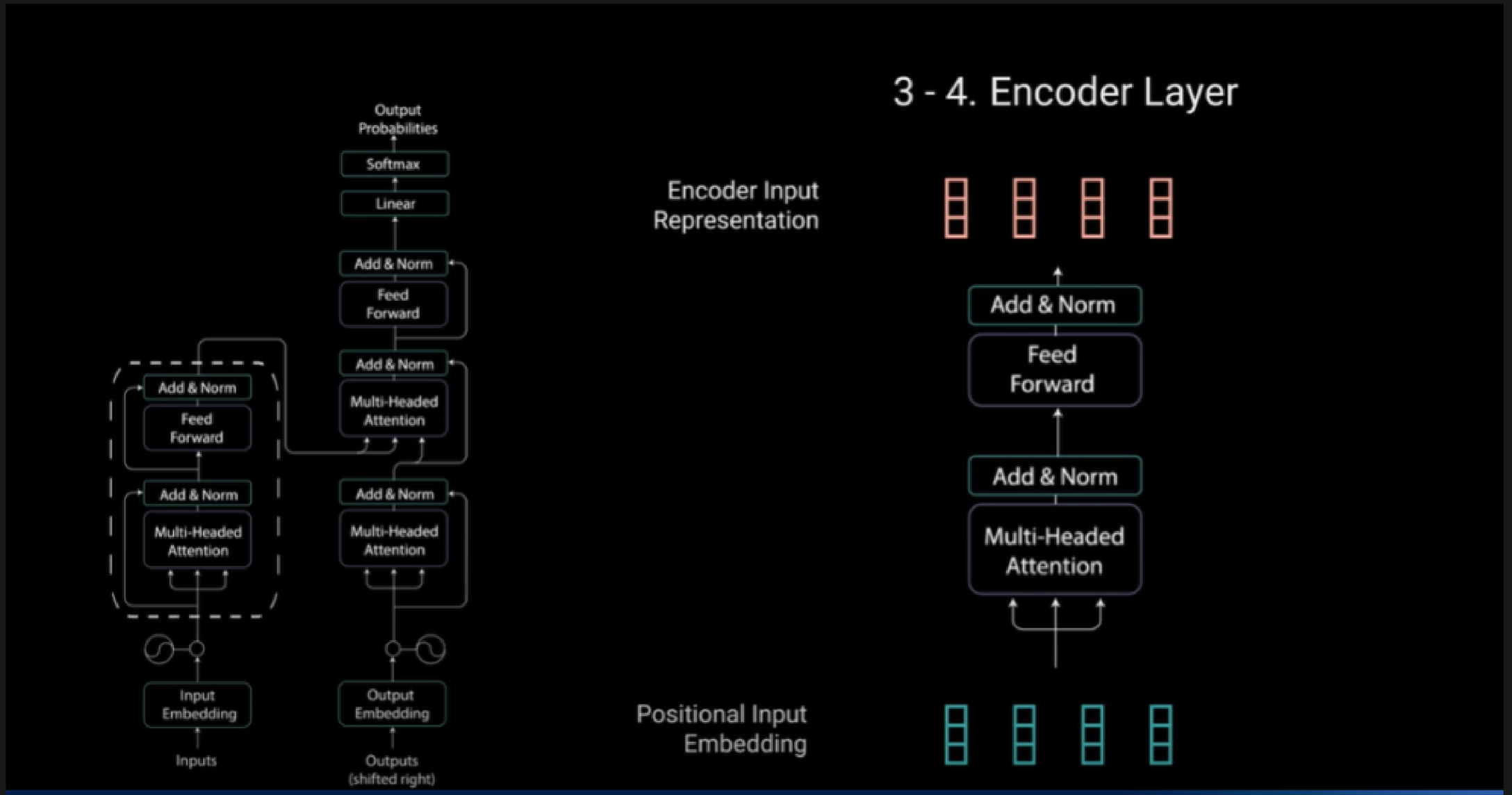
Positional Encoding

Time Step 1 2 3 4

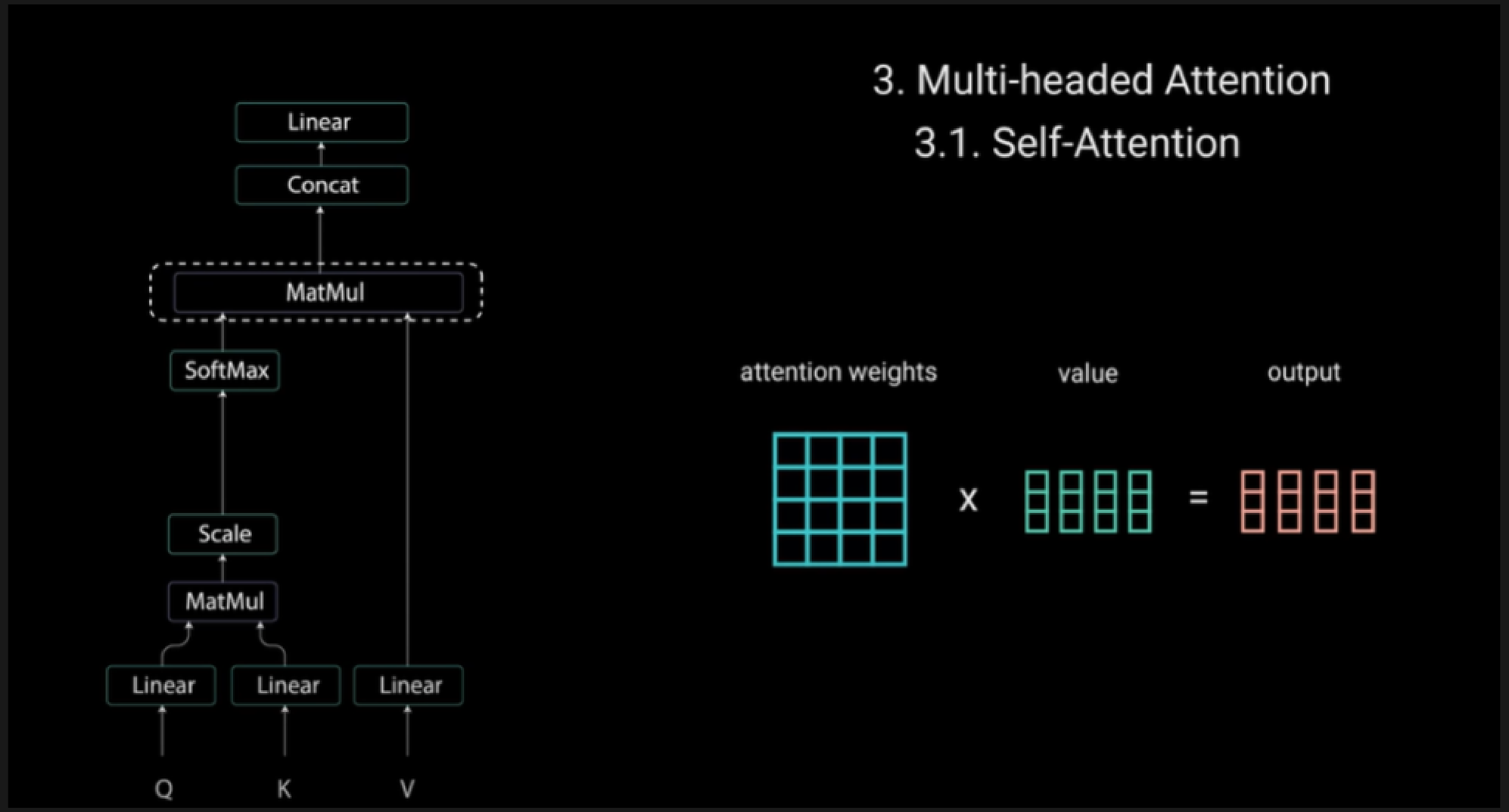
$$PE(pos, 2i+1) = cos(\frac{pos}{10000^{2i/dmodel}})$$

$$PE(pos, 2i) = sin(\frac{pos}{10000^{2i/dmodel}})$$

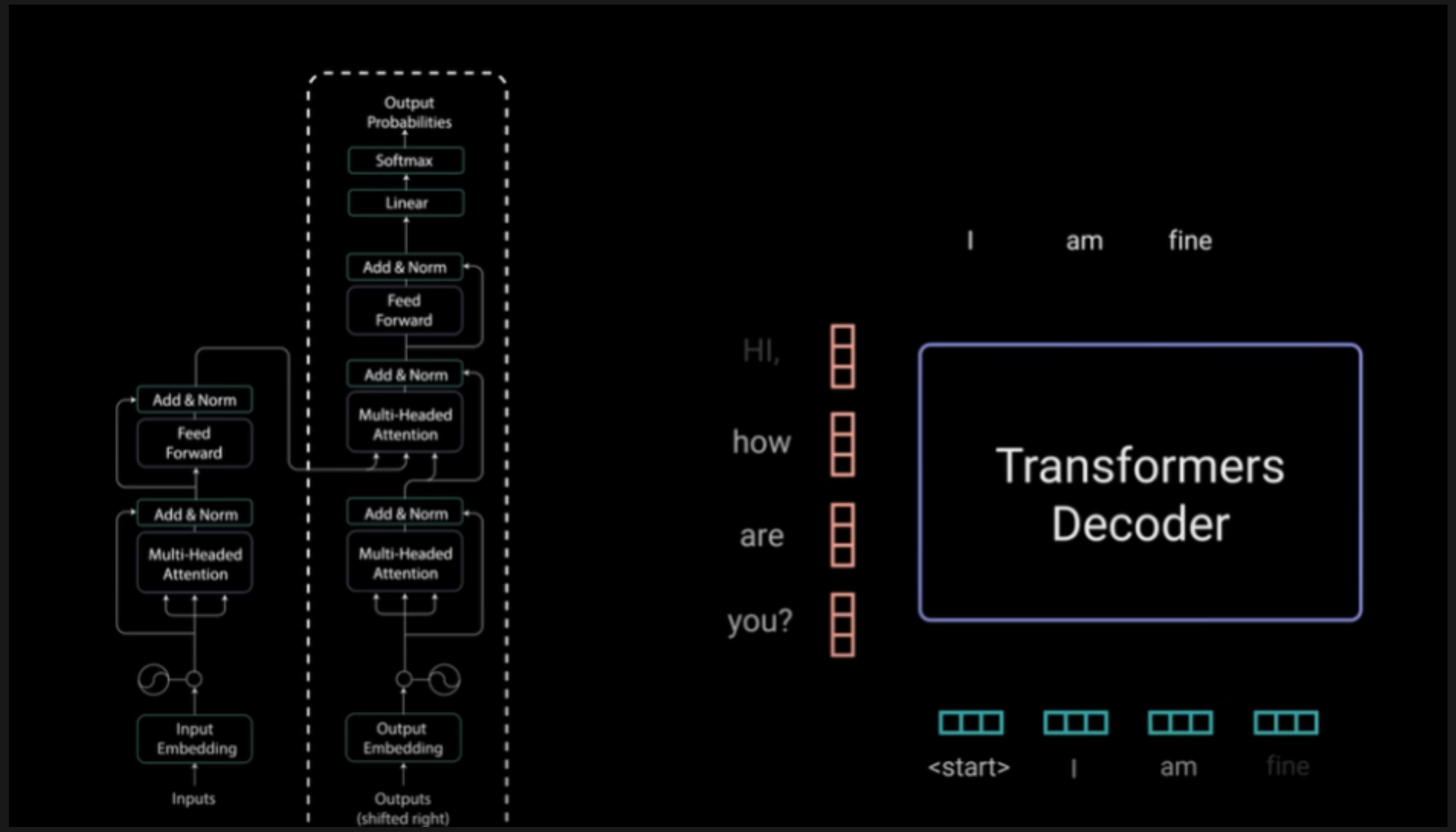
STEP -3


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Add & Norm
Multi-Headed Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)
3 - 4. Encoder Layer
Encoder Input Representation
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Positional Input Embedding

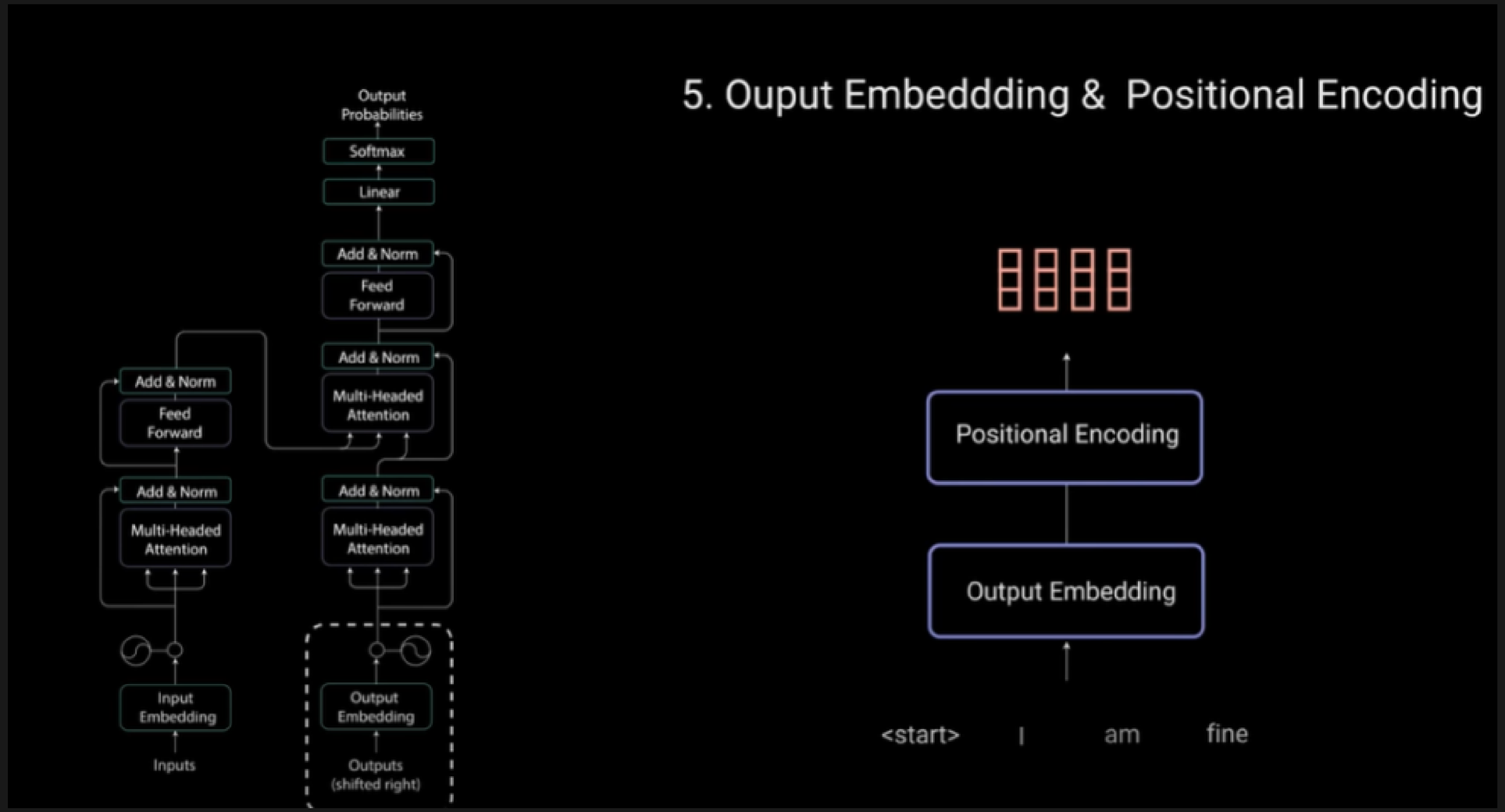
STEP - 4


Linear
Concat
MatMul
SoftMax
Scale
MatMul
Linear
Linear
Linear
Q
K
V
3. Multi-headed Attention
3.1. Self-Attention
attention weights
value
output
X
=

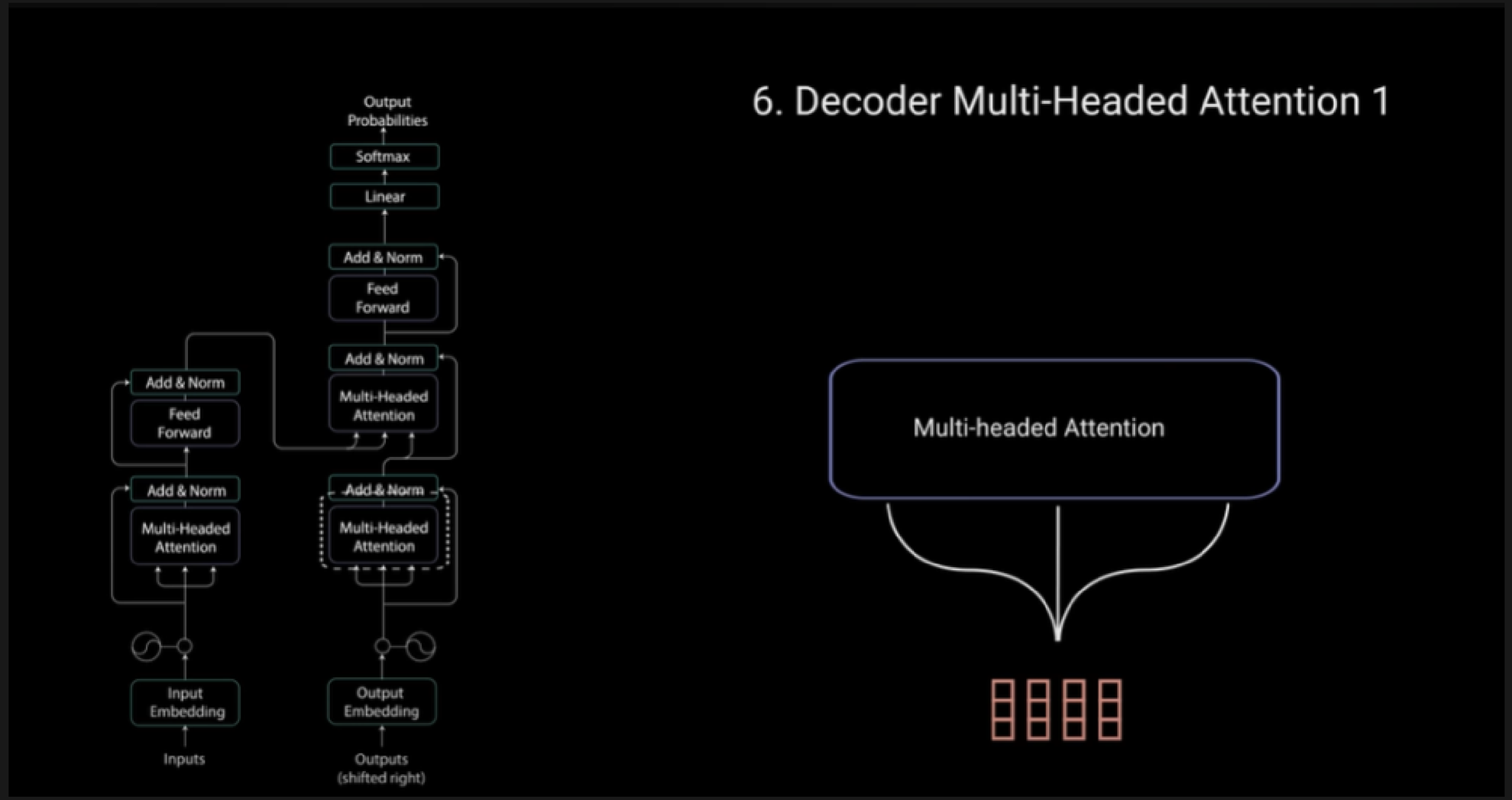
STEP - 5


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Add & Norm
Multi-Headed Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)
I am fine
HI,
how
are
you?
Transformers Decoder
<start>
I
am
fine

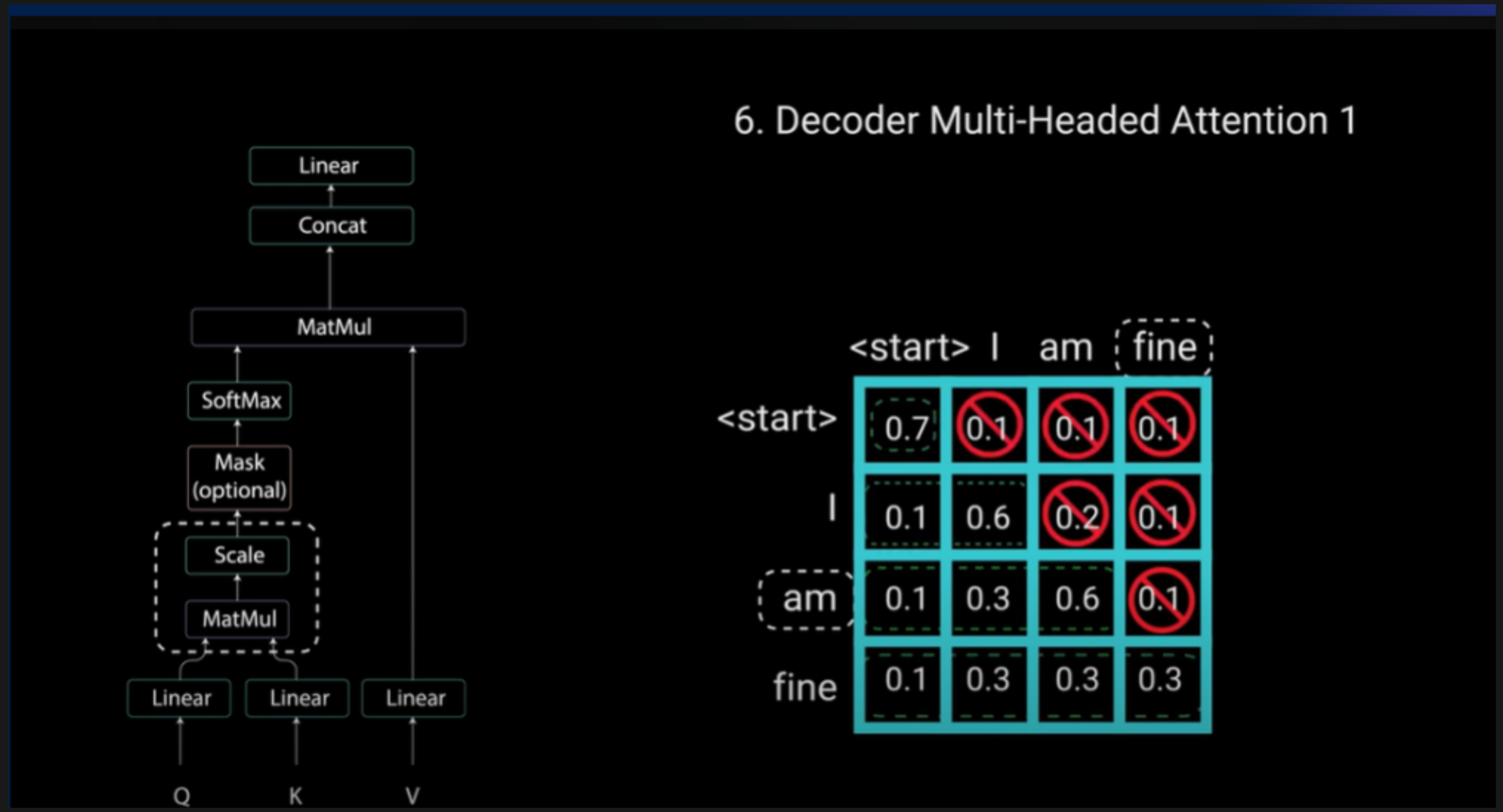
STEP - 6


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Input Embedding
Inputs
Output Embedding
Outputs (shifted right)
5. Ouput Embeddding & Positional Encoding
Positional Encoding
Output Embedding
<start>
I
am
fine

STEP -7


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Headed Attention
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)
6. Decoder Multi-Headed Attention 1
Multi-headed Attention

STEP -8


Linear
Concat
MatMul
SoftMax
Mask
(optional)
Scale
MatMul
Linear
Linear
Linear
Q
K
V
6. Decoder Multi-Headed Attention 1
<start>
I
am
fine
0.7
0.1
0.1
0.1
0.1
0.6
0.2
0.1
0.1
0.3
0.6
0.1
0.1
0.3
0.3
0.3

1

ATTENTION MECHANISM

2

LONG-RANGE DEPENDENCIES

3

PARALLELIZATION

4

VANISHING GRADIENT

5

FLEXIBLITY

WHY TRANSFORMER ?

RESULT ANALYSIS

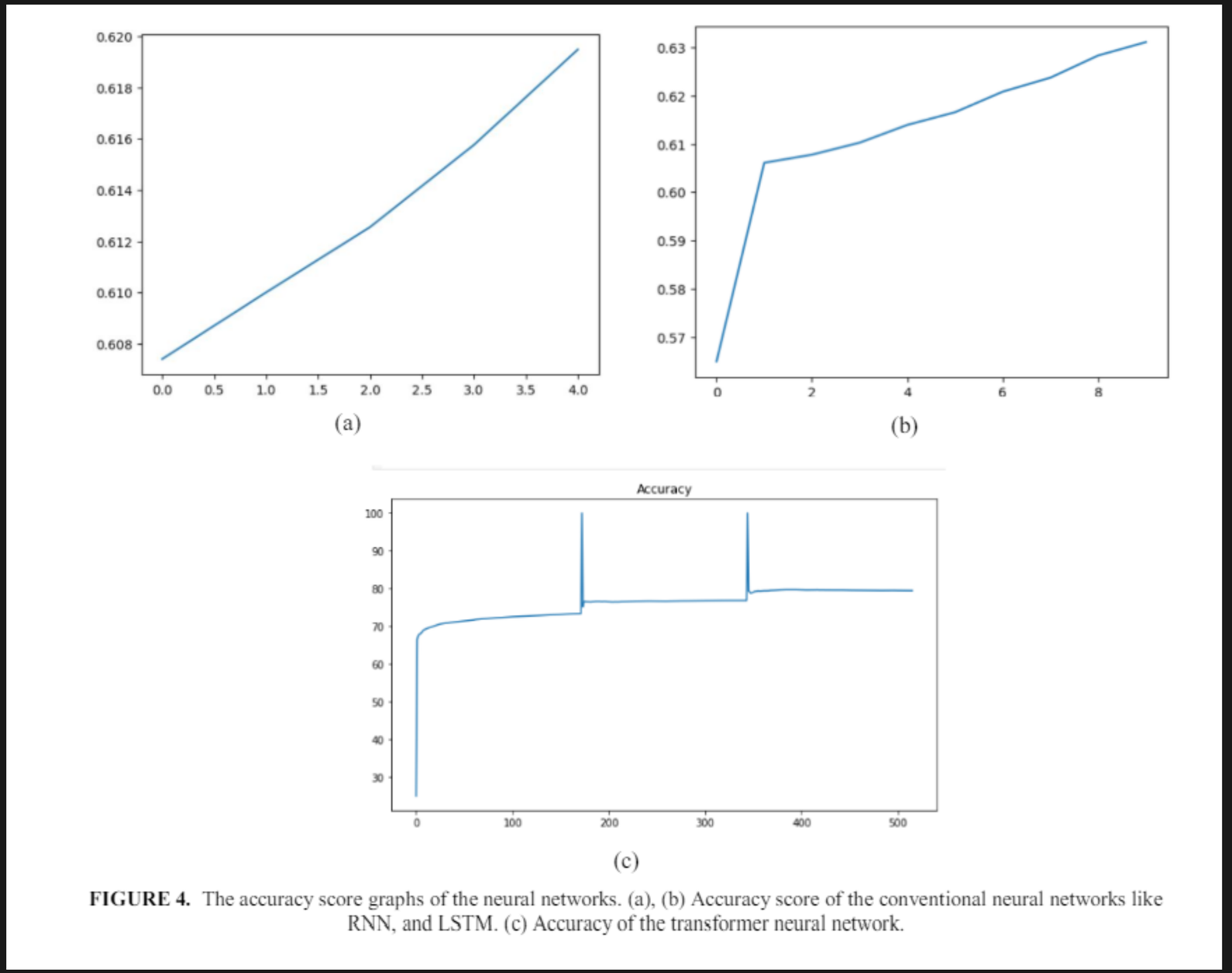


FIGURE 4. The accuracy score graphs of the neural networks. (a), (b) Accuracy score of the conventional neural networks like RNN, and LSTM. (c) Accuracy of the transformer neural network.

ACCURACY CHART

S.NO	MODEL	ACCURACY
01	SEQ2SEQ RNN	62%
02	BI-LSTM	65%
03	TNN	82%

CONCLUSION

In conclusion, the Transformer neural networks are highly eminent in handling data-intensive tasks. It exercises various steps that help the model richen the vector representations of the input sequence making the model more confident in predicting the most precise words, achieving the highest accuracy of all the neural networks. However, it’s important to understand, choosing the model based on the requirement is the wisest thing to do. For less intensive data, Seq2Seq RNN can do a better job than implementing such complex models, wasting the computational potential of the system in addition to potential underfitting threats. For medium-sized data, it’s safer to go with Bi-directional LSTM than implementing such complex models. If the data is too intensive to handle by those conventional neural networks it is wise to prefer a self-attention-based Transformer as they can do a better job not falling into the trap of long dependency problem and overfitting. In our experimentation, it was found that an attention-based transformer neural network is the best choice for highly intensive datasets.