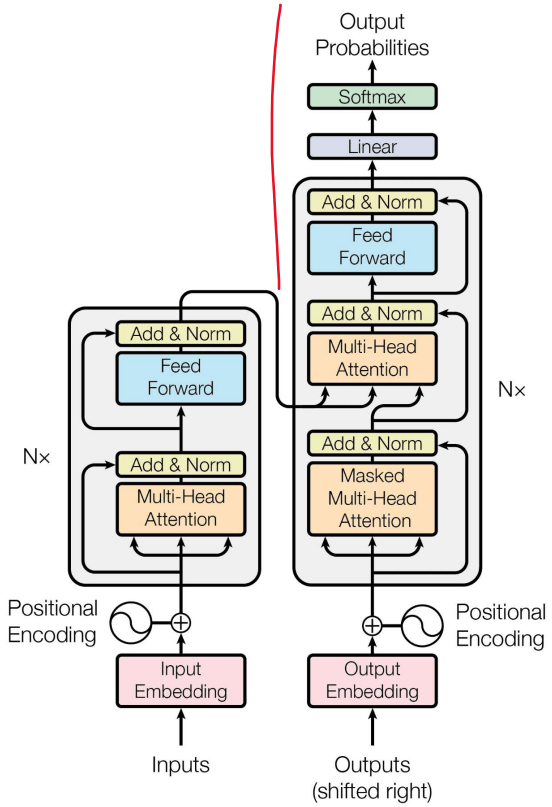
Transformers

16 July 2025 14:29

Encoder Decoder



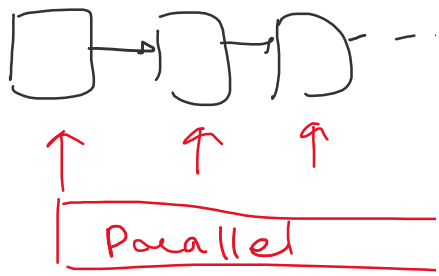
Transformers → 2017 Attention

BERT, GPT → Generative LL

OpenAI → ChatGPT

Transformer : Encoder de with self attention process NLP tasks

why transformers ?

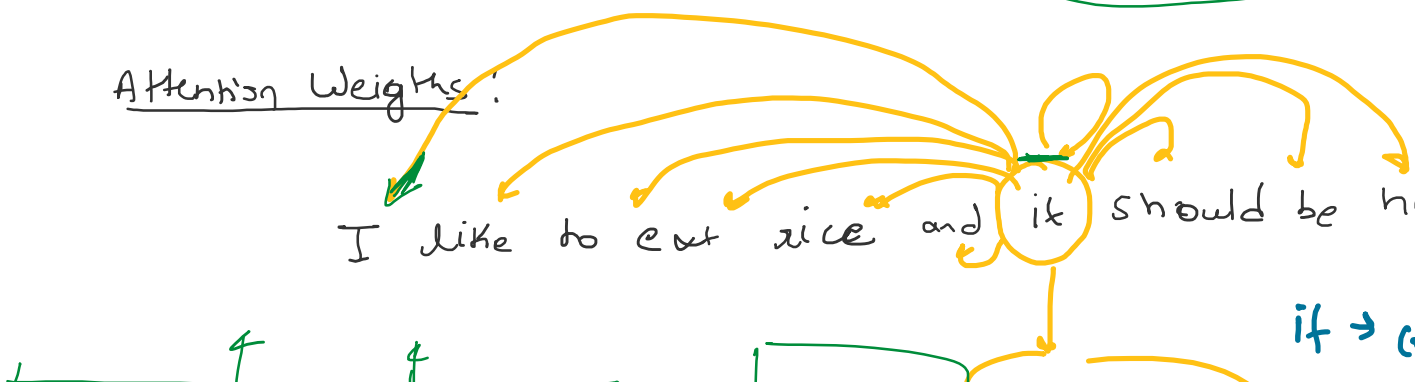


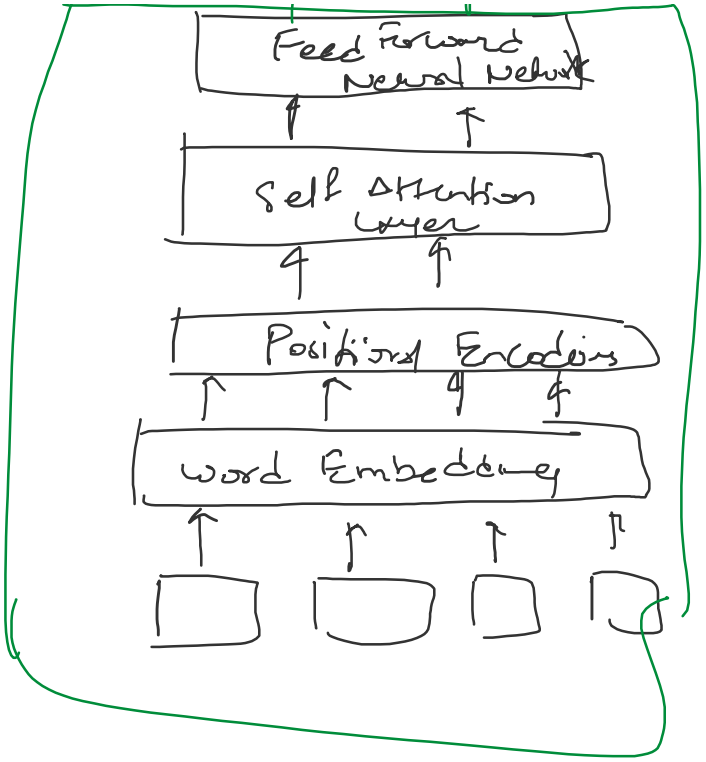
Parallel

1. Encoder Decoder
2. Word Embedding + positional Encoding
3. Multi head attention
4. Feed forward Network
5. Residual connections
6. Normalization

$i_Q \cdot I_K$

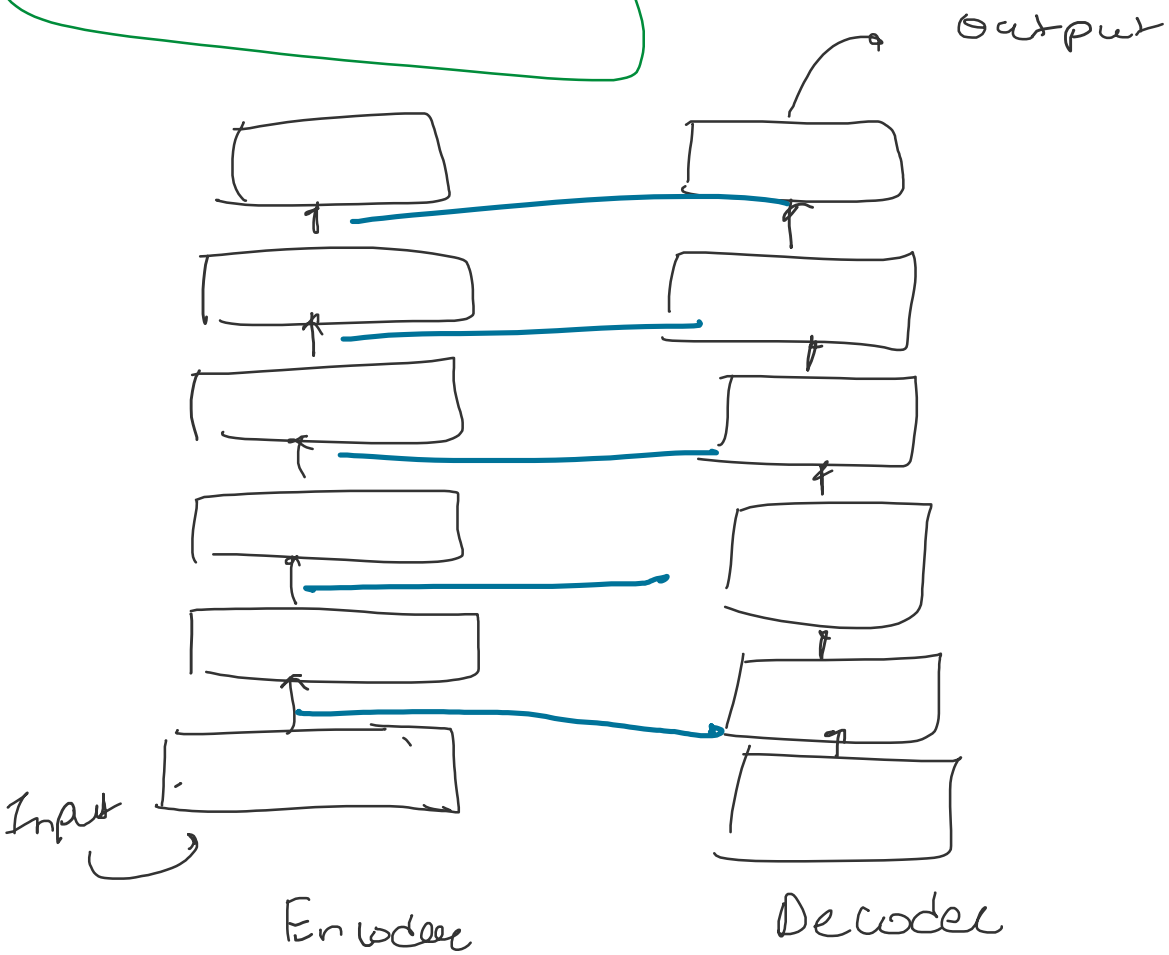
Attention Weights :





attention

Decoder



Self-Attention

(3) → Query →

→ Key

→ value

Shub likes Cricket W_Q, W_K, W_T = ?

$WE_{Shub} = [1\ 0\ 1\ 0]$

$WE_{like} = [0\ 1\ 0\ 1]$

$WE_{cricket} = [1\ 1\ 1\ 1]$

$Q_{Shub} = [1\ 0\ 1\ 0]$

2) $Q_{like} = [0\ 1\ 0\ 1]$

$Q_{cricket} = [1\ 1\ 1\ 1]$

$Shub = Score(Q_{Shub} \cdot K_{Shub}) = 2/2$

$(Q_{Shub} \cdot K_{like}) = 0/2$

$(Q_{Shub} \cdot K_{cricket}) = 2/2$

$like = Score(Q_{like} \cdot K_{Shub}) = 0/2$

$(Q_{like} \cdot K_{like}) = 2/2$

$Q_{like} \cdot K_{cricket}) = 2/2$

$Cricket = Score(Q_{cri} \cdot K_{Shub}) = 2$