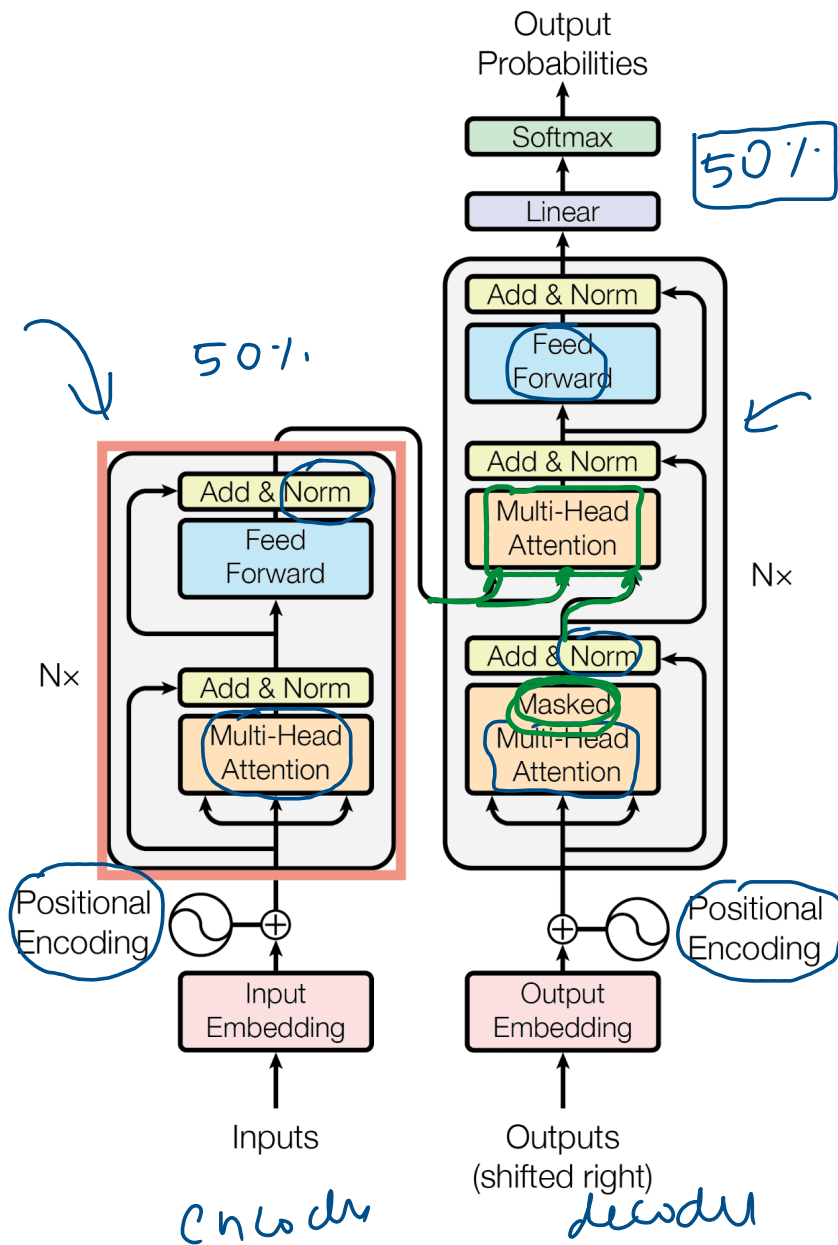


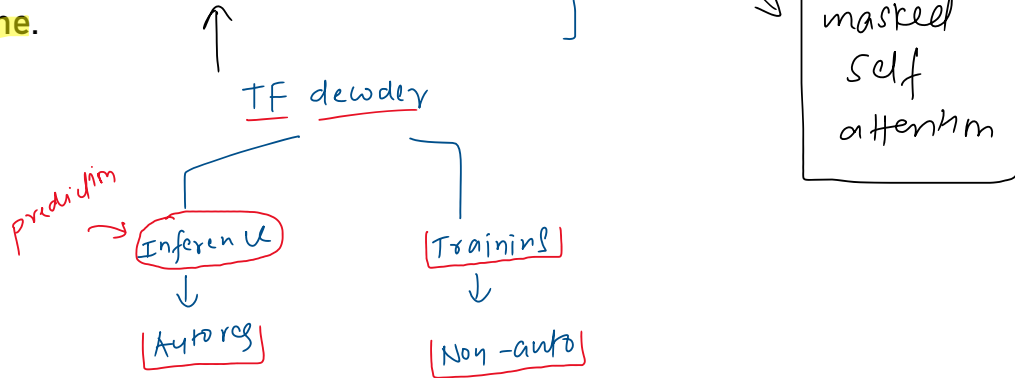
# Recap

23 July 2024 17:09

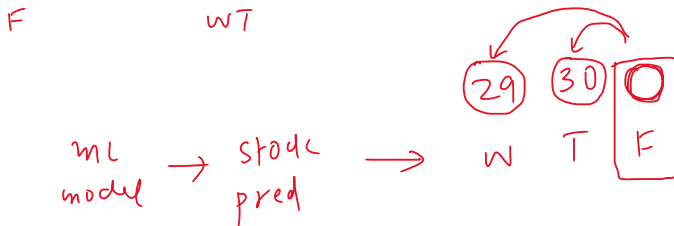


economics → Time series

The Transformer decoder is **autoregressive** at **inference** time and **non-autoregressive** at **training** time.



→ In the context of deep learning, autoregressive models are a class of models that generate data points in a sequence by conditioning each new point on the previously generated points.

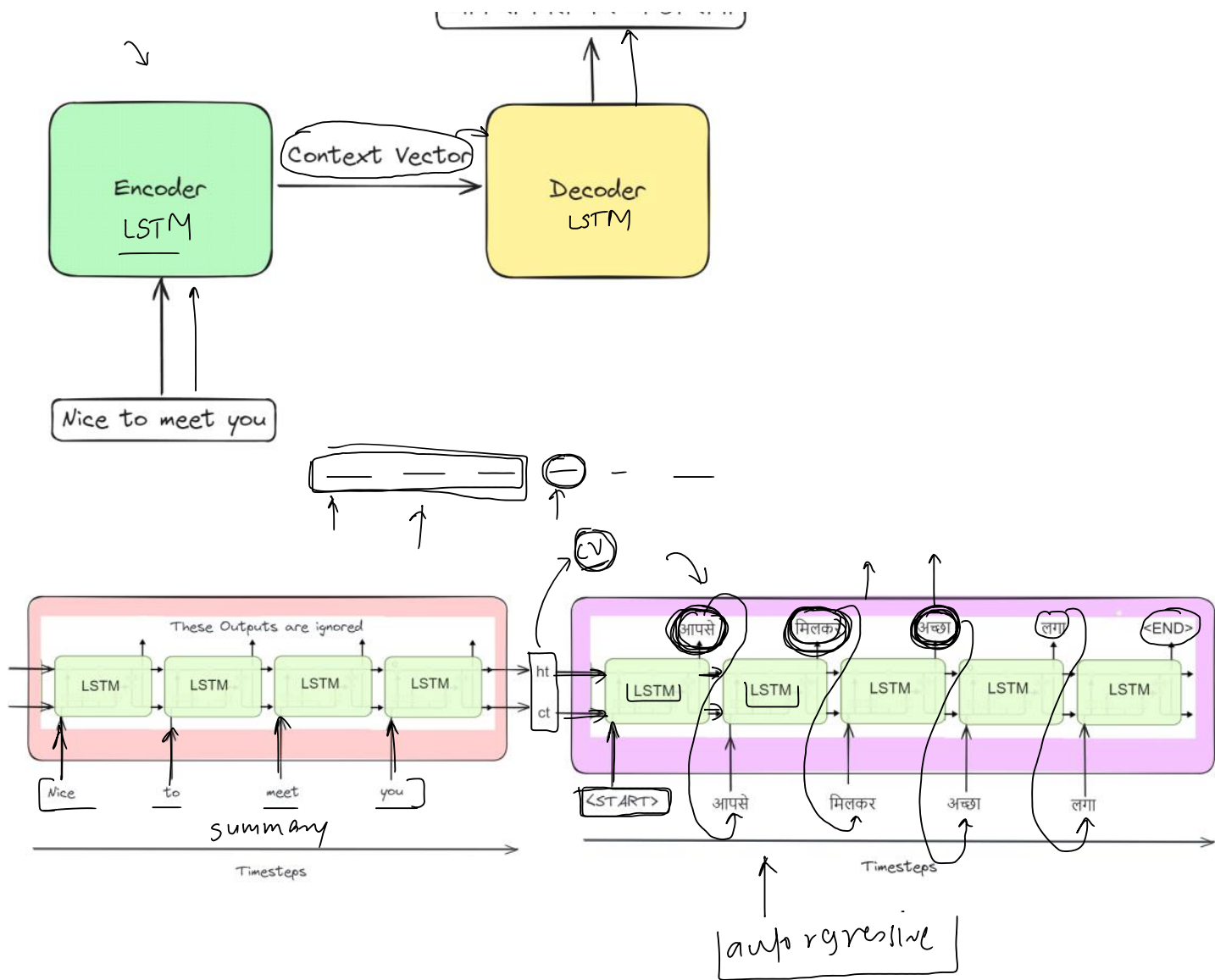


→ Encoder - Decoder → 11ya

→

आप से मिल कर अच्छा लगा

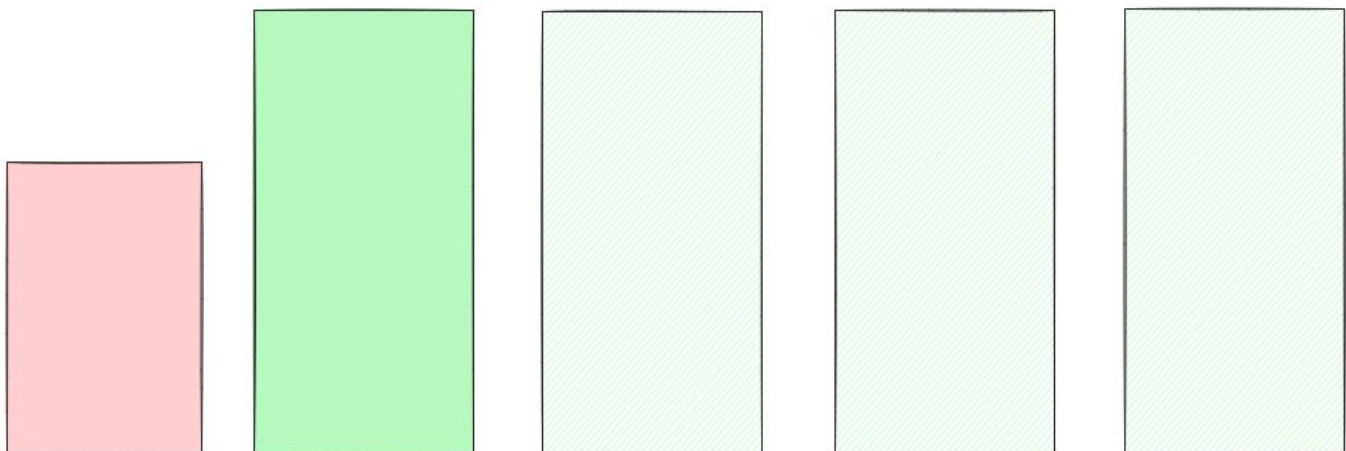
↑ ↑



**The Transformer decoder is autoregressive at inference time and non-autoregressive at training time.**

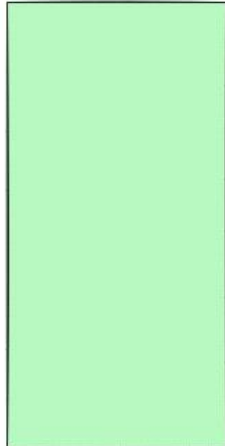
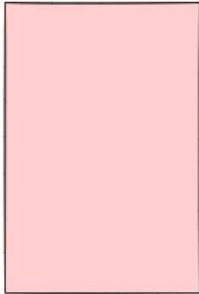
Inference

Query Sentence -> I am fine



## Training

S.No	English Sentence	Hindi Sentence
1	How are you?	आप कैसे हैं
2	Congratulations	बधाई हो
3	Thank you	धन्यवाद

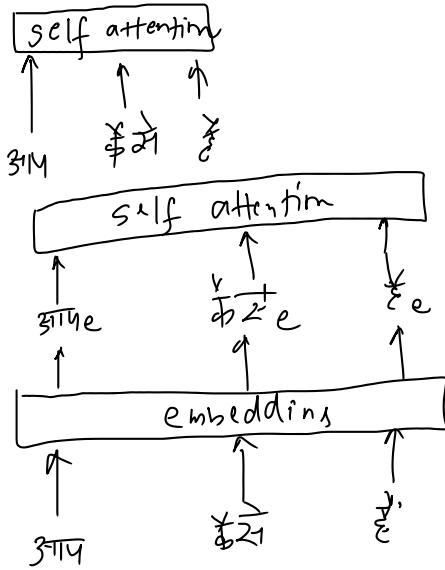


# The problem in parallelizing

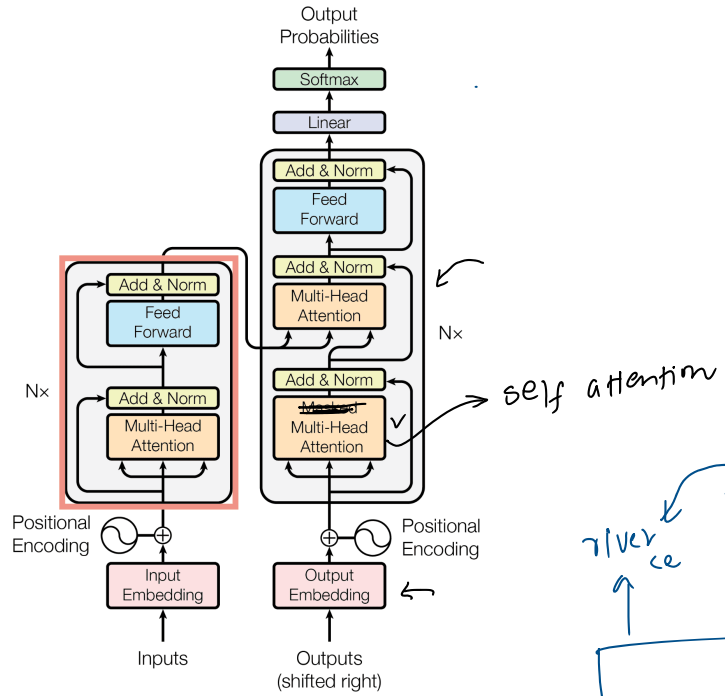
25 July 2024 22:55

S.No	English Sentence	Hindi Sentence
1	How are you?	आप कैसे हैं
2	Congratulations	बधाई हो
3	Thank you	धन्यवाद

→

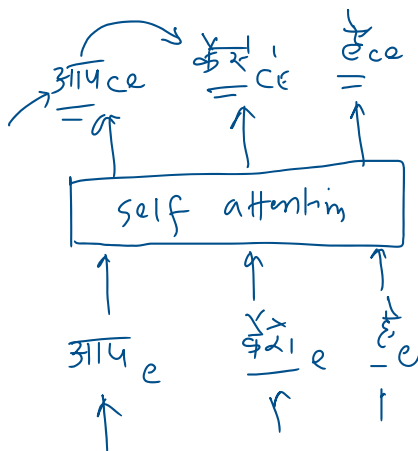


Data Leakage



river bank  
money bank

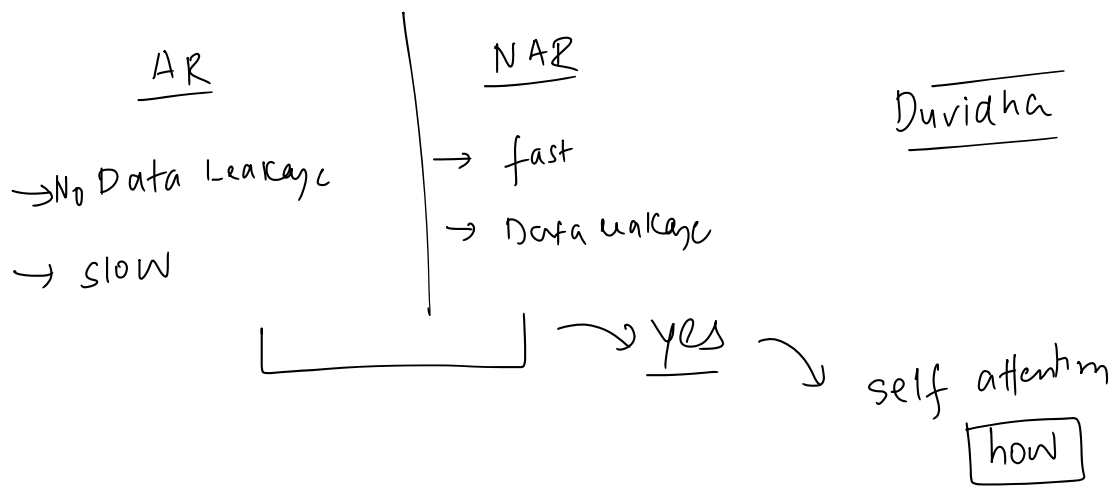
आप कैसे हैं



Cheating

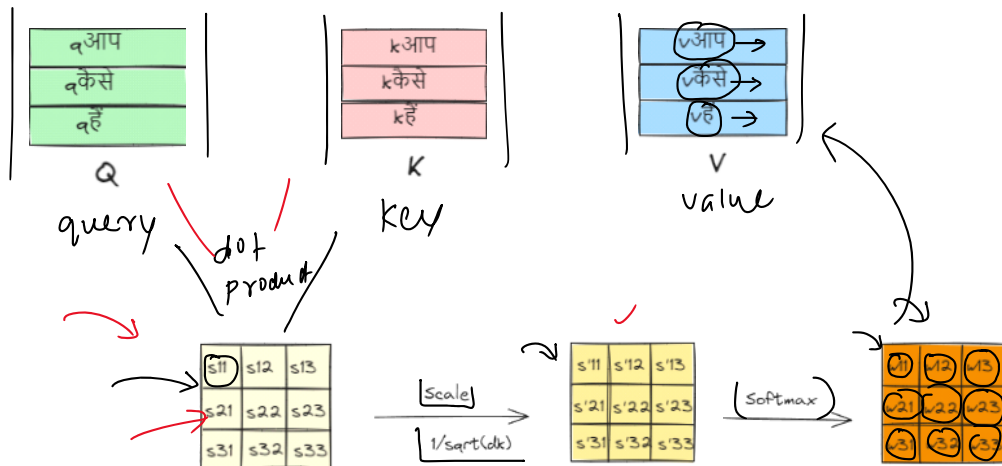
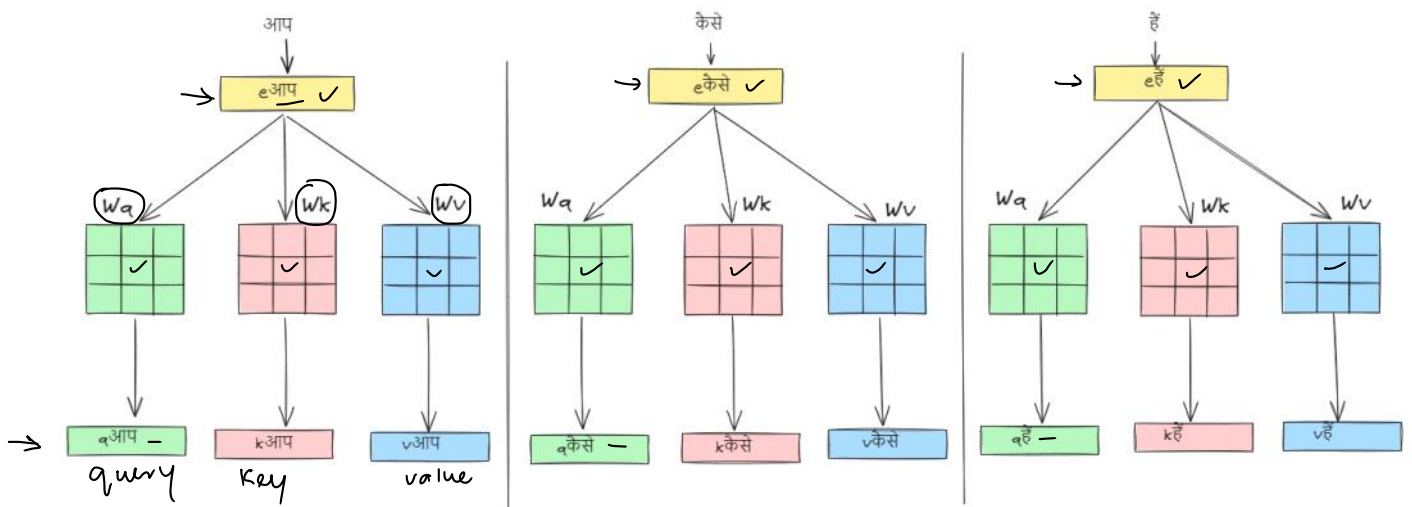
$$\begin{aligned} \text{आप}_{ce} &= 0.8 \text{आप}_e + 0.1 \text{कैसे}_e + 0.1 \text{हैं}_e \\ \text{कैसे}_{ce} &= 0.15 \text{आप}_e + 0.75 \text{कैसे}_e + 0.1 \text{हैं}_e \\ \text{हैं}_{ce} &= 0.1 \text{आप}_e + 0.2 \text{कैसे}_e + 0.7 \text{हैं}_e \end{aligned}$$

current token value  
↓  
future token value



# Finding the answer

26 July 2024 00:21



$$\begin{aligned}
 \text{आप} &= w_{11} * \text{वहाँ} + w_{12} * \text{कैसे} + w_{13} * \text{आप} \\
 \text{कैसे} &= w_{21} * \text{वहाँ} + w_{22} * \text{कैसे} + w_{23} * \text{आप} \\
 \text{वहाँ} &= w_{31} * \text{वहाँ} + w_{32} * \text{कैसे} + w_{33} * \text{आप}
 \end{aligned}$$

$$w_{12} = w_{13} = w_{23} = 0$$

$$\text{softmax}(-\infty) = 0$$

