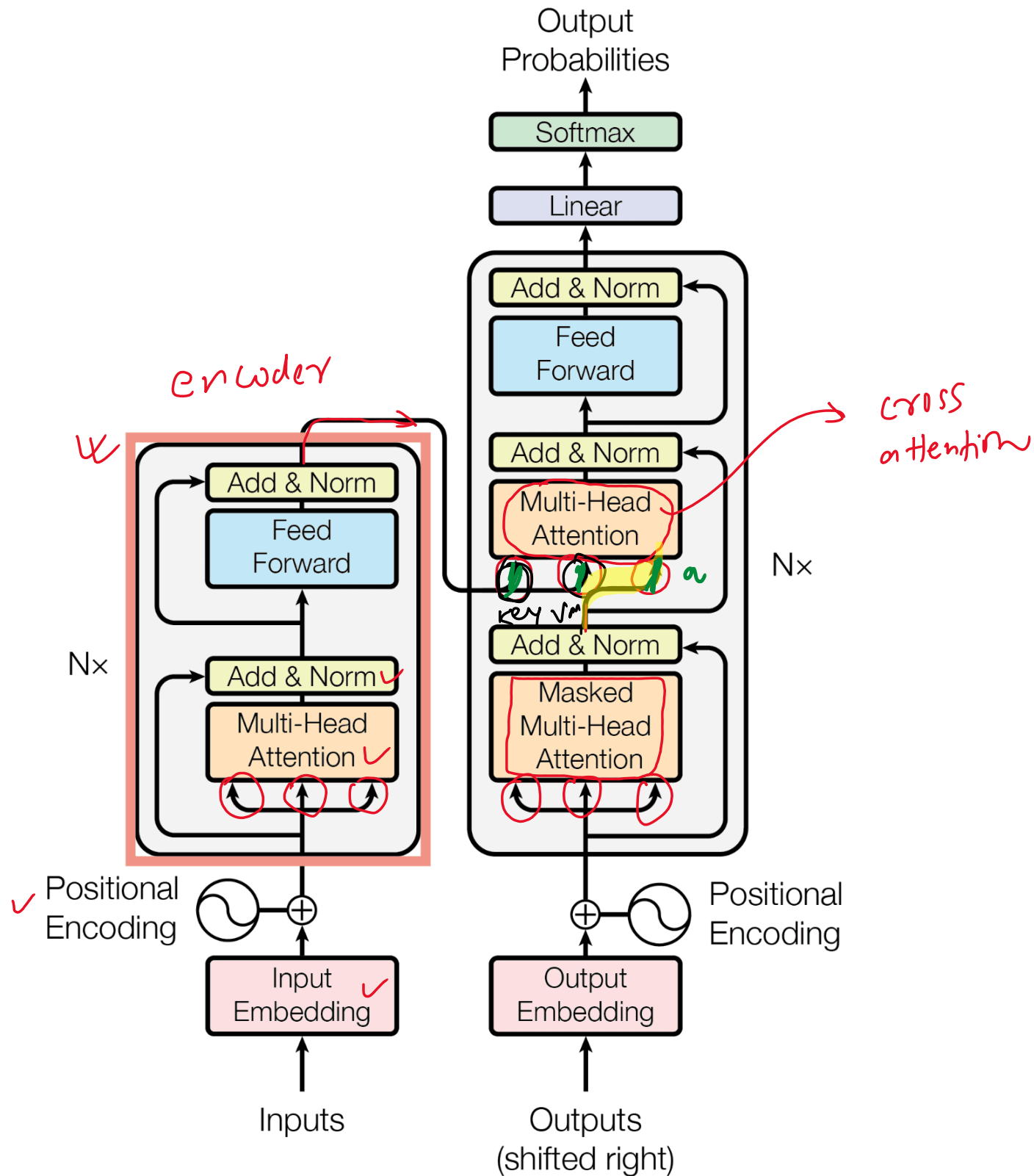


# Plan of Action

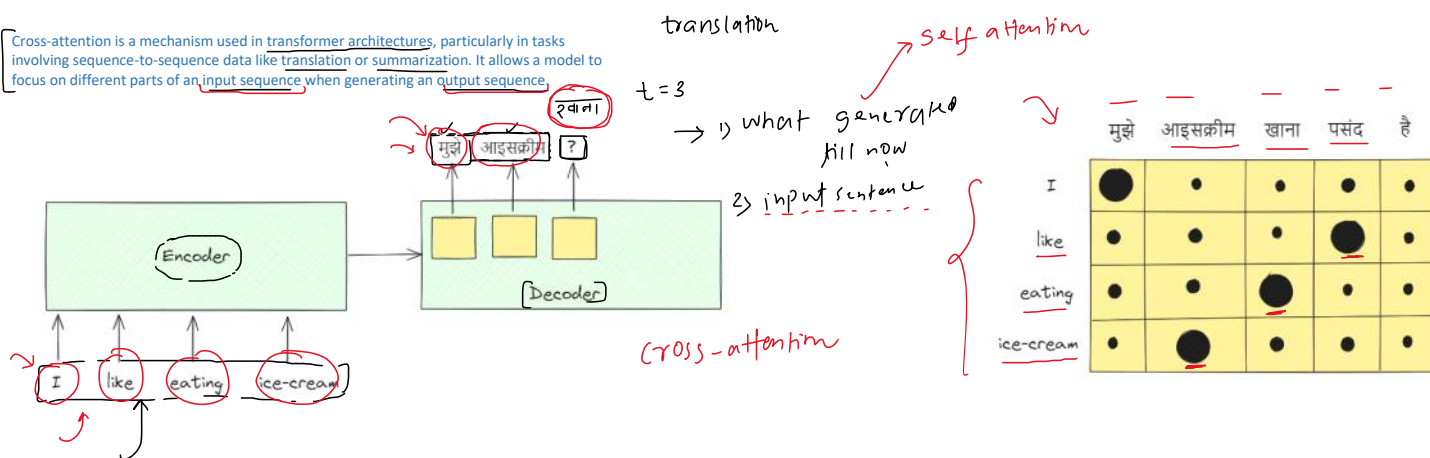
12 August 2024 17:53



## What is Cross Attention

12 August 2024 17:53

Cross-attention is a mechanism used in transformer architectures, particularly in tasks involving sequence-to-sequence data like translation or summarization. It allows a model to focus on different parts of an input sequence when generating an output sequence.



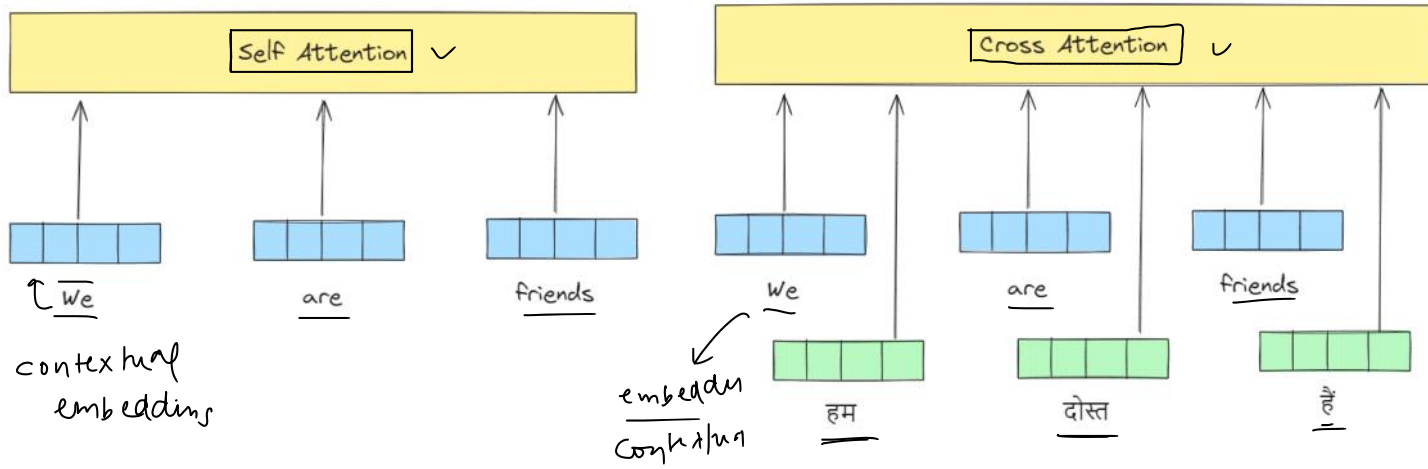
Cross Attention is conceptually very similar to Self-Attention

Self-Attention Vs Cross Attention

1. The input ✓
2. The processing ✓
3. The output ✓

## Self-Attention Vs Cross Attention (Input)

13 August 2024 08:22



# Self-Attention Vs Cross Attention (Processing)

12 August 2024 17:54

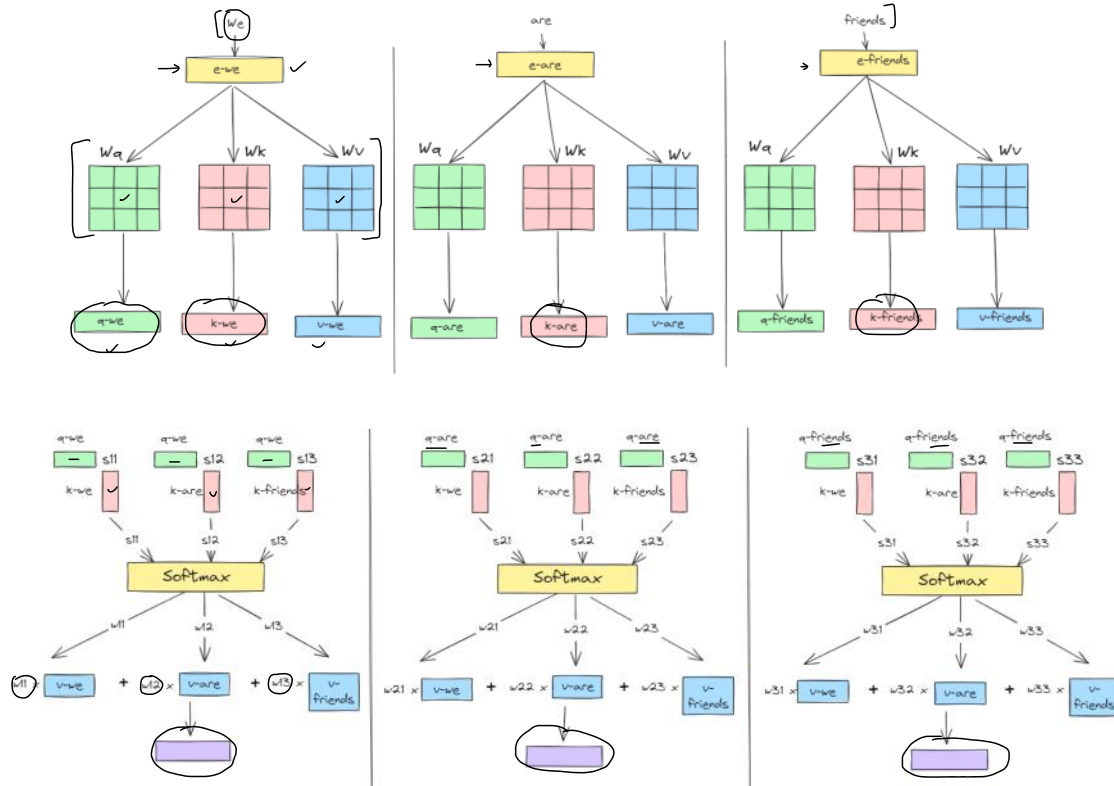


Diagram illustrating the attention weights for the sentence "We are friends".

	We	are	friends
We	●	●	●
are	●	●	●
friends	●	●	●

query → output seq  
key and value - input req

chg - bind

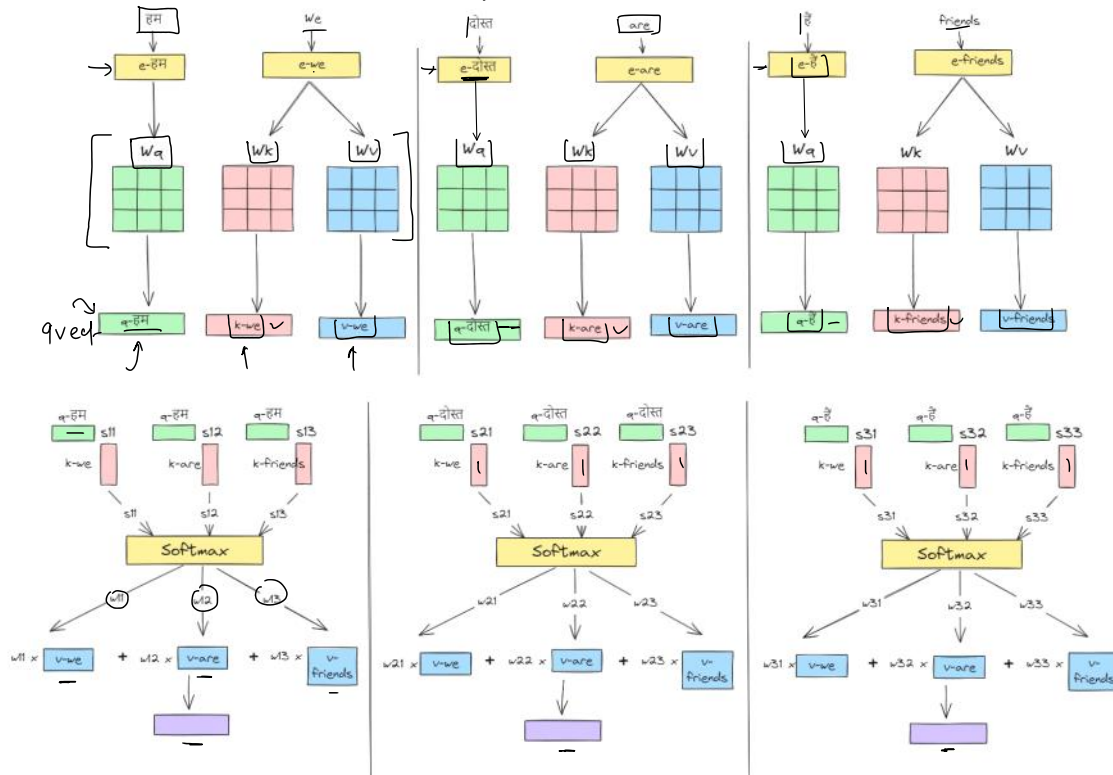
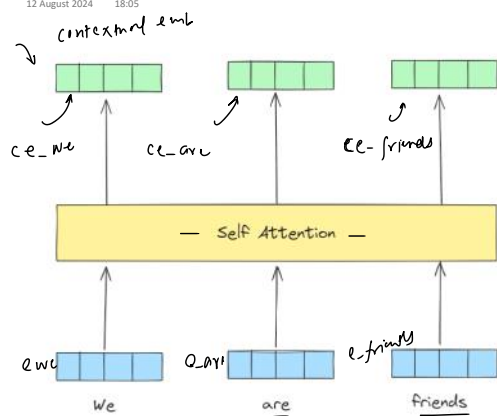


Diagram illustrating the attention weights for the sentence "हम दोस्त हैं" (We are friends).

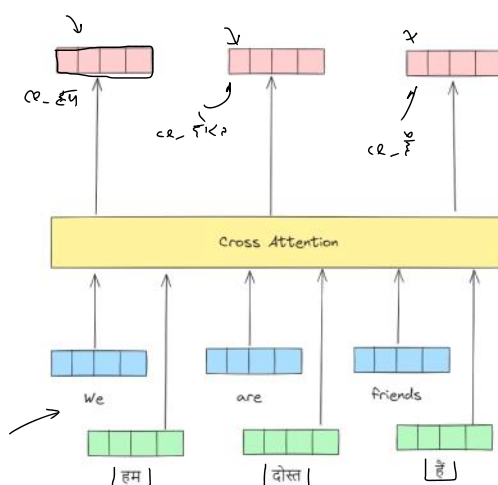
	We	are	friends
हम	●	●	●
दोस्त	●	●	●
हैं	●	●	●

# Self-Attention Vs Cross Attention [Output]

12 August 2024 18:05



$$\begin{aligned} ce\_we &= 0.8 \times e\_we + 0.1 \times e\_are + 0.1 \times e\_friends \\ ce\_are &= 0.15 \times e\_we + 0.75 \times e\_are + 0.1 \times e\_friends \\ ce\_friends &= 0.2 \times e\_we + 0.1 \times e\_are + 0.7 \times e\_friends \end{aligned}$$



$$ce\_ham = 0.5 \times e\_we + 0.3 \times e\_are + 0.2 \times e\_friends$$

$$ce\_दोस्त = 0.2 \times e\_we + 0.2 \times e\_are + 0.6 \times e\_friends$$

$$ce\_दुश्म = 0.3 \times e\_we + 0.4 \times e\_are + 0.3 \times e\_friends$$