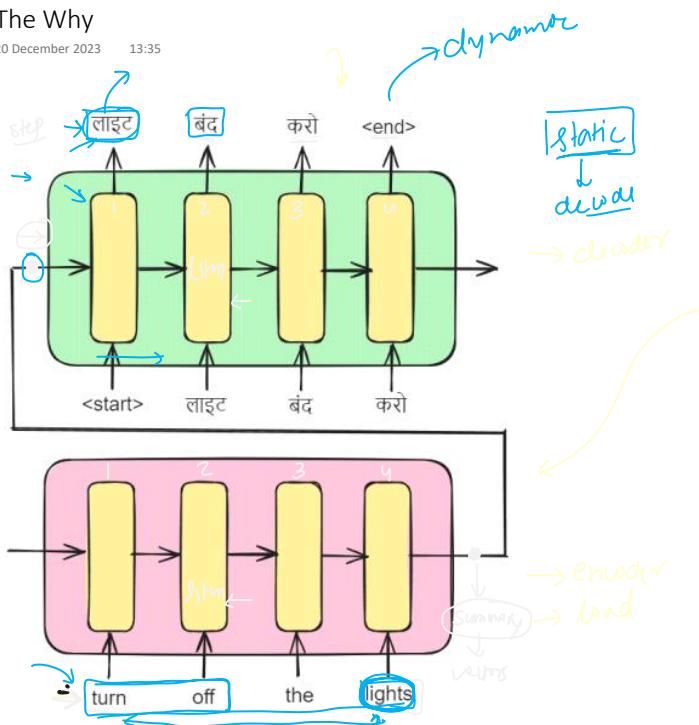


## The Why

20 December 2023 13:35



50 words

Encoder

Once upon a time in a small Indian village, a mischievous monkey stole a turban from a sleeping barber, wore it to a wedding, danced with the bewildered guests, accidentally got crowned the 'Banana King' by the local kids, and ended up leading a vibrant, impromptu parade of laughing villagers, cows, and street dogs, all while balancing a stack of mangoes on its head, creating a hilariously unforgettable spectacle and an amusing legend that the village still chuckles about every monsoon season.

Decoder

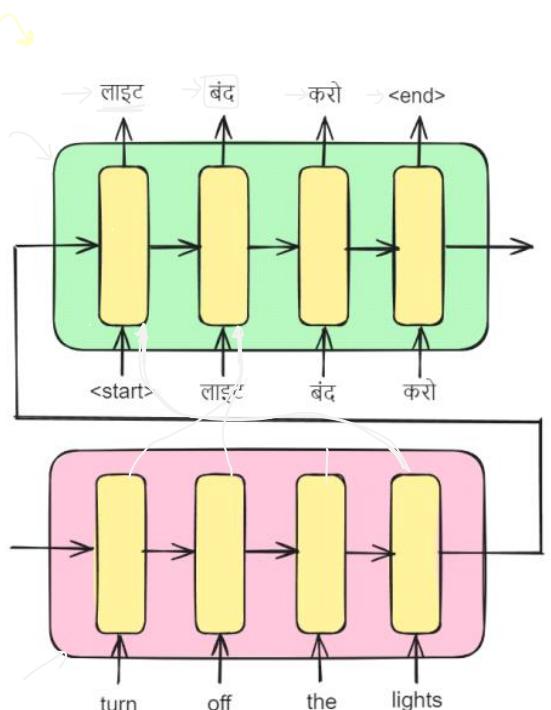
> 25 words

## The Solution

20 December 2023 17:32

Information is valuable to individual companies in determining what information part of Information Security strategy is knowledge based as part as general however intellectual and knowledge-based assets

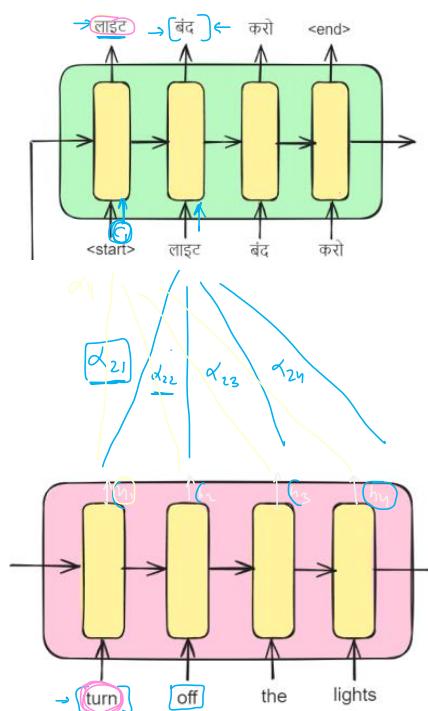
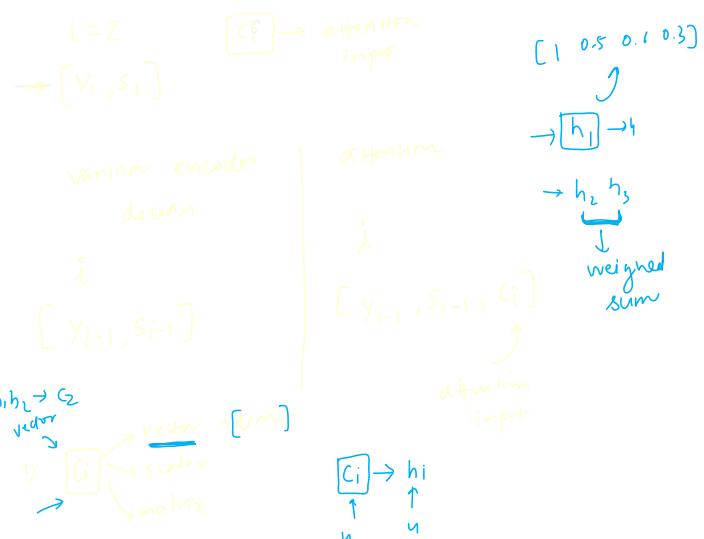
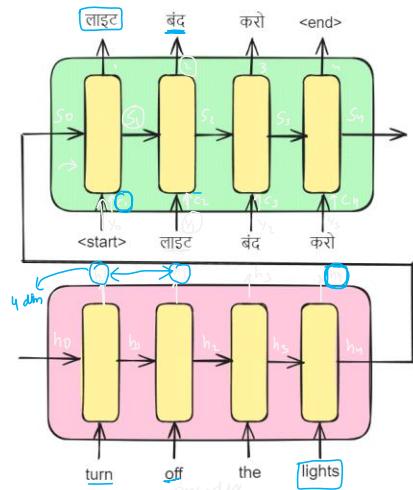
Attention



Once upon a time in a small Indian village, a mischievous monkey stole a turban from a sleeping barber, wore it to a wedding, danced with the bewildered guests, accidentally got crowned the 'Banana King' by the local kids, and ended up leading a vibrant, impromptu parade of laughing villagers, cows, and street dogs, all while balancing a stack of mangoes on its head, creating a hilariously unforgettable spectacle and an amusing legend that the village still chuckles about every monsoon season.

## The What

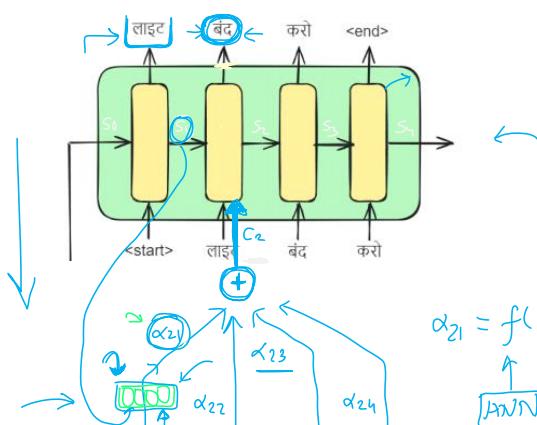
21 December 2023 06:04



$$C_i = \sum_j d_{ij} h_j$$

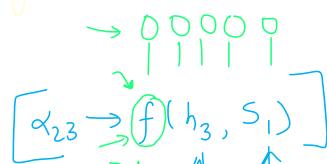
Here

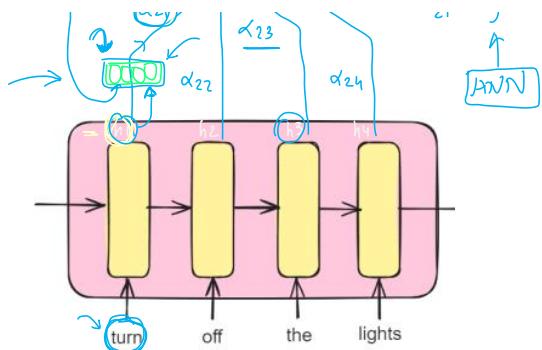
Here in this encoder decoder architecture , the decoder takes following thing at each time step : previous hidden state of at each time step decoder , current input (like token) and the contextual embedding from (i.e  $a_1 * h_1 + \dots$ ) encoder



$$\alpha_2 = f(h_1, s_1)$$

Now this is calculated using  $\text{ann}$ , which is feeded. Like if we have to calculate  $a_{21}$ , (1 for encoder time step, 2 for decoder) then  $h_2(\text{decoder}) + h_1(\text{encoder})$  wala is feeded.  $a_{21}$  is like how much we need encoder wala given that previous decoder wala is this. Using this context wala  
 $c_2 : (a_{21} * h_1 + a_{22} * h_2 + a_{23} * h_3 \dots)$   
so to produce output :  $(c_2 + h_2(\text{decoder}) + \text{prev output})$



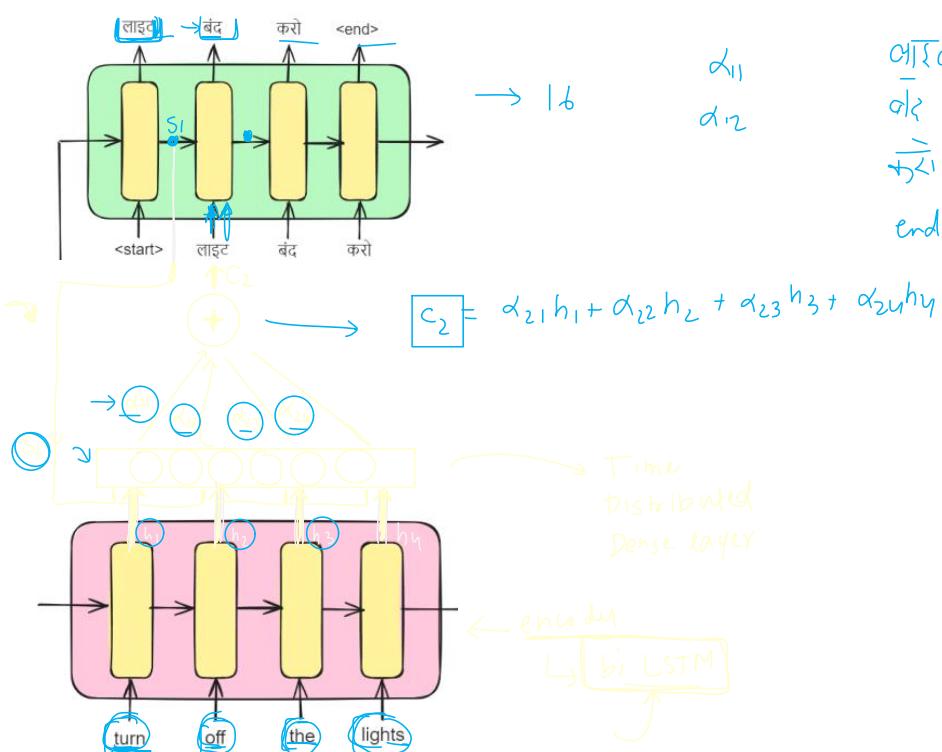


$\alpha_{21} \rightarrow f(h_1, s_1)$

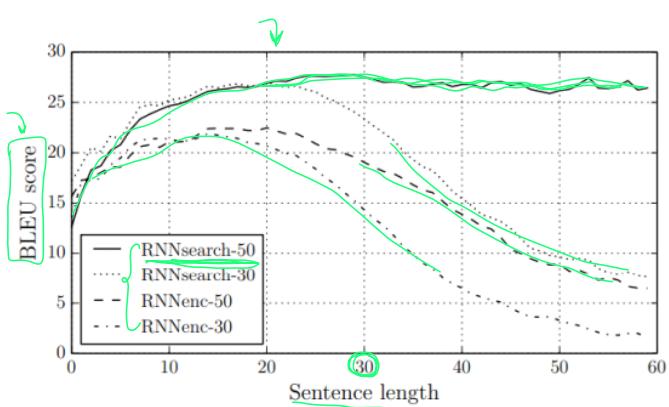
$$\alpha_{ij} = f(h_j, s_{i-1})$$

What function?

uniquely approximated

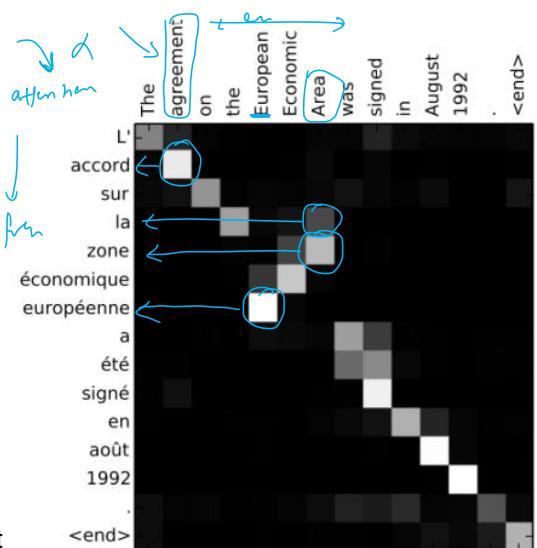


alpha : alignment score



Here like Bleu score is calculate and use of attention improves

This thing show how is like each decoder wala is influence by input wala , some requires (some two three words for context)



(a)