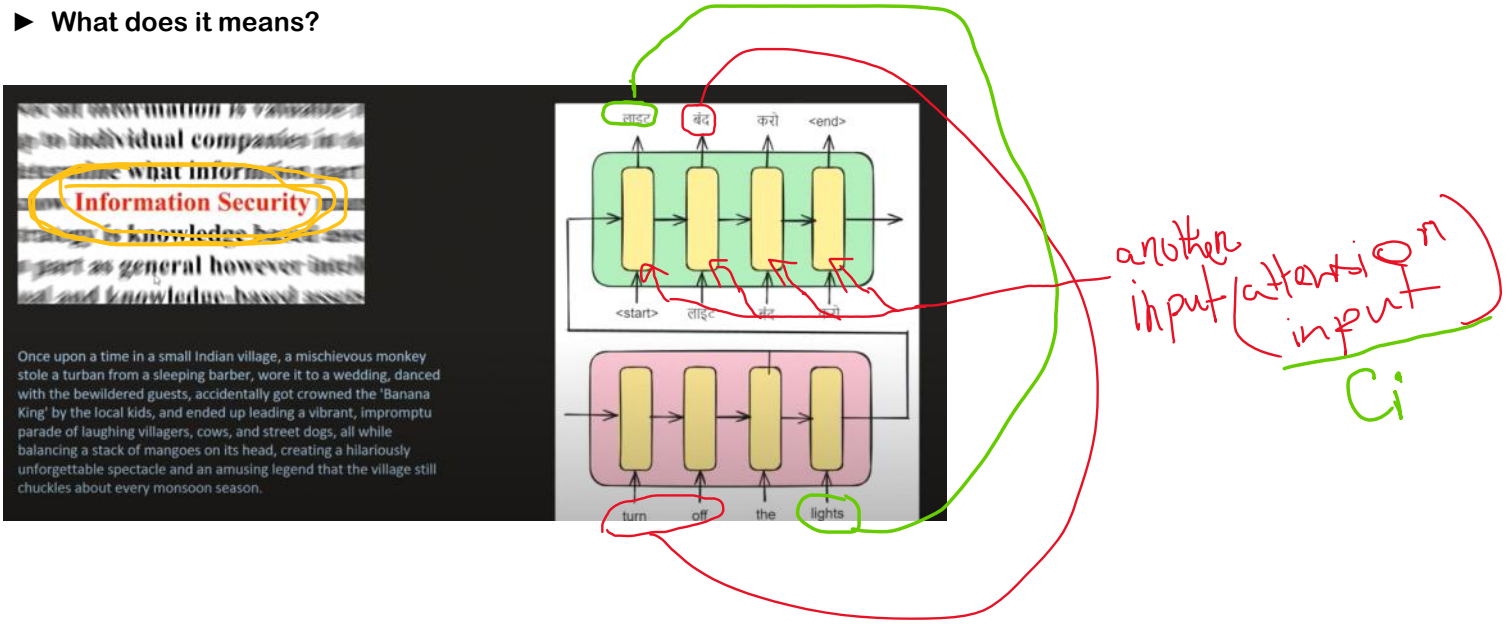
Attention Mechanism

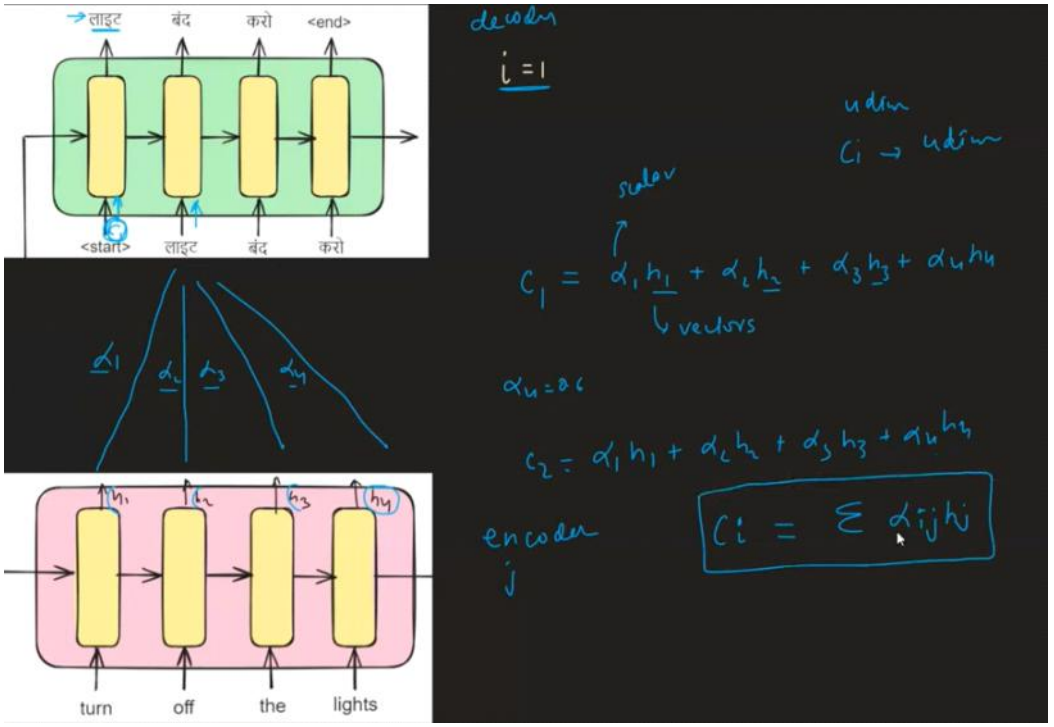
Sunday, March 9, 2025 11:45 AM

► What does it means?



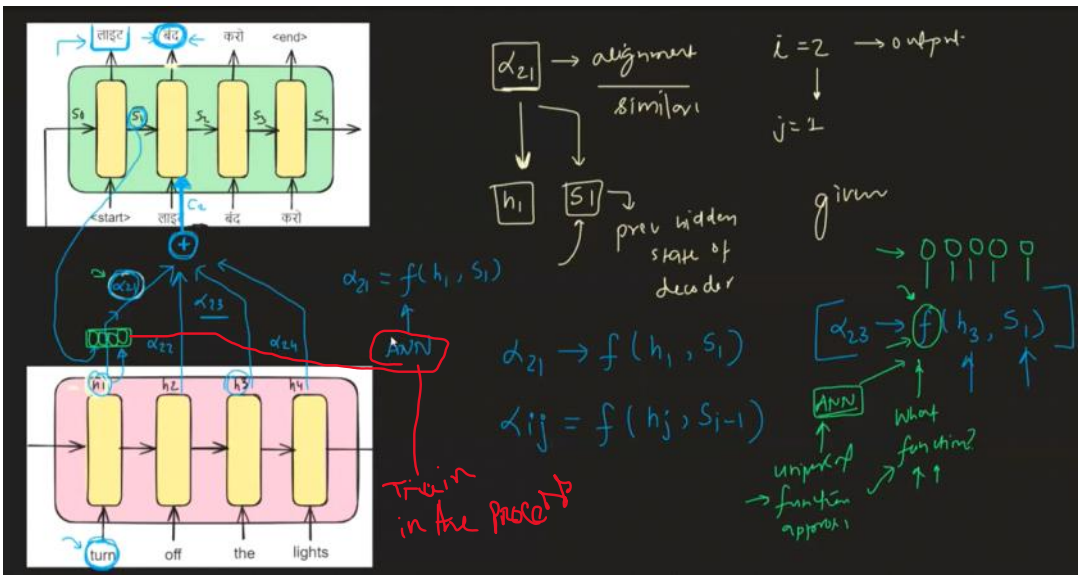
- We human don't focus on the entire sentence to translate a word
- Just focusing on a particular word can solve the problem
- Focusing on the timestep in encoder inputs can be useful to the decoder. Which timestep in important for this output?
- Here Ci is the attention vector(important for the output word) which refer from the encoder.

► How Ci get the vector?



- For each timestep there is a weight assign to that input
- Ci is the summation of the multiplication of weight and the input vector

► How Alpha(i) calculated in this process?



- Here the value of Alpha depend on "S" and "h", cause we need the value of Alpha after the last output and the depended inputs.
- But which function should be used to get better result F(h, Alpha)?
- ANN has been used in this architecture.