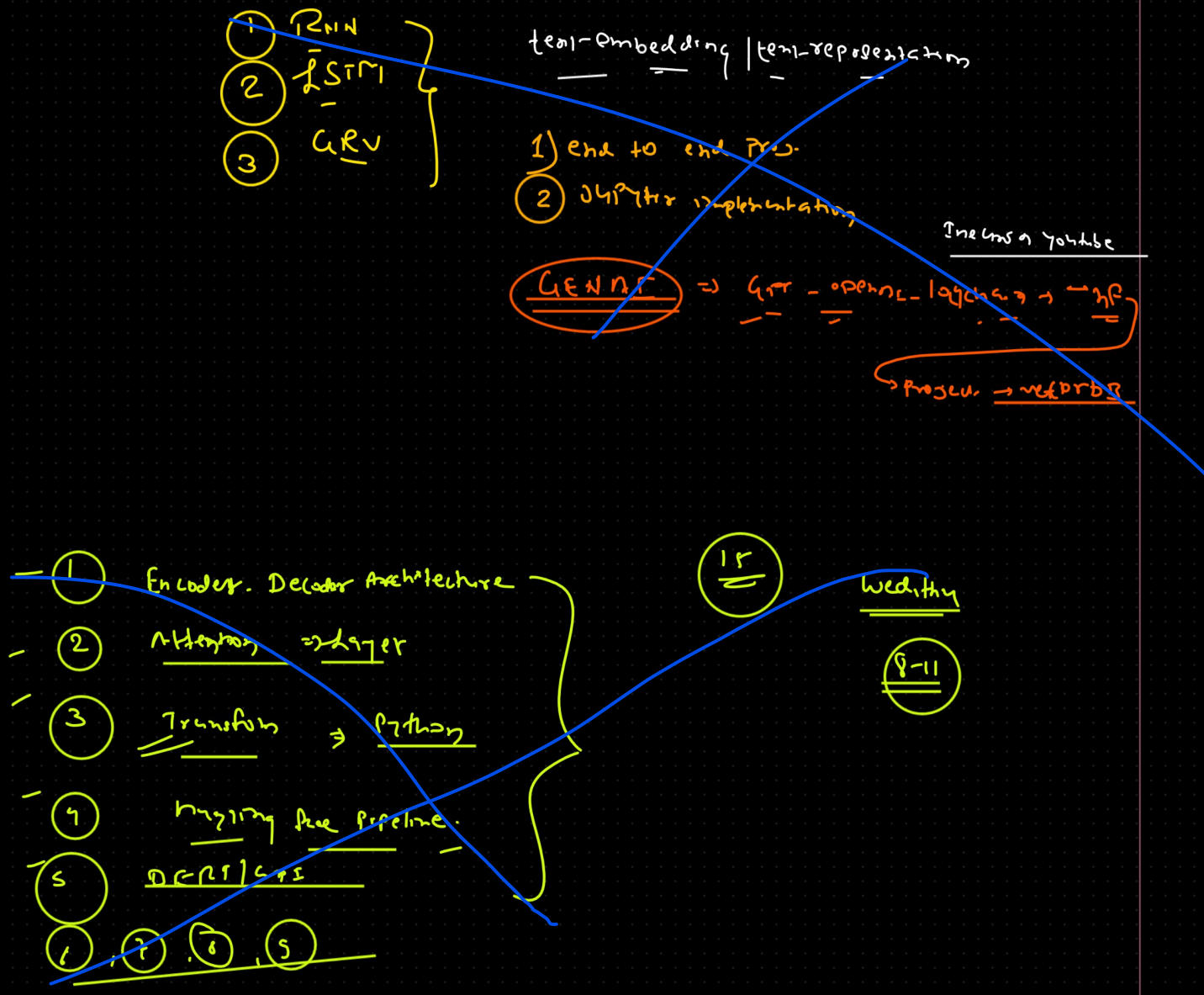



Agneda:

## Encoder-Decoder

1. introduction
2. sequence modelling Problem
3. Architecture of encoder & Decoder
4. training and test
5. Project → Understanding

## Seq to Seq Learning with NN

This is the name of the research paper published in 2014 discussing this Encoder Decoder architecture.

In last lecture we already discussed drawback of many to many seq to seq mapping while using RNN/LSTM/GRU which is fixed dimensionality. Due to this in real word scenario in case of many to many seq mapping where i/p and o/p dimensions are different if we using RNN/LSTM/GRU then they are prone to fixed dimensionality as a result we may generate output which is having missing value, wrong sequence, grammatical mistakes etc. Encoder Decoder address this stated issue.

Different evaluation metrices used in NLP:

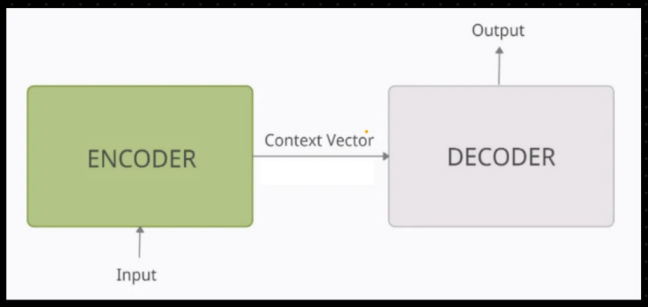
- BLEU (BiLingual Evaluation Understudy) is a metric for evaluating the quality of machine-translated text.
It's a number between 0 and 1 and measures how similar the machine-translated text(Predicted) is to a set of high-quality reference translations(Actual)

- Perplexity: In Natural Language Processing (NLP), perplexity is a way to measure the quality of a language model independent of any application

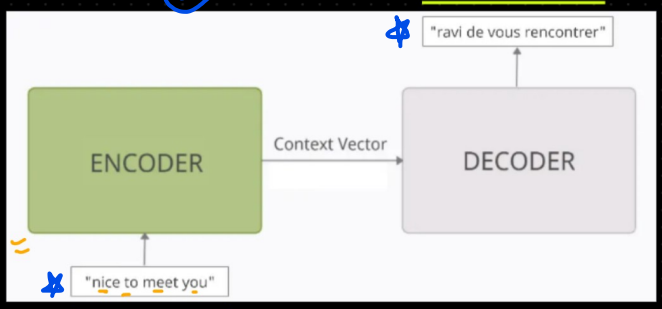
- GLUE score: GLUE, also known as General Language Understanding Evaluation, is an evaluation benchmark designed to measure the performance of language understanding models in a range of natural language processing (NLP) tasks. It provides a standardized set of diverse NLP tasks, allowing researchers and practitioners to evaluate and compare the effectiveness of different language models on these tasks.

Understanding Encoder Decoder architecture using following figures:

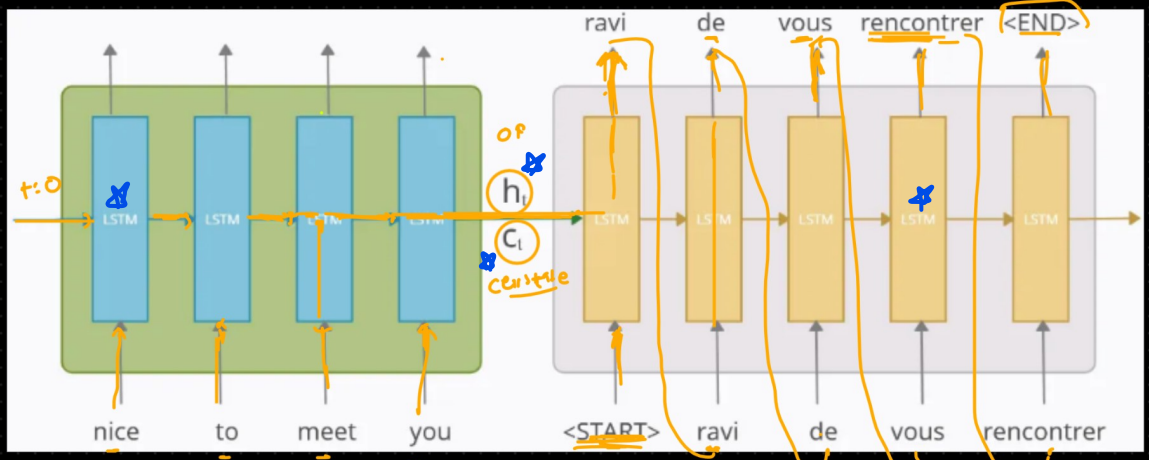
1



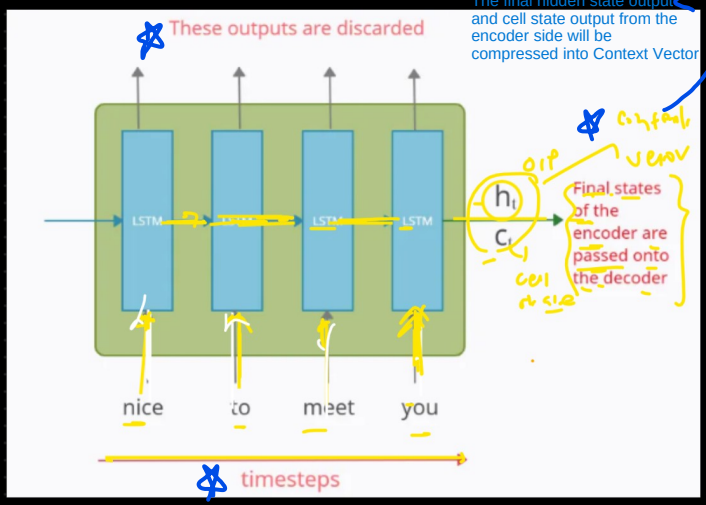

2



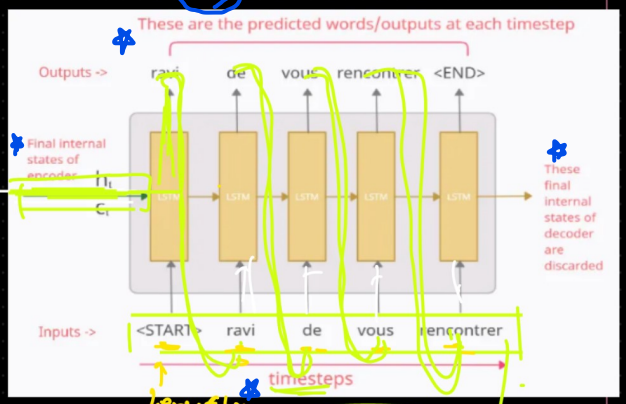

3




4 Encoder side




5 Decoder side




Unsupervised

Supervised

Supervised learning

I/P

nice to meet you

ravi de vous rencontrer

This language translation task is the supervised learning.

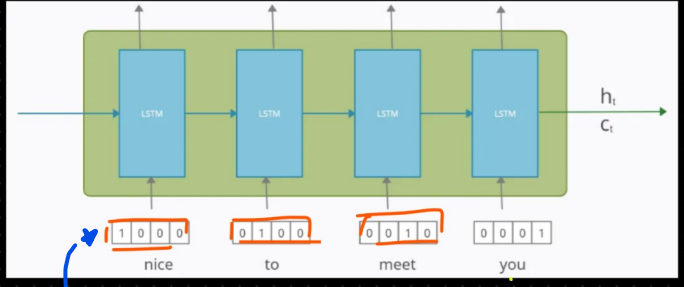
Word2vec

Onehot

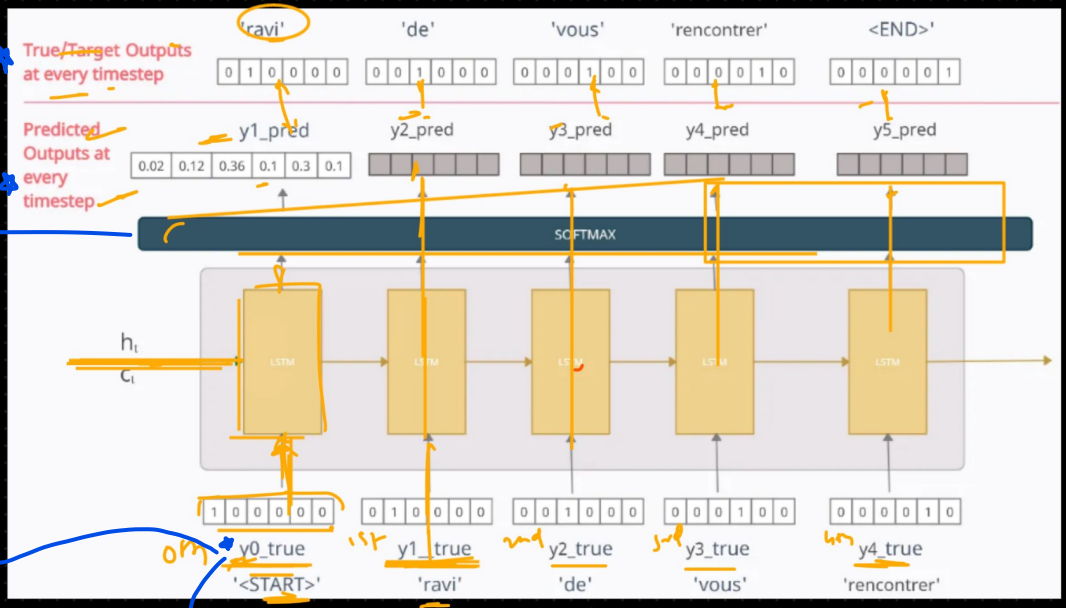
Word2vec/Elmo

Data → Vocab = Vectors

6




We also perform encoding before passing to the model. Here, for example we have performed one-hot encoding.
One can use other encoding techniques such as Word2Vec, Elmo etc.

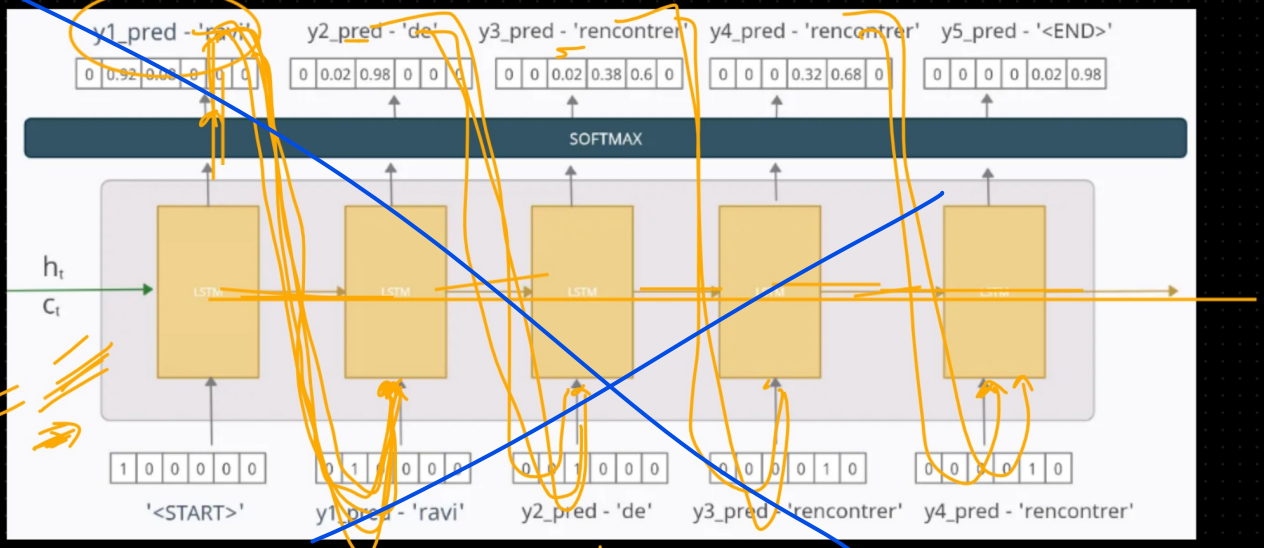



This is Decoder side

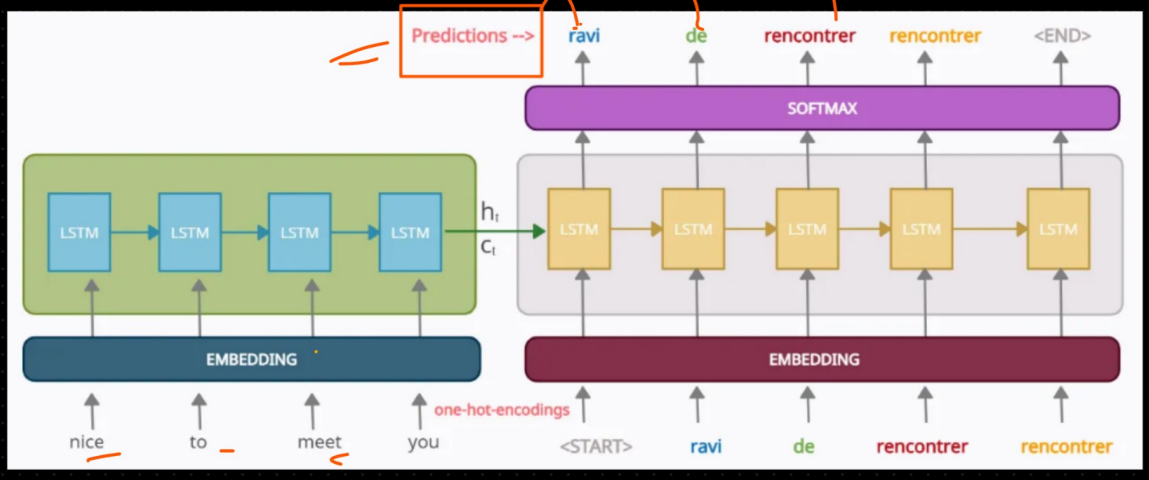
Softmax fun^ to convert the vectors into probability values

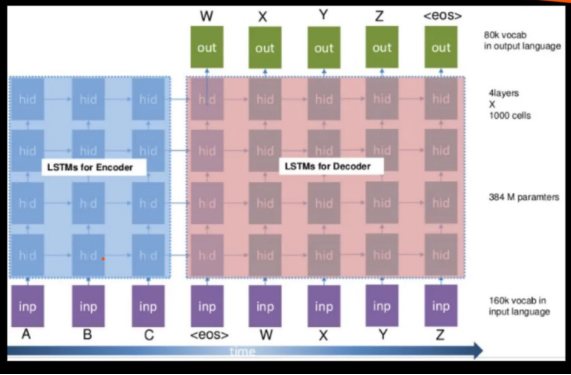
Also, observe that at the base of decoder side we are feeding the true values

y0_true means actual output at 0th timestamp. Similarly, for y1_true, y2_true .. so on.



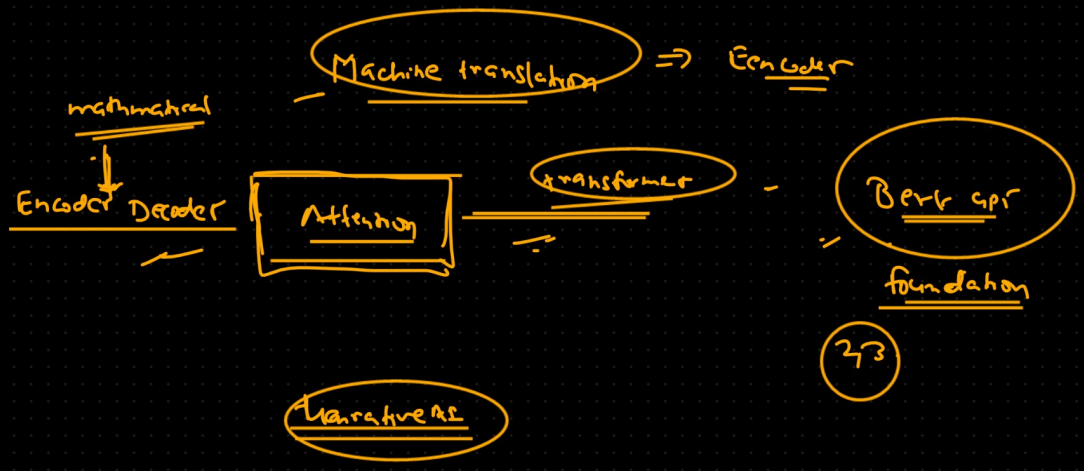

Summarized architecture of the Encoder Decoder

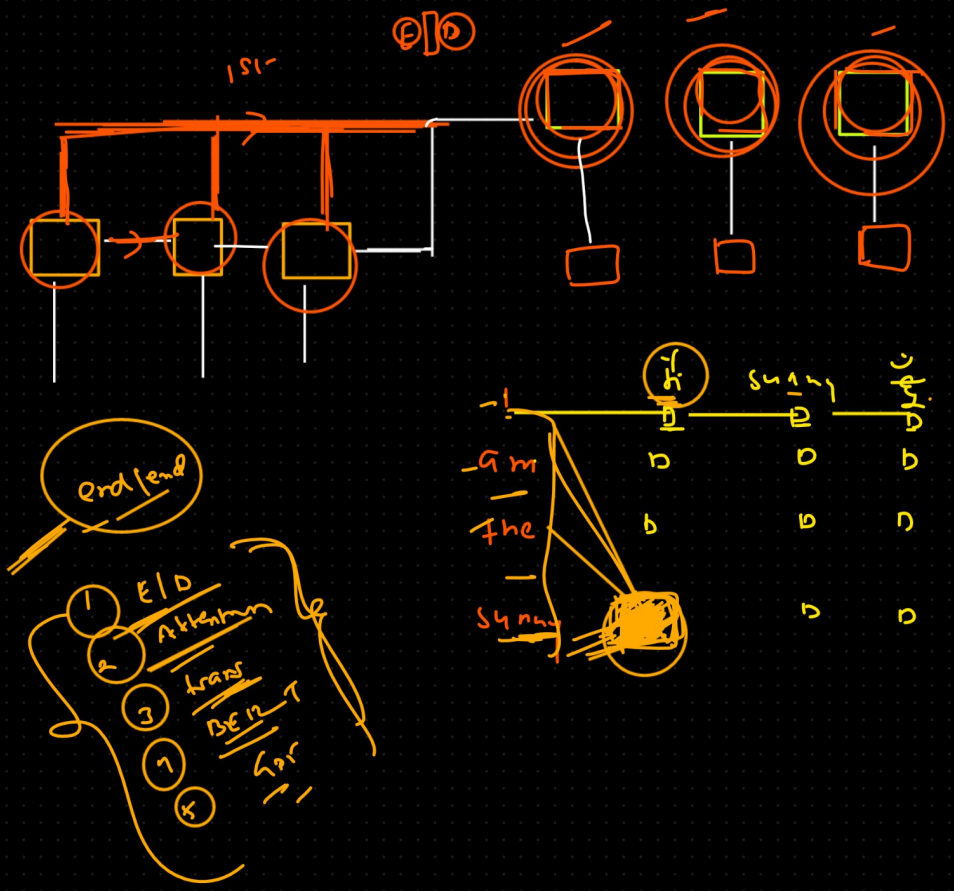



Architecture of Encoder Decoder formed through multiple layers of RNN/LSTM/GRU or Deep architecture

Encoder passes a lot of information to the decoder. Instead of passing only the last hidden state of the encoding stage, encoder passes all the hidden states to the decoder by compressing it in the form of the context vector.




Jalammar is renowned researcher in the space of NLP. Read his blogs where he has decoded different NLP research papers and concepts in a very beautiful manner.