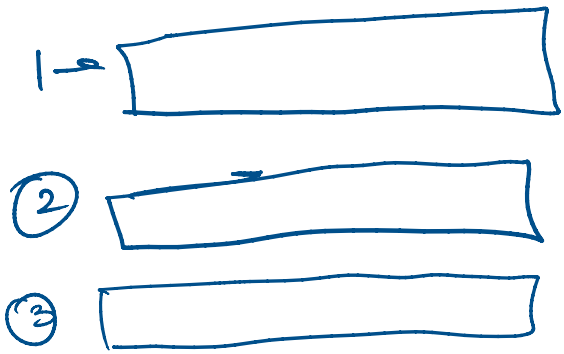
Transformers

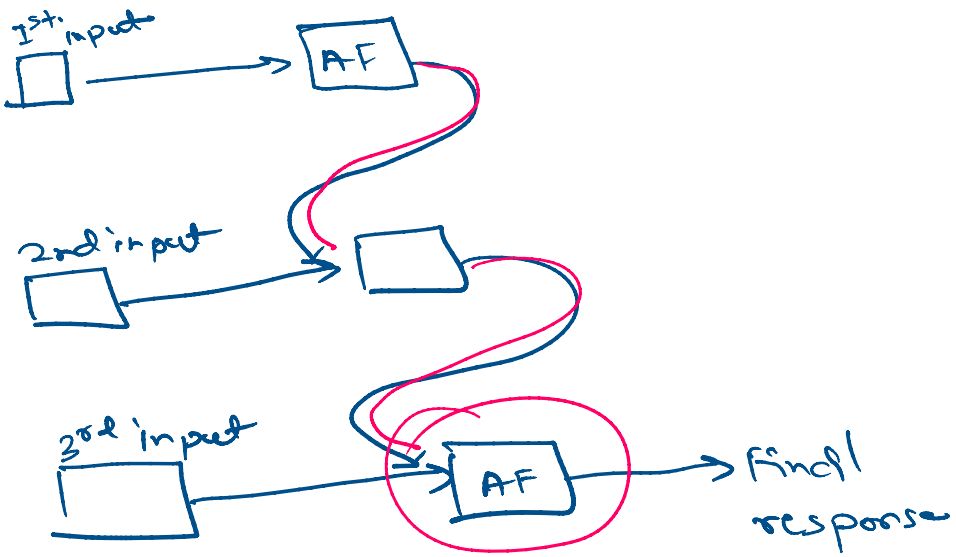
26 June 2024 07:03

RNN

Generating answers
Text summarizing
Machine Translation

RNN



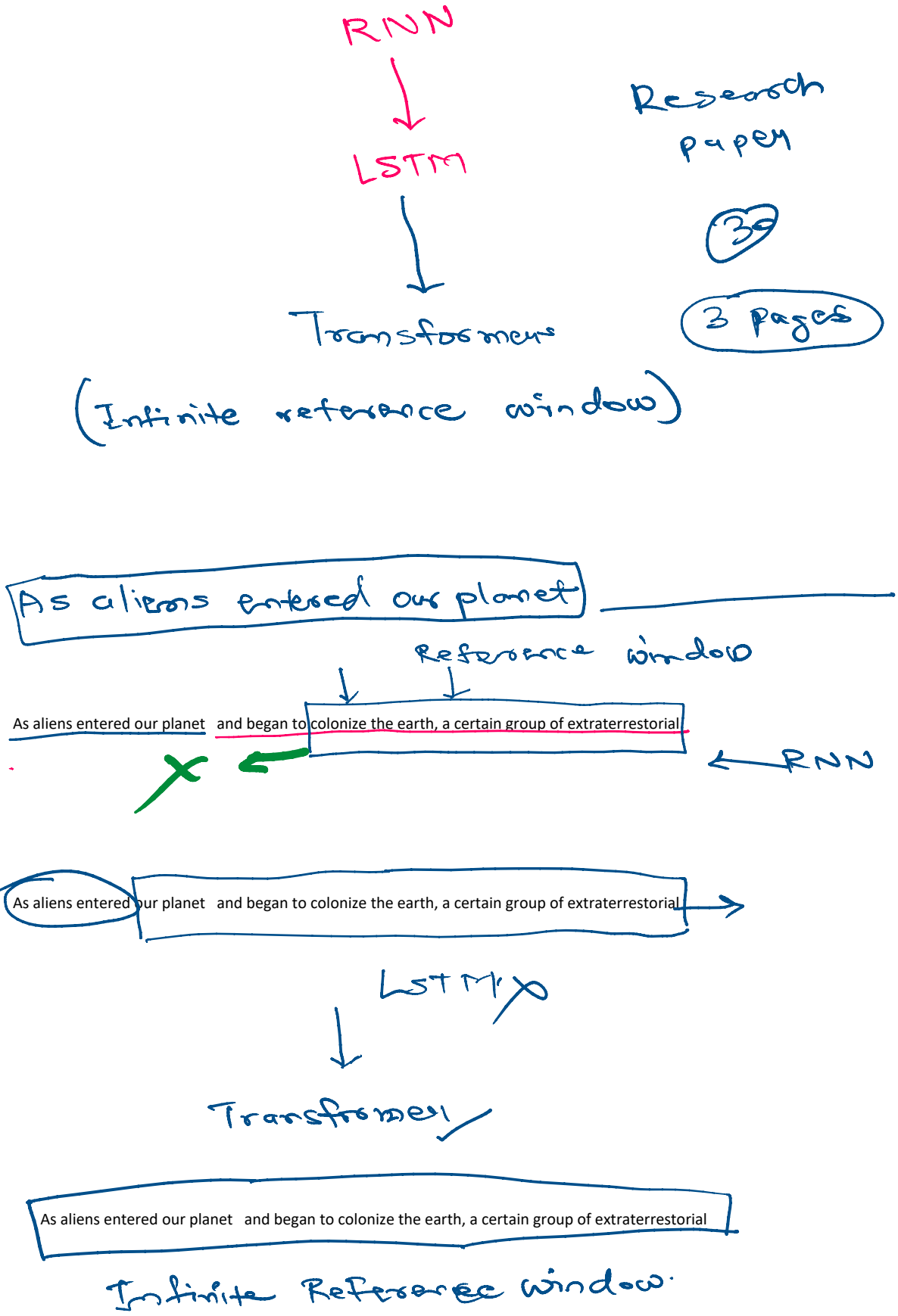


RNN

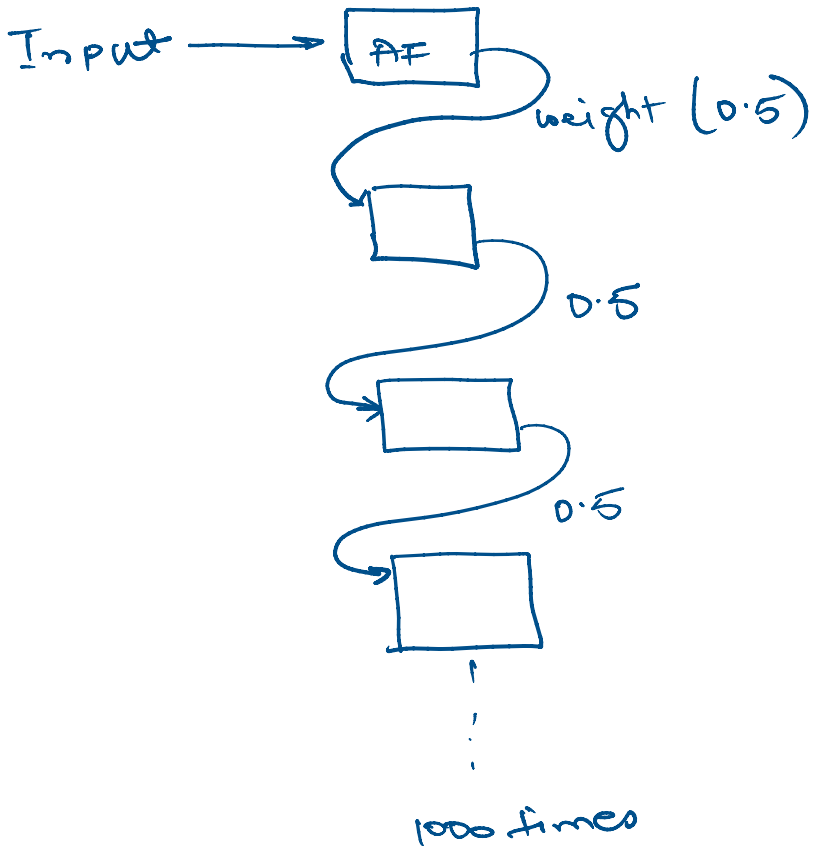
memory loss

100 lines

101, 102, 103



Why does LSTM and RNNS forget? Is it because of vanishing gradient?



Vanishing gradient

Input x 0.5 x 0.5 x 0.5 x 0.5

0.25 x 0.25

0.0625

Exploding Gradient

Input x 2 x 2 x 2 x 2 ...

2^{1000}

↓

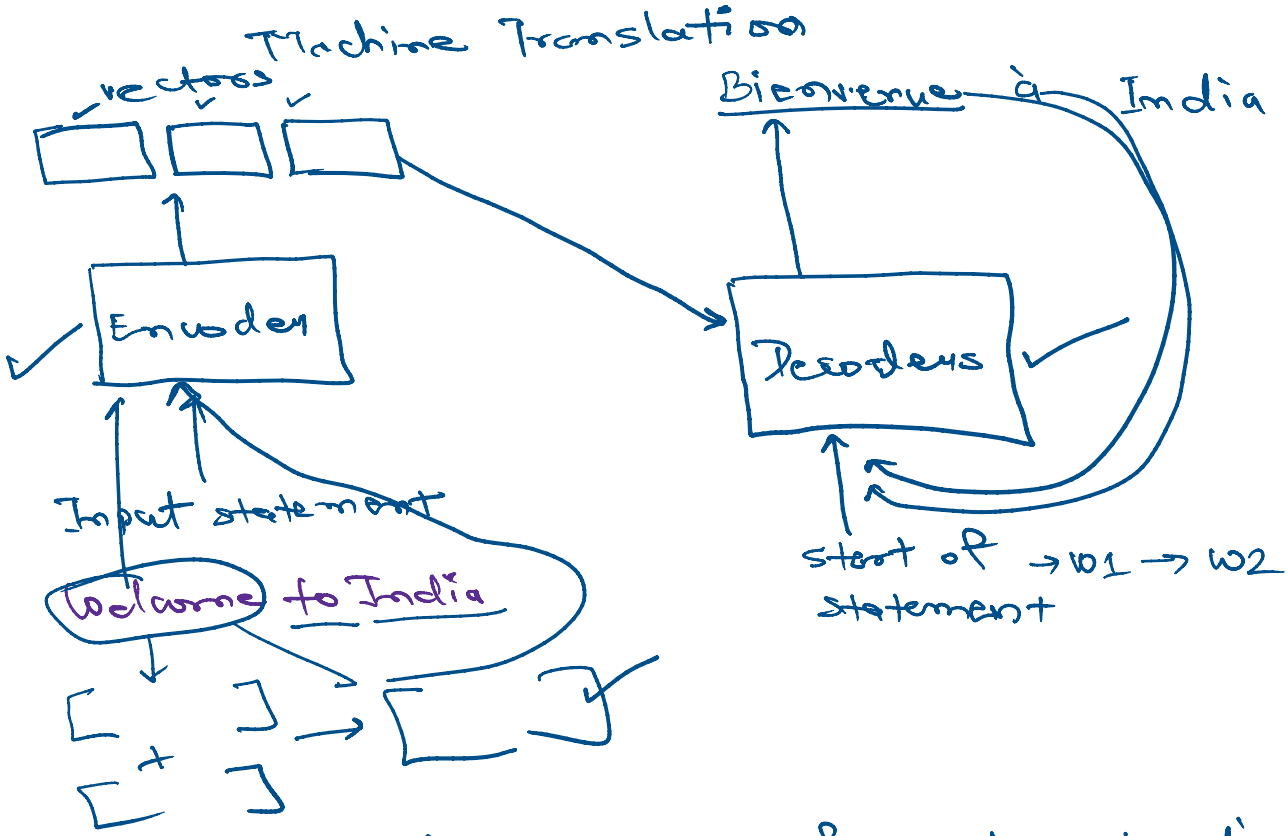
Extremely large value

Certain size

2^{50}

2^{1000}

Transformers



Encoders takes care of understanding the sequence.

Decoder generates a sequence acc to the understanding of the encoder.

Ex →

① The cat sat on the mat

↓

↓

Tokenize

② ['The', 'cat', 'sat', 'on', 'the', 'mat']

↓

③ Word Embedding

'The' : [0.2, 0.5, -0.17] → 3

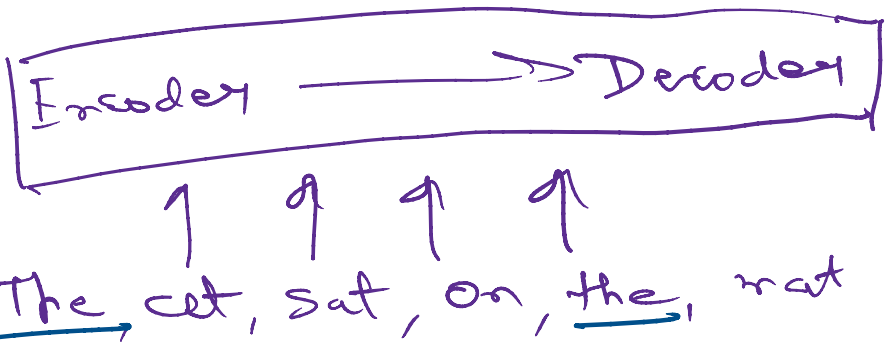
'cat' : [0.5, 0.1, 0.7]

sat : [ ]

on : [ ]

the : [ ]

mat : [ ]



Position has to be maintained.

(4) Position Encoding →