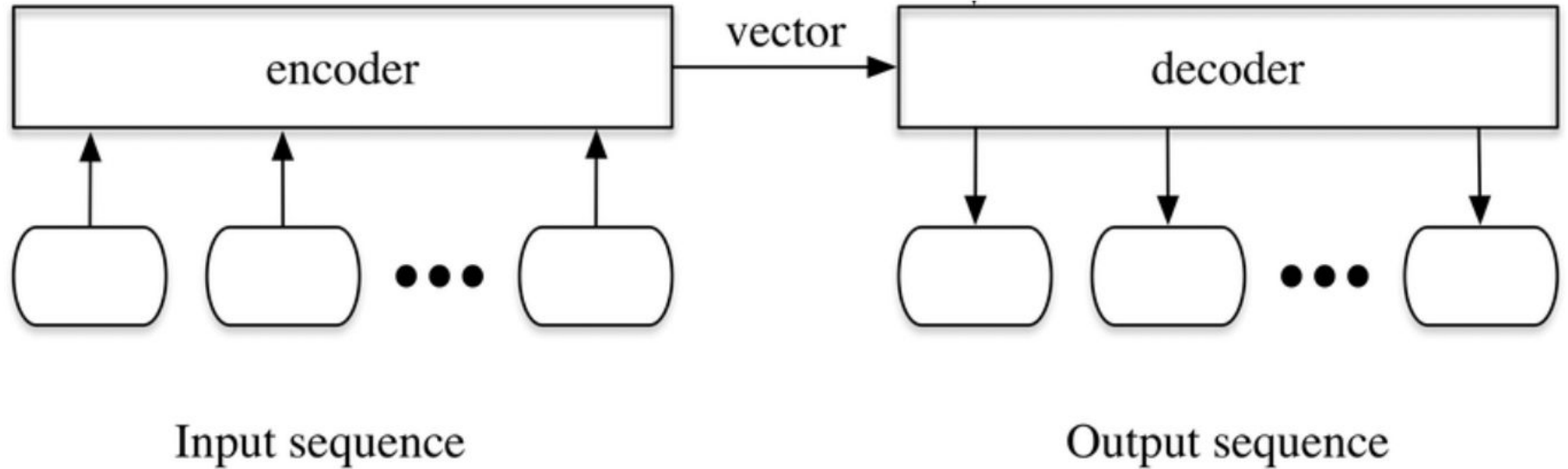


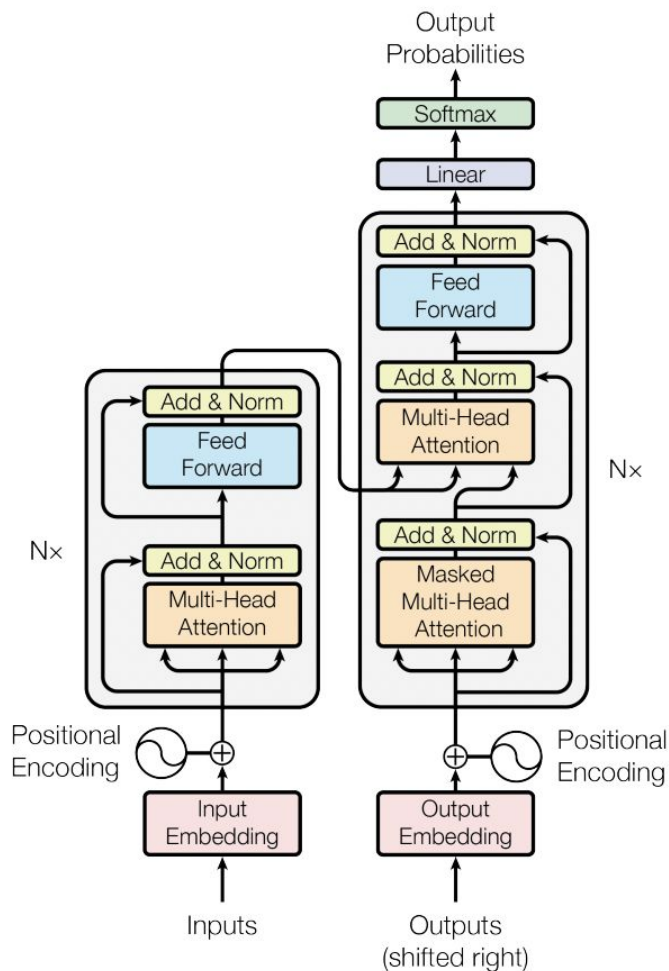
Advanced Machine Learning

Likhith Nayak

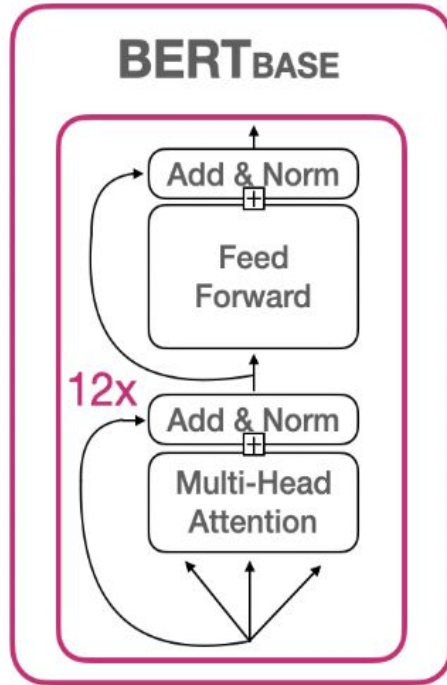
Sequence-to-sequence model



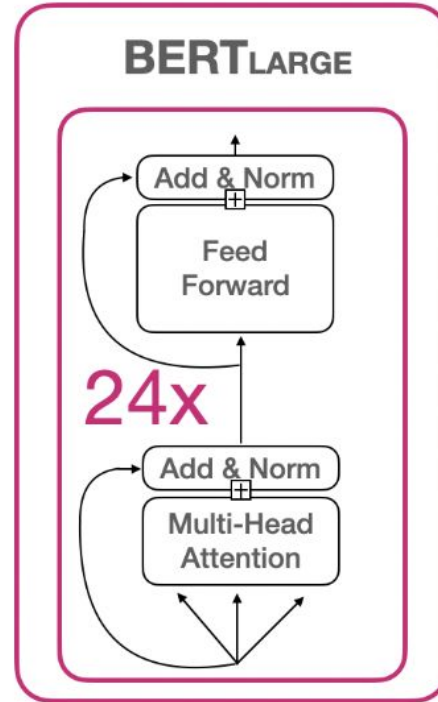
Transformer Network



BERT Architecture

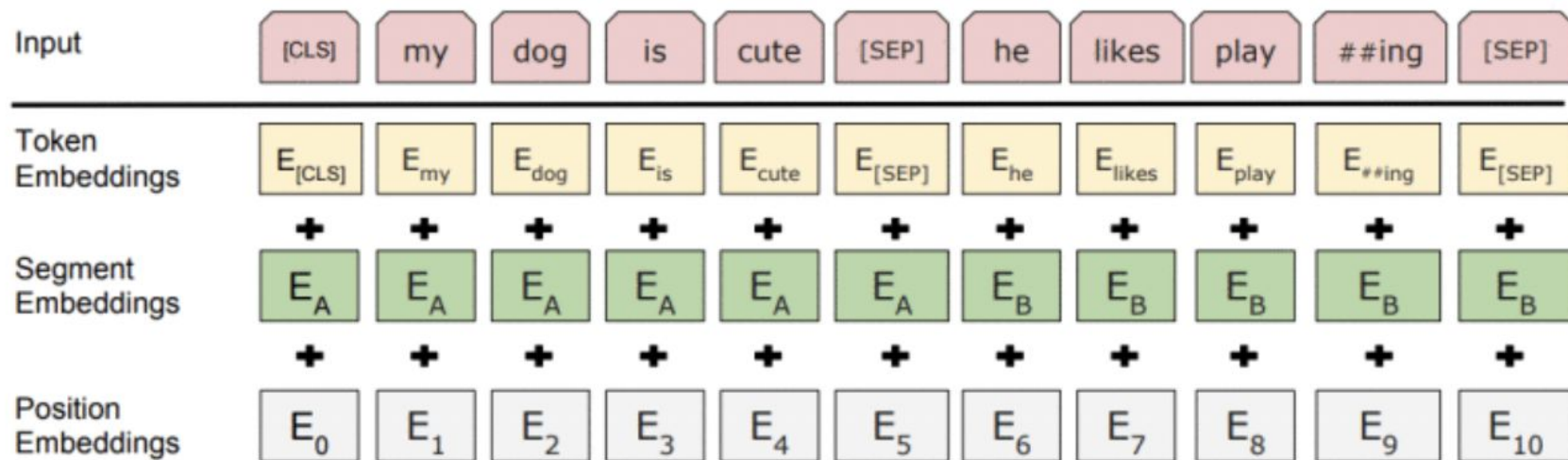


110M Parameters



340M Parameters

BERT Training



Masked LM

the man went to the [MASK] to buy a [MASK] of milk

store gallon

↑ ↑

Next Sentence Prediction

Sentence A = The man went to the store.
Sentence B = He bought a gallon of milk.
Label = IsNextSentence

Sentence A = The man went to the store.
Sentence B = Penguins are flightless.
Label = NotNextSentence

BERT Architecture

