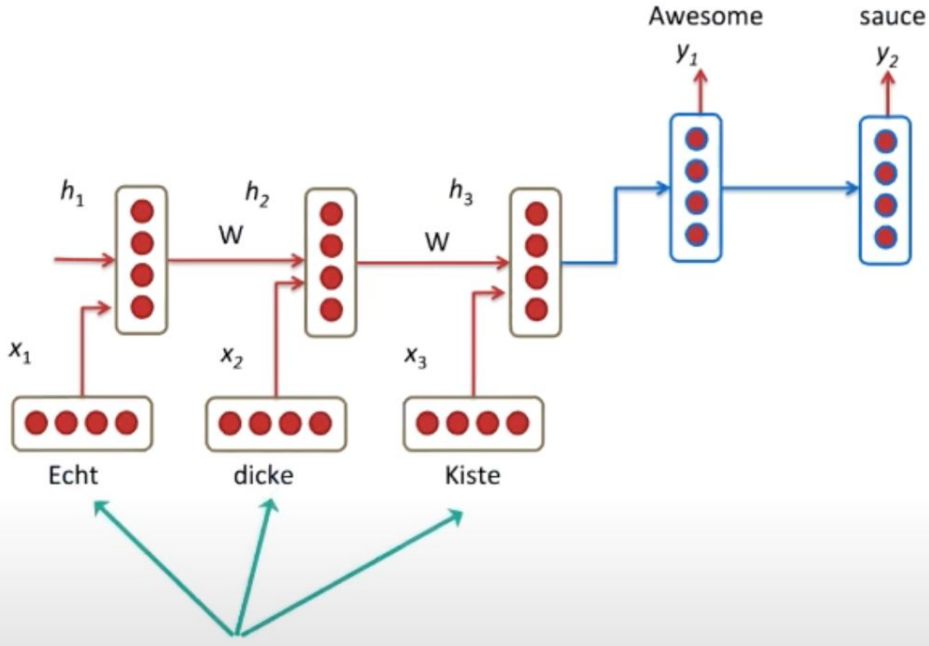
BERT

Credit : https://www.youtube.com/watch?v=xI0HHN5XKDo&t=196s

LSTM Vs Transformer

LSTM Networks

1. Slow
2. Not truly Bidirectional

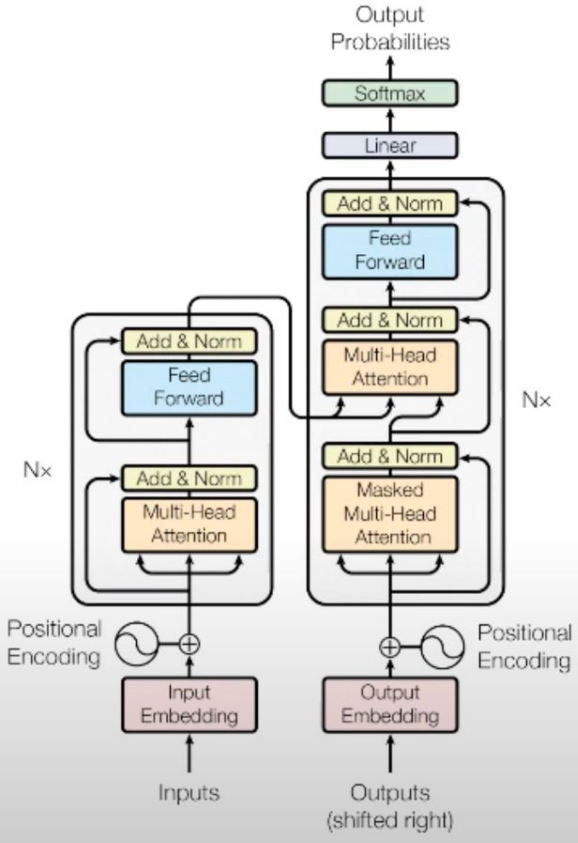


True meaning of source words not entirely captured

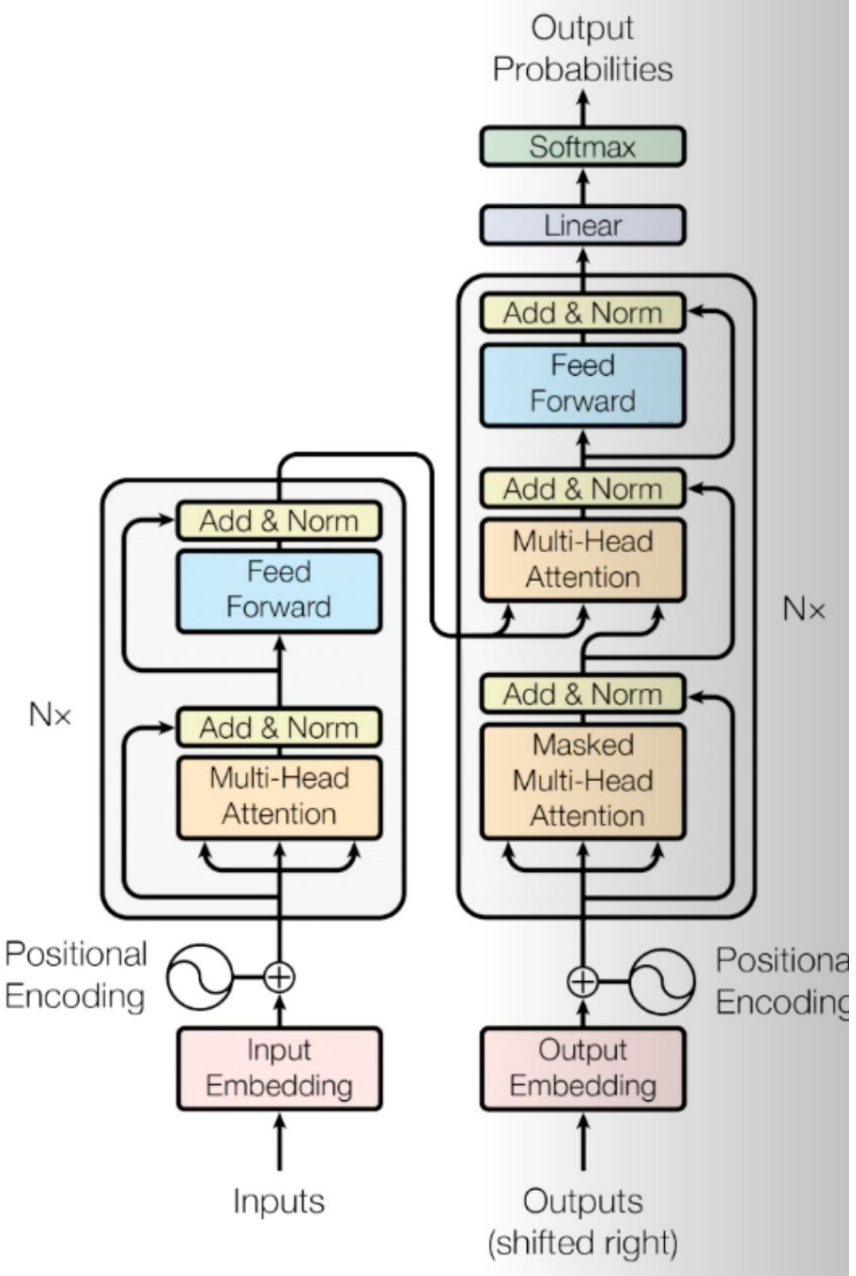
LSTM Vs Transformer

Transformer

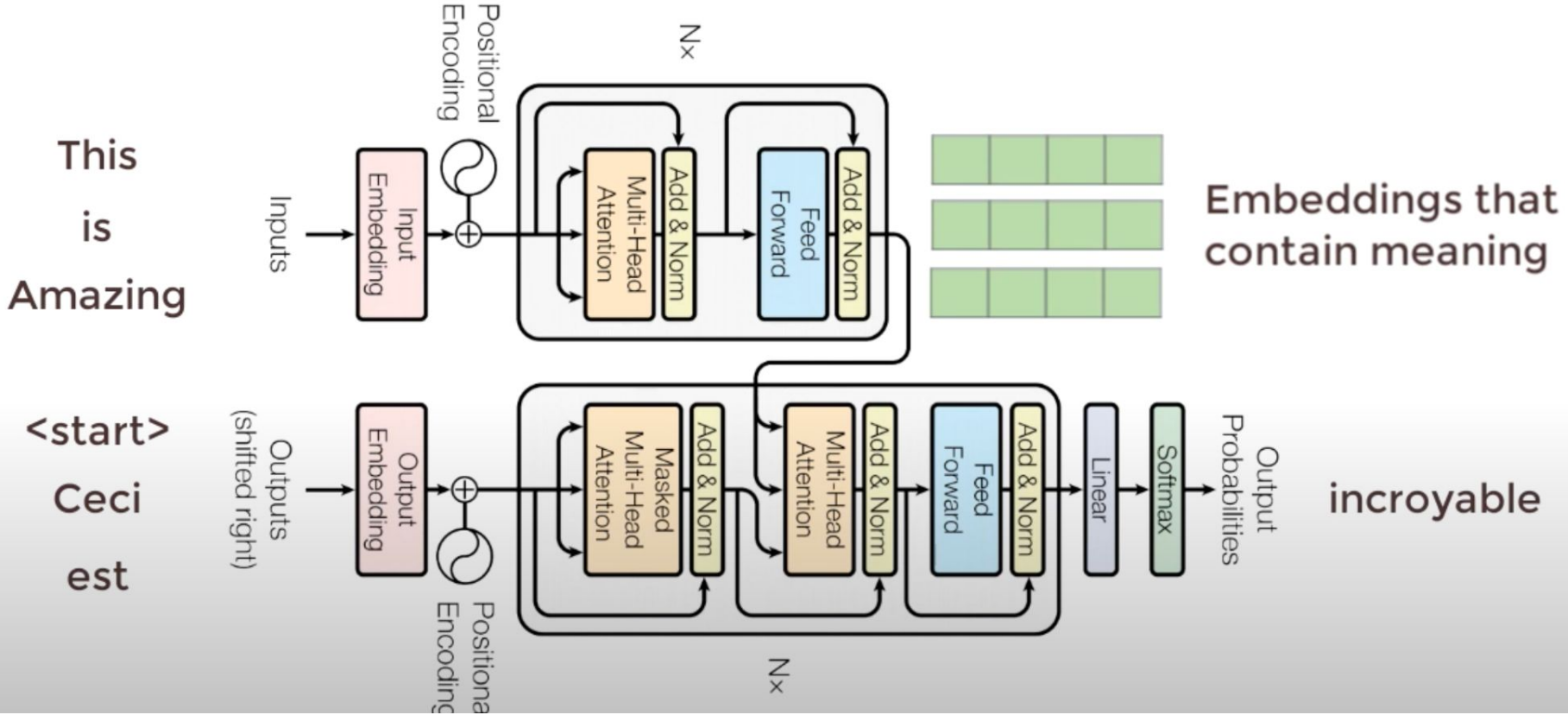
1. ~~Slow~~ Faster
2. ~~Not truly Bidirectional~~
   Deeply Bidirectional



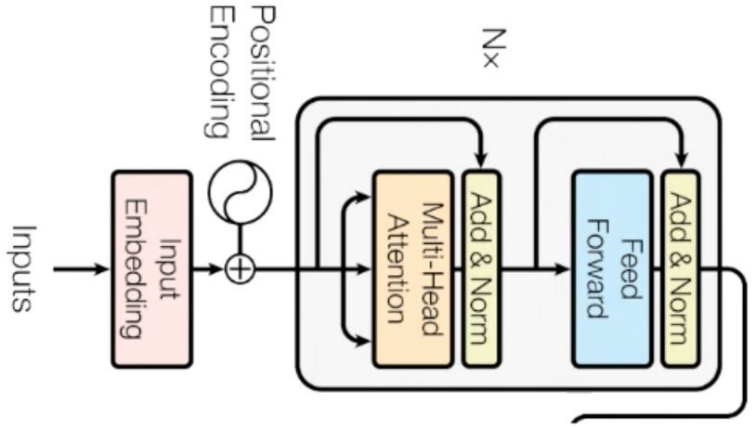
Transformer Flow



Transformer Flow

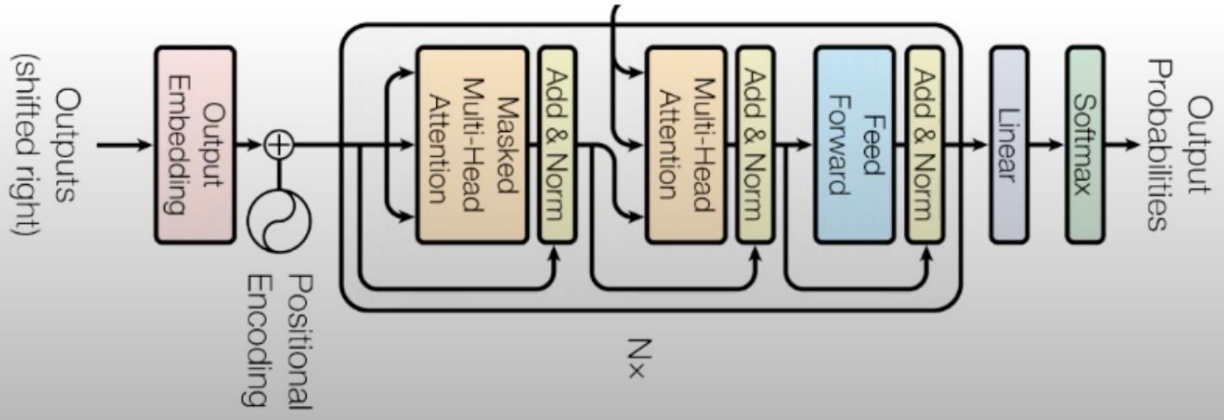


Transformer Flow



What is English? What is context?

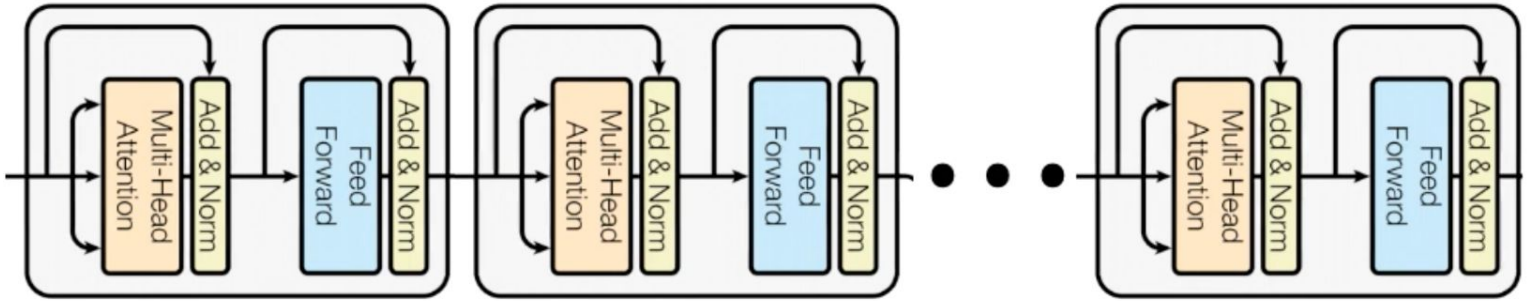
What is language!



How to map English words to French words?

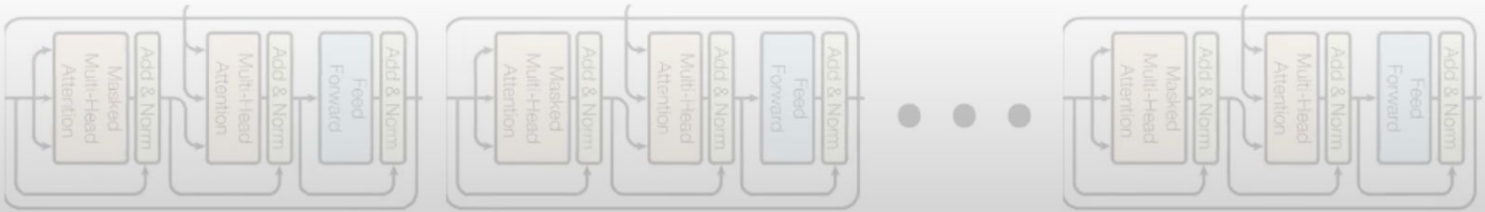
What is language!

Transformer Flow



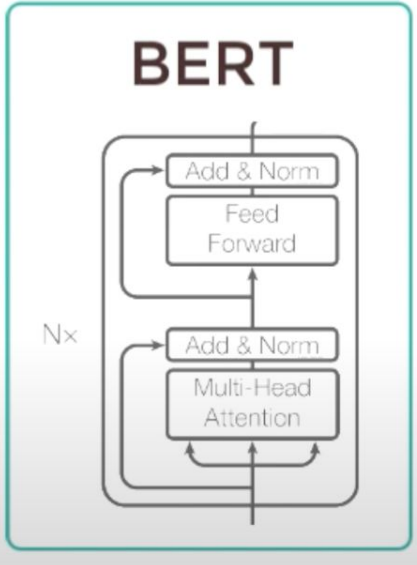
BERT

Bidirectional **E**ncoder **R**epresentation from **T**ransformers



GPT

Bidirectional Encoder Representation from Transformers



Problems to Solve

- Neural Machine Translation
- Question Answering
- Sentiment Analysis
- Text summarization

Needs Language understanding

How to solve Problems

- Pretrain BERT to understand langauge
- Fine tune BERT to learn specific task

Bidirectional Encoder Representation from Transformers

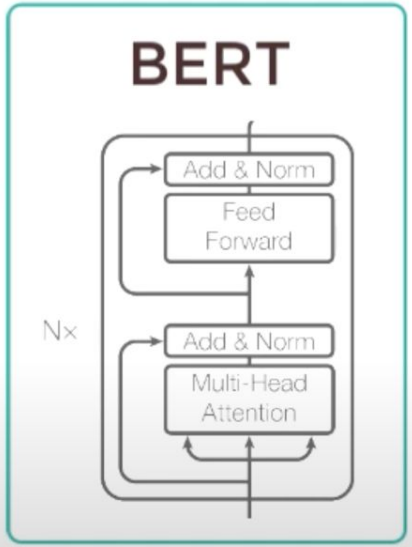
Pretraining (Pass 1) : "What is language? What is context?"

Masked Language Model (MLM)

The [MASK1] brown fox [MASK2] over the lazy dog.

Next Sentence Prediction (NSP)

A: Ajay is a cool dude.
B: He lives in Ohio



[MASK1] = quick
[MASK2]= jumped

Yes. Sentence B follows sentence A

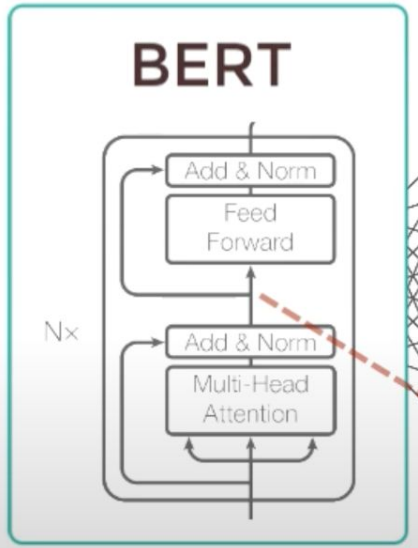
Bidirectional Encoder Representation from Transformers

Fine Tuning (Pass 1): "How to use language for specific task?"

Fine tuned Q & A

Question

Passage



Learned From scratch

Answer

Fine Tuned

FAST!

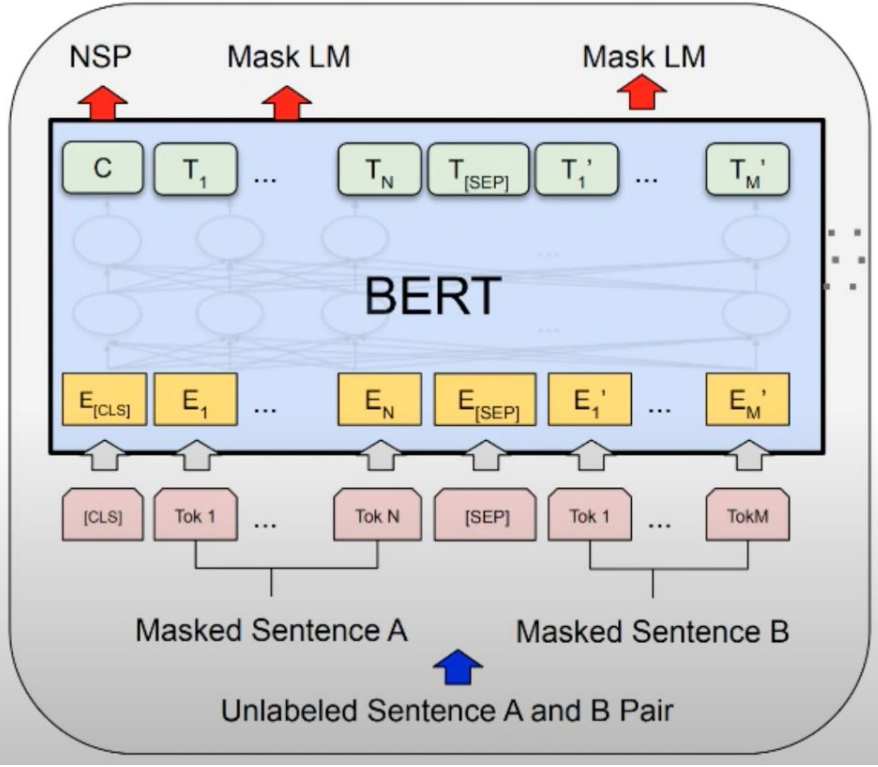
Bidirectional Encoder Representation from Transformers

Pretraining (Pass 2)

Problems to train on simultaneously:

1. Masked Language Modeling (Mask LM)

2. Next Sentence Prediction (NSP)

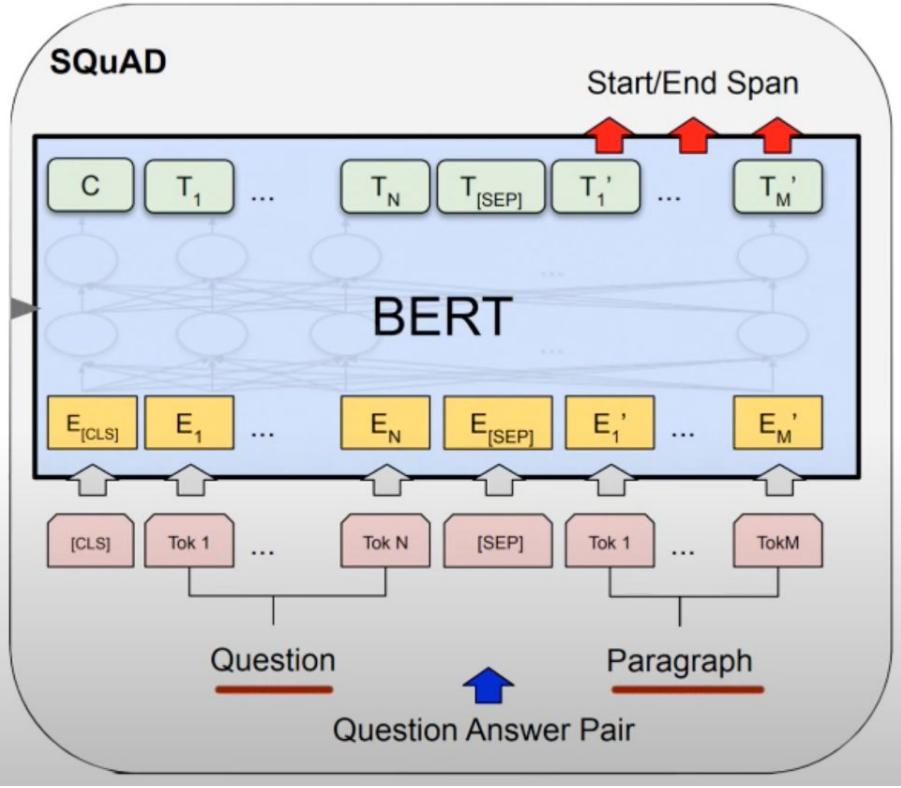


Bidirectional Encoder Representation from Transformers

Fine Tuning (Pass 2)

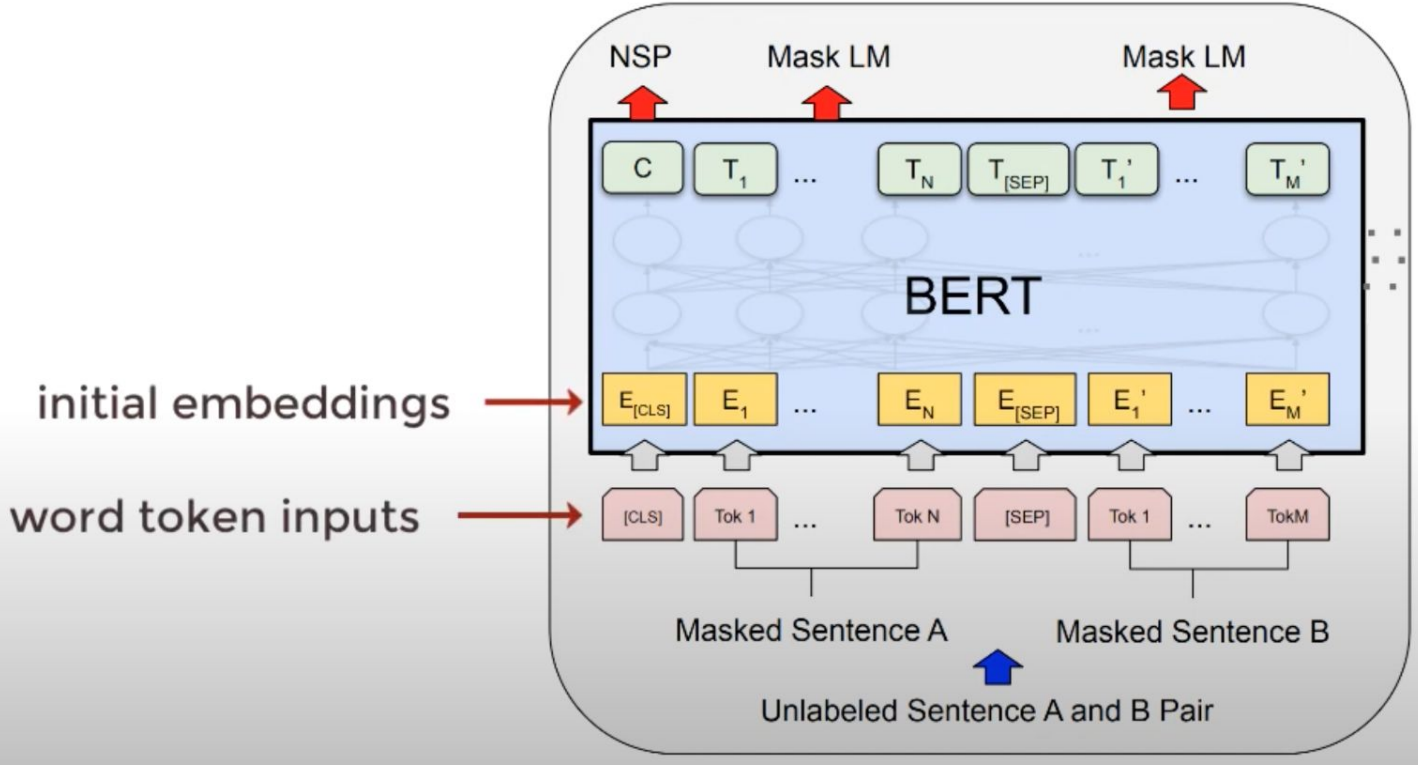
Change output to display text in which answer exists

Change inputs to take in Question, Passage



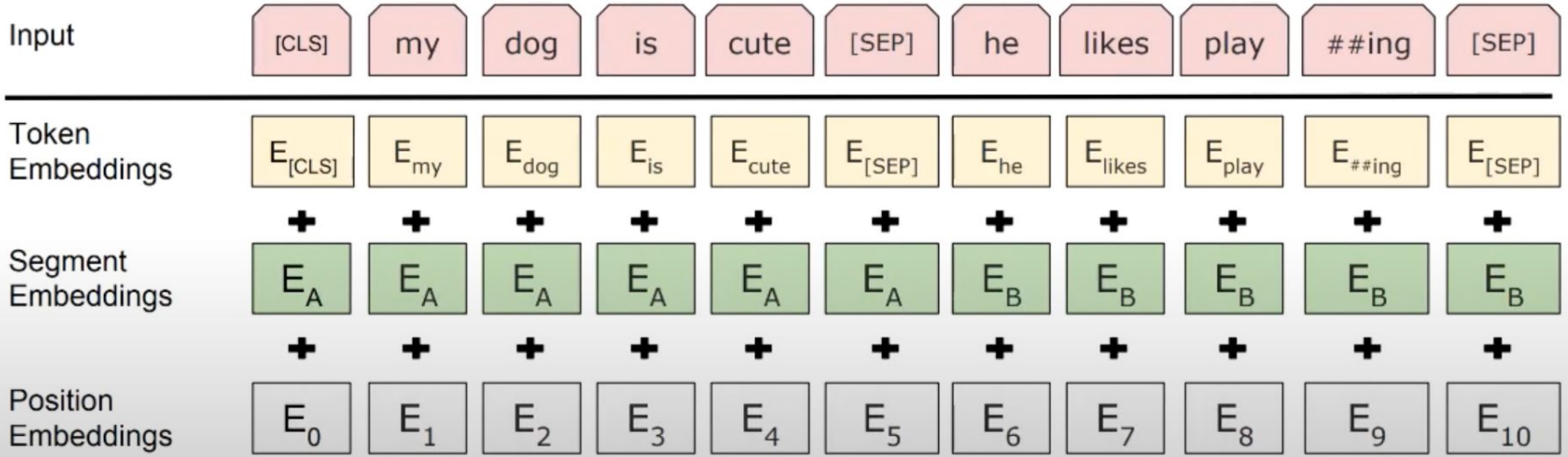
Bidirectional Encoder Representation from Transformers

Pretraining (Pass 3)



Bidirectional Encoder Representation from Transformers

Pretraining (Pass 3)

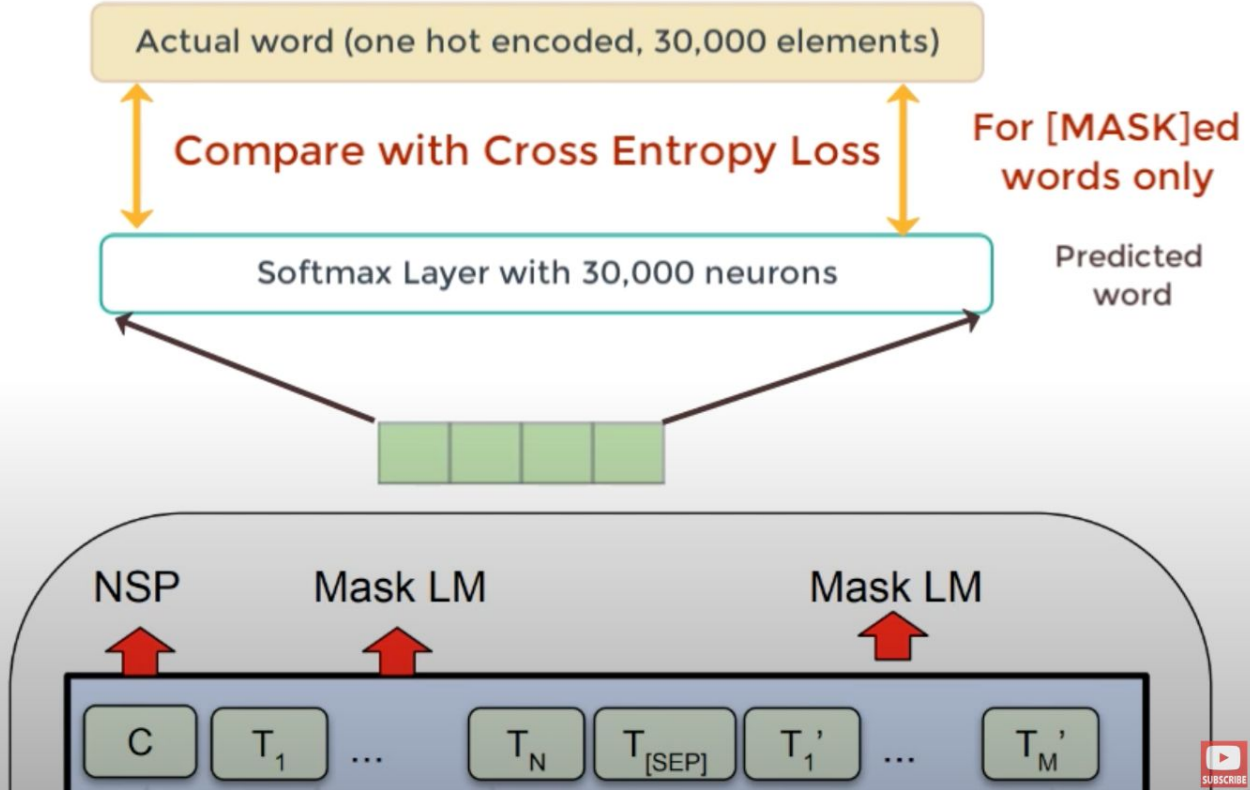


Bidirectional Encoder Representation from Transformers

Pretraining (Pass 3)

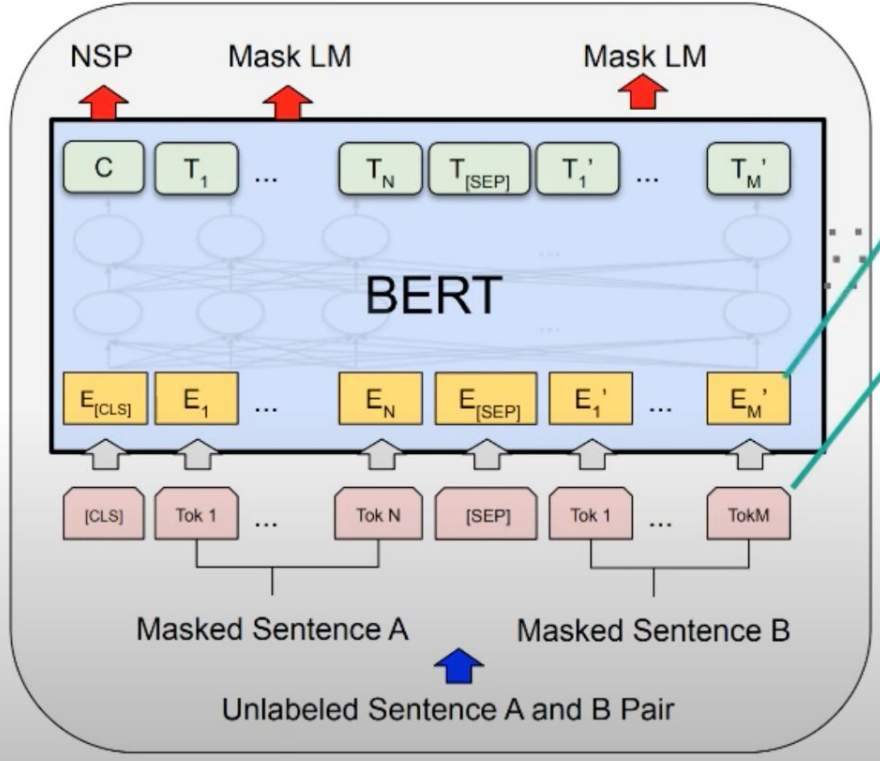
Word vectors T_i have the same size.

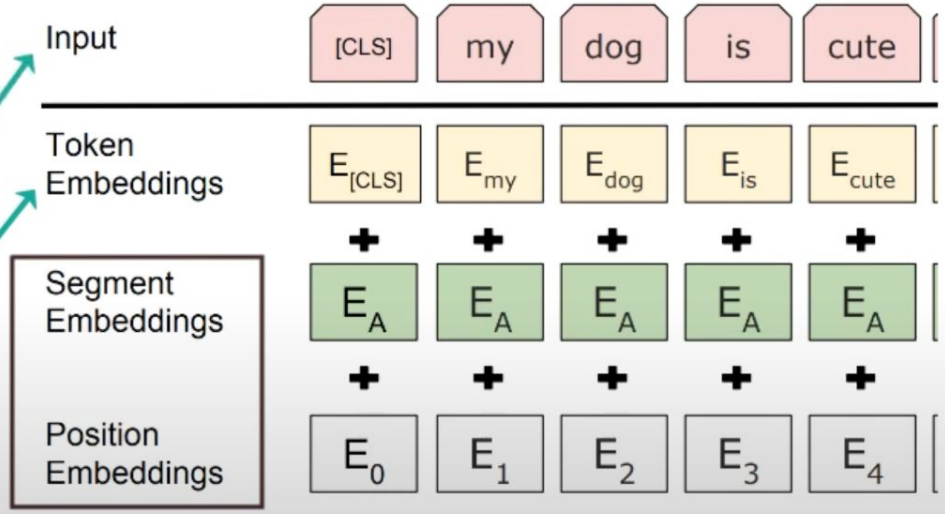
Word vectors T_i are generated simultaneously



Bidirectional Encoder Representation from Transformers

Pretraining (Summary)





Bidirectional Encoder Representation from Transformers

Pretraining (Summary)