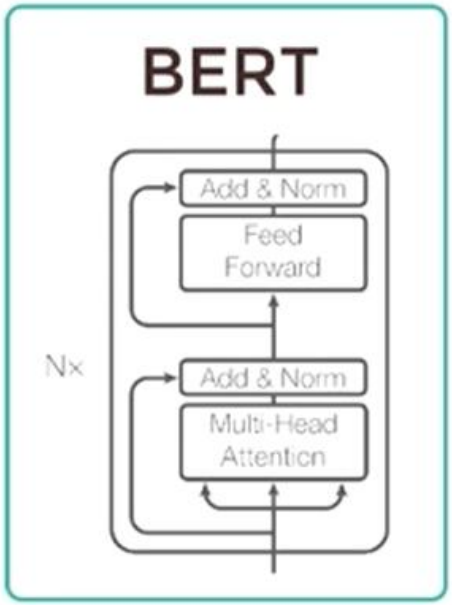
Bidirectional Encoder Representation from Transformers



Problems to Solve

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

Needs Language understanding

How to solve Problems (BERT Training)

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

Bidirectional Encoder Representation from Transformers

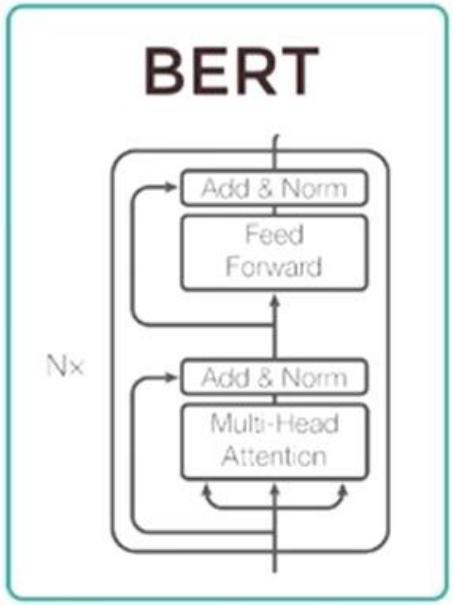
Pretraining (Pass 1) : "What is language? What is context?"

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

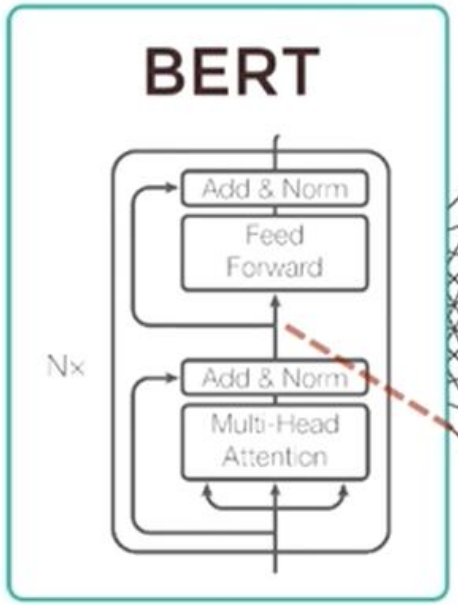
Bidirectional Encoder Representation from Transformers

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

Fine tuned Q & A

Question

Passage



Learned From scratch

Answer

Fine Tuned

FAST!

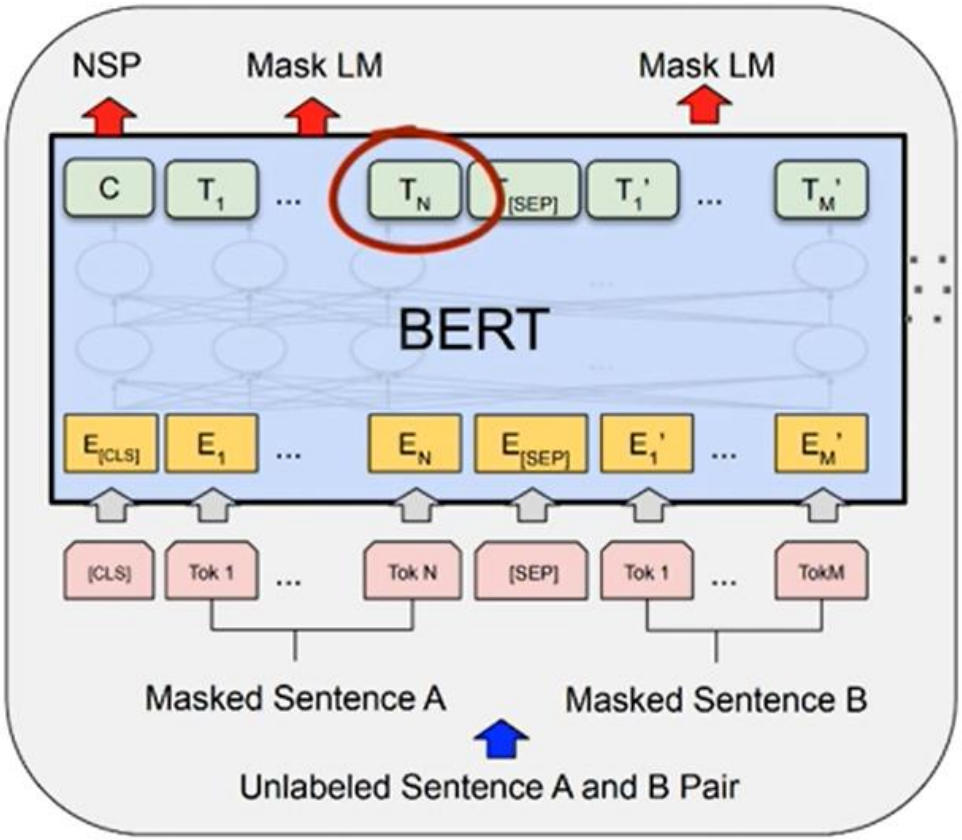
Bidirectional Encoder Representation from Transformers

Pretraining (Pass 2)

Problems to train on simultaneously:

1. Masked Language Modeling (Mask LM)

2. Next Sentence Prediction (NSP)



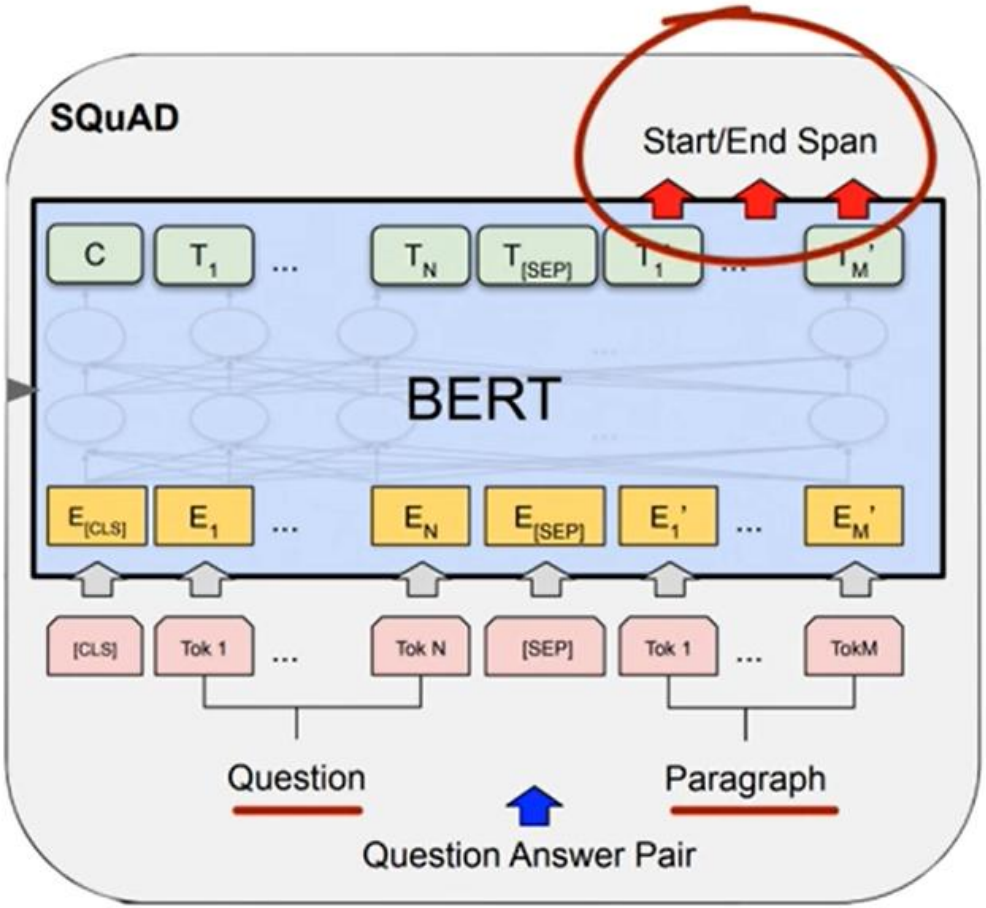
Source: BERT: Pre-training of deep bidirectional Transformers for language understanding (Devlin et al., 2019)

Bidirectional Encoder Representation from Transformers

Fine Tuning (Pass 2)

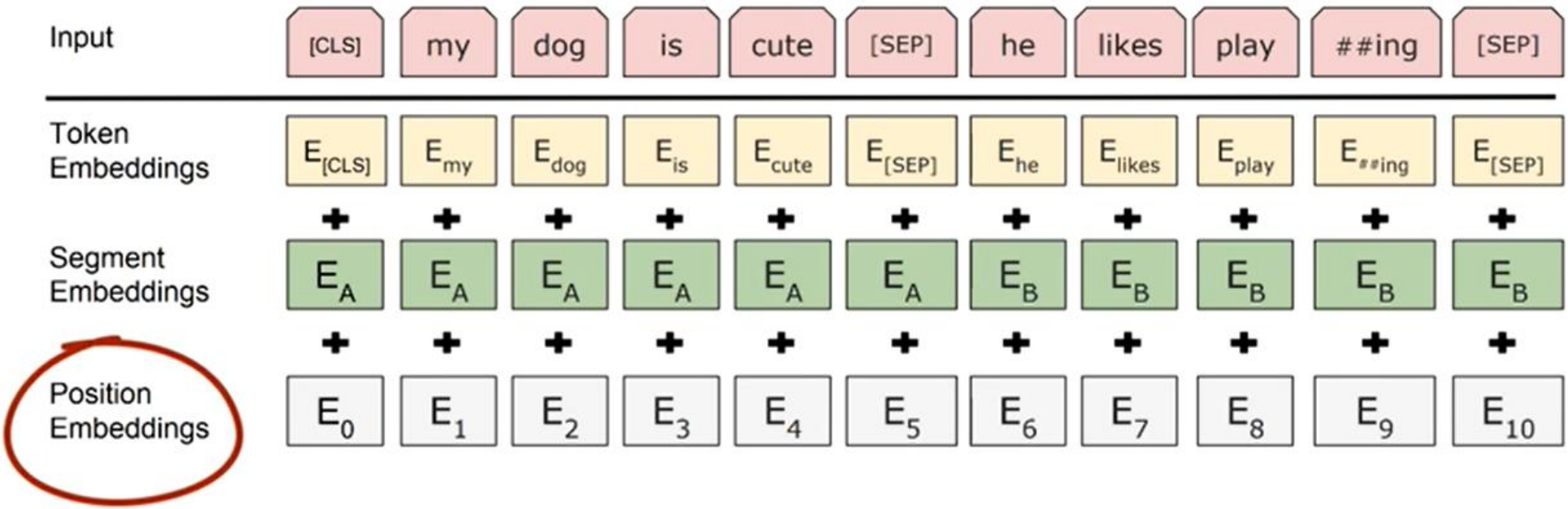
Change output to display text in which answer exists

Change inputs to take in Question, Passage



Bidirectional Encoder Representation from Transformers

Pretraining (Pass 3)

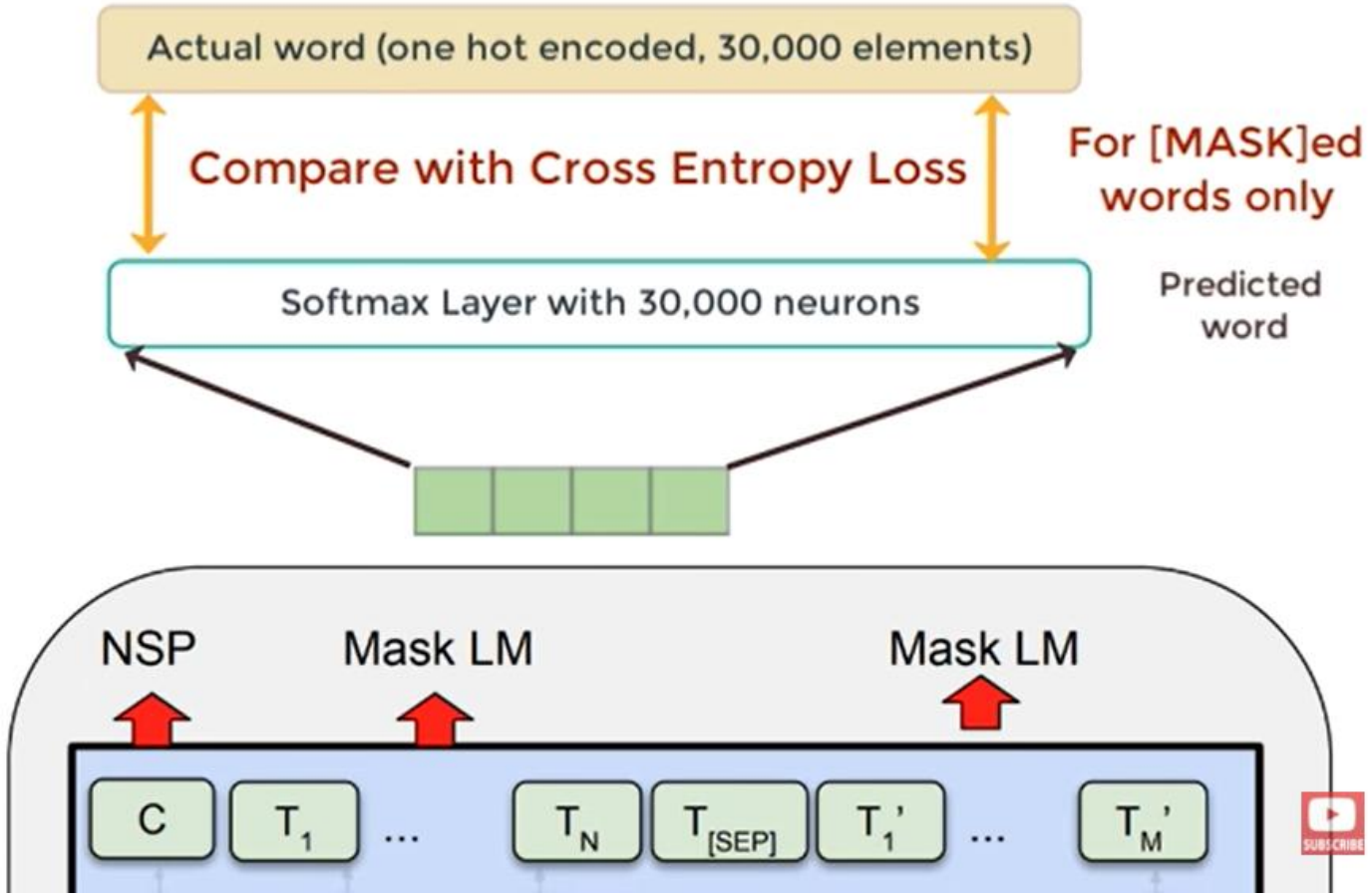


Bidirectional Encoder Representation from Transformers

Pretraining (Pass 3)

Word vectors T_i have the same size.

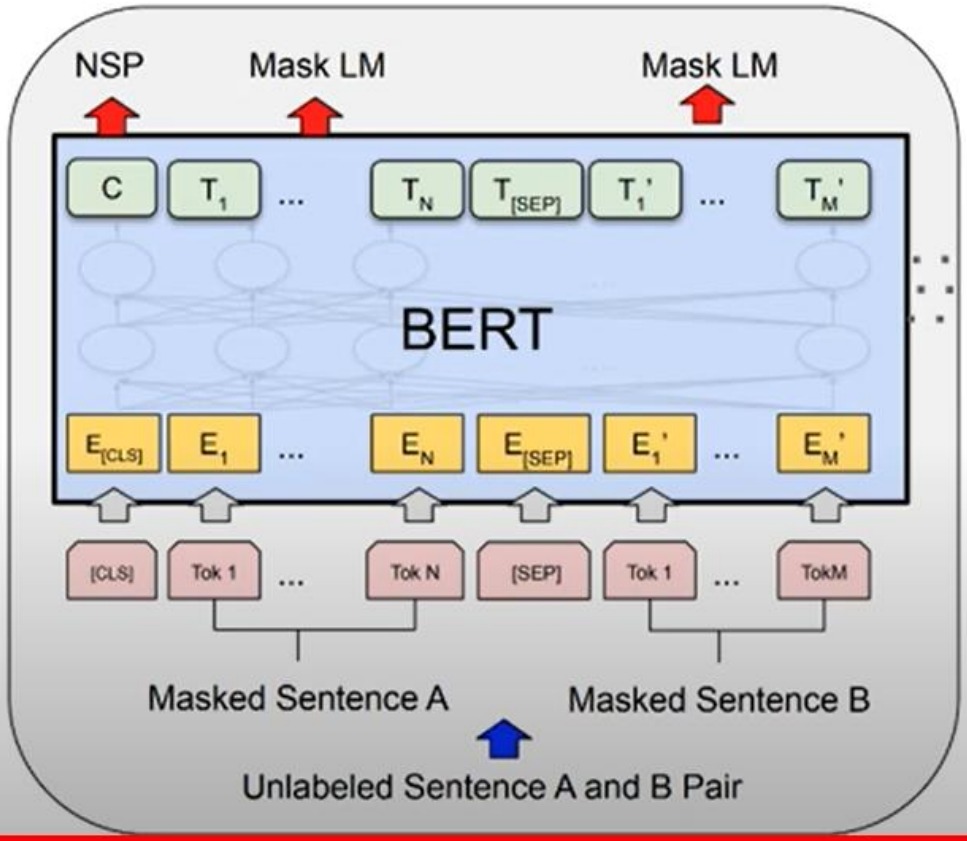
Word vectors T_i are generated simultaneously

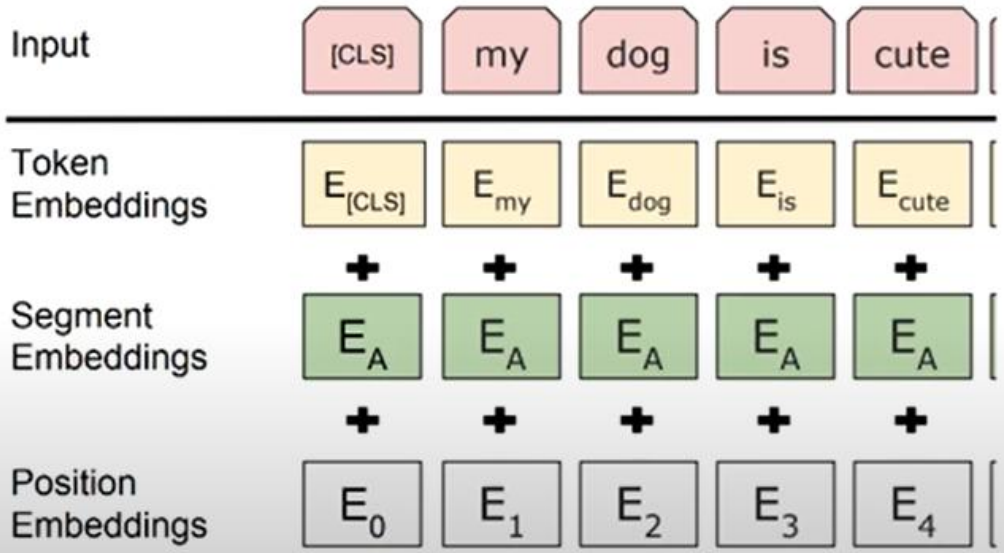


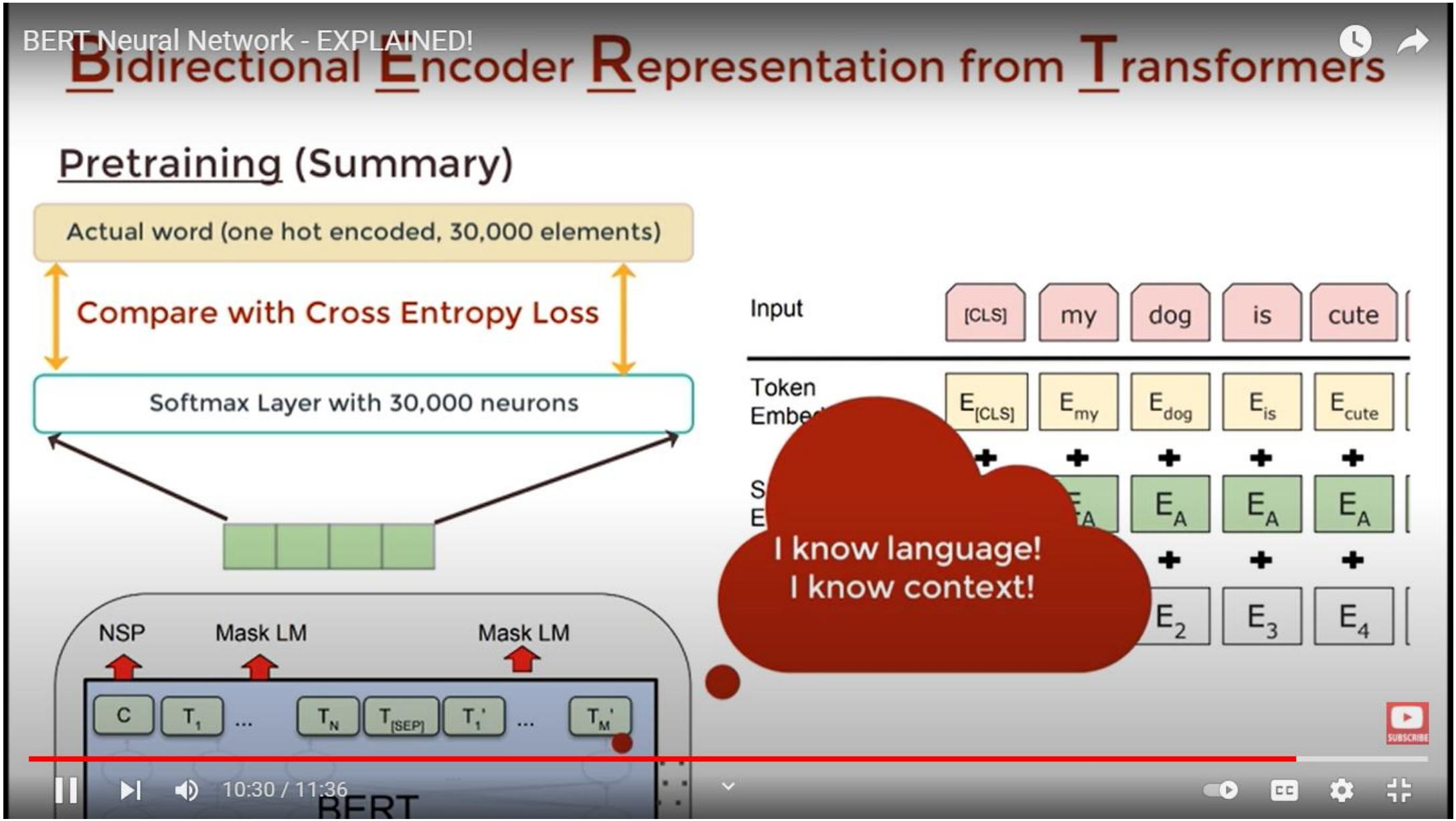
BERT Neural Network - EXPLAINED!

Bidirectional Encoder Representation from Transformers

Pretraining (Summary)

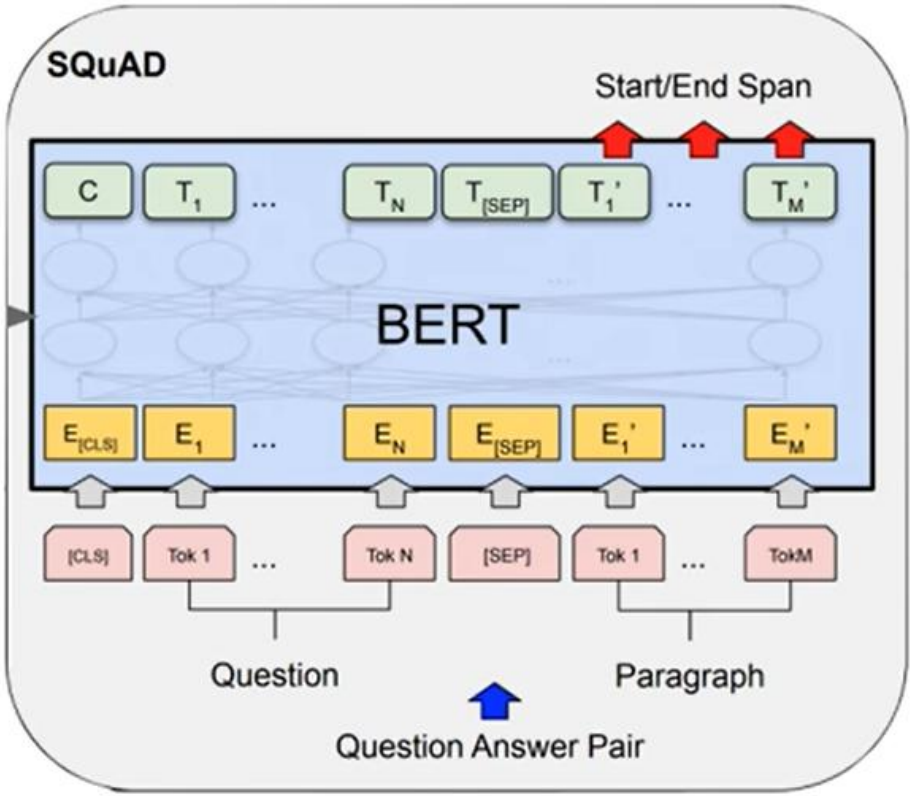





BERT Neural Network - EXPLAINED!
Bidirectional Encoder Representation from Transformers
Pretraining (Summary)
Actual word (one hot encoded, 30,000 elements)
Compare with Cross Entropy Loss
Softmax Layer with 30,000 neurons
NSP
Mask LM
Mask LM
C
T1
...
TN
T[SEP]
T1'
...
TM'
BERT
Input
[CLS]
my
dog
is
cute
Token
E[CLS]
Emy
Edog
Eis
Ecute
EA
EA
EA
E2
E3
E4
I know language!
I know context!
10:30 / 11:36

Bidirectional Encoder Representation from Transformers

Fine Tuning (Summary)

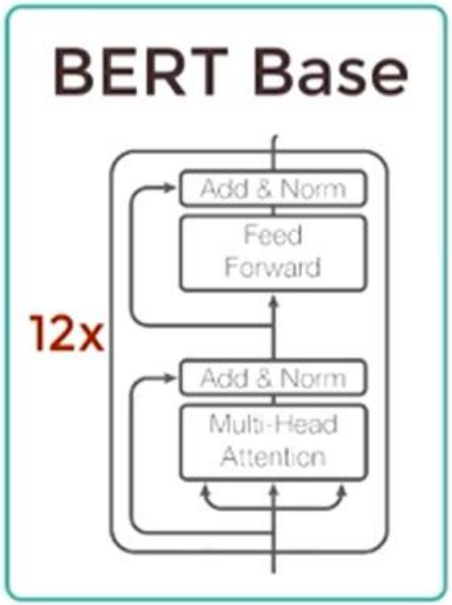


"Stanford Question & Answer Dataset"

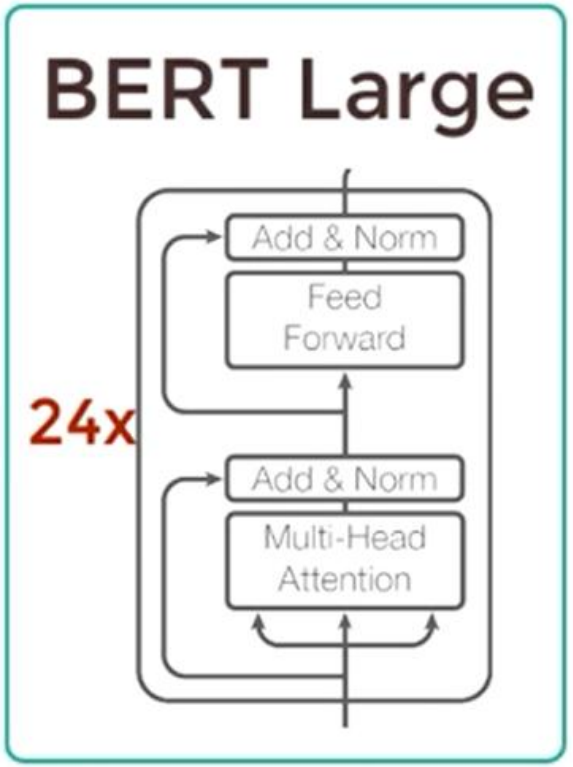
- 30 minute training on single cloud TPU with 91% F1 score.

Bidirectional Encoder Representation from Transformers

Performance



110M Parameters



340M Parameters