

# Question Answering using BERT

## Stanford Question Answering Dataset (SQuAD)

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. This region includes territory belonging to nine nations. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. States or departments in four nations contain "Amazonas" in their names. The Amazon represents over half of the planet's remaining rainforests, and comprises the largest and most biodiverse tract of tropical rainforest in the world, with an estimated 390 billion individual trees divided into 16,000 species.

**Which name is also used to describe the Amazon rainforest in English?**

Ground Truth Answers: also known in English as Amazonia or the Amazon Jungle, Amazonia or the Amazon Jungle Amazonia

Prediction: Amazonia

**How many square kilometers of rainforest is covered in the basin?**

Ground Truth Answers: 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. 5,500,000 5,500,000

Prediction: 5,500,000

**How many nations control this region in total?**

Ground Truth Answers: This region includes territory belonging to nine nations. nine nine

Prediction: nine

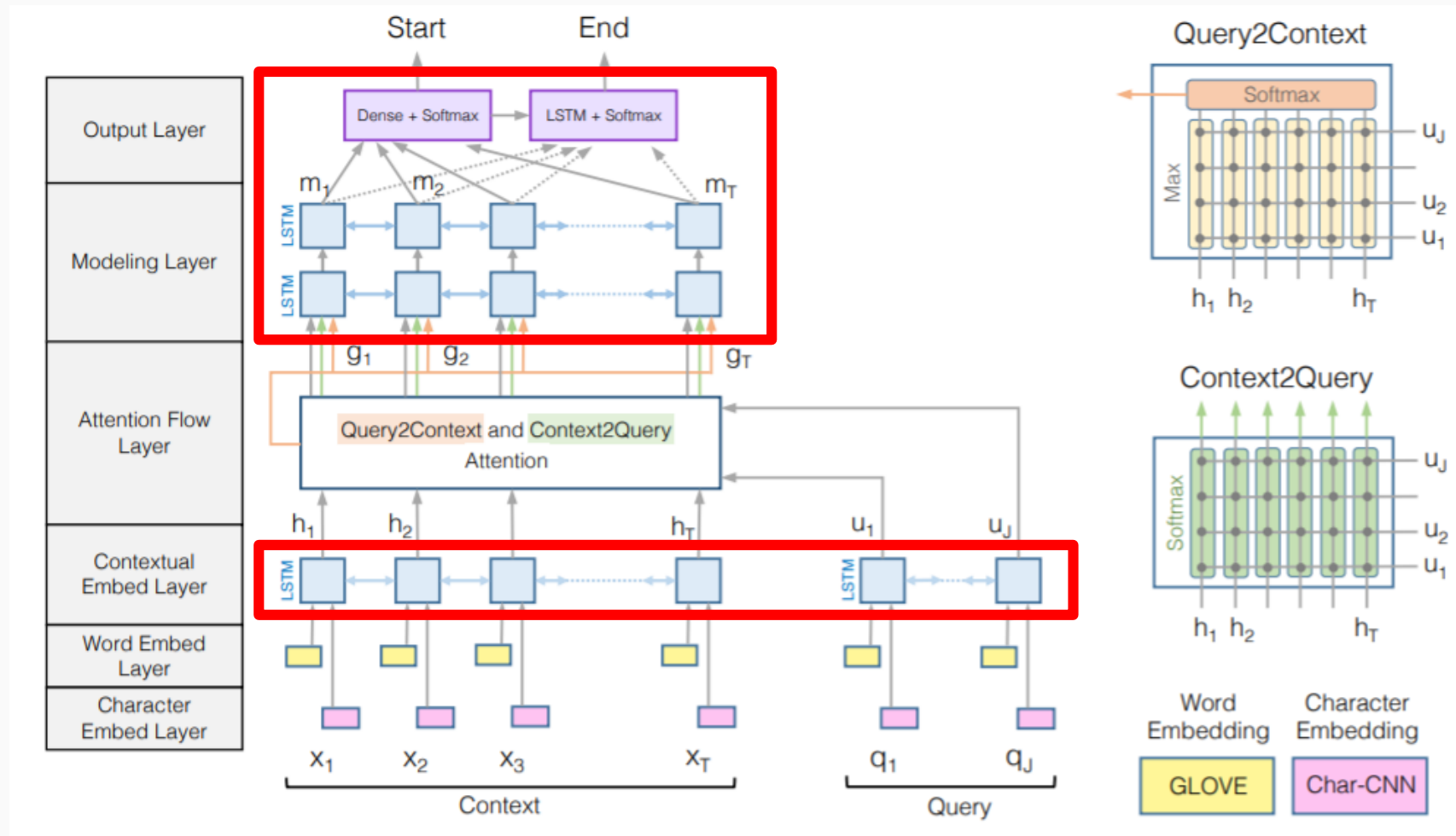
**How many nations contain "Amazonas" in their names?**

Ground Truth Answers: States or departments in four nations contain "Amazonas" in their names. four four

Prediction: four

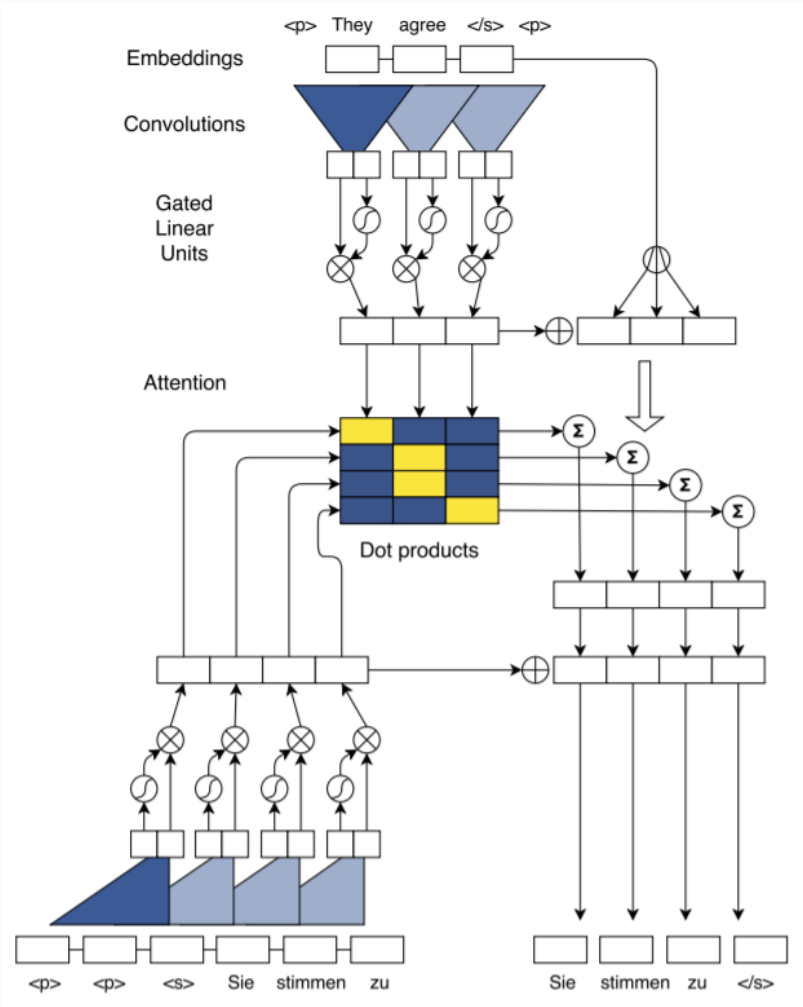
# Question Answering using BERT

## RNN based model - BiDAF

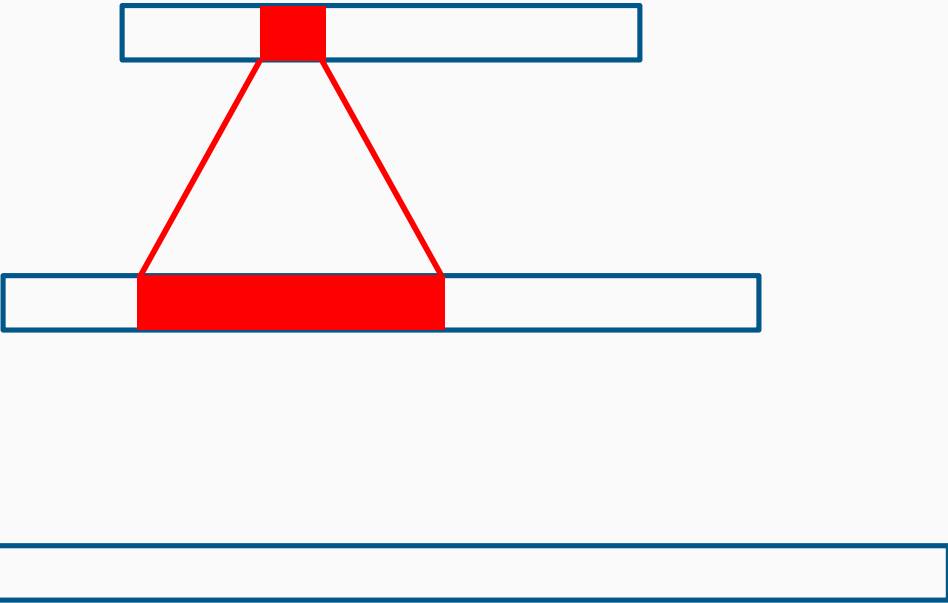
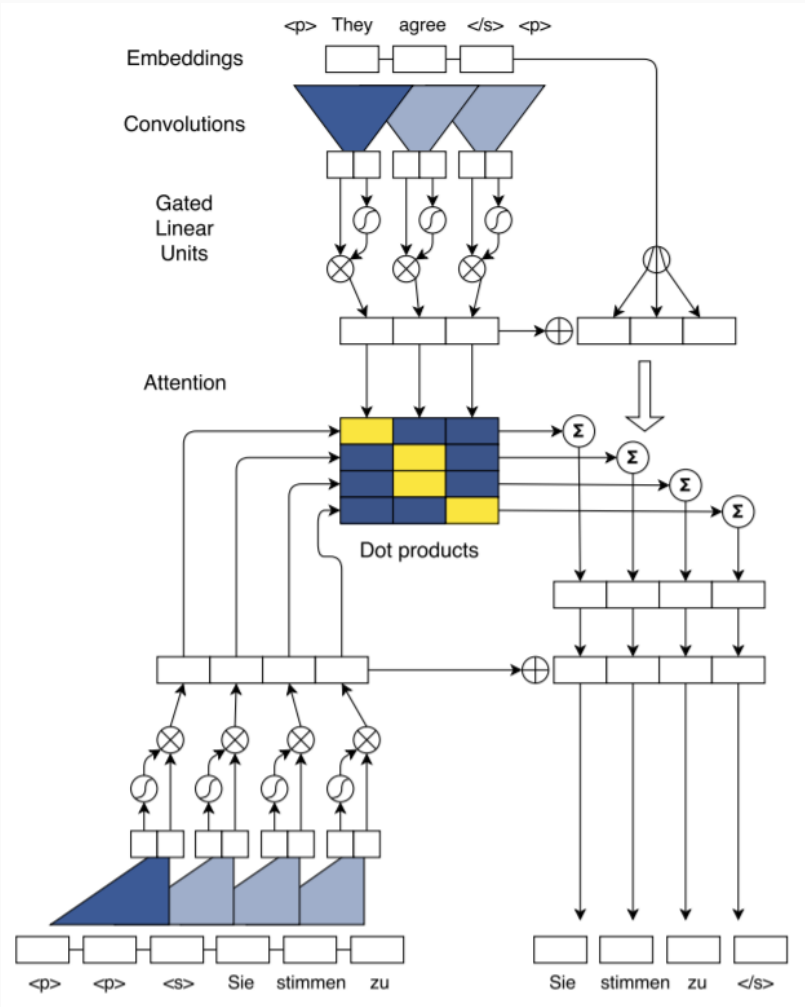


# Question Answering using BERT

## CNN based model – ConvS2S

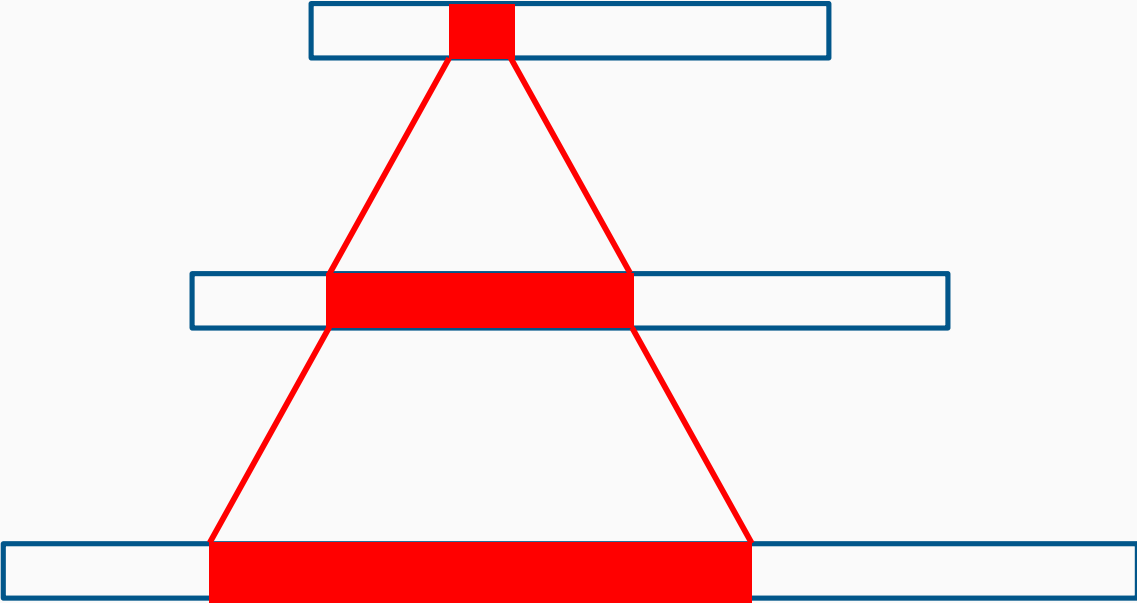
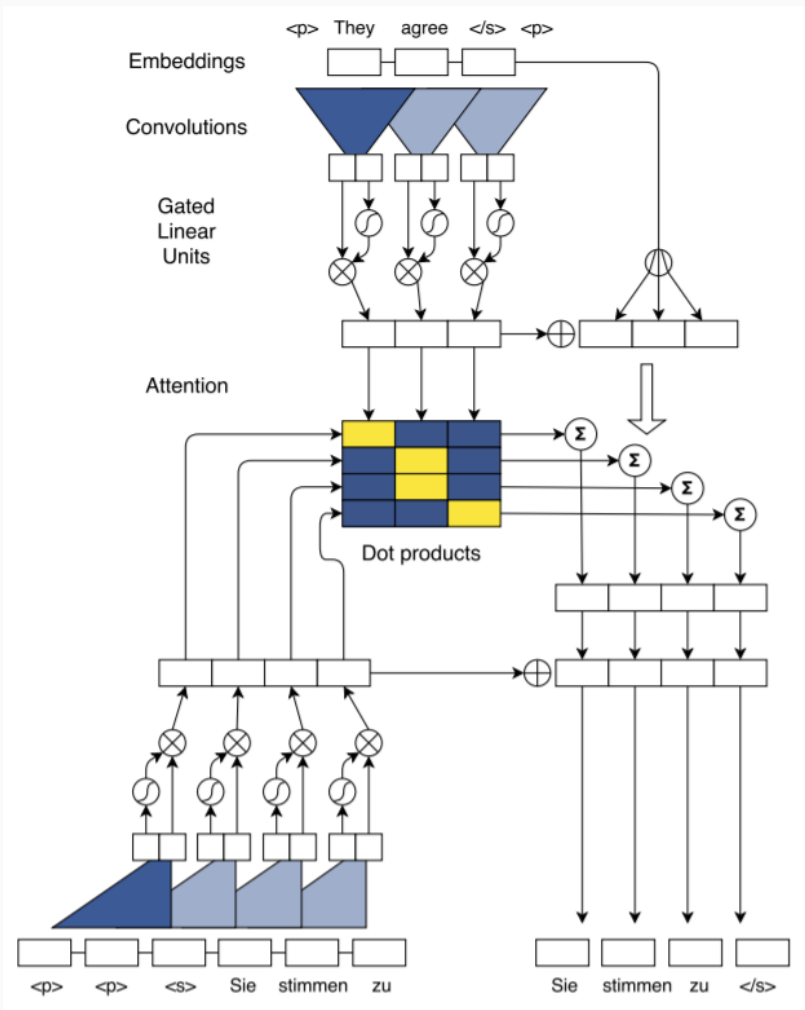


## CNN based model – ConvS2S



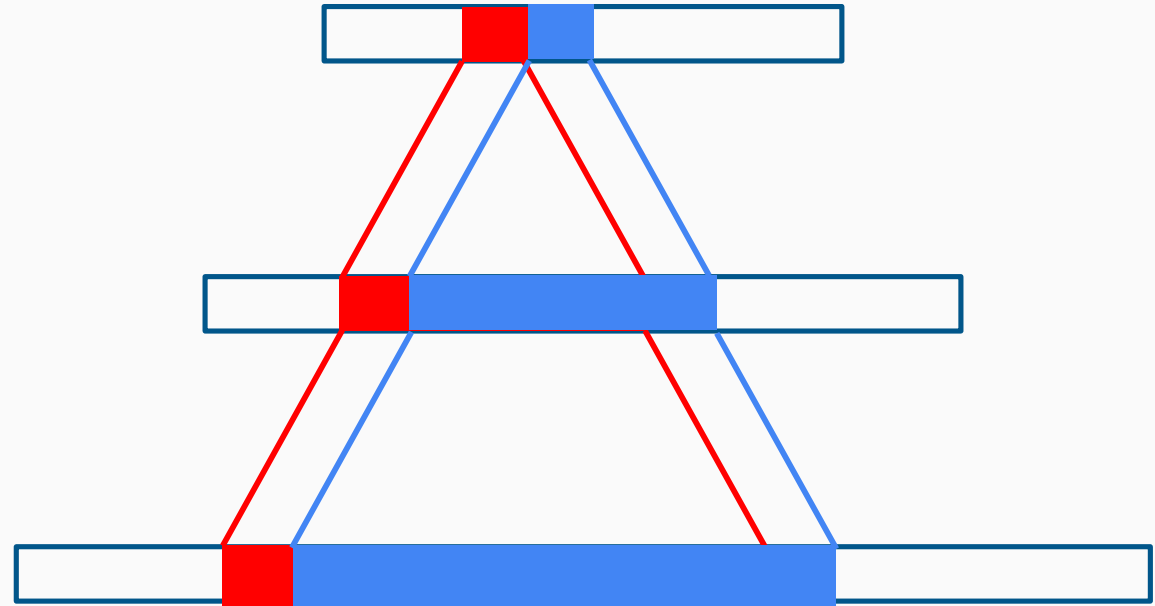
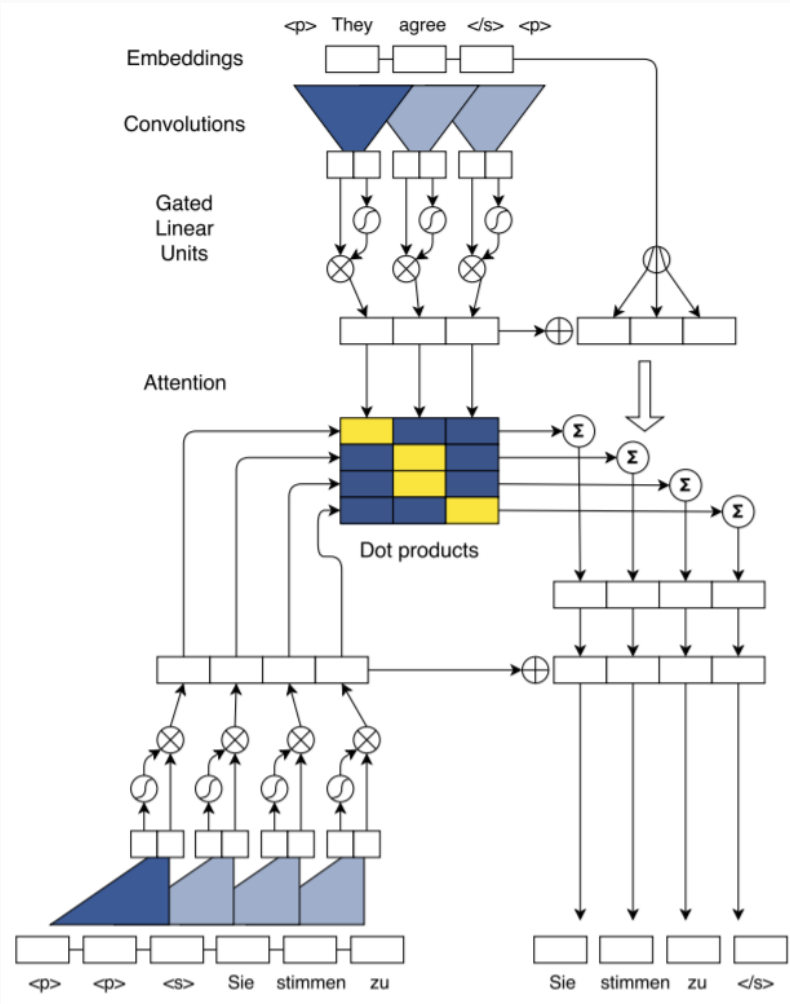
# Question Answering using BERT

## CNN based model – ConvS2S



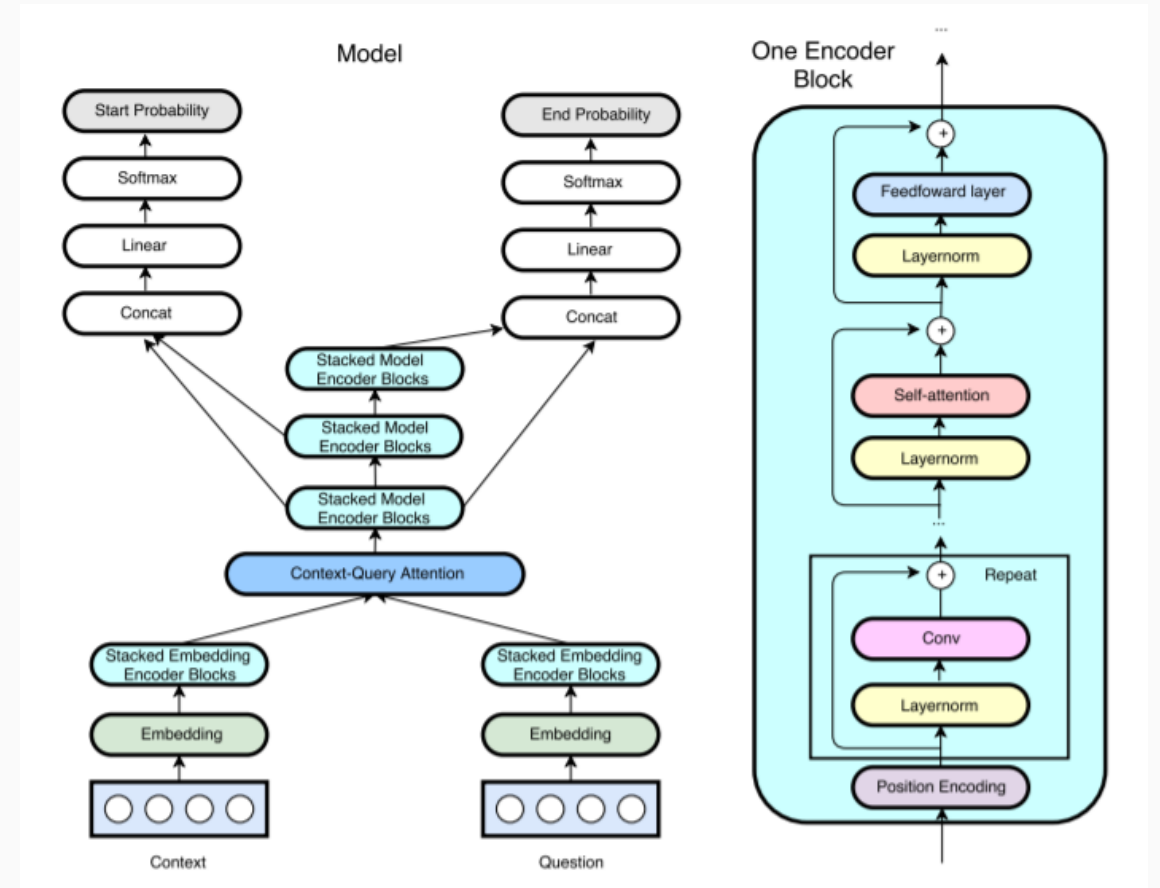
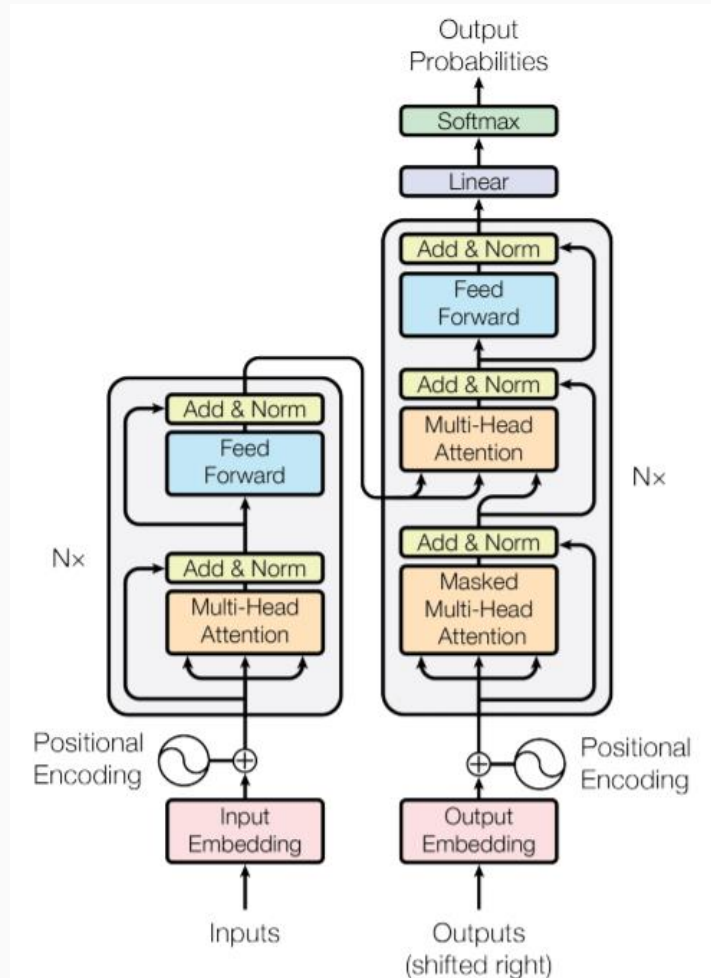
# Question Answering using BERT

## CNN based model – ConvS2S



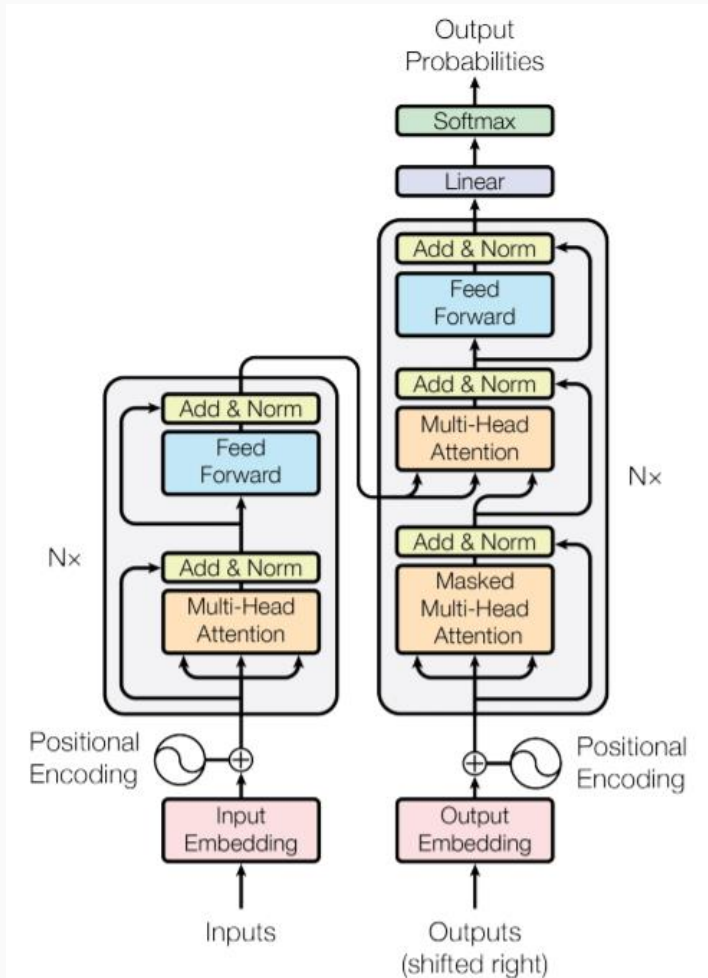
# Question Answering using BERT

## Attention based model – Transformer, QANet

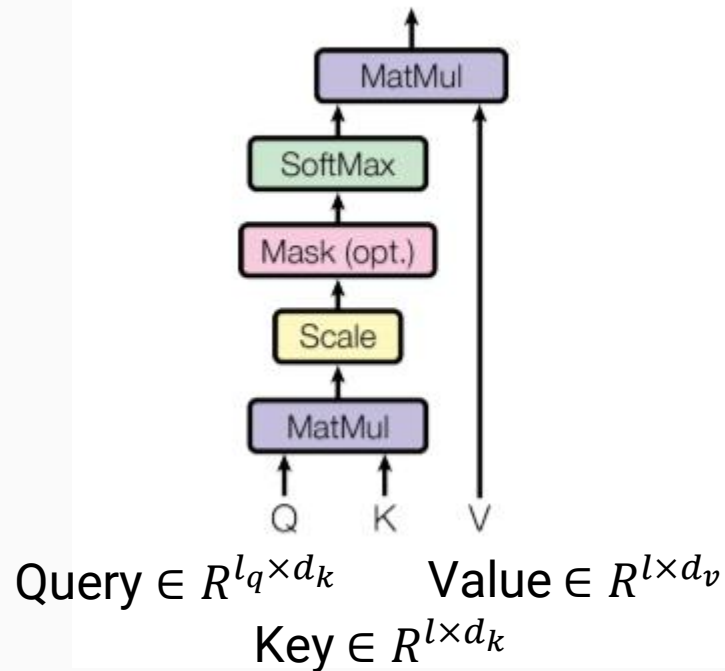


# Question Answering using BERT

## Transformer

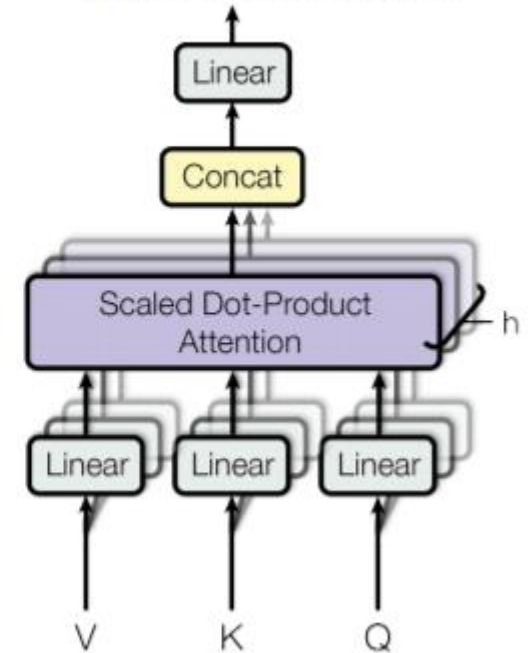


### Scaled Dot-Product Attention



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

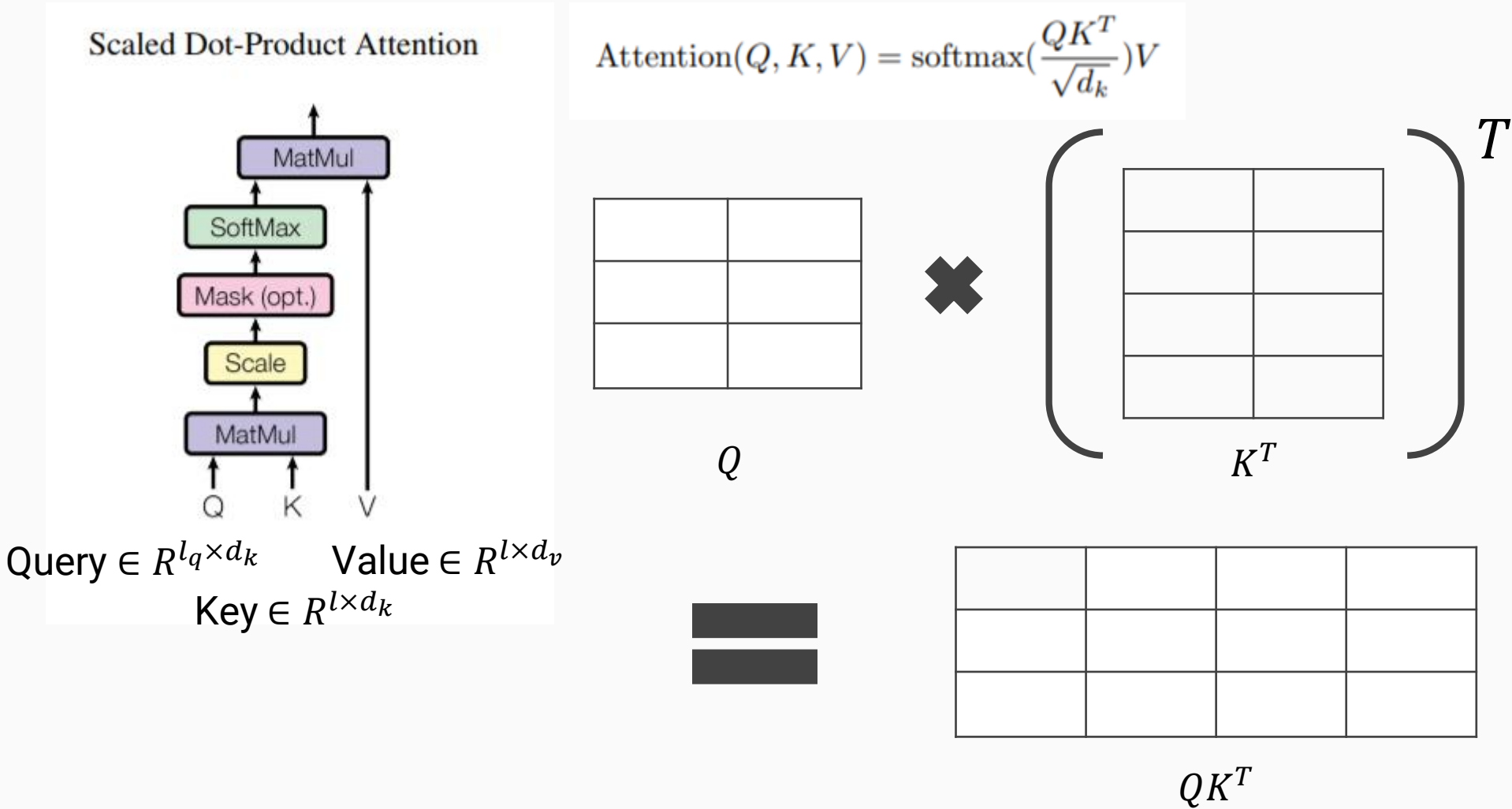
### Multi-Head Attention





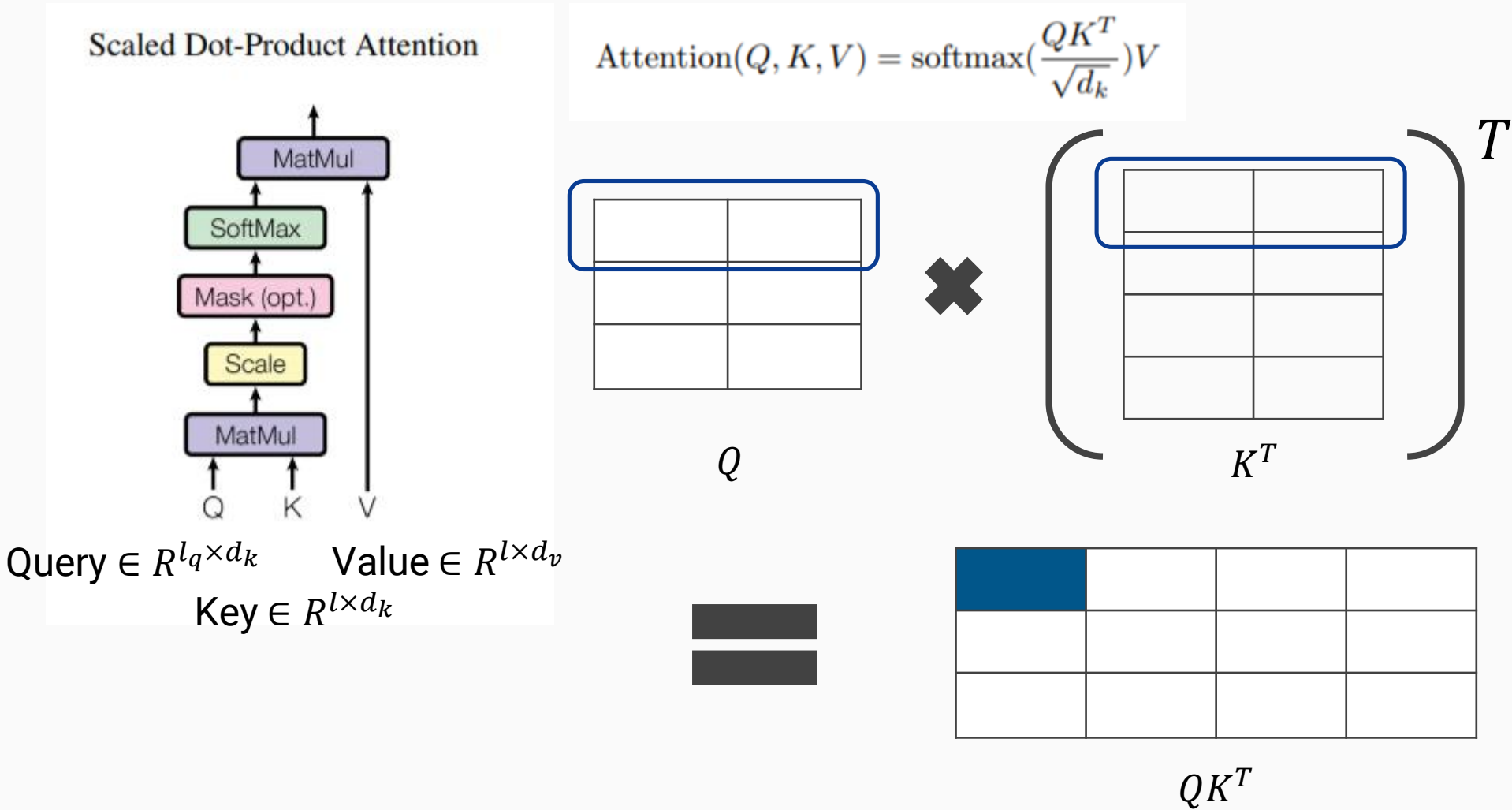
# Question Answering using BERT

## Transformer



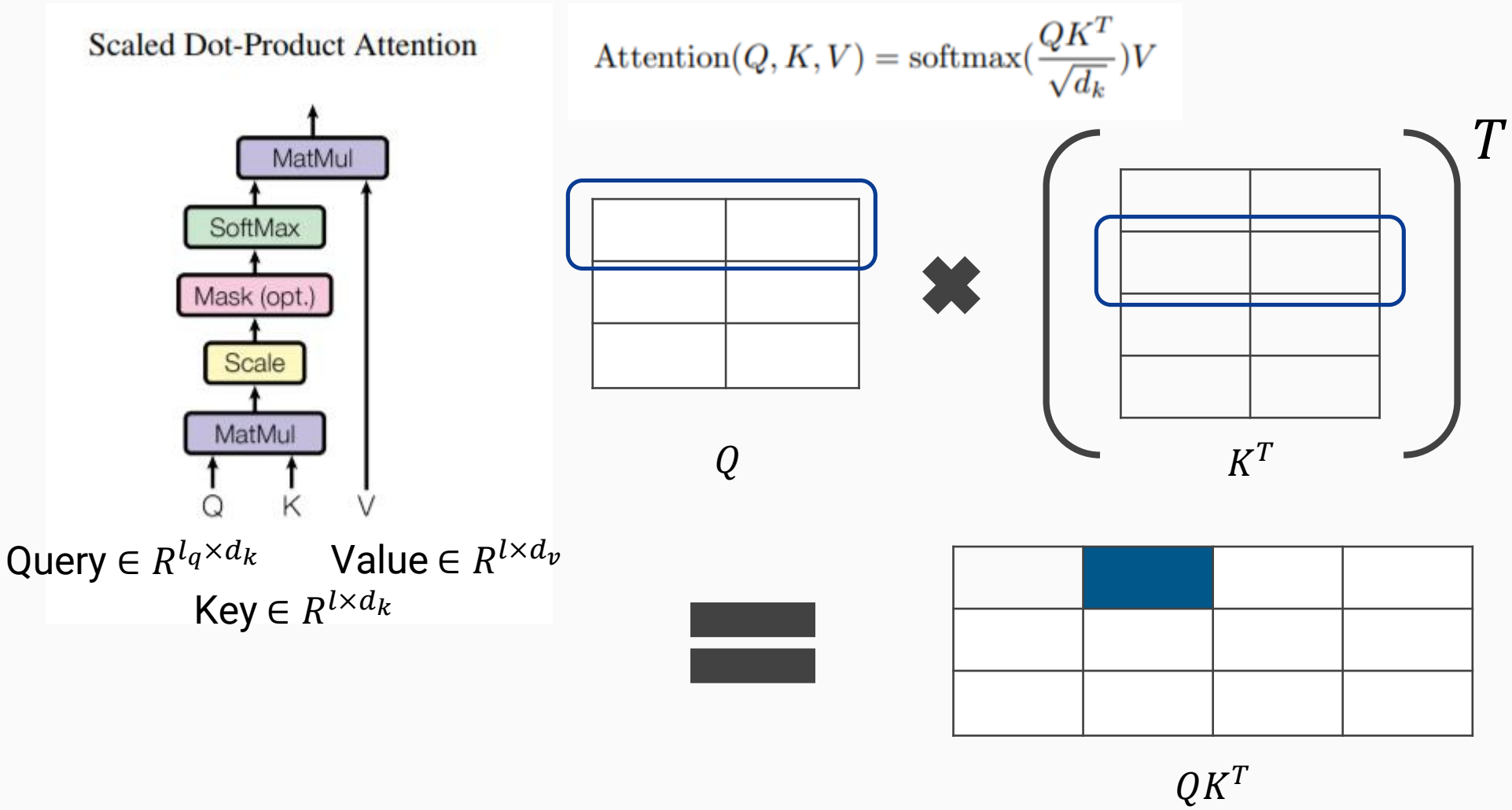
# Question Answering using BERT

## Transformer



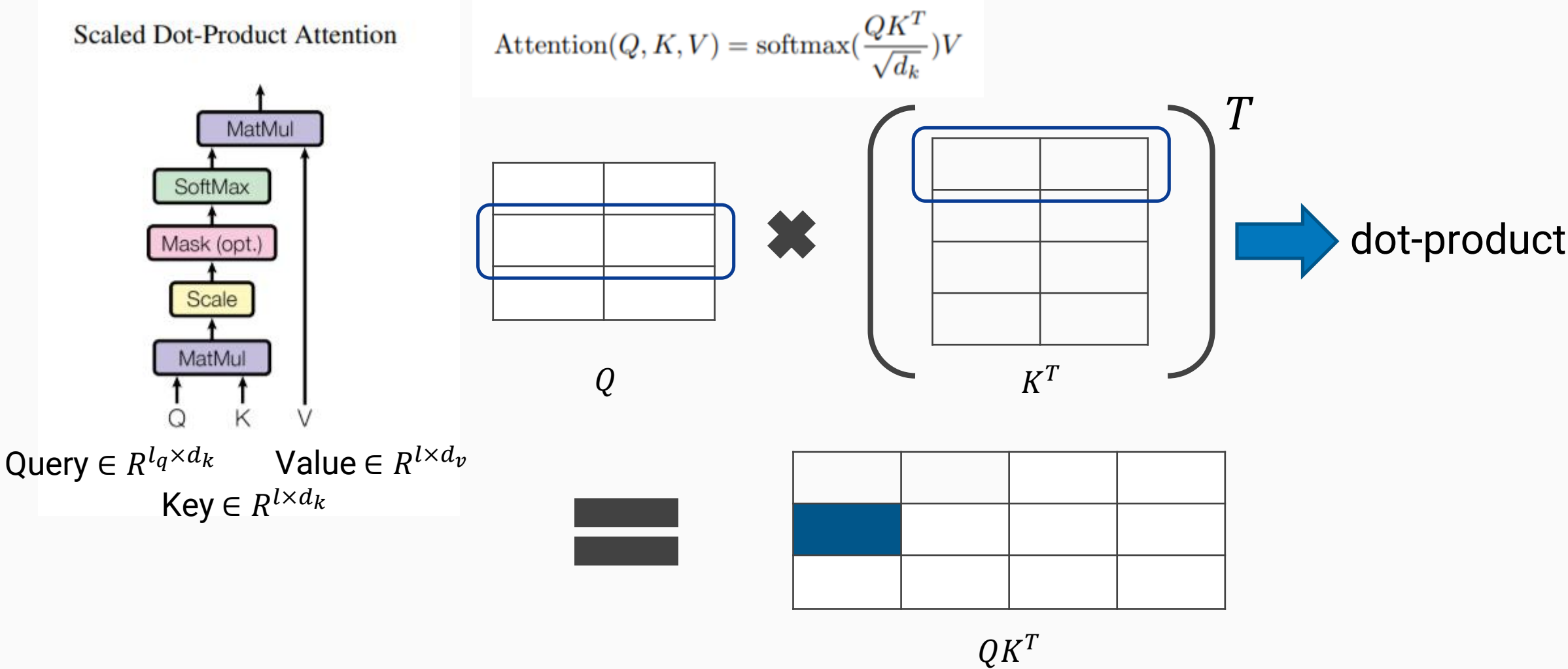
# Question Answering using BERT

## Transformer



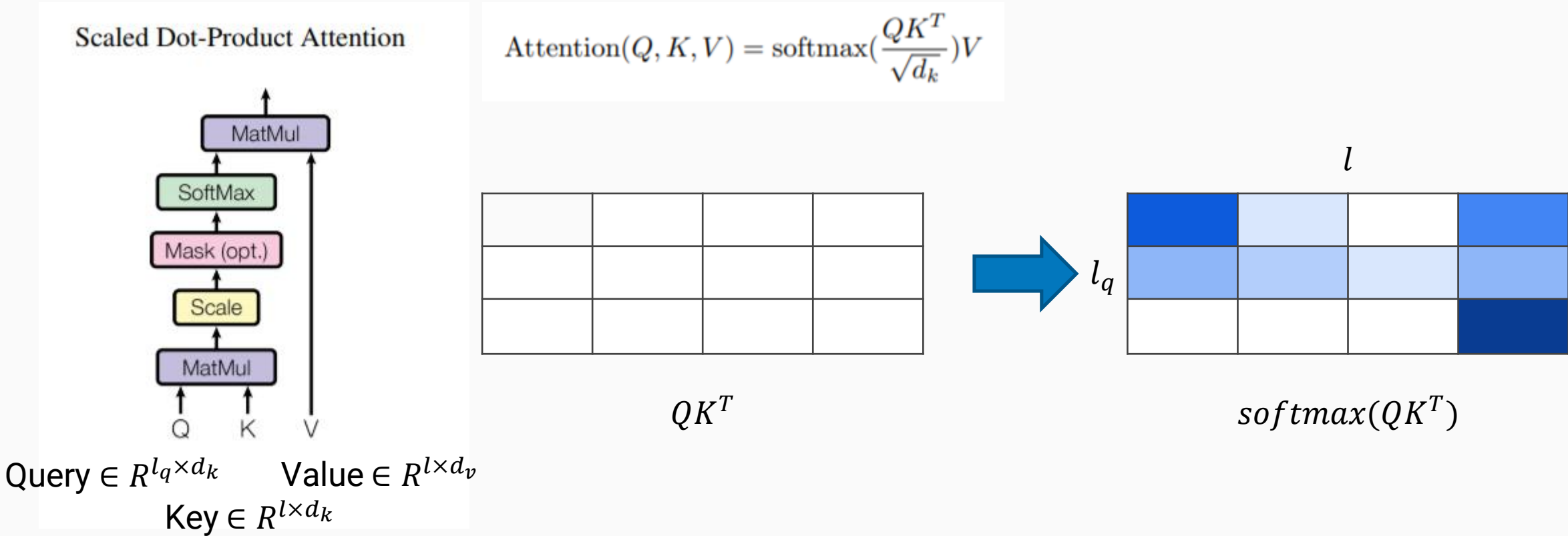
# Question Answering using BERT

## Transformer



# Question Answering using BERT

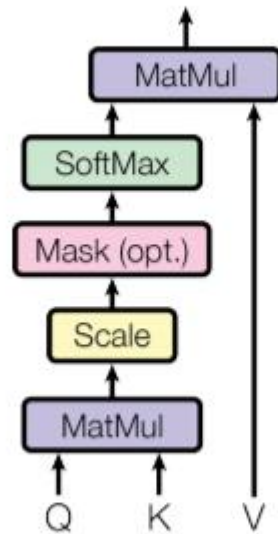
## Transformer



# Question Answering using BERT

## Transformer

### Scaled Dot-Product Attention



Query  $\in R^{l_q \times d_k}$     Value  $\in R^{l \times d_v}$   
Key  $\in R^{l \times d_k}$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Blue	Light Blue	White	Blue
Light Blue	Light Blue	Light Blue	Light Blue
White	White	White	Dark Blue

$\text{softmax}(QK^T)$

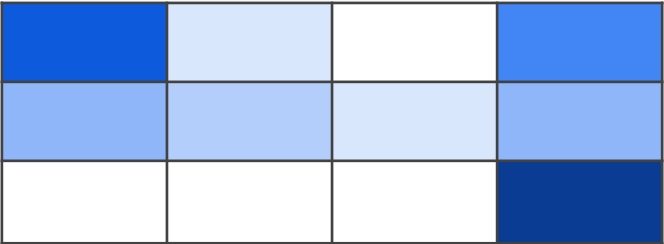


White	White	White
White	White	White
White	White	White
White	White	White

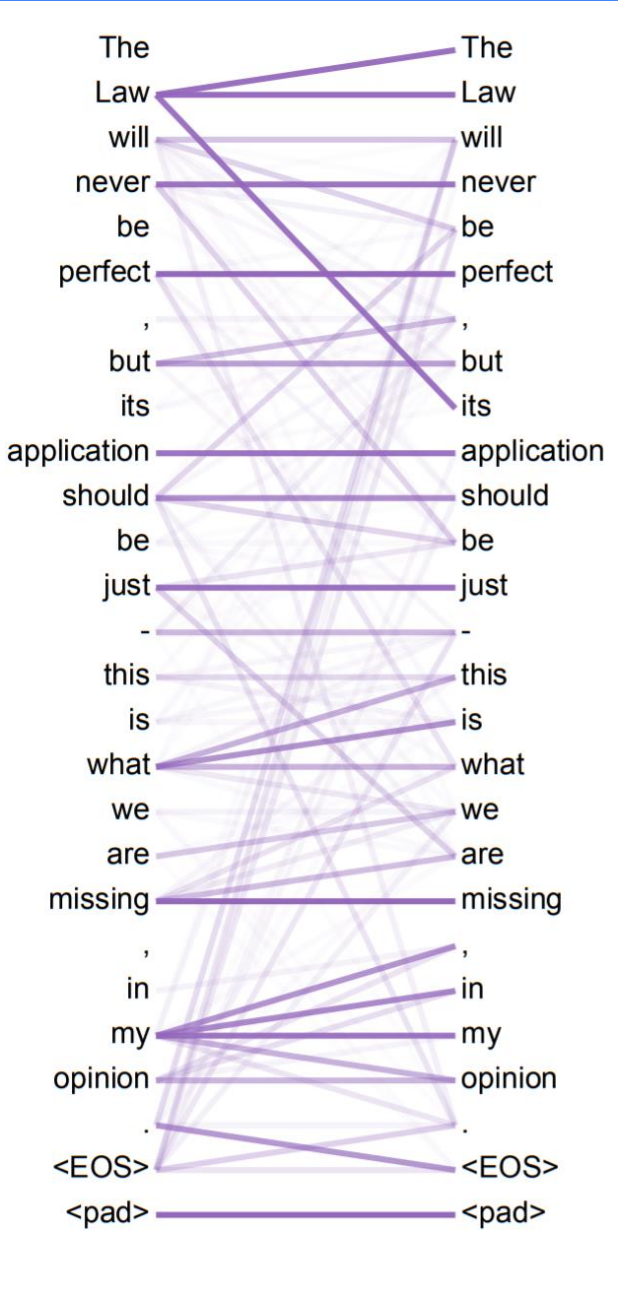
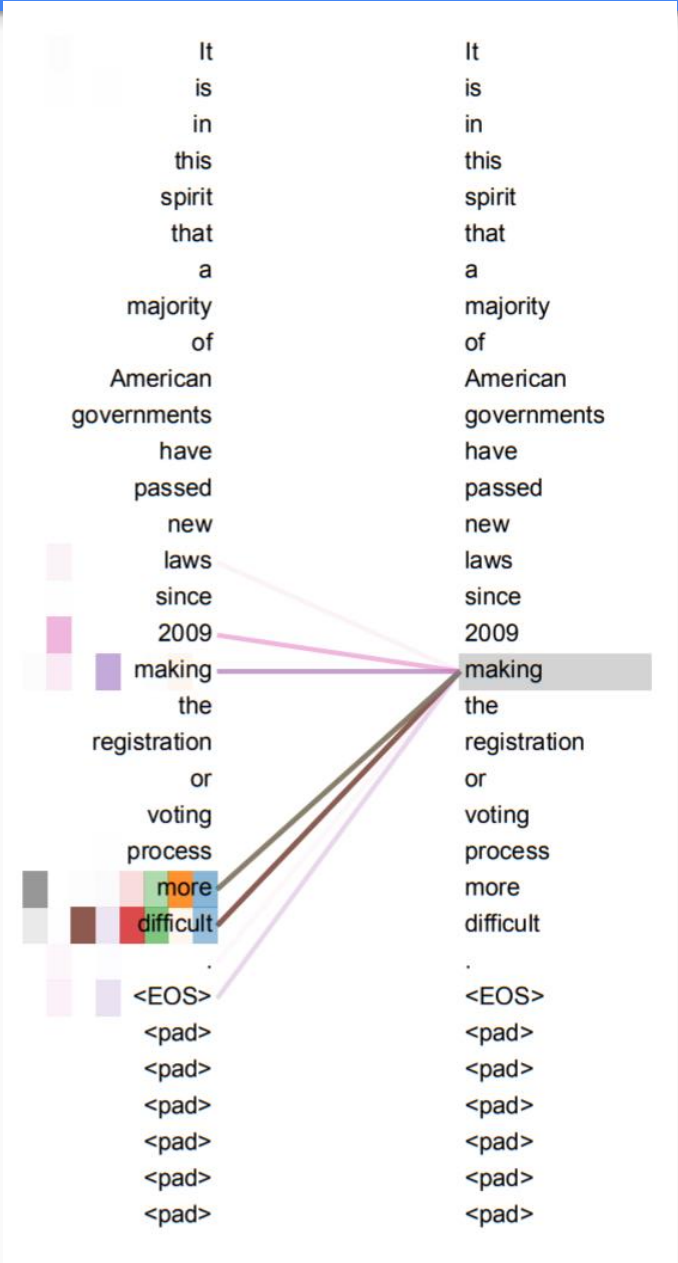
V

# Question Answering using BERT

## Transformer

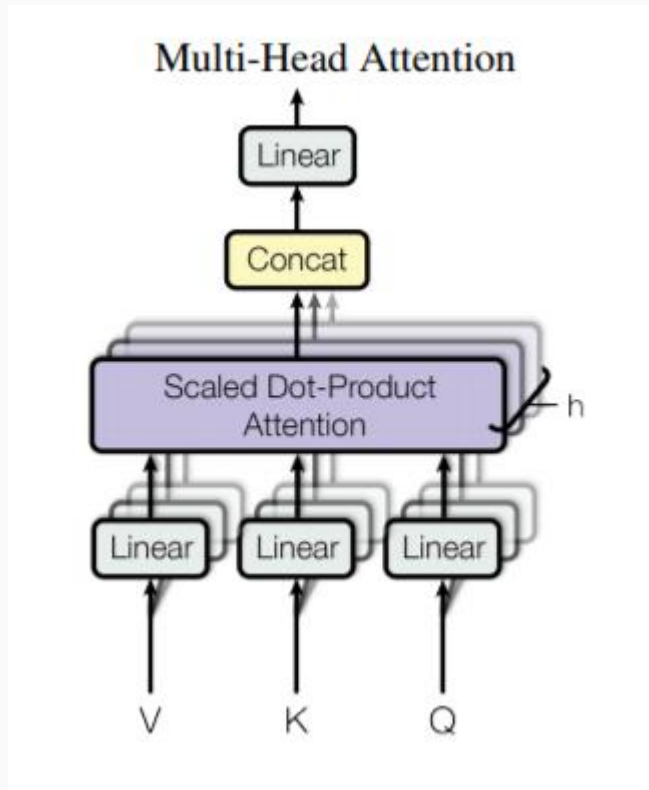


$softmax(QK^T)$



# Question Answering using BERT

## Transformer



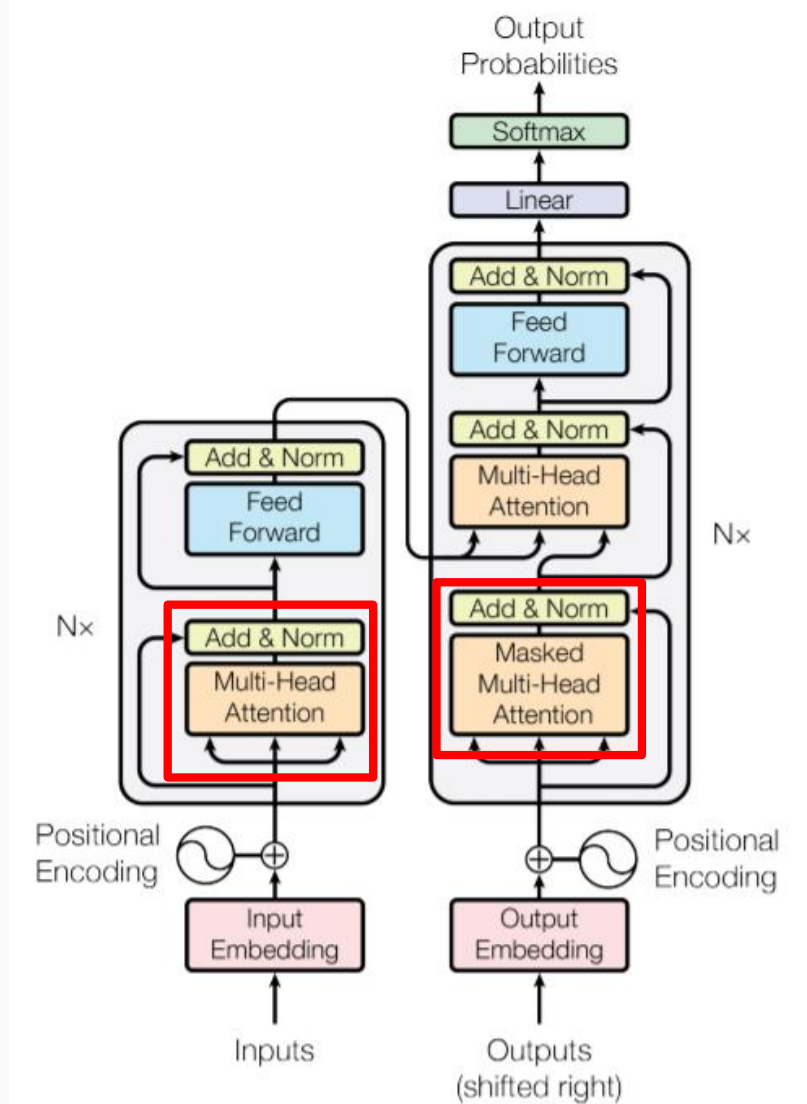
$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

where  $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$



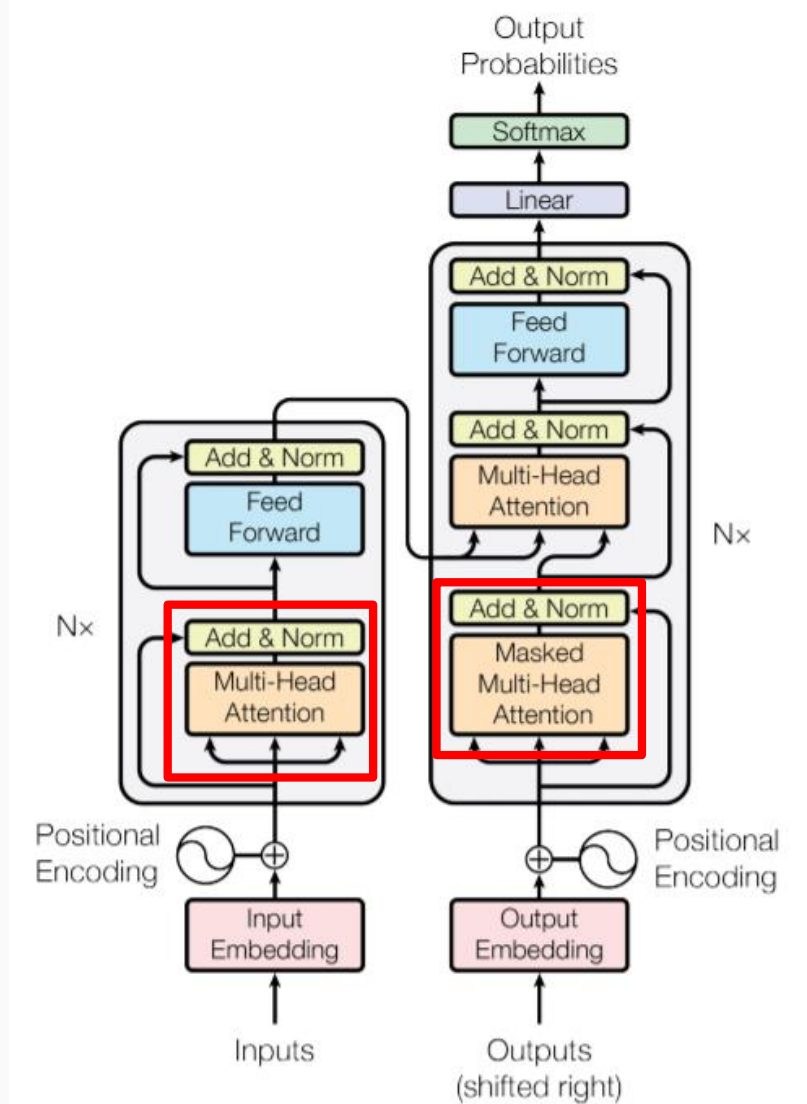
# Question Answering using BERT

## Transformer – self-attention



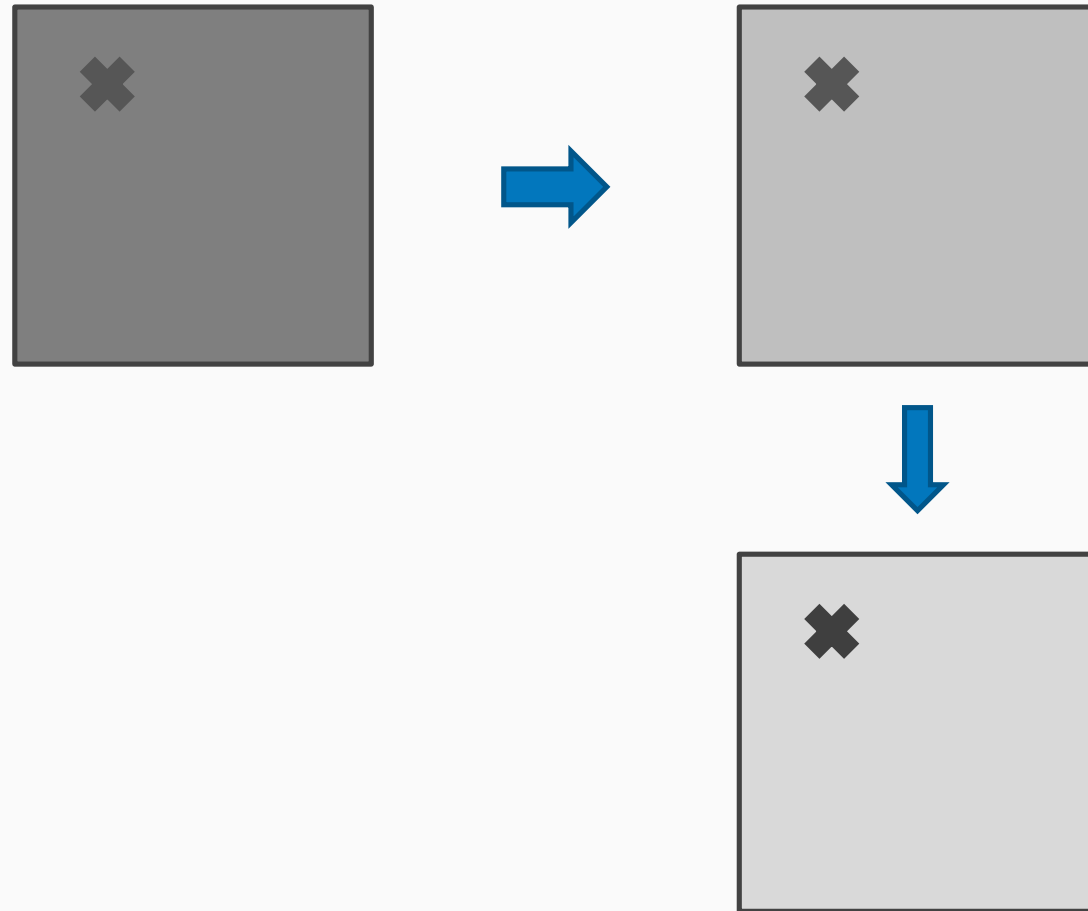
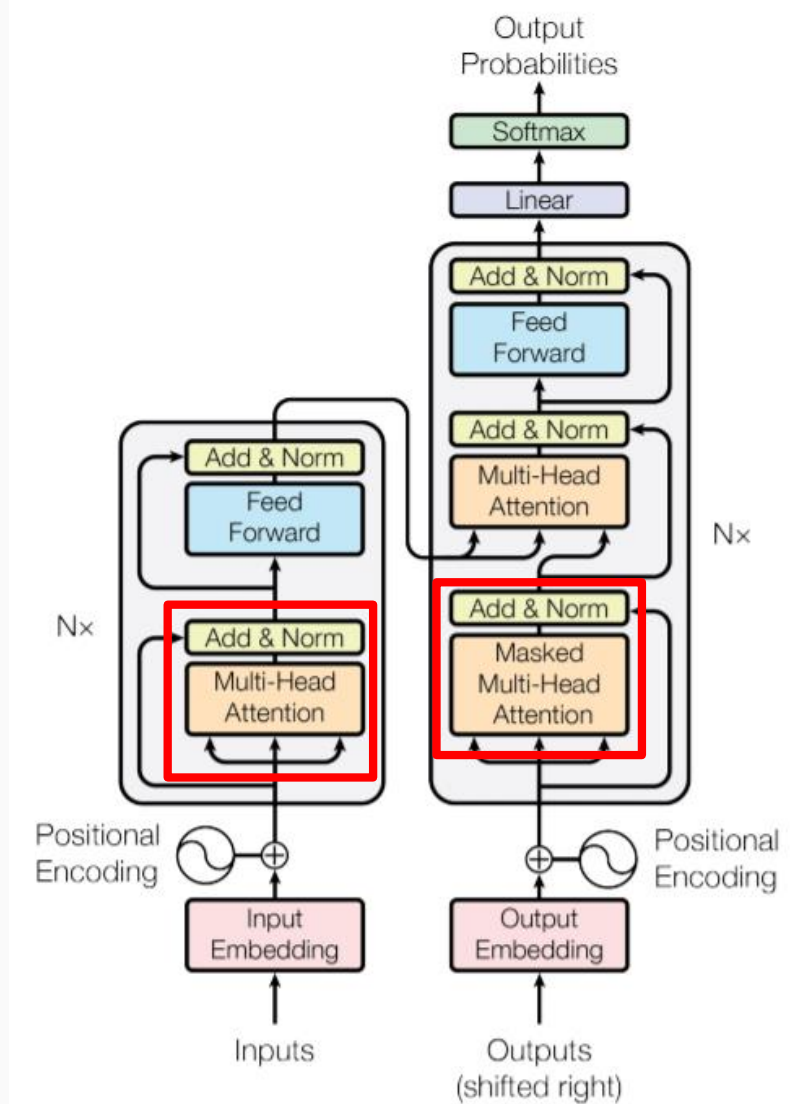
# Question Answering using BERT

## Transformer – self-attention



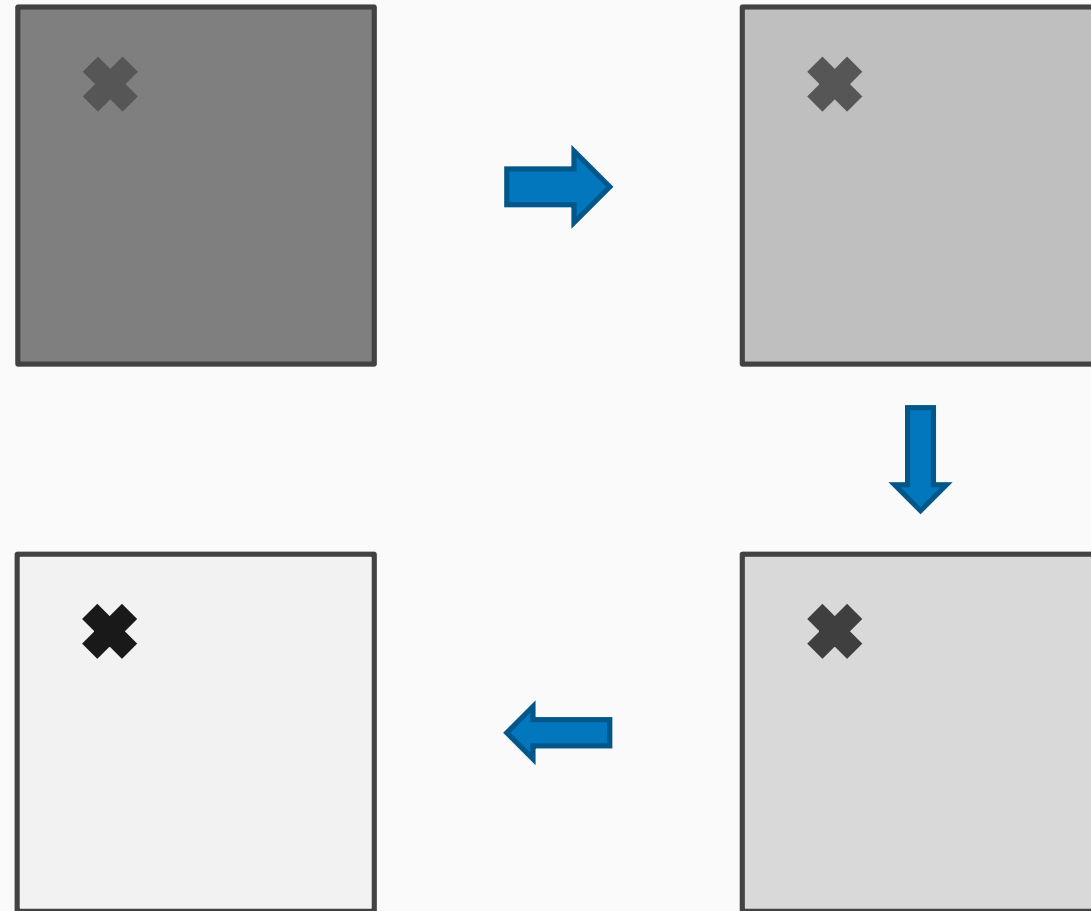
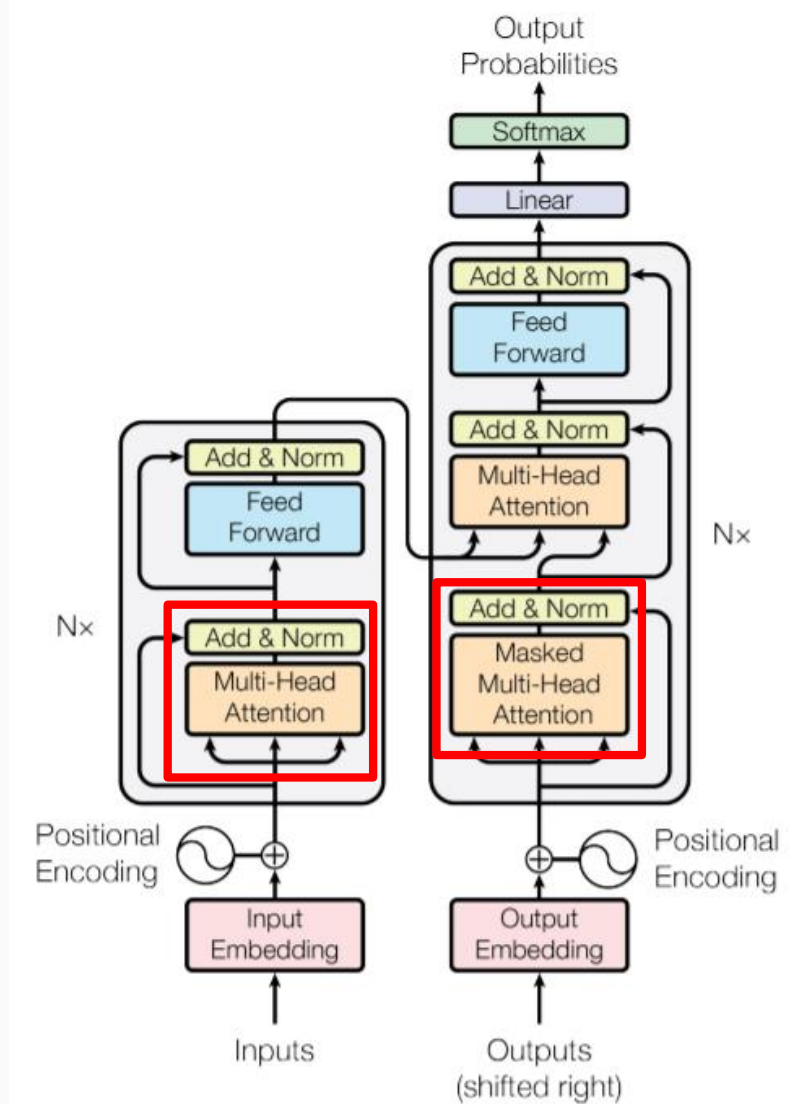
# Question Answering using BERT

## Transformer – self-attention



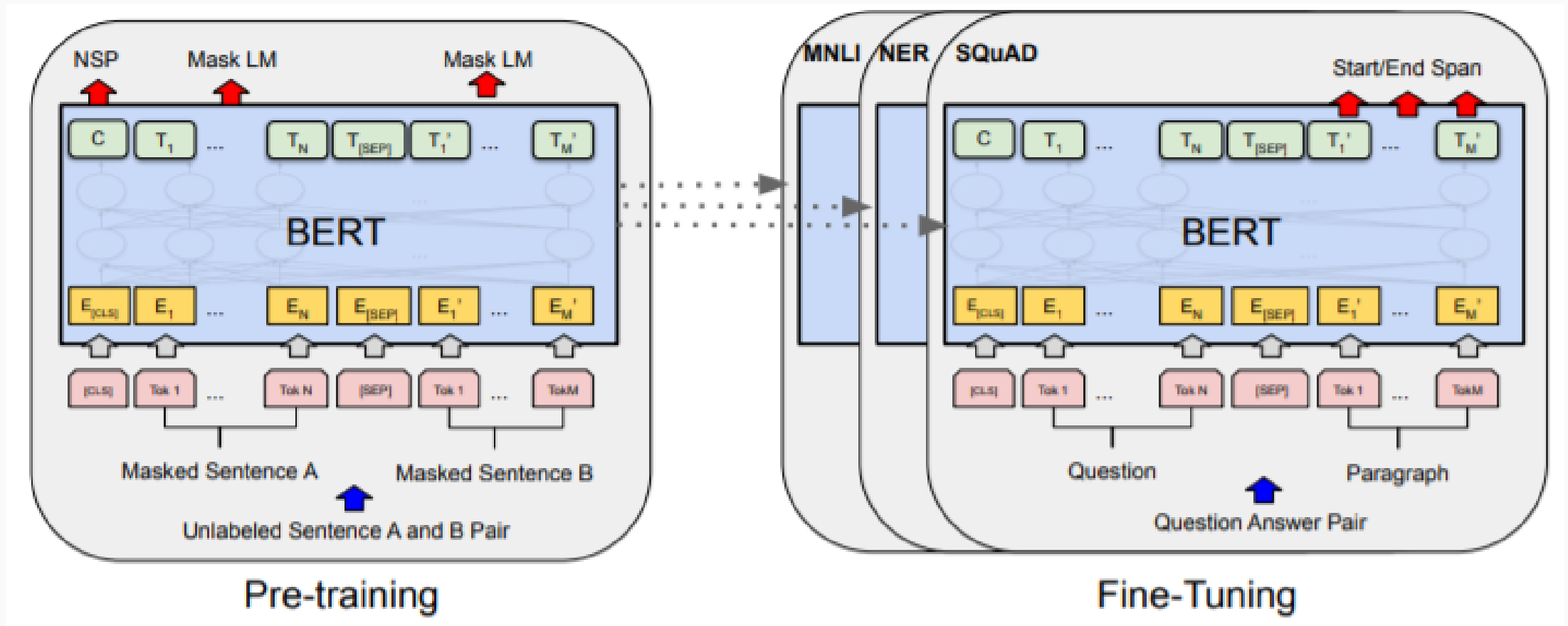
# Question Answering using BERT

## Transformer – self-attention



# Question Answering using BERT

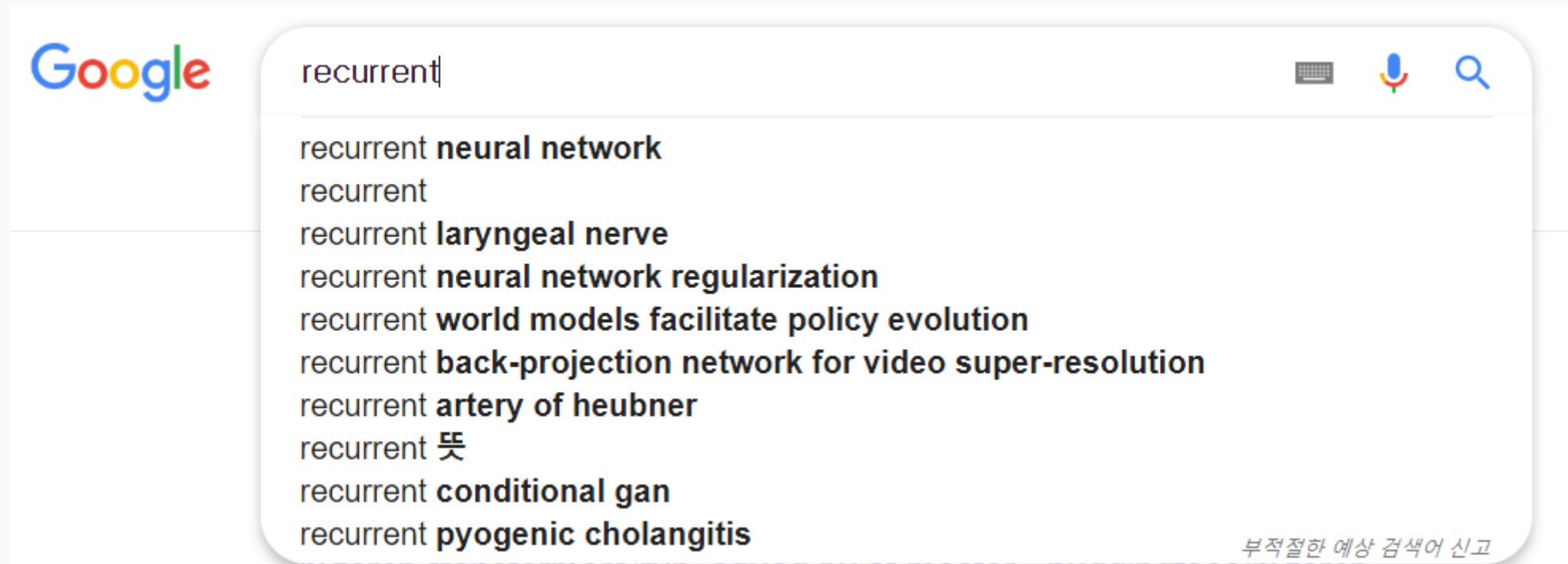
## Bidirectional Encoder Representations from Transformers (BERT)



# Question Answering using BERT

## BERT – Masked Language Model

- Goal of Language Model: Inference  $P(word_{predict} | word_{given})$



## BERT – Masked Language Model

- Goal of Masked Language Model: Inference  $P(word_i | word_{1:i-1}, word_{i+1:N})$
- Example
  - Please send a mail to me [MASK] 4 PM.
  - I really like a cute [MASK].
  - [MASK] is the capital of South Korea.
- There exist more than 1 mask in a sentence.
  - If too much mask in a sentence → hard to learn context
  - If too little mask in a sentence → need to train many time

# Question Answering using BERT

## BERT – Next Sentence Prediction

- Given two sentences  $Sentence_1, Sentence_2$ ,
- Classify whether  $Sentence_2$  is the next sentence of  $Sentence_1$  or not.
- Example

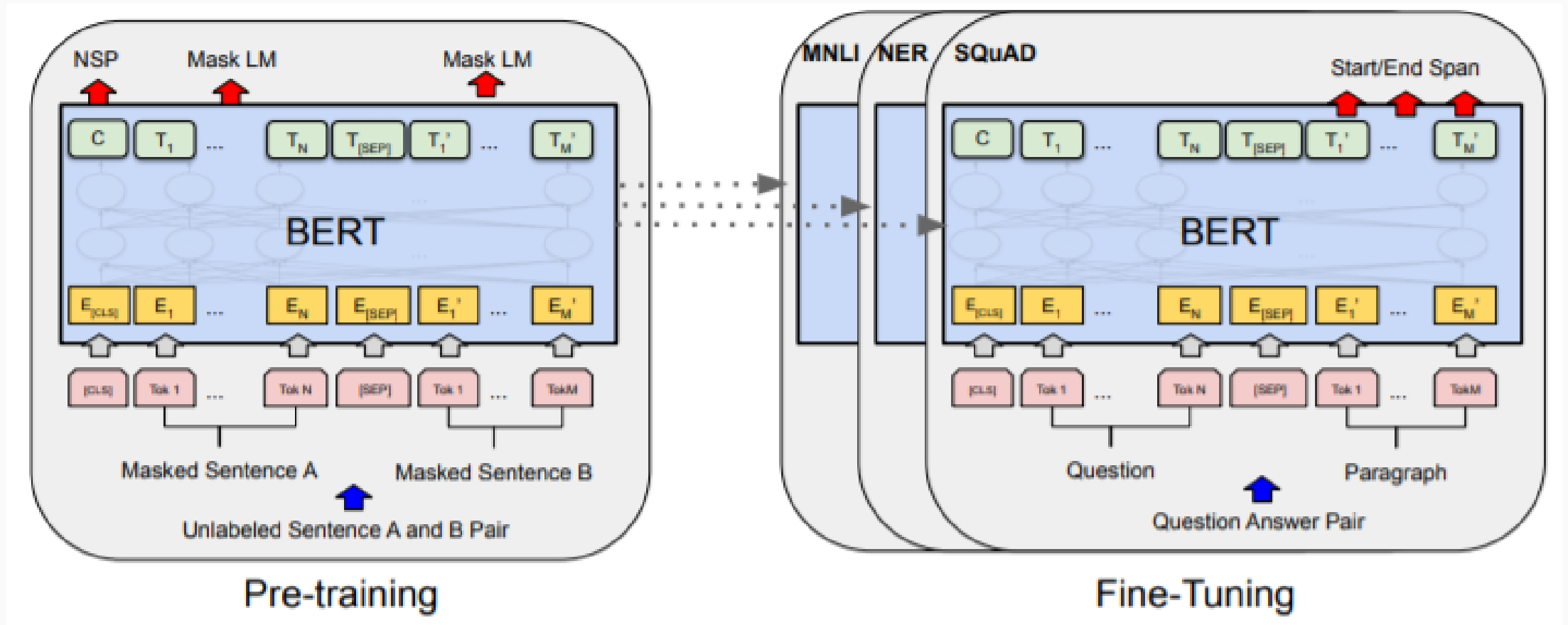
**Sentence A** = The man went to the store.  
**Sentence B** = He bought a gallon of milk.  
**Label** = IsNextSentence

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



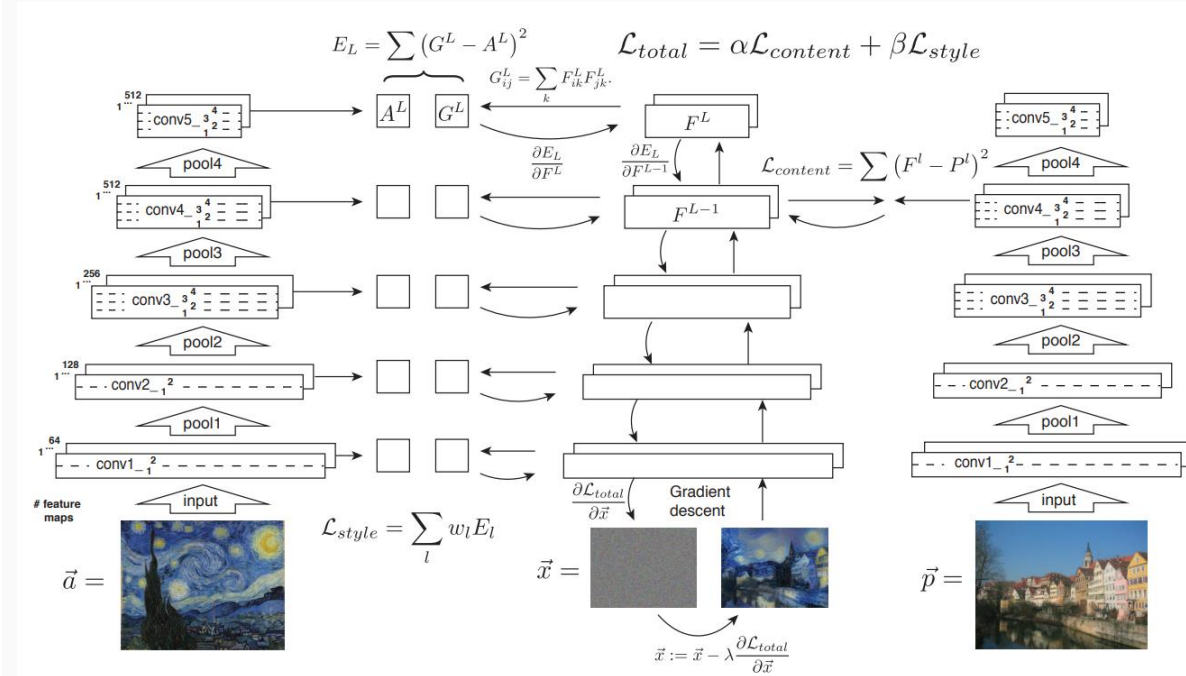
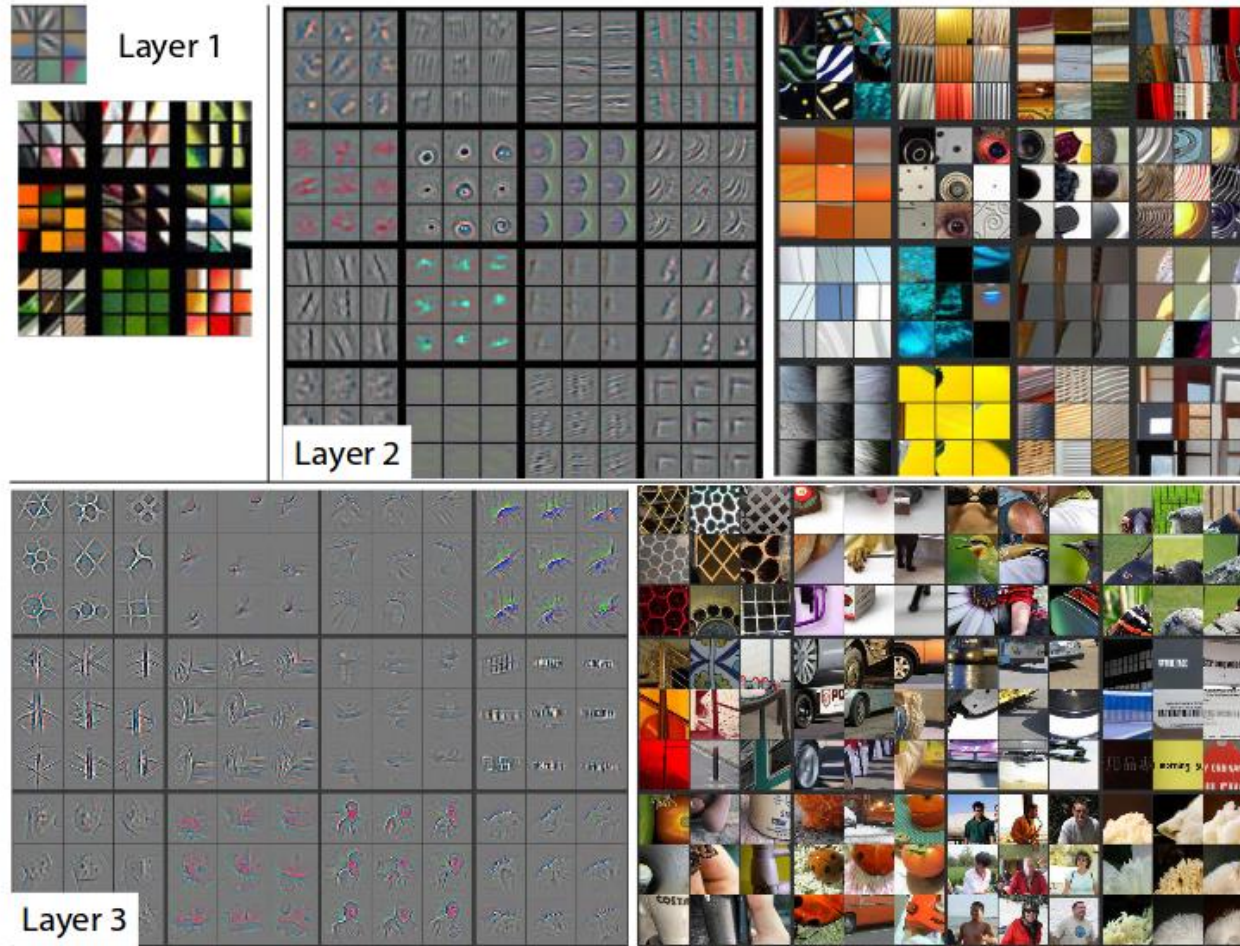
# Question Answering using BERT

## BERT – Fine-Tuning



# Question Answering using BERT

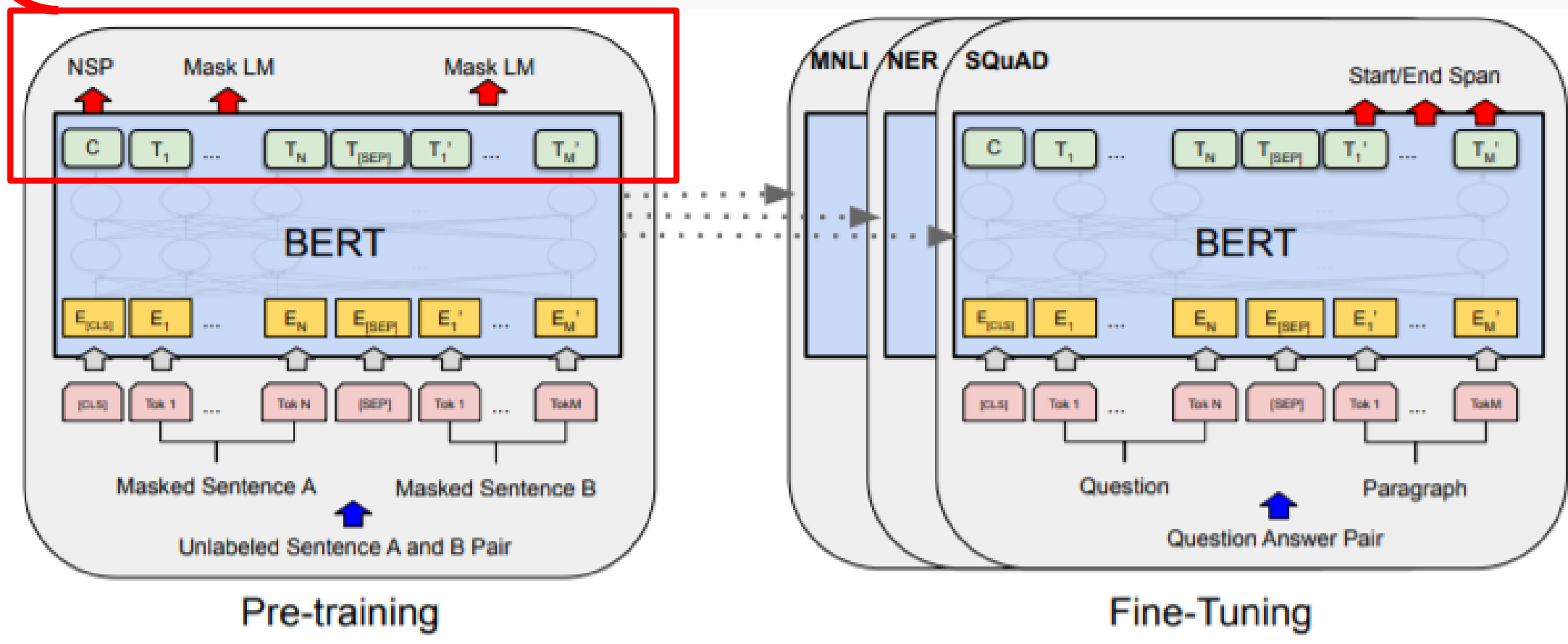
## BERT – Fine-Tuning



# Question Answering using BERT

## Bidirectional Encoder Representations from Transformers (BERT)

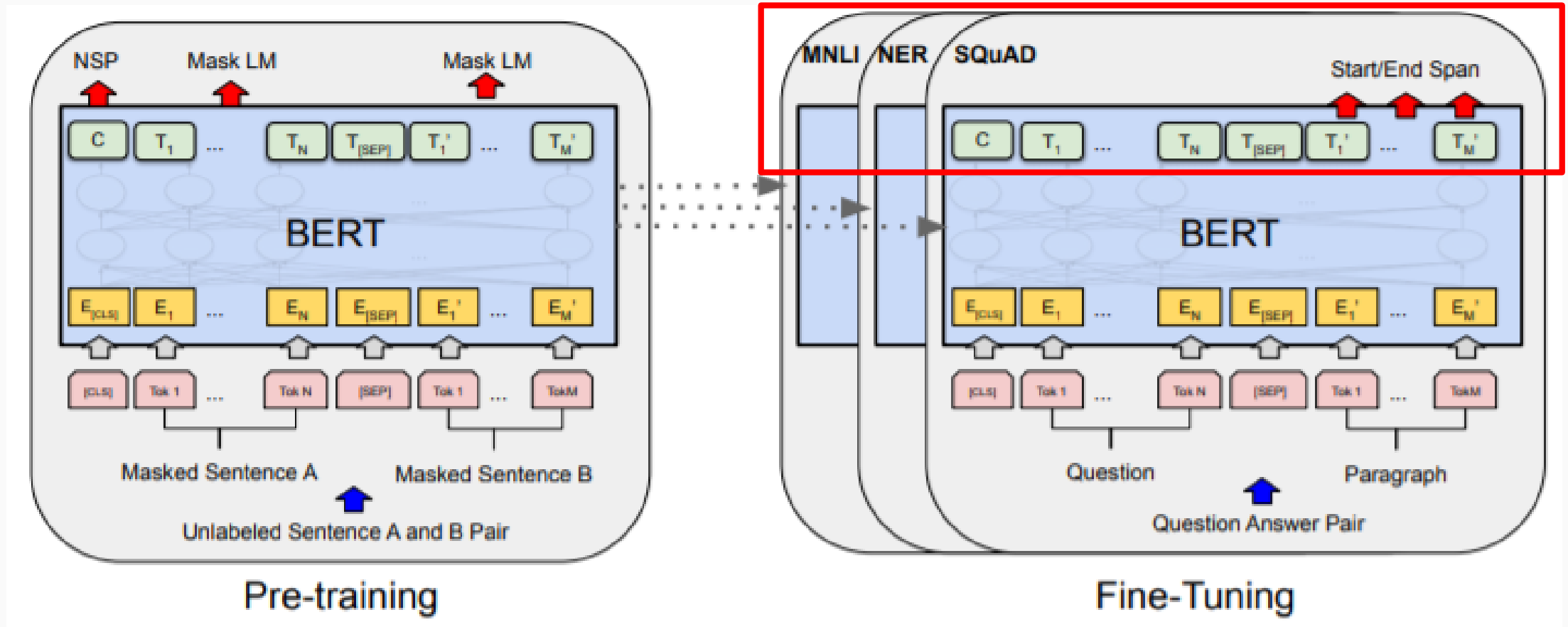
Remove Top Layer



# Question Answering using BERT

## Bidirectional Encoder Representations from Transformers (BERT)

Attach New Top Layer & Update All Parameter



# Reference

- [1] <https://rajpurkar.github.io/SQuAD-explorer/>
- [2] Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." *arXiv preprint arXiv:1611.01603* (2016).
- [3] Gehring, Jonas, et al. "Convolutional sequence to sequence learning." *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017.
- [4] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*. 2017.
- [5] Yu, Adams Wei, et al. "Qanet: Combining local convolution with global self-attention for reading comprehension." *arXiv preprint arXiv:1804.09541* (2018).
- [6] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
- [7] <https://nlp.stanford.edu/seminar/details/jdevlin.pdf>
- [8] <https://blog.nerdfactory.ai/2019/04/25/learn-bert-with-colab.html>