

BERT Architecture Overview

Classification

Token Classification

Question Answering

Contextualized Representations

Layer 4

Multi-Head
Attention



Feed
Forward

Layer 3

Multi-Head
Attention



Feed
Forward

Layer 2

Multi-Head
Attention



Feed
Forward

Layer 1

Multi-Head
Attention



Feed
Forward

Token + Position + Segment Embeddings

[CLS]

The

product

is

great

[SEP]