

BERT and its Applications



- Key Questions
- What is BERT?

Agenda

- What are the variants of BERT?
- How is BERT trained?
- What are the downstream applications of BERT?

Key Questions



What is BERT?

What are the variants of BERT?

How is BERT trained?

What are the downstream applications of BERT?

What is BERT?

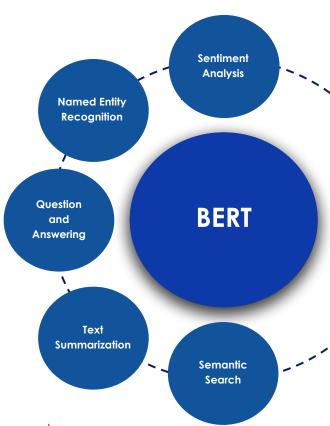


BERT stands for **B**idirectional **E**ncoder **R**epresentations **T**ransformer

It was introduced in a **research paper** by **Jacob Devlin** and his **colleagues** at Google Research in **2018**.

BERT is an **encoder only transformer** model that leverages a **stack** of **encoders** to get an deeper **understanding** of **language context**

BERT **revolutionized** Natural Language Processing (NLP) by exhibiting **state-of-the-art performance** in a variety of NLP tasks

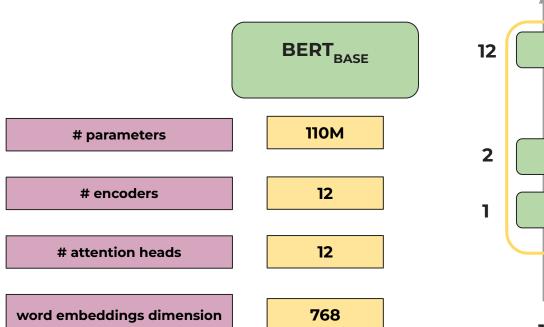


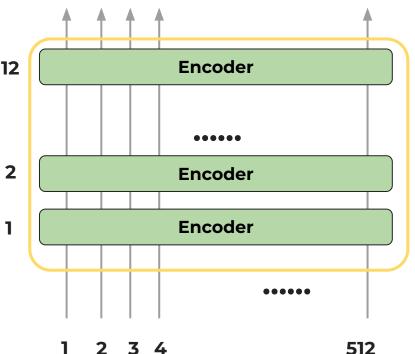
This file is meant for personal use by diegorosenberg@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action. Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What are the variants of BERT?







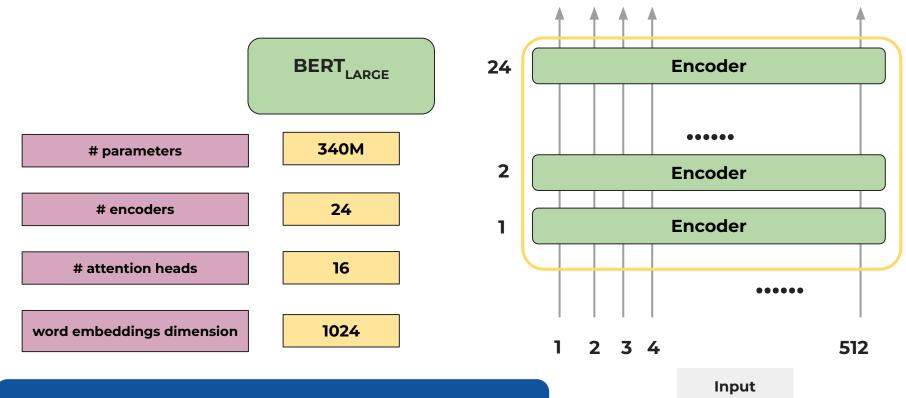
Input

BERT_{BASE} has a **limitation** on the **number of input tokens** - it can take a maximum of **512** input tokens

neant for personal use by diegorosenberg@gmail.com only.

What are the variants of BERT?





BERT_{LARGE} has a similar limitation on the number of input tokens as BERT_{RASE}

This file is meant for personal use by diegorosenberg@gmail.com only.

How is BERT trained?



BERT's training essentially consists of two stages - a **Pre-training** Stage and a **Fine-tuning** Stage

Pre-training Stage

+

Fine-tuning Stage

The model builds a foundational understanding of language

Involves exposing the model to a vast amount of text data, where it learns about word relationships and gains contextual knowledge

The model adapts its foundational understanding of language to perform better on downstream tasks

Involves adjustment of model parameters (weights) to specialize in a particular task.

BERT - Pre-training Stage



BERT's **pre-training** stage is done in **two parts**

Pre-training Stage

Masked Language Modeling (MLM)

Learns to predict masked (missing) words within sentences

It sees sentences with some words intentionally masked and learns to predict these masked words based on the surrounding context.

This task helps the model understand word relationships and meanings.

Next Sentence Prediction (NSP)

NSP focuses on understanding relationships between pairs of sentences.

It learns to predict whether a second sentence follows the first sentence in the text.

This task helps the model understand the connection and coherence between sentences in longer text

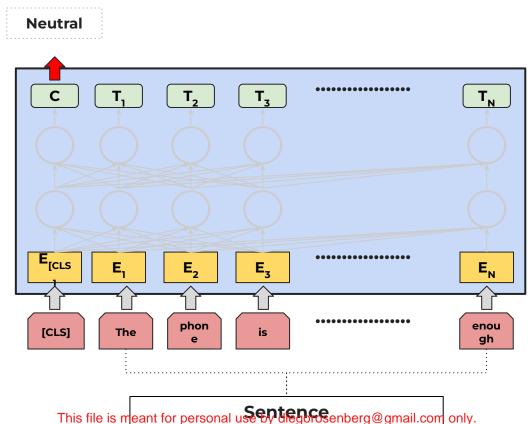
Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What are the downstream applications of BERT?



Sentiment Classification



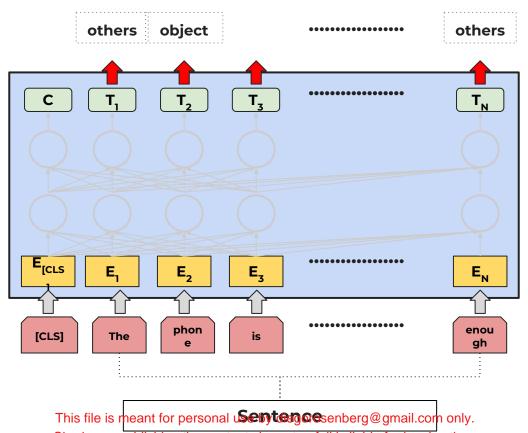
Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning, All Rights Reserved, Unauthorized use or distribution prohibited.

What are the downstream applications of BERT?



Named Entity Recognition



Sharing or publishing the contents in part or full is liable for legal action.

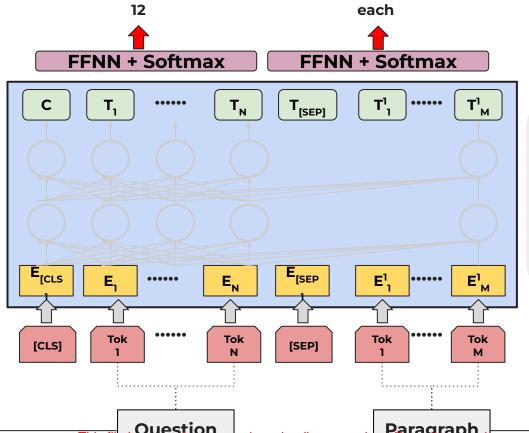
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What are the downstream applications of BERT?





How many attention heads does each transformer encoder in BERT have?



BERT consists of 12 transformer encoders with **12 attention** heads in each, totaling approximately 110 million parameters. This design allows BERT to deeply understand language nuances and relationships, making it adept at various language tasks.

This file is meant for personal use by diegorosenberg@gmail.com only

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



Happy Learning!

