

# BERT: Bidirectional Encoder Representations from Transformers

## Overview

BERT is a deep bidirectional language representation model that processes text by jointly conditioning on both left and right context in all layers. Unlike unidirectional models like GPT, BERT reads entire sequences of words to build richer contextual representations.

This feature is particularly beneficial for token-level NLP tasks, such as Question Answering and Natural Language Inference.

## Pre-Training Tasks

BERT uses two unsupervised pre-training tasks on unlabeled data to learn deep language representations:

### 1. Masked Language Model (MLM)

- Randomly masks 15% of input tokens:
  - 80% are replaced with [MASK]
  - 10% with a random token
  - 10% left unchanged
- Trains the model to predict the original token using context from both sides.
- Encourages bidirectional understanding.

# **BERT: Bidirectional Encoder Representations from Transformers**

## **2. Next Sentence Prediction (NSP)**

- Given two sentences A and B:
  - 50% of the time B is the true next sentence (IsNext)
  - 50% of the time B is a random sentence (NotNext)
- Trains the model to determine whether B logically follows A.
- Helps in learning relationships between sentences.

## **Strengths of BERT**

- Enables deep contextual understanding by considering bidirectional context.
- Highly adaptable to various downstream NLP tasks through fine-tuning.
- Proven effective for applications like sentiment analysis, QA, and language inference.