**BERT – Bidirectional Encoder Representations from Transformers**

This paper introduces BERT, a groundbreaking language representation model developed by Google AI that leverages deep bidirectional transformers. Unlike previous models that were unidirectional (like GPT, which reads text left-to-right), BERT uses a masked language modeling (MLM) approach that allows it to learn from both directions simultaneously. This bidirectionality enables BERT to understand context more effectively, making it highly adaptable across a wide range of natural language processing (NLP) tasks.

BERT is pre-trained using two objectives:

1. Masked Language Modeling (MLM): Random tokens in the input are masked, and the model is trained to predict them using context from both directions.
2. Next Sentence Prediction (NSP): The model is trained to predict whether two given sentences logically follow each other, improving its understanding of sentence relationships.

Once pre-trained, BERT can be fine-tuned with just one additional output layer for various tasks such as question answering, natural language inference, and named entity recognition—without needing task-specific architectures.

The model was released in two sizes: BERTBASE (110M parameters) and BERTLARGE (340M parameters). Both outperform prior models on several NLP benchmarks. For instance, BERTLARGE achieved state-of-the-art results on 11 tasks in the GLUE benchmark, improved F1 scores on SQuAD v1.1 and v2.0, and surpassed human performance on the SWAG common sense reasoning task.

Through comprehensive ablation studies, the paper demonstrates that both the bidirectionality and the NSP objective significantly enhance performance. It also shows that larger model sizes yield better results, even on low-resource tasks.

BERT set a new standard in NLP by combining a simple, unified architecture with effective pre-training, leading to significant improvements across a wide array of language understanding challenges.