BERT

The paper introduces **BERT (Bidirectional Encoder Representations from Transformers)**, a novel language representation model developed by Google AI. Unlike previous models that are limited to unidirectional context (left-to-right or right-to-left), BERT is designed to learn **deep bidirectional representations** by pre-training on unlabeled text using a **Masked Language Model (MLM)** and a **Next Sentence Prediction (NSP)** objective.

Key contributions include:

- **Bidirectional Pre-training**: Unlike models like OpenAI GPT, BERT uses MLM to capture context from both directions, improving understanding at both sentence and token levels.
- **Transfer Learning via Fine-tuning**: BERT can be fine-tuned with minimal changes for a wide range of NLP tasks (e.g., Question Answering, Natural Language Inference) and achieves **state-of-the-art** results on 11 benchmarks, including **GLUE, SQuAD v1.1, and SQuAD v2.0**.
- **General-purpose Model**: BERT shows that one pre-trained model can be adapted effectively to many tasks without task-specific architectures.

The release of BERT and its open-source models has had a major impact on NLP, setting new baselines and simplifying model development for various applications.