**Natural Language Processing**

# BERT

BERT, short for Bidirectional Encoder Representations from Transformers, was developed by Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova in 2018.

**Versatility**: BERT is a natural language processing (NLP) model that serves as a versatile solution for various language tasks, encompassing more than 11 common tasks like sentiment analysis and named entity recognition.

**Bidirectional Representation**: Unlike previous models, BERT pre-trains deep bidirectional representations by considering both left and right contexts in all layers, leading to a better understanding of language.

**Fine-tuning**: BERT's pre-trained representations can be fine-tuned with minimal changes to produce high-performing models for specific tasks like question answering and language inference.

**Data and Training**: BERT was trained on a massive dataset of 3.3 billion words, primarily consisting of Wikipedia (2.5 billion words) and the Google Books Corpus (800 million words), enhancing its understanding of both language and the world.

**Performance**: BERT achieves state-of-the-art results in various NLP tasks. It notably improves the GLUE benchmark to 80.4% with a 7.6% absolute improvement, achieves 86.7% accuracy on MultiNLI with a 5.6% absolute improvement, and reaches 93.2% accuracy on SQuAD 1.1, outperforming human performance by 2%.

**Ease of Use**: BERT's pre-trained model doesn't require significant architecture changes for specific NLP tasks, making it accessible and easy to implement.

# XLNet

**Model Extension**: XLNet is an extension of the Transformer-XL model, utilizing autoregressive pretraining to maximize the expected likelihood across all permutations of the input sequence factorization order.

**Generalized Autoregressive Pretraining**: XLNet employs a generalized autoregressive pretraining method, enabling bidirectional learning by maximizing the expected likelihood over all permutations of the factorization order. This formulation overcomes limitations of previous models.

**Integration of Transformer-XL**: XLNet integrates concepts from Transformer-XL, a leading autoregressive model, into its pretraining process.

**Performance**: XLNet outperforms BERT on various tasks, showing superiority in 20 tasks and achieving state-of-the-art results in 18 tasks. These tasks include question answering, natural language inference, sentiment analysis, and document ranking.

**Superiority Over BERT**: XLNet consistently outperforms BERT by a significant margin across different NLP tasks, demonstrating its effectiveness and superiority.

# PALM

**Model Description**: PALM is a dense decoder-only Transformer model with an impressive scale of 540 billion parameters. It is trained using the Pathways system, which orchestrates distributed computation for accelerators.

**Scalability**: The Pathways system enables the training of a single PALM model across multiple TPU v4 Pods, allowing for unprecedented scalability.

**Performance**: PALM achieves state-of-the-art few-shot performance across a wide range of language understanding and generation tasks. It demonstrates breakthrough capabilities in language understanding, language generation, reasoning, and code-related tasks.

**Experiments**: Experiments conducted with PALM, especially on its largest 540 billion parameter model, show a significant boost in performance across various tasks. It excels in extremely difficult tasks, including language understanding, generation, reasoning, and code generation.

# GPT-3

**Model Description**: Generative Pre-trained Transformer 3 (GPT-3) is an autoregressive language model that utilizes deep learning techniques to generate human-like text.

**Scaling for Few-shot Performance**: Researchers recognized the challenge of collecting large labeled training datasets for various tasks and proposed scaling up language models as an alternative solution to improve task-agnostic few-shot performance.

**Model Scale**: GPT-3 is a massive 175 billion-parameter autoregressive language model, which is significantly larger than its predecessors.

**Evaluation Results**: GPT-3 was trained and evaluated on a variety of natural language processing (NLP) tasks. The evaluation results demonstrate promising performance, with occasional outperformance of fine-tuned models in few-shot learning, one-shot learning, and zero-shot learning scenarios.

**Task Versatility**: GPT-3 exhibits versatility in generating text beyond human language. It can create structured text, text summarizations, and even programming code.

Sources:

<https://medium.com/ubiai-nlp/5-natural-language-processing-models-you-should-know-836958303ce3>

<https://huggingface.co/docs/transformers/model_doc/xlnet>

<https://huggingface.co/docs/transformers/model_doc/bert>

<https://palmai.tech/whitepaper>