



Contact



# TRANSFORMERS MODEL

A DEEP DIVE INTO STATE-OF-THE-ART LANGUAGE MODELS

UIT2022007 - SAJADAA SHAH

UIT2022011 - NISHIGANDHA JADHAV



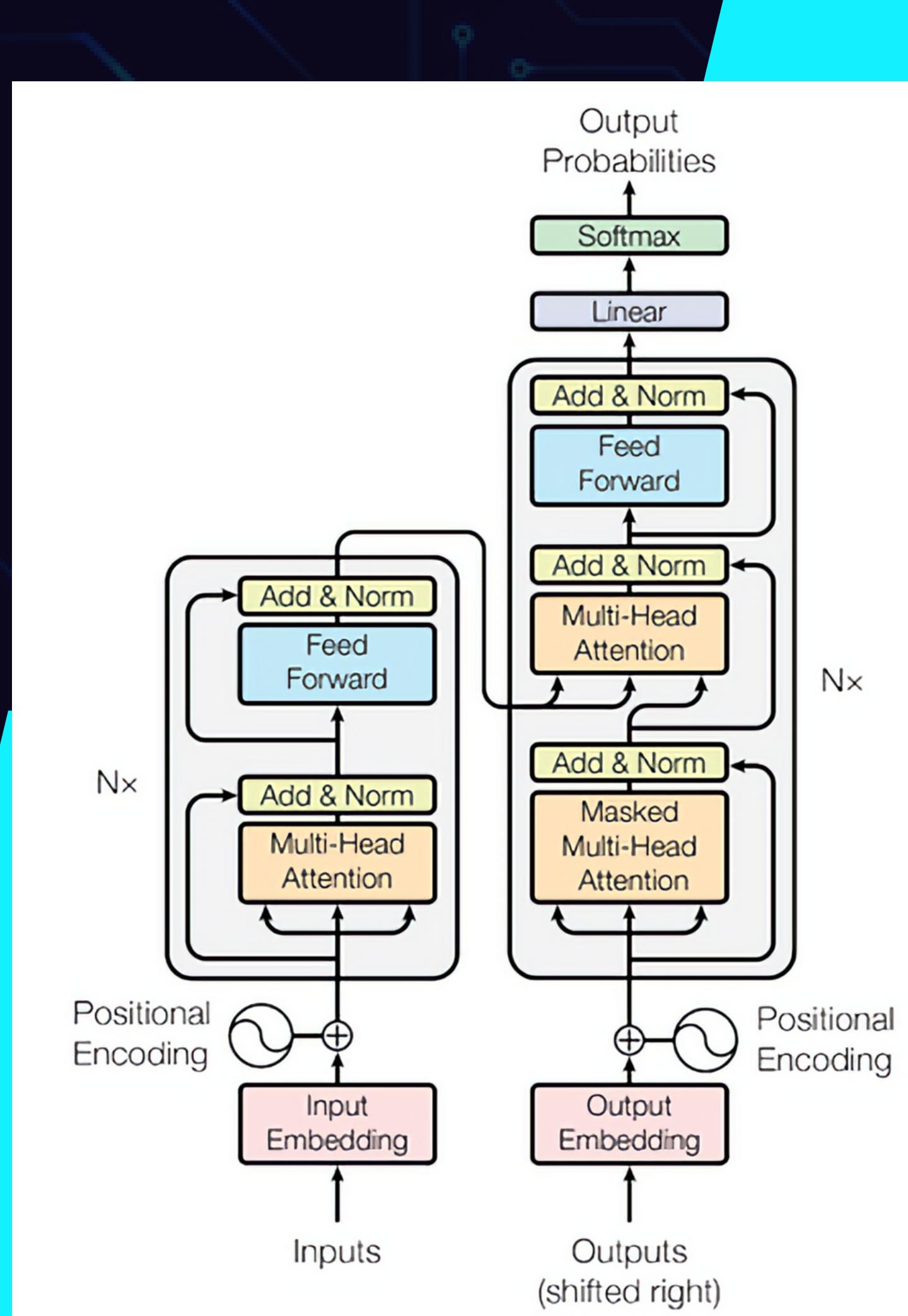
# INTRODUCTION

- **Introduction to Transformer Models in NLP:**
  - Transformer models are a type of deep learning model used for NLP tasks.
  - They excel at learning long-range dependencies between words in a sentence.
- **Encoding the Input Sentence:**
  - Transformer models encode the input sentence into a sequence of vectors.
  - This encoding is achieved using a self-attention mechanism, facilitating learning of word relationships.
- **Decoding Output Tokens:**
  - Once encoded, the model decodes the input sentence into a sequence of output tokens.
  - This decoding also employs a self-attention mechanism.
- **Key Role of Attention Mechanism:**
  - The attention mechanism allows transformer models to learn long-range dependencies.
  - It focuses on the most relevant words in the input sentence during decoding, enhancing performance in tasks like translation and summarization.

# WORK FLOW

The workflow of a Transformer model involves encoding the input sequence into numerical embeddings, applying self-attention mechanisms to capture dependencies, passing through feed-forward networks for pattern learning, and using normalization and residual connections for stability.

In tasks like translation, it utilizes both an encoder and decoder. During training, parameters are optimized, often with pre-training and fine-tuning. During inference, the model generates outputs based on input. This succinct process highlights how Transformers efficiently handle various NLP tasks.



# BERT: BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS

BERT, or Bidirectional Encoder Representations from Transformers, is a pioneering language model introduced by Google in 2018. It stands out for its ability to deeply understand language context by employing bidirectional encoding, capturing information from both left and right sides of a word. Through unsupervised pre-training with a masked language model objective, BERT learns to predict masked words in sentences, thereby gaining a nuanced understanding of language structure and semantics. This pre-trained model can then be fine-tuned on specific NLP tasks, such as sentiment analysis or question answering, achieving remarkable performance across a wide range of benchmarks. BERT's contextualized word embeddings and state-of-the-art results have revolutionized the field of NLP, driving advancements in research and practical applications alike.



# GPT

GPT, or Generative Pre-trained Transformer, is an influential language model introduced by OpenAI. Launched in 2018, GPT represents a significant milestone in natural language processing (NLP). Unlike traditional models that focus on specific tasks, GPT adopts a generative approach, capable of understanding and producing human-like text. Its architecture is based on the Transformer model, renowned for its self-attention mechanism, enabling GPT to capture intricate relationships and dependencies within language data. Pre-trained on vast amounts of text from the internet, GPT can generate coherent and contextually relevant text across a variety of domains, from writing stories to answering questions. Its flexibility and proficiency have led to widespread adoption in NLP research and applications, propelling advancements in text generation, dialogue systems, and more.

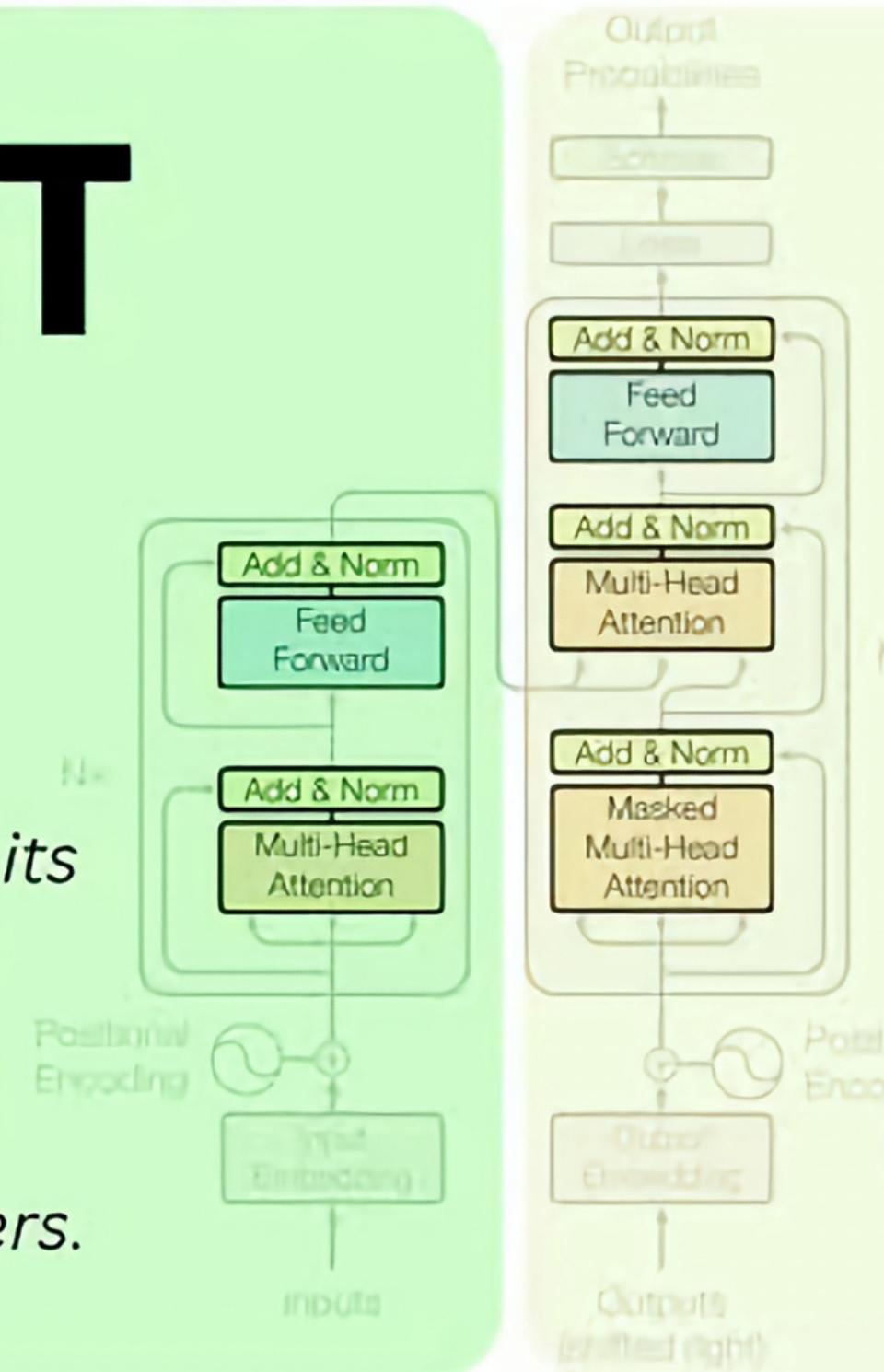


# BERT VS GPT

## BERT

Google

*use transfer learning to **continue learning** from its existing data when adding user-specific tasks and layers.*



## GPT

OpenAI

*decodes from its massive pre-learned embeddings to present output that matches user prompts. It*

Figure 1: The Transformer - model architecture.

# CONCLUSION

In conclusion, both BERT and GPT exemplify the transformative power of Transformer-based models in natural language processing. While BERT excels in understanding context and fine-tuning for specific tasks through bidirectional encoding, GPT shines in generating coherent and contextually relevant text. Together, they have revolutionized NLP research and applications, pushing the boundaries of language understanding and generation. Their versatility, state-of-the-art performance, and widespread adoption underscore their pivotal role in advancing the field, paving the way for more sophisticated and nuanced language models in the future.

