Assignment 2

الاسم: فيصل عبدالناصر عبدالموحود احمد

ID: 20210678

**GPT-4 vs. BERT: Architecture Comparison**

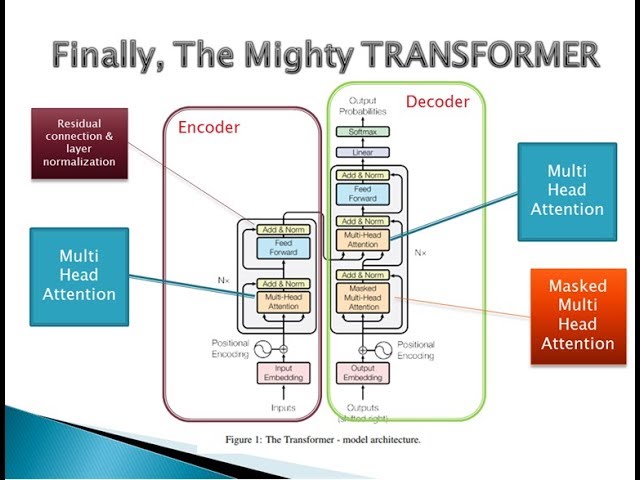
This document explains the key differences between **GPT-4** and **BERT** in simple terms, focusing on their **architecture** (how they are built and work).

**Transformers**

**Transformers are a type of neural network architecture used in artificial intelligence (AI) to understand and generate text (and other data like images or audio). They were introduced in 2017 and have since revolutionized AI tasks like translation, chatbots, and more.**

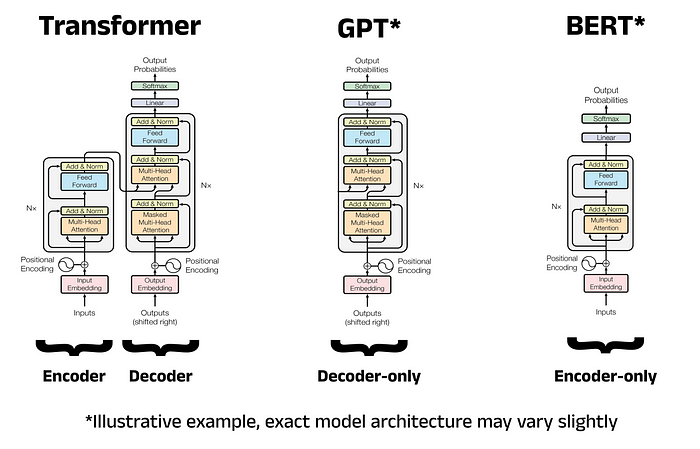
**How Do Transformers Work?**

1. **No Fixed Order: Unlike older models (e.g., RNNs or CNNs), Transformers don’t read text word-by-word. They process all words at once, making them faster and better at understanding context.**
2. **Encoder & Decoder:**
   * **Encoder: Understands the input (e.g., a sentence).**
   * **Decoder: Generates the output (e.g., a translated sentence).**
   * ***Note*: Models like BERT use only the encoder, while GPT-4 uses only the decoder.**



**Examples of Transformer Models**

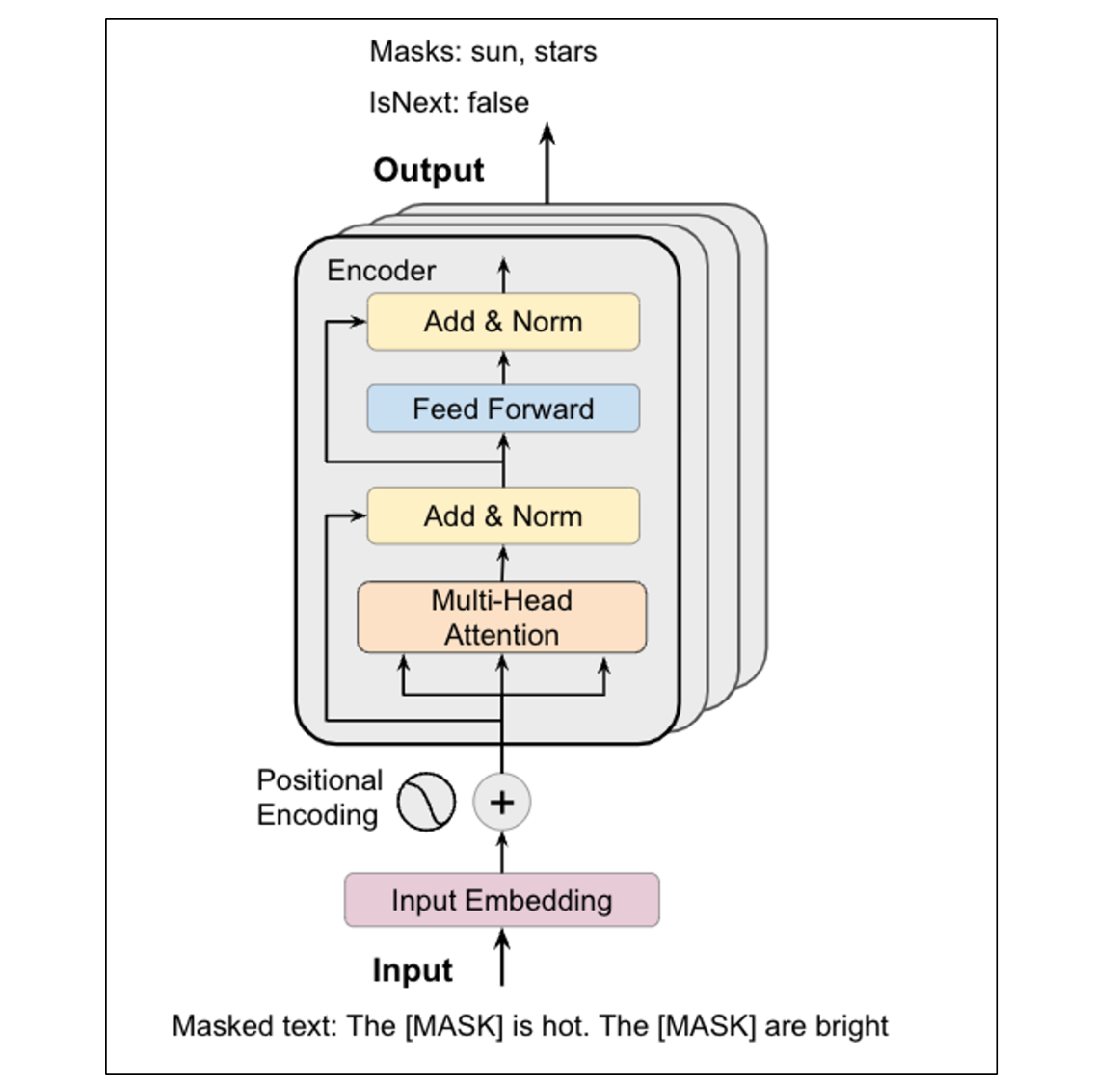
* **BERT**: Focuses on **understanding** text (e.g., answering questions).
* **GPT-4**: Focuses on **generating** text (e.g., writing stories).
* **T5**: Does both understanding and generation.



1. **What is BERT?**

**Architecture:**

* BERT is like a **smart reader** that looks at *all words in a sentence at once*.
* It uses a **bidirectional** approach:
* Imagine reading a sentence forward *and* backward to understand every word’s meaning.
* Example: For the sentence *"She went to the bank to withdraw money,"* BERT checks both "bank" (river?) and "withdraw money" (financial bank?) to decide the right meaning.



**How It Works**:

* Built as a **Transformer Encoder** (a part of the Transformer model that focuses on *understanding* text).
* Trained by guessing **masked words** (e.g., filling in blanks like "The cat sat on the \_\_\_").

**Key Features**:

* Great for tasks needing **context understanding** (e.g., answering questions, classifying text).
* Cannot **generate new text** (it only understands, doesn’t write).

**2. What is GPT-4?**

**Architecture**:

* GPT-4 is like a **creative writer** that predicts the *next word in a sentence*.
* It uses a **unidirectional** approach:
  + Reads text *left-to-right*, one word at a time.
  + Example: For the sentence *"The cat sat on the...,"* GPT-4 predicts "mat" or "roof" based on previous words.

**How It Works**:

* Built as a **Transformer Decoder** (a part of the Transformer model that focuses on *generating* text).
* Trained by predicting the **next word** in a sequence (like autocomplete).

**Key Features**:

* Excels at **text generation** (e.g., writing essays, chatbots, stories).
* Can handle many tasks without specific training (e.g., translate, summarize, code).

A diagram of a multi-head process

AI-generated content may be incorrect.

1. **Main Differences in Architecture**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Feature | BERT | GPT-4 | | Direction | **Bidirectional (reads whole sentence).** | **Unidirectional (reads left-to-right).** | | Purpose | **Understands text.** | **Generates text.** | | Training | **Guesses masked words in sentences.** | **Predicts next word in a sequence.** | | Best At | **Tasks like Q&A, sentiment analysis.** | **Tasks like writing, summarizing, chatting.** | | Limitations | **Can’t create new text.** | **May make up facts or repeat itself.** | |

**4. Simple Analogy**

* **BERT** is like a **detective**:
  + Examines all clues (words) at once to solve a mystery (understand context).
  + Example: "Is this review positive or negative?"
* **GPT-4** is like a **storyteller**:
  + Builds a story one word at a time, always looking forward.
  + Example: "Write a poem about the ocean."

**5. Which Should You Use?**

* **Choose BERT** if you need to:
  + Analyze text (e.g., classify emails, answer questions).
  + Understand context deeply (e.g., "What does 'bank' mean here?").
* **Choose GPT-4** if you need to:
  + Generate text (e.g., write articles, code, or chat).
  + Perform tasks with little to no examples (e.g., translate without training).