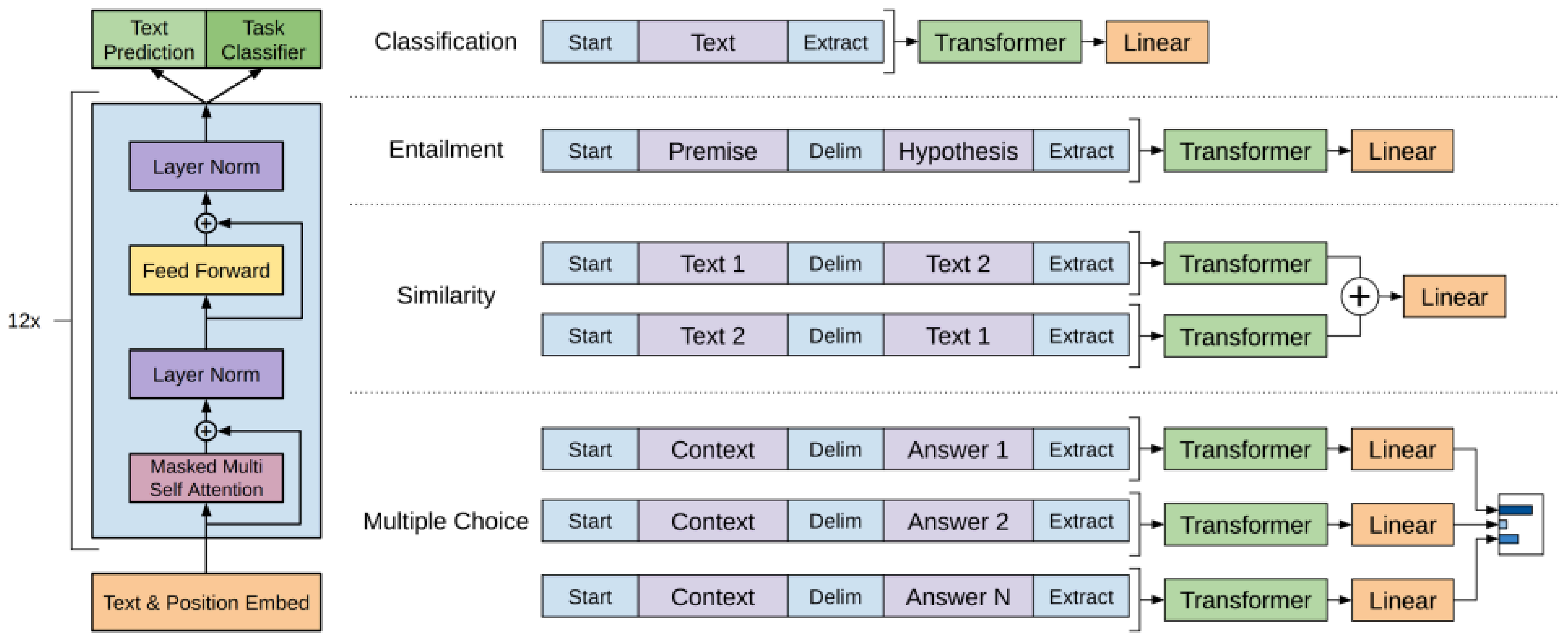
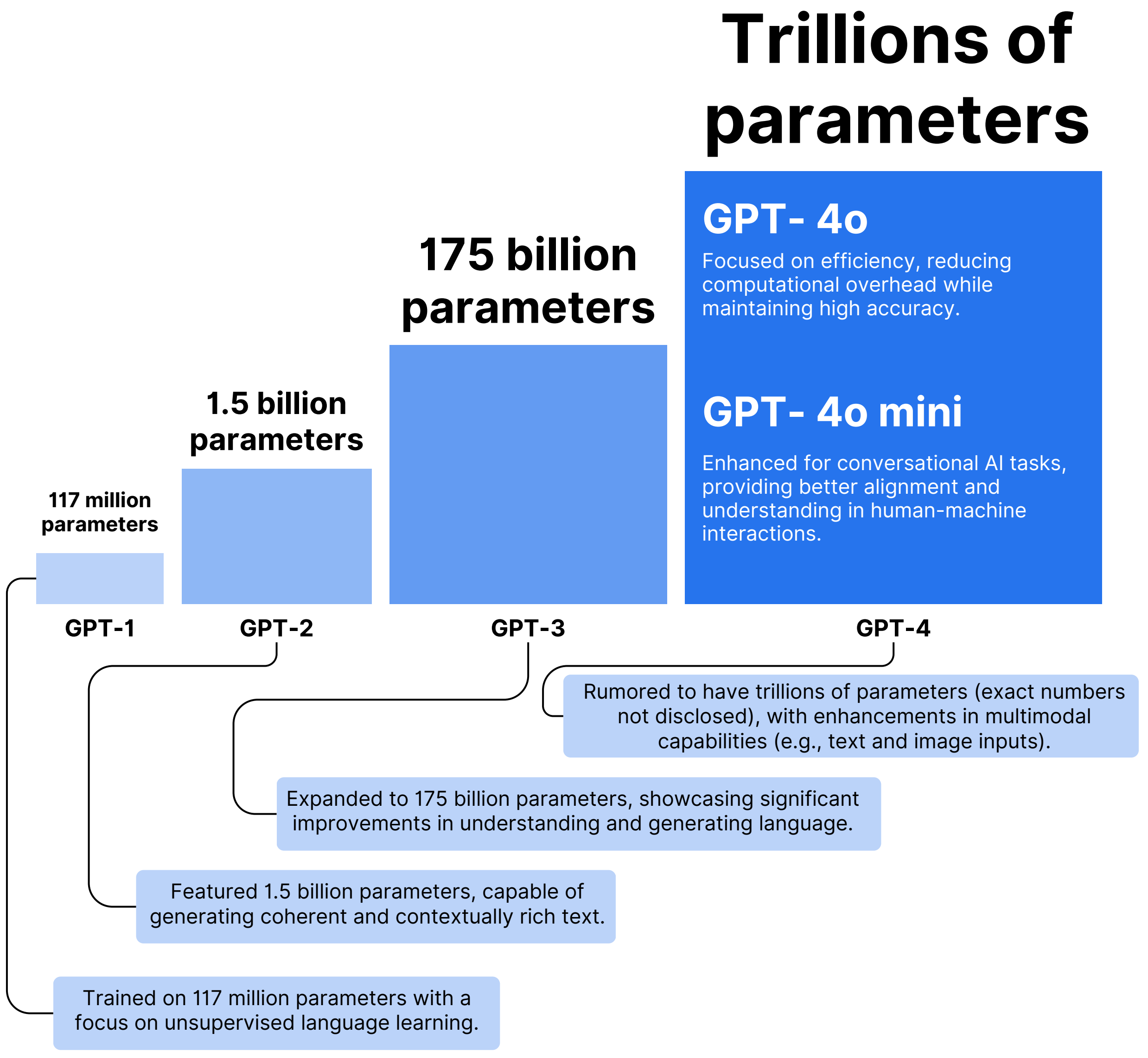
Analytics Vidhya

Mastering LLMs

Day 7: A Perfect Guide to GPT





GPT- 4o

Focused on efficiency, reducing computational overhead while maintaining high accuracy.

GPT- 4o mini

Enhanced for conversational AI tasks, providing better alignment and understanding in human-machine interactions.

Rumored to have trillions of parameters (exact numbers not disclosed), with enhancements in multimodal capabilities (e.g., text and image inputs).

Expanded to 175 billion parameters, showcasing significant improvements in understanding and generating language.

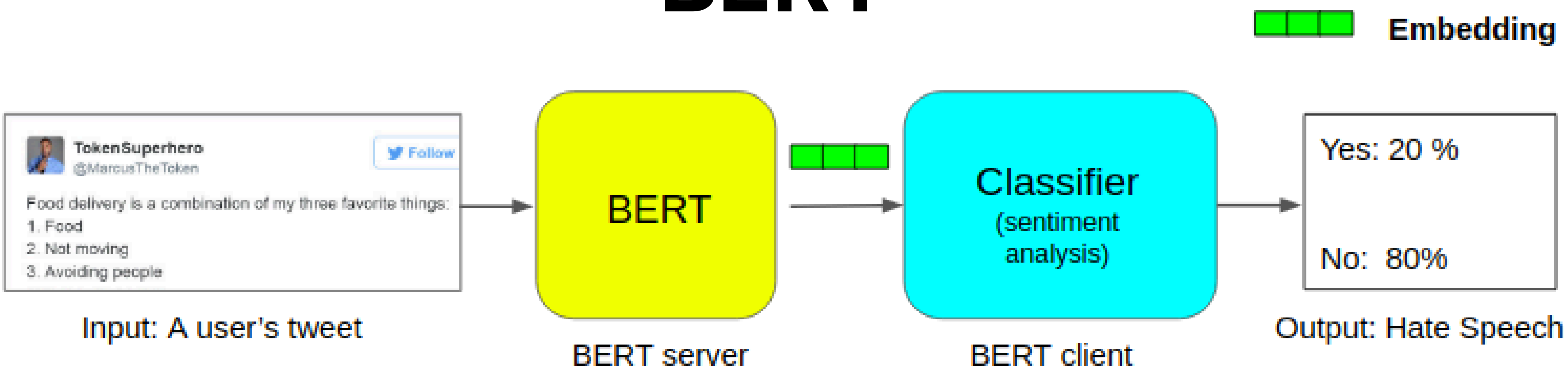
Featured 1.5 billion parameters, capable of generating coherent and contextually rich text.

Trained on 117 million parameters with a focus on unsupervised language learning.

The increase in parameters correlates with improved understanding of context, nuance, and language patterns.

Yesterday, we talked about **BERT**. Today we will cover one of the most important and transformative language model which has changed the world.

BERT



🟩🟩🟩 Embedding

TokenSuperhero
@MarcusTheToken
Follow

Food delivery is a combination of my three favorite things:
1. Food
2. Not moving
3. Avoiding people

Input: A user's tweet

BERT

BERT server

Classifier (sentiment analysis)

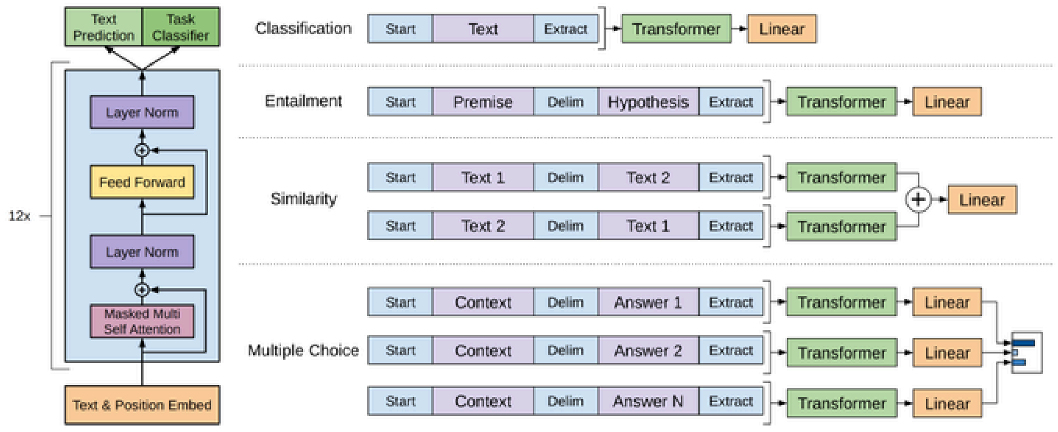
BERT client

Yes: 20 %

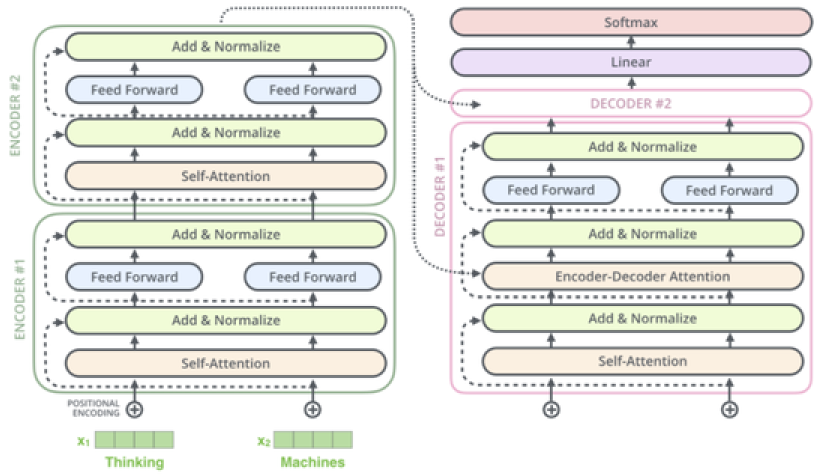
No: 80%

Output: Hate Speech

GPT



T5



ModernBERT

Answer.AI + LightOn +

Let's dive straight into

GPT →

What is GPT?

- GPT, short for Generative Pretrained Transformer, is a type of neural network model developed primarily for natural language processing (NLP) tasks.

- Created by OpenAI, GPT has become a benchmark in the AI domain for its ability to generate coherent, contextually relevant text. Leveraging the Transformer architecture, GPT is pretrained on vast datasets and then fine-tuned for specific tasks such as translation, summarization, question answering, and more.

Key Features of GPT

- **Generative Capability**: GPT can generate human-like text, complete sentences, and respond contextually to prompts.

- **Pretrained**: It is first pretrained on large-scale, diverse text corpora to capture general language patterns and structures.

- **Transformer Architecture**: The model relies on attention mechanisms to handle contextual relationships within the text efficiently.

GPT Framework

- GPT-1 uses a **12-layer decoder-only** transformer framework with masked self-attention for training the language model.

- The GPT model's architecture largely remained the same as it was in the original work on transformers. With the help of masking, the language model objective is achieved whereby the model doesn't have access to subsequent words to the right of the current word.

- GPT-1 language model was trained using the BooksCorpus dataset. BooksCorpus consists of about 7000 unpublished books which helped in training the language model on unseen data.

- This corpus also contained long stretches of contiguous text, which assisted the model in processing long-range dependencies.
- Let’s look at the training approach used for the GPT-1 model now.
- Training a GPT model consists of the following three stages:

i) Learning a high-capacity language model on a huge corpus of text (pre-training).

ii) Fine-tuning, where the model is adapted to a discriminative task with labeled data.

iii) Task-specific Input Transformations for certain tasks like question answering or textual entailment have structured inputs like triplets of documents, ordered sentence pairs, questions, and answers.

Task-Specific Input Transformations

- We can directly fine-tune the model for tasks like **text classification**. However, tasks like textual entailment, question answering, etc., that have structured inputs require task-specific customization.

- To make minimal adjustments to the model's architecture during fine-tuning, inputs to the particular downstream tasks are transformed into ordered sequences. The tokens are rearranged as follows:

    - Start and end tokens are added to the input sequences.

    - A delimiter task token is also added between the input context (prompt) and output (response)

- For tasks like question-answering (QA), multiple choice questions (MCQs), etc, multiple sequences are sent for each example.

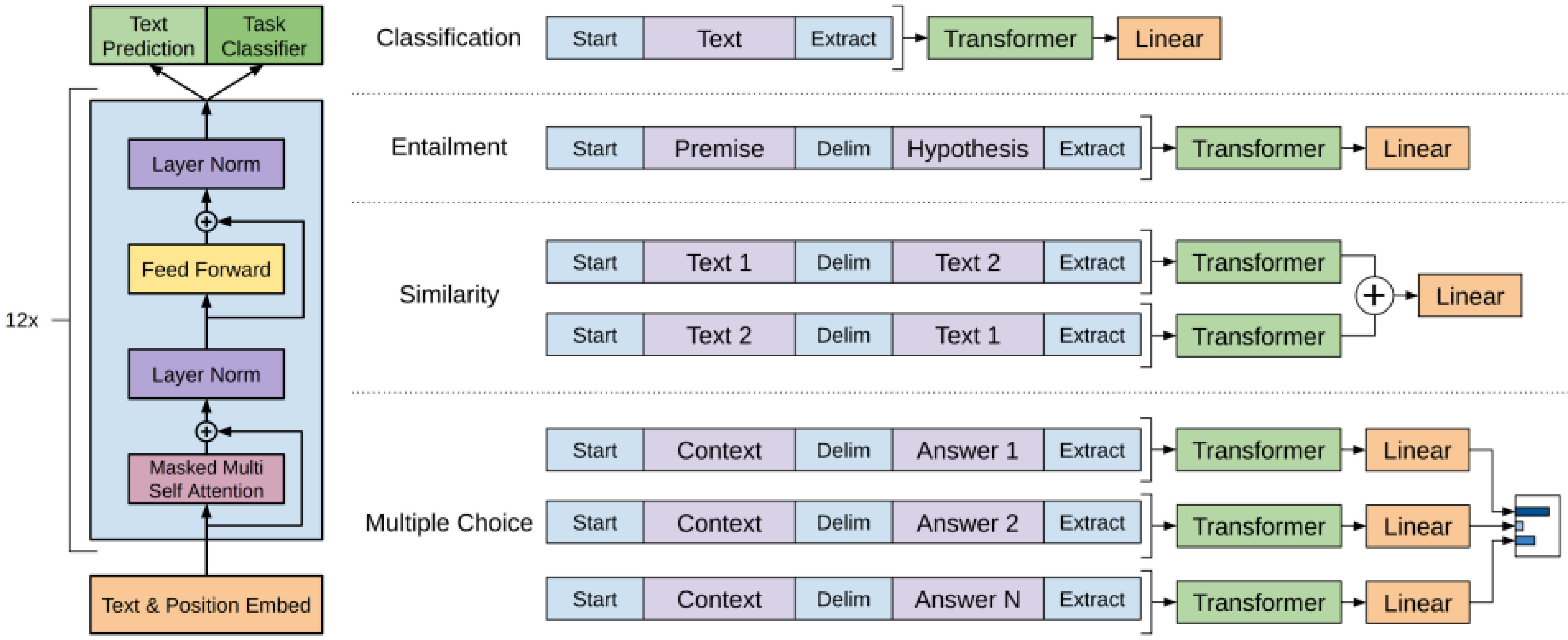
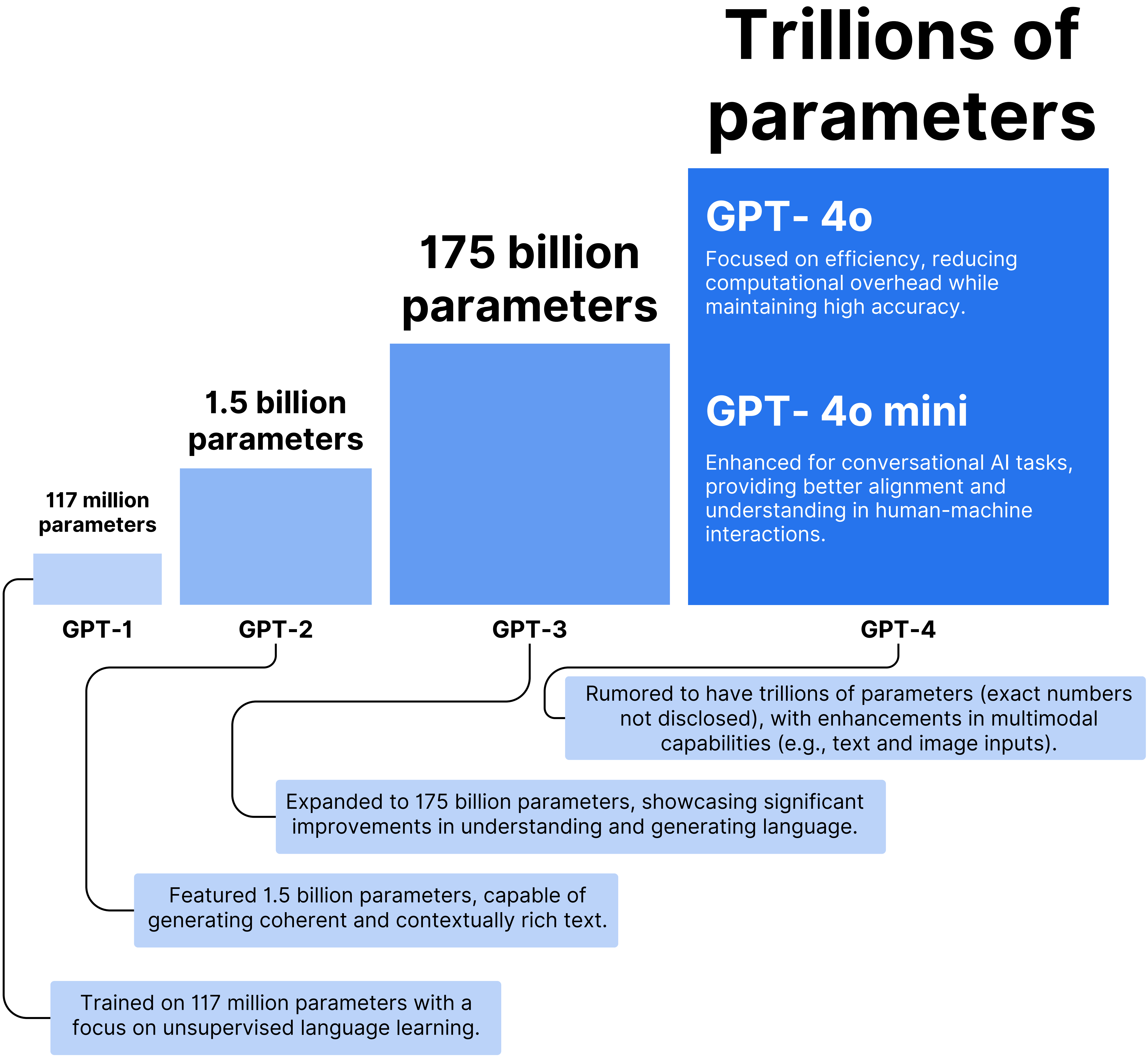


Figure 1: (Left) Generative Pre-training transformer architecture and training objectives used in this work.

(Right) Input
transformations for fine-tuning the model on different tasks.

Model Parameters

- The evolution of GPT has been marked by increasing the number of parameters, leading to better performance



GPT- 4o

Focused on efficiency, reducing computational overhead while maintaining high accuracy.

GPT- 4o mini

Enhanced for conversational AI tasks, providing better alignment and understanding in human-machine interactions.

Rumored to have trillions of parameters (exact numbers not disclosed), with enhancements in multimodal capabilities (e.g., text and image inputs).

Expanded to 175 billion parameters, showcasing significant improvements in understanding and generating language.

Featured 1.5 billion parameters, capable of generating coherent and contextually rich text.

Trained on 117 million parameters with a focus on unsupervised language learning.

Variations and Similar Models

In addition to GPT, various models and variations have been developed to address specific challenges and tasks in NLP:

- **BERT (Bidirectional Encoder Representations from Transformers)**: Unlike GPT, BERT uses a bidirectional Transformer architecture and is optimized for understanding the full context of words in both directions.
- **T5 (Text-to-Text Transfer Transformer)**: Treats all NLP tasks as a text-to-text problem, making it highly versatile.
- **XLNet**: Combines the strengths of autoregressive models like GPT and bidirectional models like BERT.
- **Megatron**-LM: Developed by NVIDIA, optimized for training massive transformer models.
- **Claude**: Developed by Anthropic, emphasizing alignment with human values and safety.
- **Gemini**: Google’s large language model known for scaling efficiency and performance.

Analytics Vidhya

Stay Tuned for **Day 8** of

Mastering LLMs