



Advanced Computer Science Course Lecture 4

Tishreen University

Computer and automatic control
engineering dept.

Master Program- 2024

1st year

Dr. Ali Mahmoud Mayya



From Transformers to GPT



Generative Pretrained Transformers (GPTs) Language Models

What is GPT? a family of neural network models that utilize the **transformer architecture**.

GPT-1

2018 by OpenAI

Uses Transformer Model

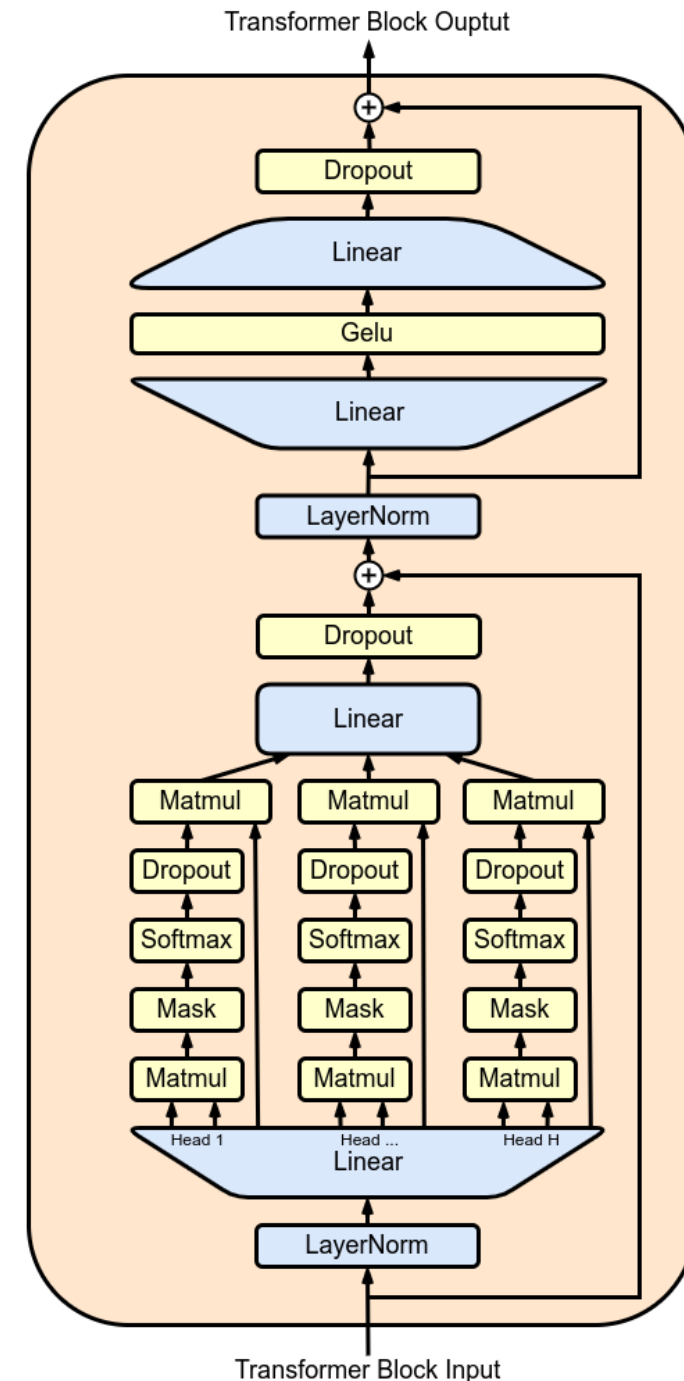
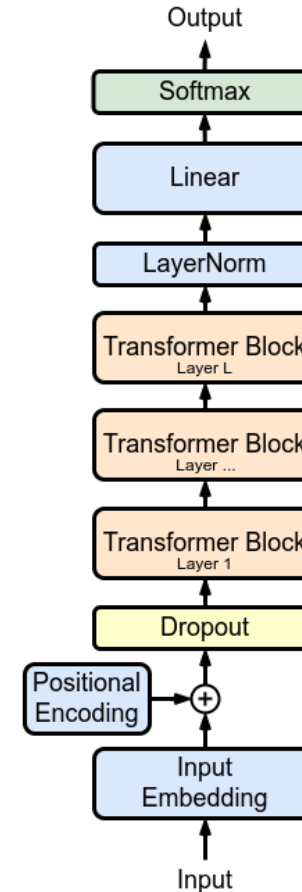
GPT-1 contains 117 million parameters

Trained on an enormous BooksCorpus dataset

Architecture: 12-layer decoder of the transformer architecture with a self-attention mechanism for training

GELU (Gaussian Error Linear Unit) activation function is used in the feed-forward layer:

- Faster than RELU (continuous derivative at 0)
- Eliminate the vanishing gradient problem.
- $GELU(x) = x * \Phi(x)$ where $\Phi(x)$ is the standard Gaussian cumulative distribution function



Generative Pretrained Transformers (GPTs)

Language Models

Unsupervised learning GPT steps:

Input embedding

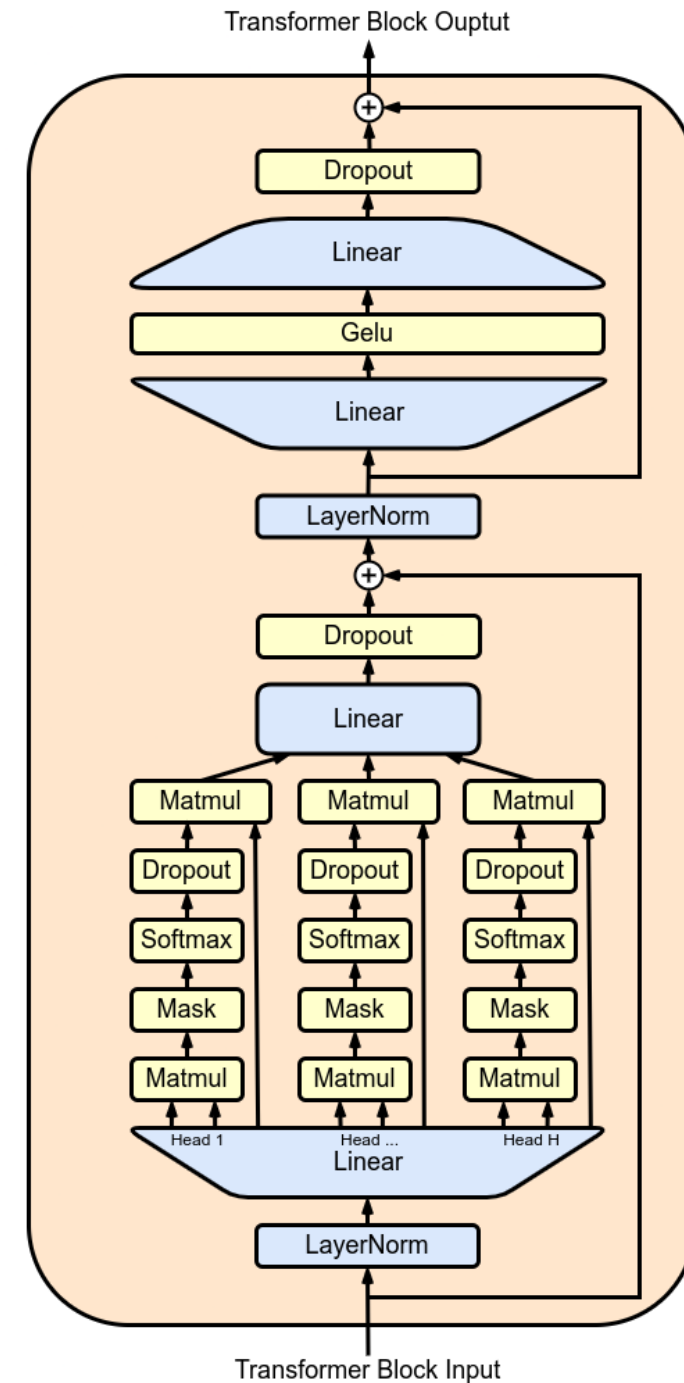
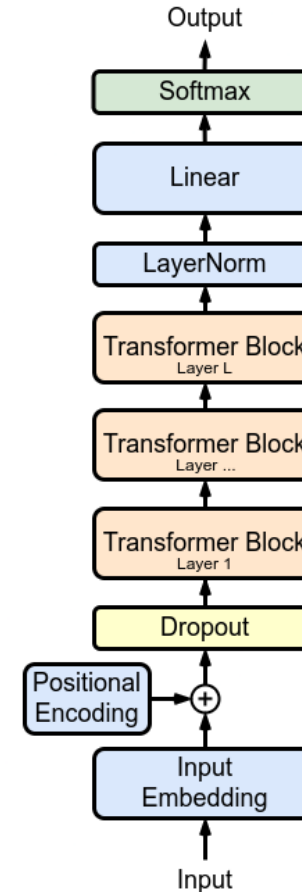
Positional encoding: give the model information about the order of the tokens in the sequence

Multi-head self-attention: each head focuses on a different aspect of the input sequence

Feedforward layer: non-linear transformation

Normalization

Output generation: The output sequence is generated one token at a time, with each token being generated based on the previously generated tokens



Generative Pretrained Transformers (GPTs)

Language Models

Training Steps

1. Unsupervised pre-training with LM objective function predict the next token based on the **context** provided by the **preceding tokens**

$$L_1(T) = \sum_i \log P(t_i | t_{i-k}, \dots, t_{i-1}; \theta)$$

T: set of tokens in unsupervised data $\{t_1, \dots, t_n\}$

K: size of context window

θ : Network parameters

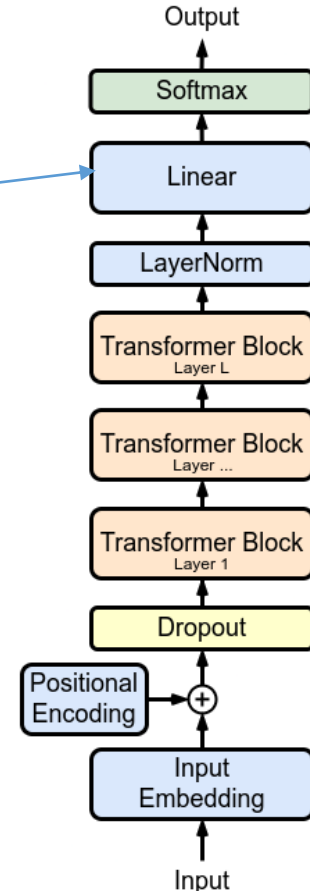
2. Supervised fine tuning for specific mission: Translation, answering questions, classifications, etc.

The aim here is to maximize the likelihood of observing label Y given tokens (x_1, \dots, x_n)

$$L_2(C) = \sum_{x,y} \log P(y | x_1, \dots, x_n)$$

C: Labeled dataset of training examples (x,y)

Supervised fine-tuning was achieved by adding a linear and a softmax layer to the transformer model to get the task labels for downstream tasks.



auxiliary learning objective

$$L_3(C) = L_2(C) + \lambda L_1(C)$$

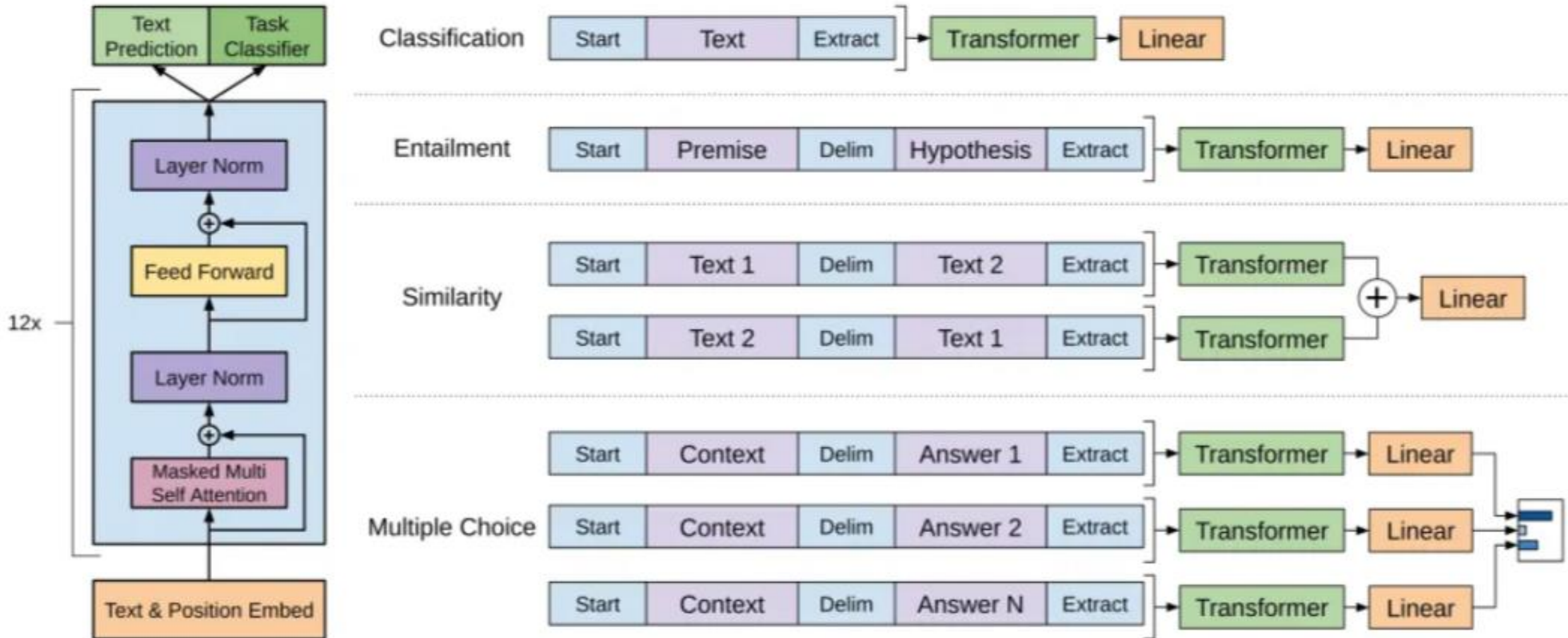
faster convergence

Generative Pretrained Transformers (GPTs)

Language Models

Training Steps

3. Task specific Input Transformations
fine-tune the model for applications like text classification



Generative Pretrained Transformers (GPTs)

Language Models

Zero-shot performance of GPT-1

GPT-1 output can be directly used on various tasks like question answering and sentiment analysis.

Generative Pretrained Transformers (GPTs)

GPT-2 paper (Language Models are unsupervised multitask learners)

Released in 2019 by OpenAI

Modifications against GPT-1

1. GPT-2 model mainly involved utilising a larger dataset and more parameters to the model
2. 1.5 billion parameters and 10 times the data of GPT-1
3. Learning objective is changed into:

$P(\text{output} | \text{input}, \text{task})$

Task conditioning is the process of teaching a model to generate several outputs for a given input depending on the task

Generative Pretrained Transformers (GPTs)

GPT-2 paper (Language Models are unsupervised multitask learners)

Zero Shot Task Transfer (GPT-2)

1. GPT-2 has the capacity to transmit zero-shot tasks, which means that it can carry out a task without being given any examples.
2. When there are no examples offered and just instructions are given, zero-shot learning is a special instance of zero-shot task transfer.

To translate from English to French, the model was given an English sentence followed by the word “French” and a prompt (:), and it was expected to understand that it was a translation task and provide the French counterpart of the English sentence.

Language is the mind of world **“French”:**



La langue est l'esprit du monde

Generative Pretrained Transformers (GPTs)

GPT-3 paper (Language models are few-shot learners)

It is a sizable language prediction and generation model created by OpenAI in 2020 that can produce *lengthy passages* of the original text.

GPT-3 can create sentences and paragraphs that essentially sound like they were written by a person.

GPT-3 is 100 times larger than GPT-2 and has 175 billion parameters. It was trained using the “Common Crawl” 500 billion word data collection.

GPT-3 is 100 times larger than GPT-2 and has 175 billion parameters. It was trained using the “Common Crawl” 500 billion word data collection.

GPT-3 is capable of carrying out additional intelligent activities and writing code snippets like SQL queries

Unfortunately, because of its 175B-parameter size, performing inference is expensive and inconvenient.

No fine tuning is required: It was trained on only **one specific task** which was predicting the next word, and thus, is an unsupervised pre-trained model.

Generative Pretrained Transformers (GPTs)

GPT-3 paper (Language models are few-shot learners)

How to train GPT-3 (we need a large text dataset)

Text: Second Law of Robotics: A robot must obey the orders given it by human beings

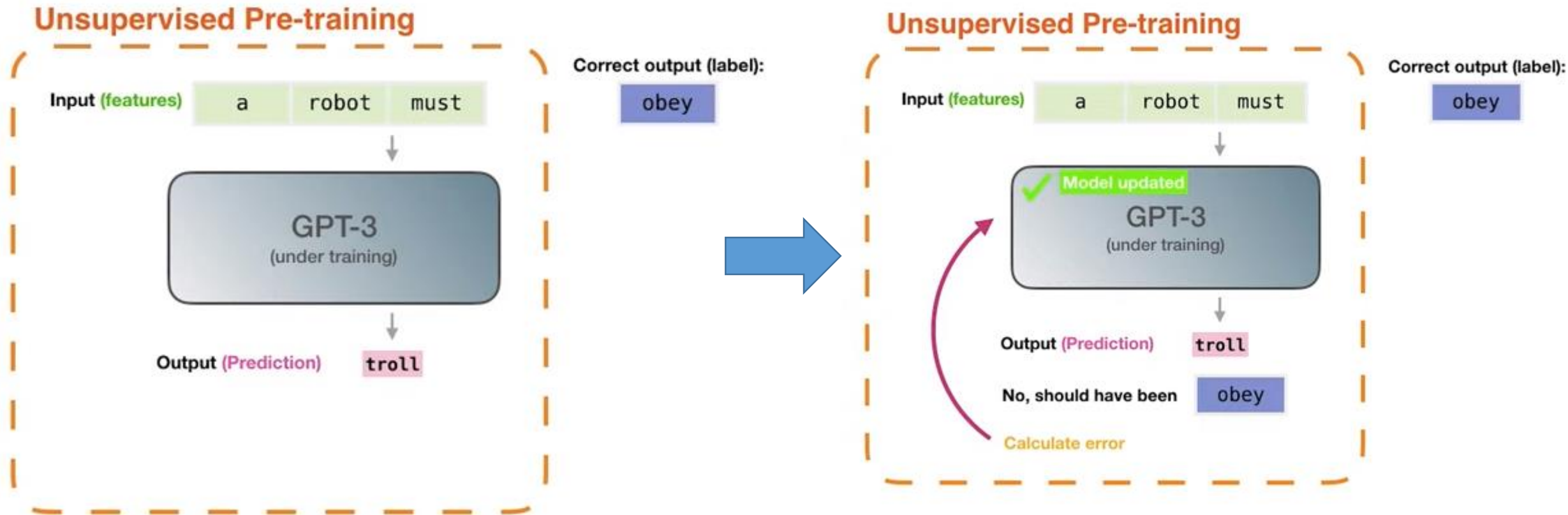


Generated training examples

Example #	Input (features)	Correct output (labels)
1	Second law of robotics :	a
2	Second law of robotics : a	robot
3	Second law of robotics : a robot	must
...		

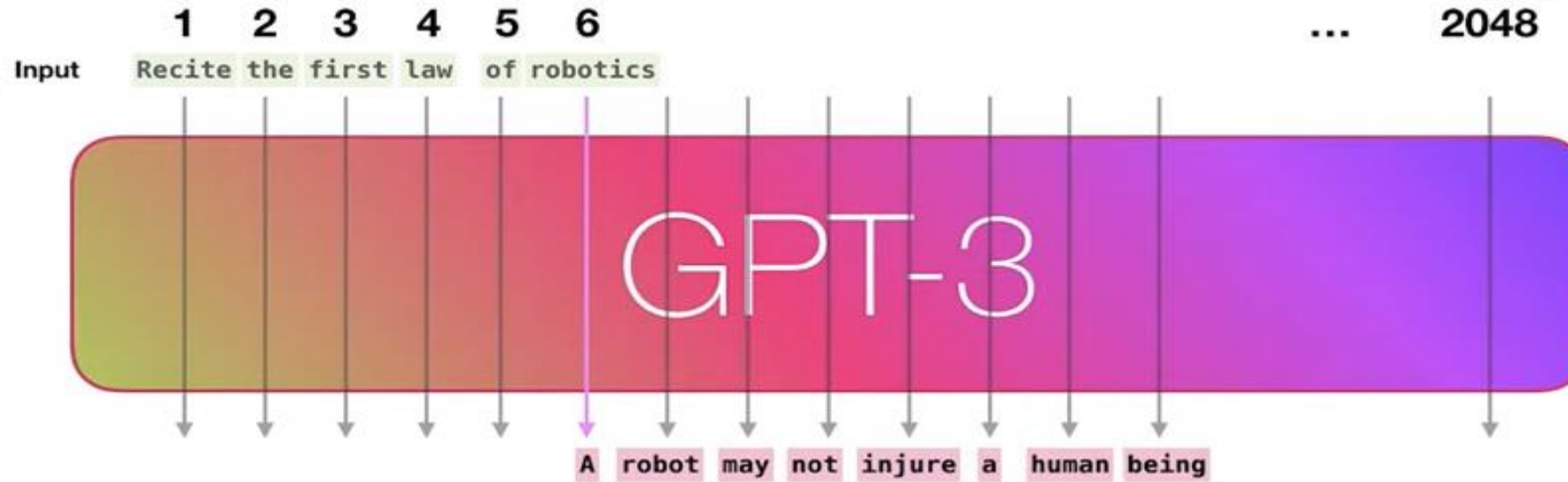
Generative Pretrained Transformers (GPTs)

GPT-3 paper (Language models are few-shot learners)



Generative Pretrained Transformers (GPTs)

GPT-3 paper (Language models are few-shot learners)



Generative Pretrained Transformers (GPTs)



ChatGPT

ChatGPT was trained on data from Wikipedia, books, websites scientific journals, etc. (until September 2021)
ChatGPT generates different responses for the same question each time (using random techniques to re-arrange the words and keeping the same context).





In two sentences, explain quantum mechanics in very simple terms



Quantum mechanics is a branch of physics that deals with the behavior of tiny particles like atoms and electrons. It tells us that these particles can act like both waves and particles, and can be in many different states at the same time.  



Quantum mechanics is a branch of physics that studies the behavior of matter and energy at a very small scale, such as atoms and subatomic particles. It is characterized by phenomena such as superposition, entanglement, and wave-particle duality, which are not easily explained by classical physics.  

Was this response better or worse?



Better



Worse



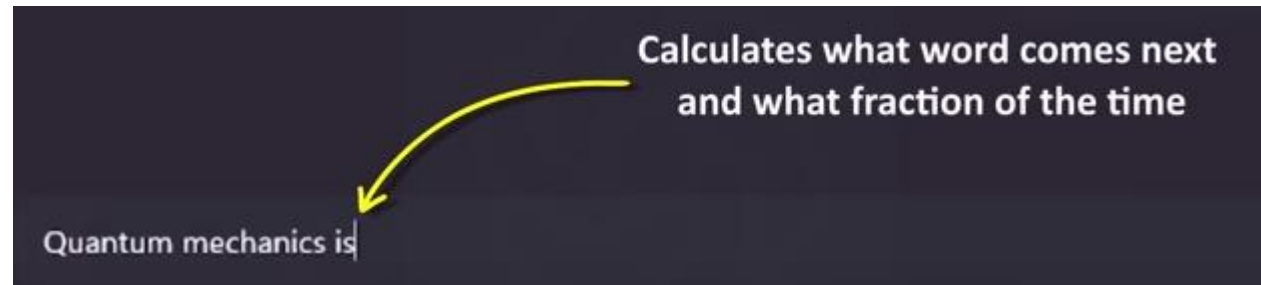
Same



Generative Pretrained Transformers (GPTs)

How can ChatGPT complete sentences?

ChatGPT Calculates what possible word comes next and what fraction of the time.
ChatGPT looks for matches in context and meaning.



Quantum mechanics is...	a	4.5%
Quantum mechanics is...	based	3.8%
Quantum mechanics is...	fundamentally	3.5%
Quantum mechanics is...	described	3.2%
Quantum mechanics is...	many	0.7%

Giving the generated text so far, what is the next word to be?

ChatGPT may not always generates the word with the highest probability meaning that each time it can select different words making the generated sentences different but tries to preserve the same meaning.
ChatGPT works not only on the “Word” level but also on the “sentence” and the “paragraph” level.

Generative Pretrained Transformers (GPTs)

Does ChatGPT always answer by completing sentences?

No, ChatGPT make a response according to the type of the text you enter (it can complete sentences, answer questions and response to orders)

Explain how quantum mechanics works



Explain how quantum mechanics works according to the Schrodinger Equation



This outputs is not by ChatGPT

How ChatGPT is trained?

Step 1: Generative Pre-Training

Step 2: Supervised Fine-Tuning

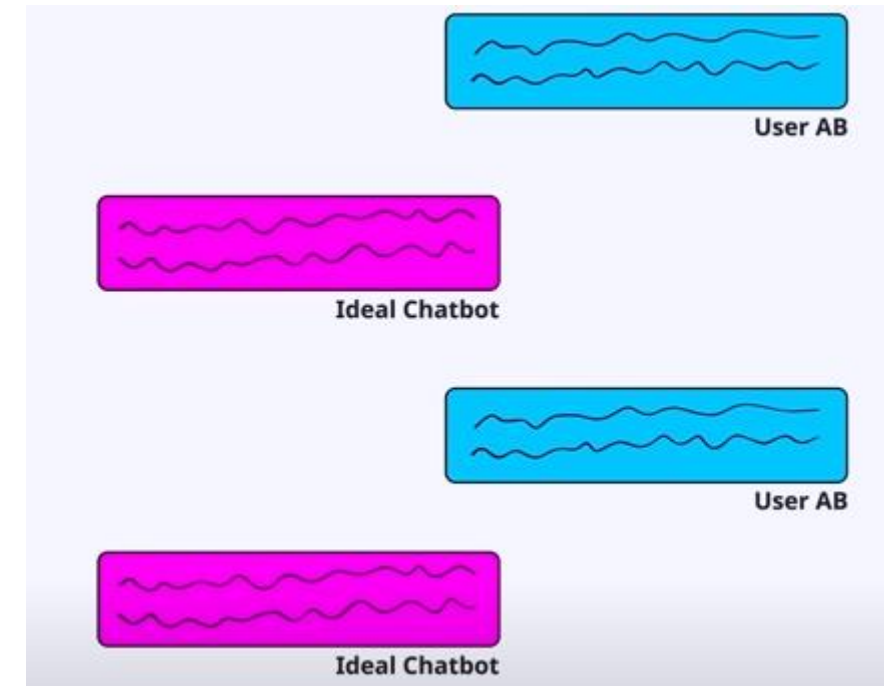
Step 3: Reinforcement Learning through Human Feedback

Generative Pretrained Transformers (GPTs)

Does ChatGPT always answer by completing sentences?

How ChatGPT is trained?

Step1: Generative Pre-Training
Train GPT on a huge dataset of chats between two sides (human and Chatbot) which are human-like conversations (this process is supervised learning one)



How ChatGPT is trained?

Step 2: Supervised Fine-Tuning

Test ChatGPT of step1 with different questions/orders/sentences etc. and get different predictions.

Each response (prediction) is evaluated and given a score by human annotator.

Teaching ChatGPT to evaluate what the best output is likely to be.

Describe an atom

- A) It's the smallest part of a substance made of electrons, neutrons, and protons. **0.95**
- B) It's a basic chemical element **0.37**
- C) It's an object made of subatomic particles **0.48**
- D) It's a ticketing service **0.12**

A > C > B > D



ChatGPT Model

How ChatGPT is trained?

Step 3: Reinforcement Learning through Human Feedback.
(Unsupervised learning step) based on the human-taught ranking system (step2).
It allows the ChatGPT model to learn pattern from much more data and many more subjects (by itself!).

Describe an atom

- A) It's the smallest part of a substance made of electrons, neutrons, and protons. **0.95**
- B) It's a basic chemical element **0.37**
- C) It's an object made of subatomic particles **0.48**
- D) It's a ticketing service **0.12**

A > C > B > D



ChatGPT Model

Generative Pretrained Transformers (GPTs)

Does ChatGPT always answer by completing sentences?

