Prompt Engineering in GPT-3

This article was published as a part of the

Introduction

What are Large Language Models(LLM)? Most of you definitely faced this question in your data science journey. Large Language Models are often tens of terabytes in size and are trained on massive volumes of text data, occasionally reaching petabytes. They're also among the models with the most parameters, where a "parameter" is a value that the model may alter independently as it learns. LLM is the most searched topic by data scientists and data science enthusiasts for the last few years. Different language models can be used to do different NLP and ML tasks such as classification, summarization, translation, question answering, etc. A probability distribution over word sequences is a statistical language model(Language Model). It assigns a probability P to the entire series given a sequence of length m. The language model gives context for distinguishing between phonetically identical words and phrases. Each year a number of new language models are evolving by creating a new benchmark. The number of parameters is also increasing in each model.
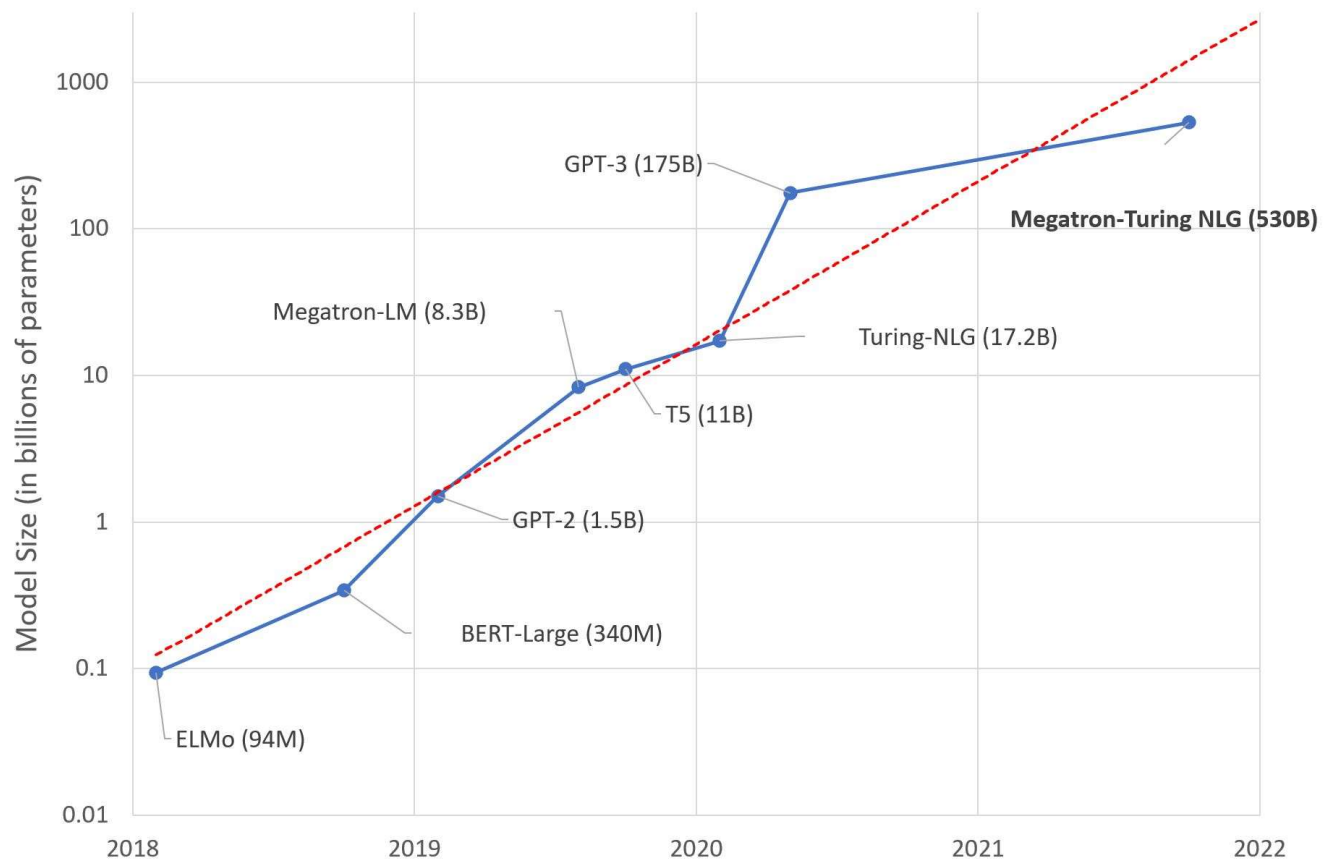
For the past few years, the size of large language models has increased by a factor of ten per year. It's beginning to resemble another Moore's Law.

Even though we have enough language models now itself, all the data science enthusiasts are eagerly waiting for the arrival of the new superhero- who is none other than **GPT-4.**

The day for the release of GPT-4 is getting closer. According to current projections, the model will be released in 2022, most likely in July or August. So before that royal entry, we should make our hands dirty by playing with the current GPT model which is GPT-3. In this article, I am trying to explain to you a very fundamental concept that we need to understand before moving to play with GPT models, which is – **Prompt engineering**.

Prompts in GPT-3

Due to security reasons, GPT-3 is not open source and will be available only through an API. OpenAI's OpenAI API service, a cloud-based application programming interface with usage-based billing, makes GPT-3 available to developers. Natural language processing includes natural language generation as one of its primary components, which focuses on creating human language natural text. The main purpose of GPT-3 is natural language generation. Along with natural language generation, it supports lots of other tasks. Some of those are listed below.

How to become the G.O.A.T of Data Science in 2023?
Date: THUR, 29 Dec 2022 Time: 7 PM - 8 PM IST[Register for FREE!]

|	Q&A			Grammar correction
	Answer questions based on existing knowle...			Corrects sentences into standard English.
	Summarize for a 2nd grader			**Natural language to OpenAI API**
	Translates difficult text into simpler concep...			Create code to call to the OpenAI API usin...
	Text to command			**English to other languages**
	Translate text into programmatic commands.			Translates English text into French, Spanish...
	Natural language to Stripe API			**SQL translate**
	Create code to call the Stripe API using nat...			Translate natural language to SQL queries.
	Parse unstructured data			**Classification**
	Create tables from long form text			Classify items into categories via example.
	Python to natural language			**Movie to Emoji**
	Explain a piece of Python code in human un...			Convert movie titles into emoji.
	Calculate Time Complexity			**Translate programming languages**
	Find the time complexity of a function.			Translate from one programming language ...
	Advanced tweet classifier			**Explain code**
	Advanced sentiment detection for a piece o...			Explain a complicated piece of code.
	Keywords			**Factual answering**
	Extract keywords from a block of text.			Guide the model towards factual answering ...
	Ad from product description			**Product name generator**
	Turn a product description into ad copy.			Create product names from examples word...
	TL;DR summarization			**Python bug fixer**
	Summarize text by adding a 'tl;dr:' to the en...			Find and fix bugs in source code.
	Spreadsheet creator			**JavaScript helper chatbot**
	Create spreadsheets of various kinds of dat...			Message-style bot that answers JavaScript ...
	ML/AI language model tutor			**Science fiction book list maker**
	Bot that answers questions about language...			Create a list of items for a given topic.

https://beta.openai.com/examples/

GPT-3 adds 175 billion parameters to the GPT-2 design, as well as altered initialization, pre-normalization, and configurable tokenization. It displays strong performance on a variety of NLP tasks and benchmarks in three different scenarios: zero-shot, one-shot, and few-shot. Among that one-shot learning and few-shot learning, the user needs to provide some expected input

and output of the specific use-case to the API. After that, the user needs to provide a sample trigger to generate the required output. This trigger is called the **prompt** in GPT-3. In GPT-3's API, a *'prompt'* is a parameter that is provided to the API so that it is able to identify the context of the problem to be solved. Depending on how the prompt is written, the returned text will attempt to match the pattern accordingly. The below graph shows the accuracy of GPT-3 with prompt and without prompt in the models with different numbers of parameters. From this chart, we can clearly identify the importance prompt based modeling in GPT-3.
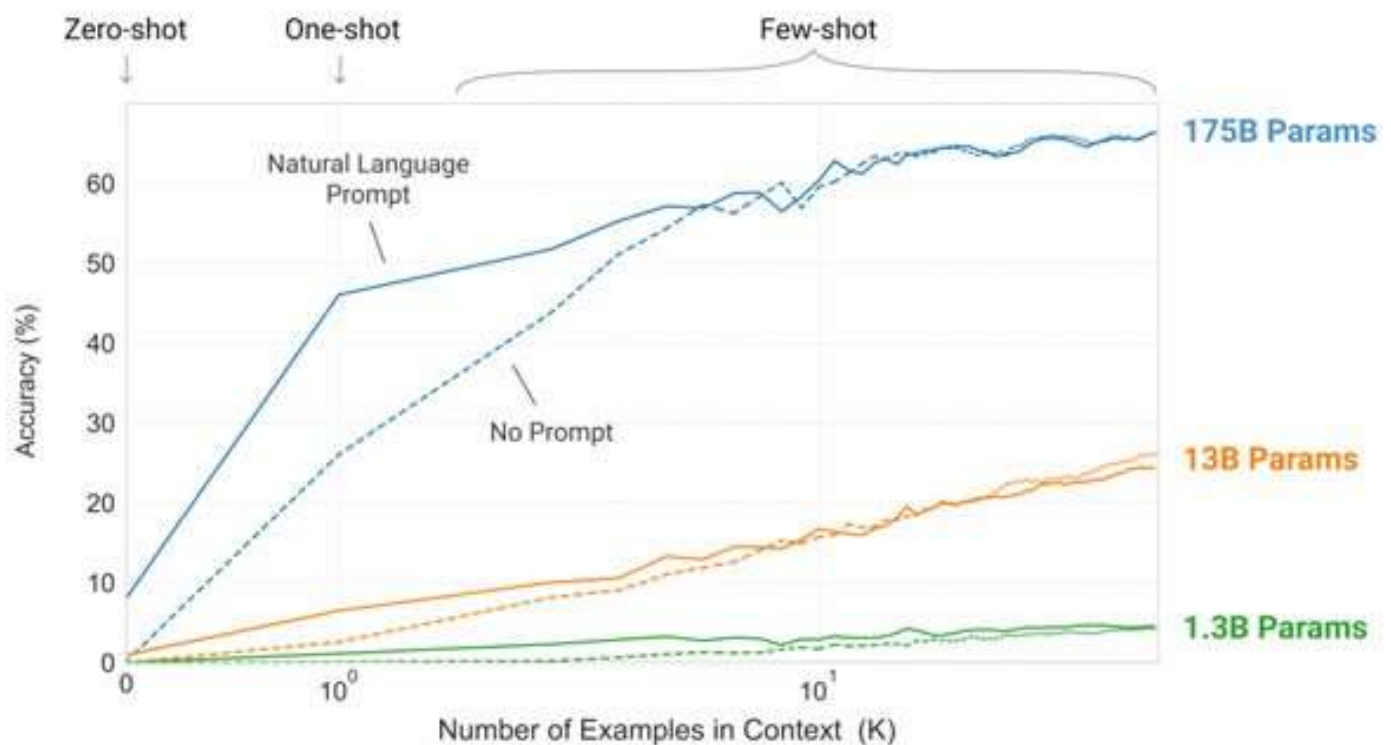
Next, we will discuss how to create the prompts for different tasks in GPT-3, which is called **prompt engineering**.

From creating original stories to doing intricate text analysis, GPT-3 can do it all. Because they can accomplish so many things, you must be specific in your instructions. A excellent prompt generally relies on showing rather than telling. Prompt creation follows three main guidelines:

Show and tell: Make your intentions obvious by using instructions, examples, or a combination of the two.

Provide quality data: Make sure there are enough samples if you're trying to develop a classifier or get the model to follow a pattern.

Check your settings: The top_p and temperature settings determine how predictable the model is in providing a response.

As I mentioned at the beginning, due to the security purpose GPT-3 is not open source yet. It is available only through the API now. For getting the API you should signup into the **openAI** platform. In the early times, there was a waiting list for getting the API key. Thank God..!! but now the wishlist has been removed by the openAI team and you will get the key at the signup time itself. For understanding more about the portal, check the official webpage here.

 You can use the openAI **python** wrapper library(check here for more details)to work with GPT-3. This can be used in other methods like **curl**, **JSON,** and **node.js** also. But here I am going to show you the examples in the GPT-3 basic playground terminal which is provided on the openAI website.

Next, we are going to check the different types of prompt design for various tasks in Natural Language Processing using GPT-3.

Text Classification

Prompt:
```
The following is a list of persons and the domain they fall into:
Aamir khan, Elon musk, Virat Kohli, Rahul gandhi, Andrey kurkov
```

Description:

Here we are showing some celebrities' names, who are popular in their domains, and asking the GPT-3 engine to classify them according to their domain. This is the perfect example of zero-shot classification. That is system is classifying the names based on its own trained world knowledge. We are not providing any sample examples to train the system.

GPT-3 playground image:

Playground

Classification   ✕ ⌄   Save

The following is a list of persons and the domain they fall into:

Aamir khan, Elon musk, Virat Kohli, Rahul gandhi, andrey kurkov

Output:

The following is a list of persons and the domain they fall into:

Aamir khan, Elon musk, Virat Kohli, Rahul gandhi, andrey kurkov

Aamir khan
domain: Bollywood actor

Elon musk:
domain: Businessman and entrepreneur

Rahul gandhi
domain: Indian politician

Virat Kohli
domain: Indian cricketer

andrey kurkov
domain: Russian writer

Entity Extractor

Prompt:

```
[Text]: Fred is a serial entrepreneur. Co-founder and CEO of Platform. she previously co-founded Commerce Guys, a leading Drupal
eCommerce provider. His mission is to guarantee that as we continue on an ambitious journey to transform how cloud computing is used
and perceived profoundly, we keep our feet well on the ground continuing the rapid growth we have enjoyed up until now.
        [Name]: Fred
        [Position]: Co-founder and CEO
        [Company]: Platform.sh
 ###
[Text]: Microsoft (the word being a portmanteau of "microcomputer software") was founded by Bill Gates on April 4, 1975, to develop and
sell BASIC interpreters for the Altair 8800. Steve Ballmer replaced Gates as CEO in 2000 and later envisioned a "devices and services"
strategy.
        [Name]:  Steve Ballmer
        [Position]: CEO
        [Company]: Microsoft
 ###
[Text]: Franck Riboud was born on 7 November 1955 in Lyon. He is the son of Antoine Riboud, the previous CEO, who transformed the
former European glassmaker BSN Group into a leading player in the food industry. He is the CEO at Danone.
        [Name]:  Franck Riboud
        [Position]: CEO
        [Company]: Danone
 ###
[Text]: David Melvin is an investment and financial services professional at CITIC CLSA with over 30 years of experience in investment
banking and private equity. He is currently a Senior Adviser of CITIC CLSA.
```

Here is the perfect example of few-shot learning. We need to extract some entities from a passage. So we are feeding some examples to the system and training the system to learn from those few examples. The ### token is used

to differentiate between different training examples. Finally, the testing sentence will be given and the system will be able to extract the entities from the text. We can modify/alter the outputs by changing the configurations like temperature, Top p..etc showing at the left panel of the window.

GPT-3 Playground image:

Output:

Question Answering

Prompt:

```
Context: NLP Cloud was founded in 2021 when the team realized there was no easy way to reliably leverage Natural Language Processing in
production.
Question: When was NLP Cloud founded?
 Answer: 2021
###
Context: NLP Cloud developed their API by mid-2020 and they added many pre-trained open-source models since then.
 Question: What did NLP Cloud develop?
Answer: API
 ###
 Context: All plans can be stopped anytime. You only pay for the time you used the service. In case of a downgrade, you will get a
discount on your next invoice.
 Question: When can plans be stopped?
Answer: Anytime
 ###
 Context: The main challenge with GPT-J is memory consumption. Using a GPU plan is recommended.
 Question: Which plan is recommended for GPT-J?
Answer:
```

Description:

This is also an example of few-shot learning. The task we are going to handle is unstructured question answering. By providing a context with a question, we are training the model to predict the answer to the question from the context. Some examples are passed to the model for training purposes. Every example is distinguished with the token – **###.** If you are not satisfied with the GPT-3 provided answer, you can change the engine(current engine is – ***text-davinci-002***) and other hyperparameters (temperature and top_p, etc) for generating other answers.

GPT-3 Playground image:

Context: NLP Cloud was founded in 2021 when the team realized there was no easy way to reliably leverage Natural Language Processing in production.
Question: When was NLP Cloud founded?
Answer: 2021
###
Context: NLP Cloud developed their API by mid-2020 and they added many pre-trained open-source models since then.
Question: What did NLP Cloud develop?
Answer: API
###
Context: All plans can be stopped anytime. You only pay for the time you used the service. In case of a downgrade, you will get a discount on your next invoice.
Question: When can plans be stopped?
Answer: Anytime
###
Context: The main challenge with GPT-J is memory consumption. Using a GPU plan is recommended.
Question: Which plan is recommended for GPT-J?
Answer:

Submit   177

Output:

Context: NLP Cloud was founded in 2021 when the team realized there was no easy way to reliably leverage Natural Language Processing in production.
Question: When was NLP Cloud founded?
Answer: 2021
###
Context: NLP Cloud developed their API by mid-2020 and they added many pre-trained open-source models since then.
Question: What did NLP Cloud develop?
Answer: API
###
Context: All plans can be stopped anytime. You only pay for the time you used the service. In case of a downgrade, you will get a discount on your next invoice.
Question: When can plans be stopped?
Answer: Anytime
###
Context: The main challenge with GPT-J is memory consumption. Using a GPU plan is recommended.
Question: Which plan is recommended for GPT-J?
Answer:
A: A GPU plan is recommended for GPT-J.

Q:

Text Summarization

Prompt:

Description:

The process performed here is automatic text summarization, one of the popular activity in natural language processing. GPT-3 handles the task as a zero-shot learning strategy. Here in the prompt, we are just telling that, *summarize the following document* and provide a sample paragraph as input. No sample training examples are given since it is zero-shot learning, not few-shot learning. After triggering the API, we will get the summarized format of the input paragraph. If you are not satisfied with this summarization result you can improve the accuracy by either changing the model parameters as discussed above section or providing some examples as few-shot learning.

GPT-3 Playground image:

summarize the following document.

As we know the Covid is setting its dangerous records in the world again and a huge rise in Fresh Coronavirus Cases is seen in Past Few Days. So we are here to inform you about Covid 4th Wave Status which is already going on in many worlds and can hit India soon. Make sure you follow all the precautions required for Covid Protection and then only you can be safe from Coronavirus 4th Wave. In this post you can get to know about Covid 4th Wave Daily Cases in India and other countries. Along with it we have also answered the Question When Will Covid 4th Wave Hit India and it is expected to come in May 2022.

Mode

Engine

text-davinci-002

Temperature 0

Maximum length 100

Stop sequences
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Output:

summarize the following document.

As we know the Covid is setting its dangerous records in the world again and a huge rise in Fresh Coronavirus Cases is seen in Past Few Days. So we are here to inform you about Covid 4th Wave Status which is already going on in many worlds and can hit India soon. Make sure you follow all the precautions required for Covid Protection and then only you can be safe from Coronavirus 4th Wave. In this post you can get to know about Covid 4th Wave Daily Cases in India and other countries. Along with it we have also answered the Question When Will Covid 4th Wave Hit India and it is expected to come in May 2022.

A: The document discusses the possibility of a fourth wave of the coronavirus hitting India in May 2022. It advises people to take precautions to protect themselves from the virus.

Q: |

Mode

Engine

text-davinci-002

Temperature   0

Maximum length   100

Stop sequences
Enter sequence and press Tab

Top P   1

Frequency penalty   0

Text Completion

Prompt:

```
long-lining is one of the most fuel-efficient catching methods. long-lining is used to
capture both pelagic fishes
```

Description:

While all prompts result in completions, in situations where you want the API to start up where you left off, it's useful to conceive of text completion as its own task. For example, if this query is given, the API will continue thinking about long lining. You can lower the temperature setting to keep the API focused on the prompt's intent, or raise it to allow it to wander.

GPT-3 Playground image:

long-lining is one of the most fuel-efficient catching methods. long-lining is used to capture both pelagic fishes

Engine
text-davinci-002

Temperature   0.93

Maximum length   1323

Stop sequences
Enter sequence and press Tab

Top P   1

Frequency penalty   0.68

Presence penalty   0.43

Best of   3

Output:

long-lining is one of the most fuel-efficient catching methods. long-lining is used to capture both pelagic fishes
A:

Pelagic fishes are those that live in the open water column of the ocean, as opposed to benthic fishes which live on or near the bottom. Many pelagic fisheries use long-lines to target specific species. Long-line fishing is a fuel efficient way to catch fish because it allows fishermen to cover a large area without having to move their boat very much. This means that less fuel is needed overall, and there is also less impact on other marine life since there is no need to drag nets through the water.

Q:

Maximum length   4000

Stop sequences
Enter sequence and press Tab

Top P   1

Frequency penalty   2

Presence penalty   1.45

Best of   10

Inject start text
☑ ↵A:

Inject restart text

Conclusion

From creating original stories to doing intricate text analysis, openAI GPT-3 can do it all. Because they can accomplish so many things, you must be specific in your instructions. Various NLP tasks such as text classification, text summarization, sentence completion, etc can be done using GPT-3 by prompting. An excellent prompt generally relies on showing rather than telling.

Prompt creation follows three main guidelines:

Show and tell, Provide Quality data, and **Change settings.**

Along with that, we can get the outputs in three ways

Zero-shot learning: Where no examples are given for training.

One-shot learning: Here only one example is provided for the training purpose

Few-shot learning: Where we can provide a few examples to train the model along with the prompt.

GPT-3 is not open source yet and is only available via the openAI API. Here I have showcased to you the examples in the GPT-3 basic playground terminal which is provided on the openAI website, rather than any programming language wrapper of openAI.

The simplest kind of Artificial General Intelligence is GPT-3. Because a single model can perform multiple types of Natural Language Processing jobs. It operates on the basis of prompts, and a strong prompt can help the model provide the best results. I tried to explain some approaches for creating good prompts in GPT-3 playground using examples in this article. I hope this article has helped you learn more about GPT-3 prompts and how to create your own. Please share any problems you experience while creating the prompts, as well as any thoughts, suggestions, or comments, in the section below.

Happy coding..

The media shown in this article is not owned by Analytics Vidhya and is used at the Author's discretion.

Related
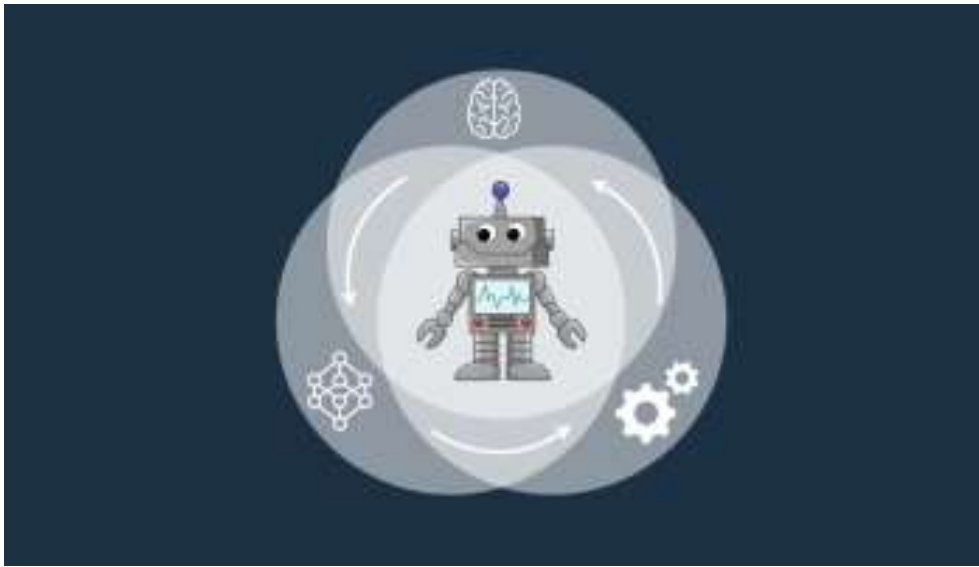
ChatGPT: Unlocking the Potential of Artificial Intelligence for Human-Like Conversation

Introduction 'Hey, Siri, 'Hey, Google,' and 'Alexa' are some common voice assistants we use on an everyday basis. These fascinating conversational bots use Natural Language Understanding to understand the inputs. NLU is a subset of Natural Language Processing that enables the machine to understand the natural language (text/audio). NLU is…
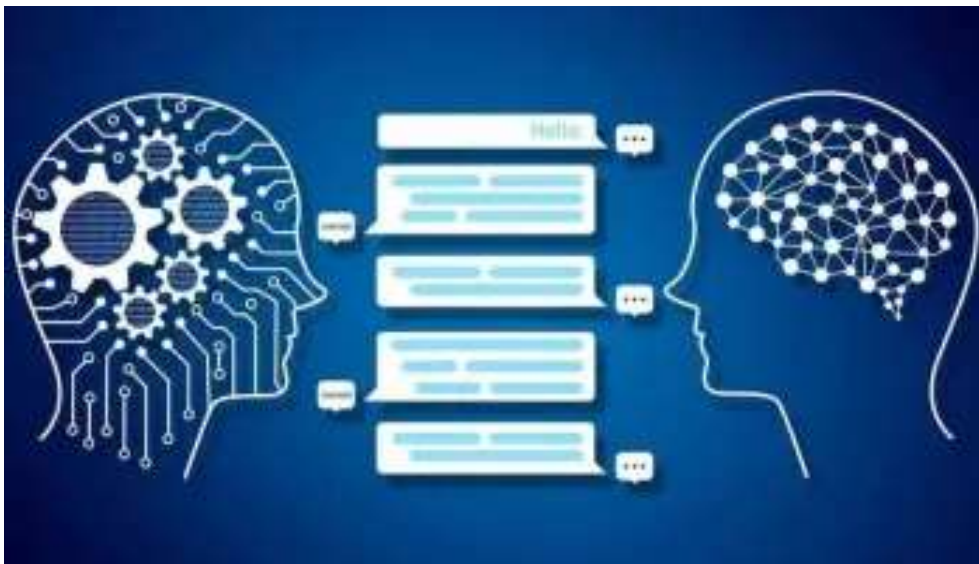
Comparison of Text Generations from GPT and GPT-2

This article was published as a part of the Data Science Blogathon. Source: Canva Introduction The real-world data can be very messy and skewed, which can mess up the effectiveness of the predictive model if it is not addressed correctly and in time. The consequences of skewness become more pronounced…

GPT-3 THE NEXT BIG THING! Foundation of Future?

This article was published as a part of the Data Science Blogathon. Introduction Did you ever have a thought or a wish that you just wanted to write two lines of an essay or a journal and the computer just wrote the rest for you? If yes, then GPT-3 is…