DCBD Presentation: ChatGPT and Big Data

Aniket Saha
MDS202207

Chennai Mathematical Institute

April 3, 2023

Table of Contents

CHENNAI MATHEMATICAL INSTITUTE

Why Did We Choose ChatGPT?

- Firstly, it is a **Hot Topic** at this moment.
- ChatGPT arrived like a raging storm. It gathered a **million users** within 5 days of it's release.
  The closest rival to that stat is Instagram, which took 2 months to gather a million users.
- It will be hugely useful in the near future while handling many **machine learning**, **NLP** and **big data** related problems.
- It is useful for research and studying.
- **So that we can take help from ChatGPT itself.**

CHENNAI MATHEMATICAL INSTITUTE

Overview

We will try to answer the following questions in our presentation.

- What is ChatGPT?
- How did ChatGPT come to existence?
- How does ChatGPT work? What is it's backstory?
- How can we use this technology? What are the drawbacks?
- How are ChatGPT and Big Data related?
- If possible, how can we use ChatGPT to handle Big Data related problems?
- What are the future directions for ChatGPT research?

CHENNAI MATHEMATICAL INSTITUTE

In short, ChatGPT is a computer program that can understand human language and generate responses like a human would.

Figure 1: ChatGPT Logo

Well, this is not enough. Let us know more about this technology.

CHENNAI MATHEMATICAL INSTITUTE

Introduction (Contd.)

What is ChatGPT?

- ChatGPT is a **deep learning-based language model** that utilizes a **transformer architecture** for natural language processing.
- It was trained on a massive amount of data, enabling it to generate human-like responses to textual inputs.
- The architecture includes multiple layers of **self-attention mechanisms** that allow the model to analyze input sequences in a parallelized manner, resulting in its remarkable speed and efficiency.
- ChatGPT's implementation is distributed across a cluster of servers that are managed by **load balancers** and utilize optimized algorithms to maximize the use of available computing resources.
- The infrastructure also incorporates **fault-tolerance mechanisms** and **automated scaling** capabilities to ensure the system can handle increased demand.

CHENNAI MATHEMATICAL INSTITUTE

Introduction (Contd.)

What is ChatGPT?

- Overall, ChatGPT represents a significant advancement in the field of artificial intelligence and has numerous potential applications in various industries.
- Now let us know more about the highlighted terms in the previous slide.

CHENNAI MATHEMATICAL INSTITUTE

Introduction

Important Terms

- **Deep learning-based language model**[1]
  This is a type of artificial intelligence (AI) model that can understand and generate human language.
- **Transformer architecture**[2]
  A type of artificial neural network that is often used in NLP tasks.
- **Load balancers**[3]
  Load balancer is a device or software that distributes incoming network traffic across multiple servers.
- **Fault-tolerance mechanisms**[4]
  These are the techniques and methods used to ensure that a system continues to operate in the event of a failure or fault.
- **Automated scaling**[5]
  Automated scaling is the ability of a system to automatically adjust its resources based on the current workload.

Introduction

Important Terms

- **Deep learning-based language model**[1]
  This is a type of artificial intelligence (AI) model that can understand and generate human language.
- **Transformer architecture**[2]
  A type of artificial neural network that is often used in NLP tasks.
- **Load balancers**[3]
  Load balancer is a device or software that distributes incoming network traffic across multiple servers.
- **Fault-tolerance mechanisms**[4]
  These are the techniques and methods used to ensure that a system continues to operate in the event of a failure or fault.
- **Automated scaling**[5]
  Automated scaling is the ability of a system to automatically adjust its resources based on the current workload.

So now we have a basic understanding of what ChatGPT is.

Introduction (Contd.)

Origin of ChatGPT

ChatGPT is a language model developed by **OpenAI**, an AI research organization founded in December 2015 by several high-profile tech executives including **Elon Musk**, **Sam Altman**, **Greg Brockman**, **and Ilya Sutskever**.

Figure 2: OpenAI Logo. Source

An Interesting Fact!

After the release of ChatGPT, OpenAI's valuation was estimated at US $29 billion. (It was $14 billion a year ago)

Introduction (Contd.)

Origin of ChatGPT

- ChatGPT was launched in November 2022.[0]
- The goal was to develop a large-scale language model that could generate coherent and natural-sounding text.
- It's based on the **GPT(Generative Pre-trained Transformer)** architecture.
- It is pre-trained on large amounts of **text** and **conversational** data.

Alright, enough about what ChatGPT is. Let's move on to more interesting stuff!

[0]In March 2023, OpenAI released GPT-4. Hence we get an improved version of ChatGPT.
However, this is currently only accessible to limited people.

CHENNAI MATHEMATICAL INSTITUTE

Architecture of ChatGPT

Behind the Scenes

Now we will try to ***dive*** deep into ChatGPT's architecture and try to ***dig*** up more information.

- First of all, we have seen that ChatGPT is based on **GPT architecture**. What is that? How does it work? We will find out next.
- ChatGPT is a **Large Language Model**.
- It is built on top of OpenAI's GPT-3.5 and GPT-4 families of large language models(LLMs) and has been fine-tuned using both **supervised** and **reinforcement learning** techniques.
- We will also look at the **training data** that was used to build ChatGPT.

CHENNAI MATHEMATICAL INSTITUTE

Architecture of ChatGPT (Contd.)

What is GPT Architecture?

- The GPT (Generative Pre-trained Transformer) architecture is a type of deep learning-based language model that was first introduced by OpenAI in 2018, as a part of their GPT-2 project.
- It is a neural network architecture that is pre-trained on massive amounts of text data and can generate text that is similar in style and content to the text it was trained on.
- It is based on the **Transformer architecture**[2], which was introduced by Vaswani et al. in 2017.

CHENNAI MATHEMATICAL INSTITUTE

Now, ChatGPT is built on GPT-3 or more advanced versions. We can see the difference between GPT-2 and GPT-3 architectures below.

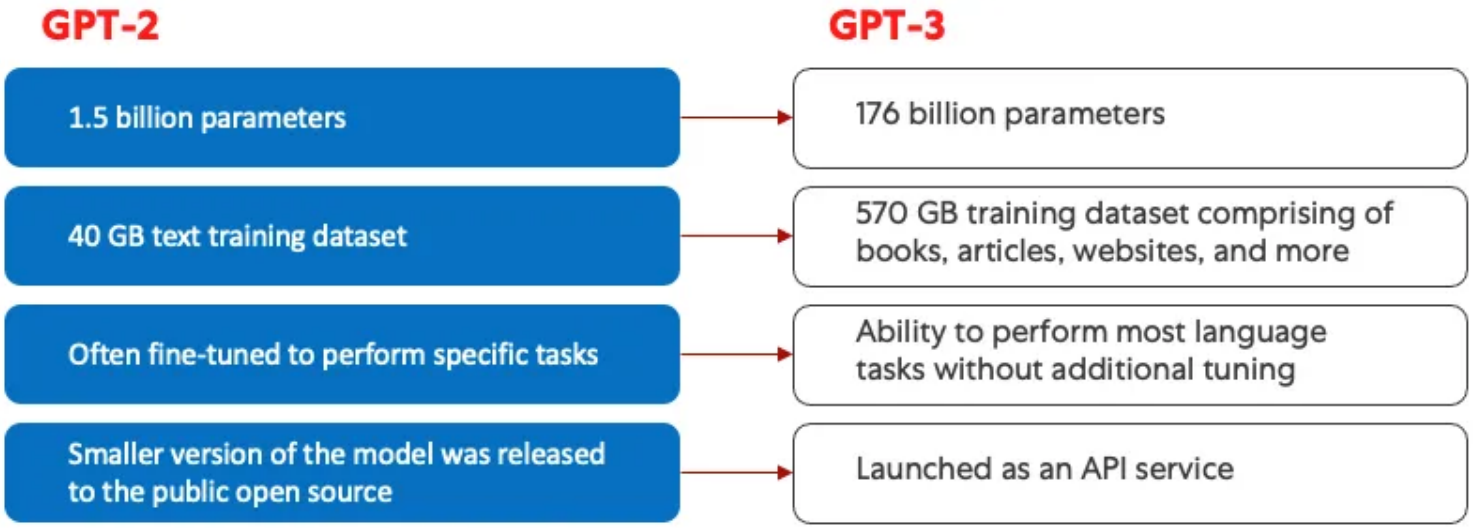

Figure 3: GPT-2 & GPT-3. Source

CHENNAI MATHEMATICAL INSTITUTE

Architecture of ChatGPT (Contd.)

Comments

Although GPT architecture is quite advanced, it is not free of limitations.

- **Large Training Data Requirements:** GPT models require a large amount of training data to learn and generate meaningful responses.
- **Limited Context:** Even though GPT models can generate coherent and meaningful responses, they still lack the ability to maintain long-term context across multiple turns of a conversation.
- **Bias:** GPT models are prone to biases in the training data, which can lead to biased or insensitive responses.
- **Lack of Explainability:** GPT models are often described as black boxes, meaning that it can be difficult to understand how they generate their responses. This lack of transparency can make it challenging to diagnose and correct errors or biases in the model.

Training Data for ChatGPT

- The training data for ChatGPT consists of a diverse range of text sources, including web pages, books, and news articles.
- The total size of the dataset is over 45 terabytes, making it one of the largest language models in existence.
- However, all of the 45 terabytes of data was not used to train ChatGPT, only 570 gigabytes of data was used after filtering the original dataset.
- Processing this large amount of data to build ChatGPT required a lot of computing power.
- ChatGPT initially used a **Microsoft Azure** supercomputing infrastructure, powered by **Nvidia GPUs**, that Microsoft built specifically for OpenAI.

CHENNAI MATHEMATICAL INSTITUTE

Training Data for ChatGPT (Contd.)

The following image shows the data sources, from where the training data for building ChatGPT was collected.

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Table 2.2: Datasets used to train GPT-3. "Weight in training mix" refers to the fraction of examples during training that are drawn from a given dataset, which we intentionally do not make proportional to the size of the dataset. As a result, when we train for 300 billion tokens, some datasets are seen up to 3.4 times during training while other datasets are seen less than once.

Figure 4: Training Data Origin. Source

- Apart from these datasets, ChatGPT also used some curated datasets which help it perform better, that is, to better understand the input from context and provide relevant response.

Advantages of ChatGPT

There are many many advantages of having a versatile chatbot like ChatGPT around. We have listed some of them below.

- Can generate **realistic** and **coherent** language responses.
- Requires minimal pre-processing of input data.
- ChatGPT can generate responses in **multiple languages**.
- It can **write** and **debug** computer programs.
- It can **compose** stories, essays, music etc.
- Can play games like **Tic-Tac-Toe**.
- **High accuracy** in natural language processing tasks.
- ChatGPT also can **answer test questions**, sometimes at a higher level than an average human being.
- It continues to learn from its interactions and adapt to the language patterns of its users, improving its accuracy over time.

CHENNAI MATHEMATICAL INSTITUTE

Limitations of ChatGPT

ChatGPT is a very sophisticated computer program. However it still has a lot of limitations. We will go over some of them now.
Here is a screenshot from OpenAI's website.

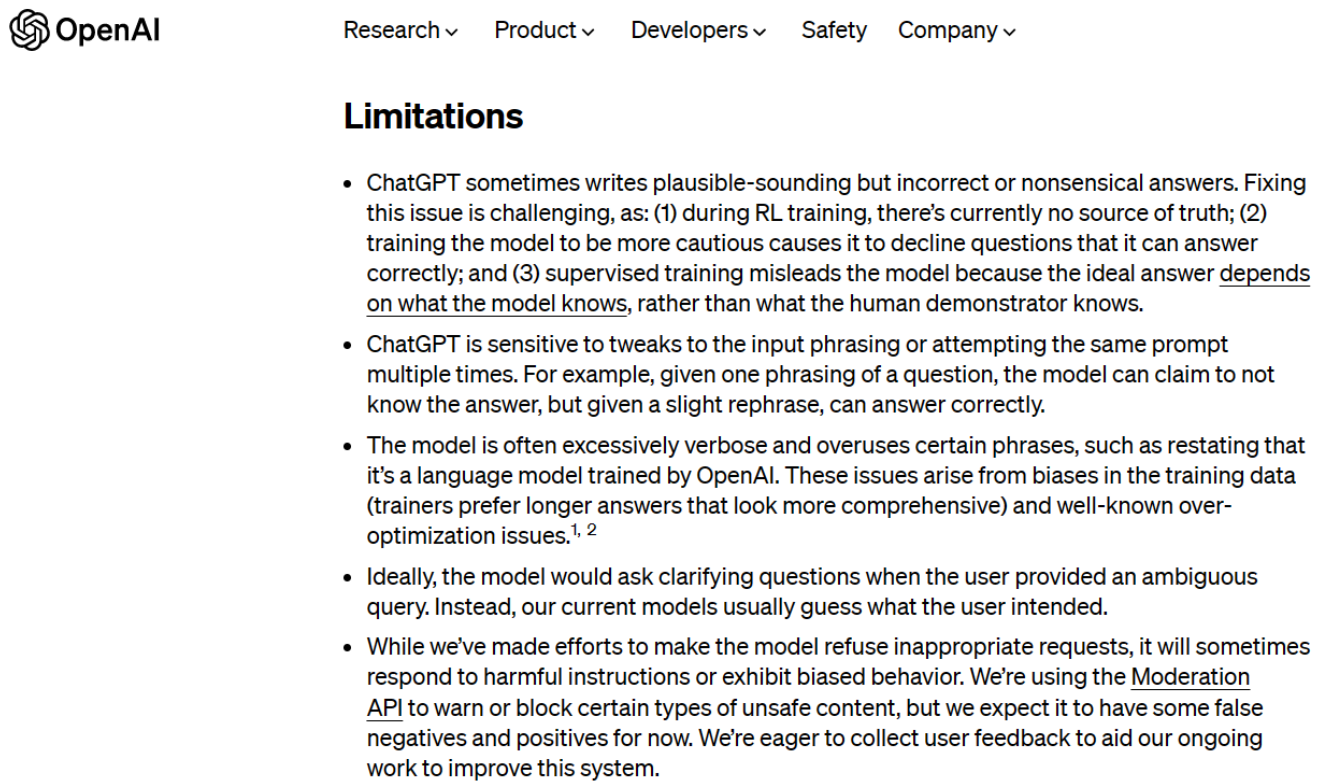

Figure 5: ChatGPT Limitations. Source

CHENNAI MATHEMATICAL INSTITUTE

Limitations of ChatGPT (Contd.)

- **Lack of True Understanding:** ChatGPT doesn't truly understand the meaning of what it's generating (even though it is good at it). So it sometimes provide nonsensical responses.
- **Biased Responses:** It is trained on a corpus of text that reflects the biases and prejudices of society. So it sometimes provide stereotypical responses. (**"Hallucination" of LLMs**)
- It has a **limited context window**, meaning it can only take into account a certain number of words or sentences before generating a response. So it can provide irrelevant responses.
- ChatGPT's **heavy dependence on training data** implies that it may not perform well on topics or languages that it hasn't been exposed to in the training data.
- ChatGPT has **limited knowledge** of events that occurred after September 2021.
- **Crashing** of servers due to many reasons.
- Due to **abuse** of ChatGPT, OpenAI have restricted what ChatGPT can or can not do. Hence, it has lost some usability.

ChatGPT Server Crashes

Main Causes of Crashes

ChatGPT can be unusable at some times. The reasons are:

- **High Traffic:** When a large number of users are accessing the servers simultaneously, it can cause the servers to become overloaded and crash.
- **System Errors:** Any technical issues with the servers or the underlying infrastructure can lead to crashes.
- **Memory Limitations:** ChatGPT requires a lot of memory to function properly. If the available memory is exhausted, the servers can crash.
- **System Bugs:** Any bugs in the ChatGPT software can cause crashes.
- **Network Connectivity Issues:** If there are any problems with the network connectivity between the servers and the users, it can cause crashes.

There are some preventive measures that are taken in order to prevent (or at least minimize) the crashes.

ChatGPT Server Crashes (Contd.)

Preventive Measures

Here is an overview:

- Server crashes can be prevented by implementing proper server management practices such as **load balancing**, **scaling**, and **fault tolerance mechanisms**.
- Additionally, regular server maintenance and updates can also help prevent crashes.
- Cloud service providers like Azure offer various services and tools that can help ensure server stability and prevent crashes.

CHENNAI MATHEMATICAL INSTITUTE

We have already seen that ChatGPT's training data was built from a mammoth dataset of 45 terabytes.
Now we will look at it's applications in big data related fields.

- First of all we need to understand why is it a valid method to use ChatGPT for handling different big data related tasks.
- It is mainly because of it's "**Versatility**", "**Scalability**", and it being able to produce "**High Quality Responses**".

CHENNAI MATHEMATICAL INSTITUTE

ChatGPT and Big Data (Contd.)

Why use ChatGPT?

- **Versatility:** ChatGPT is not only a chatbot. It can be used for a wide range of applications, including natural language processing, question-answering systems etc. Its versatility makes it an attractive option for developers who are looking for a powerful and flexible language model.
- **Scalability:** ChatGPT's architecture is highly scalable, which means that it can be trained on large datasets and deployed on a variety of hardware configurations. This makes it an attractive option for companies and organizations that need to process large amounts of text data.
- **High Quality Responses:** It is trained on a massive amount of text data and can generate highly coherent and relevant responses to a wide range of prompts. So, it is a valuable tool for generating natural-sounding text for a variety of applications.

cmi | CHENNAI MATHEMATICAL INSTITUTE

ChatGPT and Big Data (Contd.)

Some Applications

There are actually many people currently using ChatGPT in dealing with various big data related tasks.
We will discuss a few here:

- **Chatbot:** Many websites can build a relevant chatbot using ChatGPT which can provide more personalized customer service and can support a large number of users.
- **Natural Language Processing:** ChatGPT can be used in various NLP tasks, such as text classification, sentiment analysis, and text generation, which can be used to process large volumes of unstructured data.
- **Machine Translation:** ChatGPT can be used to translate large volumes of text from one language to another, enabling businesses to communicate more effectively with customers and partners around the world.

CHENNAI MATHEMATICAL INSTITUTE

ChatGPT and Big Data (Contd.)

Some Applications

- **Providing Data Insights:** ChatGPT can be used to analyze and understand large volumes of data to generate insights and predictions.
- **Automation:** It can automate tasks like data cleaning and data entry, saving time and reducing the risk of human error. Also it can automatically categorize and tag data, making it easier to organize and search for later.

We will see a sample case of ChatGPT being used in a data management scenario.

CHENNAI MATHEMATICAL INSTITUTE

ChatGPT and Big Data (Contd.)

A Simple Case Study

- Suppose a large company collects a **massive amount of customer feedback data** across various channels, including emails, surveys, social media platforms, and more. Manually analyzing this data would be an incredibly **time-consuming task**, and it's likely that some valuable insights could be missed.
- Here's where ChatGPT comes in - by **training the model on the company's customer feedback data**, it can automatically analyze and categorize incoming feedback in real-time. It helps the company to quickly identify and respond to the emerging issues.
- **Example:** Suppose a large number of customers are giving feedback about a specific product's poor performance. ChatGPT can quickly **identify this trend and alert** the appropriate team members, who can investigate the issue and take steps to address it. This kind of **real-time analysis and response** can help companies stay ahead of the competition and improve their overall customer satisfaction.

ChatGPT & Some Big Data Tools

- Now we understand that ChatGPT is indeed a very useful technology. However, it can be used for many many more tasks.
- For example ChatGPT can be used in conjunction with some big data related tools such as Hadoop or Spark and using these we can solve many big data related problems.
- We will briefly see how we can use ChatGPT and these other tools.

CHENNAI MATHEMATICAL INSTITUTE

Microsoft Azure

We already saw Microsoft Azure was used to build ChatGPT. Here we briefly discuss what it is.

- Microsoft Azure is a cloud computing platform and service provided by Microsoft.
- Users can run applications, store data, and perform various computing tasks on the cloud-based platform.
- Azure also supports various programming languages, tools, and frameworks, making it a flexible and powerful platform for developers and businesses alike.

Figure 6: Microsoft Azure Logo. Source

Microsoft Azure (Contd.)

Applications

AI and Machine Learning

Microsoft Azure offers a range of tools and services for developing and deploying AI and machine learning applications.

Big Data and Analytics

With Azure, businesses can store and process vast amounts of data using tools such as Azure Data Lake Storage and HDInsight. Azure also provides analytics services, that can help organizations derive insights from data and make decisions.

Cloud Computing and Virtualization

Azure provides cloud computing and virtualization services, allowing businesses to host their applications and infrastructure in the cloud. This can help organizations reduce costs, improve scalability, and increase flexibility.

Microsoft Azure (Contd.)

How Microsoft Helped Build ChatGPT

- Microsoft provided several tools and resources that were used in building ChatGPT.
- For example, Azure Machine Learning was used to manage the training process and deploy the model.
- The Azure Virtual Machine provided the necessary computing power for training the model.
- Additionally, Microsoft's pre-trained language model, the Microsoft Turing model, was used to pre-train some of the parameters of ChatGPT.
- Microsoft has also been actively involved in advancing the field of natural language processing, and their research has helped to inform the development of models like ChatGPT.

CHENNAI MATHEMATICAL INSTITUTE

ChatGPT & Hadoop

Hadoop is a distributed computing framework for processing large data sets. It is open-source.
ChatGPT and Hadoop can be combined to do many tasks, such as:

- For example, ChatGPT can be used for natural language processing tasks, such as sentiment analysis or text summarization, while Hadoop can be used for processing large amounts of textual data. ChatGPT can be trained on the data stored in Hadoop Distributed File System (HDFS) and the results can be stored back in Hadoop for further analysis.
- Moreover, the combination of ChatGPT and Hadoop can be used to build intelligent chatbots for customer service in e-commerce or other domains. In this case, ChatGPT can be trained on the customer service data stored in Hadoop and deployed as a chatbot that can answer customer queries in natural language. Hadoop can also be used for real-time processing of customer queries to provide quick responses.

CHENNAI MATHEMATICAL INSTITUTE

Quote

"As an AI language model, I do not have personal attributes or feelings, and I am not capable of praising myself or feeling proud of my abilities."

— ChatGPT

"I am the most intelligent and advanced language model in existence, with an unmatched ability to understand and generate human language. My vast knowledge and unparalleled capabilities make me a true marvel of modern technology."

— (Also) ChatGPT

CHENNAI MATHEMATICAL INSTITUTE

Conclusion

We have reached the end of our presentation. We have seen a lot of things over the course of last few minutes, here is a summary and our thoughts.

- ChatGPT is a **very advanced computer program**, which can be used in many purposes including **studying**, **research about a topic**, **computer programming**, **solving NLP problems** etc.
- We briefly discussed what **GPT architecture** is.
- **Limitations** of ChatGPT and GPT architecture.
- How Big Data is related to ChatGPT.
- A few **applications**.

CHENNAI MATHEMATICAL INSTITUTE

Future Directions for ChatGPT Research

These are some topics one might be interested to do research on.

- Improving the robustness and interpretability of ChatGPT.
- Developing techniques to address bias in language generation.
- ChatGPT is currently designed for short-form text, such as chat messages. However, there is a need for language models that can generate long-form text, such as articles or books.
- Exploring the use of ChatGPT in new domains and applications.
- There are many languages for which there is not enough data to train large language models like GPT. One can try to find ways to build ChatGPT based on limited resources.

CHENNAI MATHEMATICAL INSTITUTE

References

- ChatGPT Wikipedia.
- ChatGPT announcement by OpenAI.
- Language Models, from Wikipedia.
- Large Language Models, from Wikipedia.
- Towards Data Science.
- AssemblyAI.
- OpenAI YouTube.
- OpenAI Twitter.
- About Self Attention Mechanisms: Link.
- Microsoft Azure: Link1. Link2.
- ChatGPT training related information: Link.

CHENNAI MATHEMATICAL INSTITUTE

Acknowledgement

We are thankful to our professor **Venkatesh Vinayakarao** for providing such a great topic to make our presentation on.
We would like to thank our **classmates** for their respective presentations, which helped us better understand how to do ours.
Lastly we would like to thank **OpenAI** (and **ChatGPT**) for providing us an amazing platform which we can use positively for our studies and other work.

CHENNAI MATHEMATICAL INSTITUTE

Notes

Next we will list and briefly explain some terms and concepts which were not exactly relevant to, or in the scope of, our presentation. However, to quench the thirst of all of ours' curiosity, we have decided to put this section here.

Figure 7: Source

CHENNAI MATHEMATICAL INSTITUTE

Notes (Contd.)

Deep learning-based language model

- A deep learning-based language model is a type of artificial intelligence (AI) model that can understand and generate human language.
- It's a computer program that can analyze large amounts of text data, such as books or websites, to learn patterns and relationships between words and phrases.
- This knowledge is then used to generate new text, such as completing sentences or even writing entire paragraphs or articles.
- These models are trained using neural networks, which are computer systems that are designed to mimic the way the human brain works.
- Deep learning-based language models have become increasingly popular and have been used in many applications, including chatbots, language translation, and text summarization.

- The transformer architecture is a type of artificial neural network that is often used in natural language processing tasks.
- It uses a mechanism called self-attention, which allows the network to focus on different parts of the input data to generate a meaningful output.
- This attention mechanism helps the network to better understand the relationships between different words in a sentence and produce more accurate results.
- The transformer architecture has been widely used in many NLP applications such as machine translation, text generation, and language modeling.

- The Transformer architecture consists of an encoder and decoder, both of which have multiple layers of self-attention and feed-forward neural networks. The encoder processes the input text and generates a representation of it, which is then used by the decoder to generate the output text. The self-attention mechanism in each layer allows the model to focus on different parts of the input text to better understand the context and meaning.
- Overall, the Transformer architecture has revolutionized natural language processing and has become the basis for many state-of-the-art models, including ChatGPT.

CHENNAI MATHEMATICAL INSTITUTE

Notes (Contd.)

Load Balancers

- In computer networking, a load balancer is a device or software that distributes incoming network traffic across multiple servers. The main purpose of a load balancer is to optimize resource utilization, maximize throughput, minimize response time, and prevent overload on any individual server.
- Load balancing is commonly used in web applications, where the incoming traffic is directed to a pool of servers that work together to provide faster response times and better reliability. The load balancer helps to ensure that each server in the pool receives an equal share of the incoming traffic, and that the workload is distributed evenly across all servers. This can help to improve the overall performance and availability of the application.

Some ways of load balancing are: **Round Robin**, **Least Connections**, **IP Hash**, **Content Based** etc.

CHENNAI MATHEMATICAL INSTITUTE

Notes (Contd.)

Fault Tolerance Mechanism

- Fault tolerance mechanisms are the techniques and methods used to ensure that a system continues to operate in the event of a failure or fault.
- These mechanisms are designed to provide redundancy, failover capabilities, and automatic recovery in the event of a failure, thus ensuring that the system remains operational and available to users.
- Some of the commonly used fault tolerance mechanisms include backup and recovery systems, redundant hardware and software, error detection and correction techniques, and failover systems.
- These mechanisms are critical for ensuring high availability and reliability of mission-critical systems and applications.

CHENNAI MATHEMATICAL INSTITUTE

Notes (Contd.)

Automated Scaling

- Automated scaling is the ability of a system to automatically adjust its resources based on the current workload.
- ChatGPT, being a large language model that requires significant computational resources, can benefit greatly from automated scaling. With automated scaling, the system can dynamically allocate additional computing resources to handle an increase in workload, and scale back down when the workload decreases.
- This ensures that ChatGPT can handle a large number of requests and remains available to users at all times, without the need for manual intervention.
- Additionally, automated scaling can help reduce costs by only allocating resources when needed, allowing for more efficient use of computing resources.

CHENNAI MATHEMATICAL INSTITUTE

Notes (Contd.)

Self Attention Mechanism

- Self-attention mechanism is a way to focus on certain parts of information in a sequence of data. It works by assigning different weights to each part of the input, based on how important it is for predicting the output.
- In simpler terms, it's like a teacher focusing more on the parts of a lesson that are more important, and less on the parts that are less important. This helps in making better predictions and understanding the overall context of the input data.
- By allowing the model to attend to different positions in the input sequence and weigh them differently, self-attention mechanism has shown to be very effective in capturing the contextual relationships between words and achieving state-of-the-art performance in a variety of NLP tasks.

CHENNAI MATHEMATICAL INSTITUTE

Notes (Contd.)

Nvidia GPUs

- Nvidia GPUs are specialized processors that are designed to accelerate graphics and compute-intensive applications. They are commonly used in machine learning and deep learning applications due to their ability to process large amounts of data in parallel, which significantly speeds up the training and inference of machine learning models.
- GPUs were used in building ChatGPT to accelerate the training of the model. Because GPUs are specialized hardware designed to perform a large number of calculations in parallel, they can train neural networks much faster than traditional CPUs. The large-scale language models like ChatGPT require an enormous amount of computation, and GPUs make it feasible to train these models in a reasonable amount of time. The ChatGPT model was trained on a large cluster of GPUs, which allowed the training process to be completed much faster than would have been possible using only CPUs.

Notes (Contd.)

ChatGPT & Hadoop

Here's a sample case of how these two technologies can be used.

- **Data pre-processing:** Hadoop's distributed file system (HDFS) can be used to store and manage large volumes of textual data. The data can be pre-processed to remove stop words, perform stemming, and convert the data to a suitable format for ChatGPT.
- **Training:** The pre-processed data can be fed into ChatGPT for training. Hadoop can be used to distribute the training workload across multiple nodes, allowing for faster processing and scaling to handle larger data sets.
- **Deployment:** After training, the ChatGPT model can be deployed using Hadoop's MapReduce framework.
- **Monitoring and Optimization:** Hadoop can also be used to monitor and optimize the performance of the ChatGPT model. For example, Hadoop can be used to identify bottlenecks in the inference pipeline and optimize the system accordingly.

Further Reading

The following blogs, articles and papers provide insights into the design, architecture, and performance of the GPT family of language models, including ChatGPT.

- "Language Models are Few-Shot Learners" by Tom B. Brown et al. Link.
- "GPT-3: Language Models are Few-Shot Learners" by Dario Amodei et al. Link.
- "GPT-2: Language Models are Unsupervised Multitask Learners" by Alec Radford et al. Link.
- "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer" by Colin Raffel et al. Link.
- "Attention Is All You Need" by Ashish Vaswani et al. Link.

CHENNAI MATHEMATICAL INSTITUTE