



Document Information

Analyzed document	chatgpt_analysis[1] (1).docx (D167336828)
Submitted	5/17/2023 5:35:00 PM
Submitted by	Priyanka
Submitter email	priyankasharma.cse@cumail.in
Similarity	2%
Analysis address	priyankasharma.cse.cumail@analysis.urkund.com

Sources included in the report

	URL: https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286 1ce5fca96286 Fetched: 5/17/2023 5:35:00 PM	 4
---	---	---

Entire Document

ChatGPT Analysis

Atul Yadav Department of CSE, Chandigarh University, Punjab, India atul26092001@gmail.com

Anshika Mittal Department of CSE, Chandigarh University, Punjab, India mittalanshika28@gmail.com Priyanka Sharma Department of CSE, Chandigarh University, Punjab, India priyankasharma.cse@cumail.in

Vigy Vigy Department of CSE, Chandigarh University, Punjab, India Vigyavigy555@gmail.com

Aishwary Krishna Department of CSE, Chandigarh University, Punjab, India aishwary1969@gmail.com

Abstract—This review paper presents an extensive and comprehensive analysis of ChatGPT, a language model developed by Open AI based on the GPT-3.5 architecture. The paper begins with an overview of the model's technical specifications and its training process, which involved massive amounts of text data from the internet. The review then delves into ChatGPT's performance on a wide range of natural language processing tasks, including language translation, question answering, and text completion. The review paper provides a detailed analysis of how ChatGPT works, and its architecture and later discussed its advantages and disadvantages which also act as features and limitations of ChatGPT, comparing its performance to other state-of-the-art language models. It highlights the impressive capabilities of the model, such as its ability to generate coherent and contextually appropriate responses in conversation with humans. However, the review also notes areas where ChatGPT falls short, such as its tendency to generate biased and sometimes offensive content due to the training data's biases or supervised learning. The "ChatGPT Analysis" review paper provides an in-depth and insightful analysis of one of the most advanced Natural Language Processing NLP models currently available. It highlights both the impressive capabilities and the limitations of the model, while also addressing the ethical implications of its use and some future works. Overall, the paper offers valuable insights into the current state of NLP research and the potential for language models to advance the field in the future. Keywords- Transformer, Generative Pre-trained Transformer, NLP, Large Language Models, Reinforcement Learning

I. Introduction Chatbots have attracted a lot of research attention because of the development of machine learning and natural language processing (NLP), and they are now being employed for both commercial and non-commercial purposes. One of the kind of chatbots is ChatGPT which is a tool based on GPT-3 model. ChatGPT is a language-based model which uses artificial intelligence technology to generate natural or human-like responses for a given input. It is helpful and beneficial in different fields and sectors from natural language processing to content creation and understanding the scenario to provide suitable predictions about it. In this review paper, we will discuss its origin, how it works and what is its architecture, and its study by different authors and their future work. So, let's discover how it is changing our lives and what its future is.

ChatGPT was developed by Open AI that utilizes Generative Pre-Training transformer (GPT) technology to fulfill a range of text-based requests. GPT (Generative Pre-training) is a family of large language models developed by Open AI that are based on transformer architecture. The transformer architecture is made up of multiple layers, which are designed or built to process and generate text. GPT models were developed using a combination of machine learning techniques and artificial intelligence. The models were trained on large datasets using techniques such as unsupervised learning and supervised learning, reinforcement, or reward-punishment system and also need a training process required for significant computational resources. The models themselves are implemented in software and are designed to run on a variety of hardware platforms. ChatGPT is capable of understanding and interpreting user requests, generating appropriate responses quickly and efficiently in nearly natural human-like language, and completing advanced tasks such as Writing Assistance, Language Translation, General Knowledge, Personal Development, addressing productivity issues, etc. In this review paper how ChatGPT works, as well as the potential impacts of this technology on various industries, are discussed. This can be useful for tasks such as answering questions, translating text, or generating content. Fig.1 explains how ChatGPT works. (Fig.1 is a screenshot taken from the official website of OpenAI[2]). Beginning with its reaction to our attempt to address the history of chatbots, we also requested ChatGPT to address some of the questions.

II. Related work In this work, we presented the Transformer, the first sequence transduction model based entirely on attention, replacing the recurrent layers most commonly used in encoder-decoder architectures with multi-headed self-attention. For translation tasks, the Transformer can be trained significantly faster than architectures based on recurrent or convolutional layers. On both WMT 2014 English-to-German and WMT 2014 English-to-French translation tasks, achieving a new state-of-the-art. In the former task, the best model outperforms even all previously reported ensembles. They planned to extend the Transformer to problems involving input and output modalities other than text and to investigate local, restricted attention mechanisms to efficiently handle large inputs and outputs such as images, audio, and video. Making generation less sequential is another research goal of ours. The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best-performing models also connect the encoder and decoder through an attention mechanism. The authors proposed a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. This model shows that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.[1]

This study of language generation introduced a framework for achieving strong natural language understanding with a single task-agnostic model through generative pre-training and discriminative fine-tuning. By pre-training on a diverse corpus with long stretches of contiguous text our model acquires significant world knowledge and the ability to process long-range dependencies which are then successfully transferred to solving discriminative tasks such as question answering, semantic similarity assessment, entailment determination, and text classification, improving the state of the art on 9 of the 12 datasets we study. Using unsupervised pre-training to boost performance on discriminative tasks has long been an important goal of Machine Learning research. This work suggests that achieving significant performance gains is indeed possible, and offers hints as to what models (Transformers) and data sets (text with long-range dependencies) work best with this approach. We hope that this will help enable new research into unsupervised learning, for both natural language understanding and other domains, further improving our understanding of how and when unsupervised learning works.[2]

This paper discusses the historical overview of chatbots and the technology behind Chat Generative Pre-Training Transformer. In addition, we highlight some of the important limitations of the current version of ChatGPT. In this paper, we provided a historical overview of the development of chatbots. In addition, we looked at the significant technological developments that enabled the emergence and success of ChatGPT. We then described the potential of ChatGPT in several domains and applications. In healthcare, ChatGPT can potentially be used for medical screening, answering general questions, and exam preparation. In education, ChatGPT can be used in tutoring and detecting plagiarism. ChatGPT can also aid researchers with writing, summarizing information, and translating. [3]

This study gave a brief historical summary of chatbot development. We also took a close look at the important technology advancements that made ChatGPT possible and successful. The potential of ChatGPT in many areas and applications were then discussed in the article. ChatGPT has the potential to be employed in the healthcare industry for medical screening, general question answering, and test preparation. ChatGPT can be utilized in the educational setting for tutoring and plagiarism detection. Additionally, ChatGPT can help researchers with writing, information summaries, and translation. But there are a lot of ethical and privacy issues with ChatGPT that need to be addressed. For instance, some users have complained about racial and gender prejudice in ChatGPT's responses⁷. Furthermore, because of its efficacy, ChatGPT could be utilized in schools for dishonest goals like cheating. ChatGPT highlights moral concerns regarding plagiarism and copyright in research. Concerning privacy issues, ChatGPT has been trained using more than 300 billion words, which might include sensitive data about internet users⁸. Last but not least, ChatGPT keeps becoming better with user engagement, although prompts including private information, such as contact information, may be processed and even learned by the model.^[4] This review paper comprehends about the integration of Chat Generative Pre-Trained Transformer (ChatGPT) in education. ChatGPT is AI technology that is used to generate conversational interactions with end users and is based on deep learning algorithms. According to this research, ChatGPT is not considered a high-risk technology in relation to cheating. ChatGPT is expected to provide a trigger for educational discussions on assessment design and academic integrity. To examine the potential impact on learning and assessment two chemistry-focused modules of 1st and 2nd year of pharmaceutical sciences are used. These investigate the responses generated by ChatGPT in correspondence to the end-of-year examinations. It was initiated that for the knowledge and understanding-based questions, ChatGPT was capable of generating the responses but for the application of knowledge and interpretation with non-text information, there was a technology barrier. In the Future, it should focus on training using a larger dataset so that it is capable of answering questions more effectively.^[5] This research paper discusses the benefits and potential pitfalls of using AI-based ChatGPT. This paper also states that ChatGPT is the NLP model that applies generative artificial intelligence techniques to algorithm-generated responses to question prompts. It provides algorithm-generated conversational responses to text-based prompts. ChatGPT can offer enhanced customer engagement, improved service, personalization, social interaction, and communication practice, and insights into customer behavior. GPT-3 is free of cost. The potential pitfalls of ChatGPT involve concerns about misinformation, lack of content, privacy concerns, and ethical considerations, and security. This paper concludes by outlining the potential research agenda for the future in the area of ChatGPT and customer studies. As millions of consumers have become users of ChatGPT, there are potential avenues for future research on topics like the newness of AI tools such as ChatGPT and their implications, context, theories, and research methods suggested in this research area.^[6]

This research paper gives indications about when Open AI's ChatGPT known as GPT-3 was used as the interviewee instead of a human to check the impacts of its impacts on higher education and academics. ChatGPT was asked several questions based on the impacts of the AI-based machine learning models and OpenAI's ChatGPT on the hospitality and tourism industry and education. Basically, this paper represents the ChatGPT as the interviewee. It helps instructors to delegate tedious tasks such as grading and focus more on intellectual tasks. Students can also use ChatGPT to improve their vocabulary and generate new, brainstorming, and creative ideas. It also states that over-reliance on ChatGPT can diminish the critical thinking of students as well as educational inequalities. It cannot substitute human intelligence and intellect as originality and novelty lack in responses generated by ChatGPT. Whereas the tourism and hospitality industry can take advantage of ChatGPT such as personalized services, content creation, and many more. It is suggested to look for ways to use chatGPT in the field of tourism and hospitality. AI tools can facilitate their cognitive tasks through automation and allocate their human resource for more empathetic services.^[7]

According to this paper, AI offers transformative potential in the field of education. This study tells about how ChatGPT can answer questions related to science education and how educators can utilize ChatGPT in their science pedagogy. The research applies a self-study methodology to investigate the technology. The main ethical concerns include its potential impacts, issues related to content modification, novelty, originality, and risk of copyright infringement. It is important for educators to responsibly use ChatGPT, prioritize critical thinking, science quizzes, rubrics, design units, narration, and expectations should be clear. Educators should critically evaluate AI tools and adapt them to use by making modifications according to their teaching methodology. It is only for assistance in certain work to save some time and generate new responses. For Future work transparency and clearer guidelines about the use of AI platforms in research are suggested for advancing scientific knowledge.^[8]

This article provides a comprehensive overview of the future of nursing through the ChatGPT model, developed by Open AI. This chat transcripts with the ChatGPT to offer key trends and potential developments in the field of nursing, like increasing use of technology and tools for the diagnoses of patients, robotics in patients' health care, integration of AI, as well as insights into the implications of these changes for nurses, patients, and healthcare. However, the future of nursing is evolving continuously so nurses need to keep abreast of the latest technological advancements and adapt to them while keeping in mind the quality care. For the future of nursing integrating technology and robots into nursing practices can improve patient outcomes, increase efficiency and reduce errors in care delivery.[9]

According to Md. Asraful Haque- ChatGPT is not a programming language, but rather a machine learning model that has been trained on a large dataset and can generate responses to prompts in a way that is similar to how a human might respond. ChatGPT can automate many tasks that previously required human intervention. It has the potential to revolutionize the way many tasks are completed by improving efficiency, accuracy, personalization, accessibility, and innovation. The ChatGPT, a powerful conversational tool trained by OpenAI is a revolutionary model in the field of artificial intelligence and natural language processing. It has received a lot of attention because of its potential to automate a variety of tasks and possibly have an impact on sectors like translation, customer service, and content creation. It uses GPT-3 (Generative Pre-training Transformer 3) language model to process user queries. GPT-3 has been trained on a very large dataset, which includes a wide range of texts from the internet and other sources. This has given it a broad knowledge base and has allowed it to generate responses to a wide range of prompts that are coherent and human-like. GPT-3 is one of the largest and most powerful language models to date, and it has the ability to perform a wide range of natural language processing tasks.[10]

In this research, of the authors proposed BioGPT, a generative pre- trained Transformer language model for biomedical text generation and mining. We adopted GPT-2 as our backbone model and pre-trained on 15M PubMed abstracts corpus. We carefully designed and investigated the prompt and the target sequence format when applying pre-trained BioGPT to downstream tasks. We applied the pre-trained BioGPT to biomedical NLP tasks: end-to-end relation extraction task, 10 Renqian Luo et al. question answering task, document classification task, and text generation task. BioGPT achieves SOTA results on three end-to-end relation extraction tasks and one question-answering task. It also demonstrates better biomedical text generation ability compared to GPT-2 on the text generation task. And as future work of plans to train larger-scale BioGPT on larger-scale biomedical data and apply it to more downstream tasks. Pre-trained language models have attracted increasing attention in the biomedical domain, inspired by their great success in the general natural language domain. Among the two main branches of pre-trained language models in the general language domain, i.e., BERT (and its variants) and GPT (and its variants), the first one has been extensively studied in the biomedical domain, such as BioBERT and PubMedBERT.[11]

In this study, the authors examined the advantages and disadvantages of ChatGPT, as well as its limitations and features. It also discusses the impact of ChatGPT on academics, cyber security, customer support, software development, jobs, and information technology, as well as its potential applications for researchers and scholars. Implemented through a deep neural network architecture that consists of several layers of transformers. Due to his work and contribution opportunities for ChatGPT to be used in fields such as education, healthcare, and mental health therapy, where conversational agents can be used to provide support and guidance to people in need and Potential to transform the way we interact with technology and make our lives easier and more efficient. [12] By utilizing NLP (Natural Language Processing) skills, this research offers ten real-world examples from several fields of environmental research. ChatGPT excels in basic information retrieval and idea explication because it is a large language model (LLM). An LLM's internal workings show that it may provide falsified or incorrect information. Additionally, there is a lack of accountability in decision-making, as well as a cost of relying on chatGPT. Prompt engineering and in-context learning are two methods utilized in chatGPT. While quick engineering includes creating model inputs, such as questions and statements, to get better results, in-context learning allows the agent to learn and adapt in real-time There is little doubt that ChatGPT will revolutionize society and streamline or automate a variety of tasks, including research. Such advancements should be welcomed and utilized to achieve our objectives, but we must also exercise caution to stay out of traps and recognize our limitations. It is crucial to remember that a number of publications, like the American Chemical Society, have stated categorically that AI tools like ChatGPT do not qualify as authors. Any use of artificial intelligence (AI) techniques for the creation of text or images should be disclosed in the article. More modern LLMs (like GPT-4) will make them more trustworthy and capable of doing more challenging tasks, which may assist with some of the current issues. However, we use chatgpt and other technologies differently.[13] The author of this study collected tweets from the first month after ChatGPT, a cutting-edge AI chatbot, was released. A total of 233,914 English tweets were investigated using the latent Dirichlet allocation (LDA) topic modeling method to see what ChatGPT is capable of. The studies revealed three key themes: news, technology, and reactions. Additionally, the author specified five functional domains: essay writing, question-answering, code-writing, prompt-writing, and code-writing. The analysis also showed that ChatGPT has the potential to have both positive and negative effects on people and technology. A new technological environment, the pursuit of artificial intelligence, and four major challenges that must be addressed as a result of this AI progress are outlined by the author in his conclusion. The possible negative effects on technology and people, however, are perhaps far more crucial and will need greater investigation in the future. The results of this investigation lead to the identification of four major concerns. The next stage of employment progression. ChatGPT and other cutting-edge AI technologies will probably have an influence on the evolution of occupations in the future. Although this progress may be advantageous to individuals, there are concerns involved, such as the loss of jobs for white-collar and creative workers. Furthermore, it appears that time is passing more quickly than expected. Even "safe" occupations like AI programmers, trainers, and analysts might be threatened by ChatGPT.[14]

This research aims to highlight the general-purpose potential of LLMs and their possible implications for US workers. Previous literature demonstrates the impressive improvements of LLMs to date. Their findings confirm the hypothesis that these technologies can have pervasive impacts across a wide swath of occupations in the world and that additional advancements supported by LLMs, mainly through software and digital tools, can have significant effects on a range of economic activities. However, while the technical capacity of LLMs to make human labor more efficient appears evident, it is important to recognize that social, economic, regulatory, and other factors will influence actual labor productivity outcomes. As capabilities continue to evolve, the impact of LLMs on the economy will likely persist and increase, posing challenges for policymakers in predicting and regulating their trajectory. Further research is necessary to explore the broader implications of LLM advancements, including their potential to augment or displace human labor, their impact on job quality, impacts on inequality, skill development, and numerous other outcomes. By seeking to understand the capabilities and potential effects of LLMs on the workforce, policymakers and stakeholders can make more informed decisions to navigate the complex landscape of AI and its role in shaping the future of work.[15]

When a large language model is trained on a sufficiently large and diverse dataset it is able to perform well across many domains and datasets. GPT-2 zero-shots to state-of-the-art performance on 7 out of 8 tested language modeling datasets. The diversity of tasks the model is able to perform in a zero-shot setting suggests that high-capacity models trained to maximize the likelihood of a sufficiently varied text corpus begin to learn how to perform a surprising amount of tasks without the need for explicit supervision. The paper demonstrates that language models begin to learn these tasks without any explicit supervision when trained on a new dataset of millions of web pages called WebText. When conditioned on a document plus questions, the answers generated by the language model reach 55 F1 on the CoQA dataset - matching or exceeding the performance of 3 out of 4 baseline systems without using the 127,000+ training examples. The capacity of the language model is essential to the success of zero-shot task transfer and increasing it improves performance in a log-linear fashion across tasks. Our largest model, GPT-2, is a 1.5B parameter Transformer that achieves state-of-the-art results on 7 out of 8 tested language modeling datasets in a zero-shot setting but still under its WebText. Samples from the model reflect these improvements and contain coherent paragraphs of text. These findings suggest a promising path towards building language processing systems that learn to perform tasks from their naturally occurring demonstrations.[16]

This work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. In this paper, the authors scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches. GPT-3 is trained on an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and tests its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. At the same time, we also identify some datasets where GPT-3's few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora. Finally, we find that GPT-3 can generate samples of news articles in human evaluators have difficulty distinguishing from articles written by humans. We discuss the broader societal impacts of this finding and of GPT-3 in general. The authors presented a 175 billion parameter language model which shows strong performance on many NLP tasks and benchmarks in the zero-shot, one-shot, and few-shot settings, in some cases nearly matching the performance of 40 state-of-the-art fine-tuned systems, as well as generating high-quality samples and strong qualitative performance at tasks defined on-the-fly. They documented roughly predictable trends of scaling in performance without using fine-tuning. They also discussed the social impacts of this class of models. Despite many limitations and weaknesses, these results suggest that very large language models may be an important ingredient in the development of adaptable, general language systems.[17] This technical report presents GPT-4, a large multimodal model capable of processing image and text inputs and producing text outputs. Such models are an important area of study as they have the potential to be used in a wide range of applications, such as dialogue systems, text summarization, and machine translation. As such, they have been the subject of substantial interest and progress in recent years. One of the main goals of developing such models is to improve their ability to understand and generate natural language text, particularly in more complex and nuanced scenarios. To test its capabilities in such scenarios, GPT-4 was evaluated on a variety of exams originally designed for humans. In these evaluations, it performs quite well and often outscores the vast majority of human test takers. For example, on a simulated bar exam, GPT-4 achieves a score that falls in the top 10% of test takers. This contrasts with GPT-3.5, which scores in the bottom 10%. This report focuses on the capabilities, limitations, and safety properties of this model. GPT-4 is a Transformer-style model pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF)[17]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.[18]

III. Architecture Of ChatGPT The Transformer model, which was first presented in the publication "Attention is All You Need" by Vaswani et al. (2017) [1], serves as the foundation for ChatGPT's design. The Transformer model is a neural network architecture created to process and produce output data sequences from sequences of input data, such as phrases or paragraphs of text. the design of OpenAI's GPT (Generative Pre-trained Transformer). For natural language processing (NLP) applications including language translation, text summarization, and chatbot building, it is one of the most sophisticated and potent models available. A set of Transformer models trained by OpenAI for language modeling tasks is known as generative pre-trained transformers. The BERT model predates the first GPT model. Unlike BERT, which depended on creative training objectives, OpenAI trained its models to anticipate the next word (Language Modelling). The original GPT model included 117 million parameters and significantly altered several tasks' state-of-the-art figures. But it wasn't until GPT-2, with its 1.5 billion parameters, that it really began to get notice from the general public.

100%

MATCHING BLOCK 1/4

W

ChatGPT is a spinoff of InstructGPT, which introduced a novel approach to incorporating human feedback into the training process to better align the model outputs with user intent. Reinforcement Learning from Human Feedback (RLHF) is described in depth in OpenAI's 2022 paper Training language models to follow instructions with human feedback and is simplified below.

Figure 1. Rollout, evaluation, and optimization [17]
Step 1: Supervised Fine Tuning (SFT) Model The initial step in the process was to

70%

MATCHING BLOCK 2/4

W

fine-tune the GPT-3 model by employing 40 contractors to provide a supervised training dataset, in which each input has a

predetermined outcome from which the model may learn. The Open API was used to gather inputs, or prompts, from real user entries. After that, the labelers responded appropriately to the prompt, producing a predictable output for each input. This new, supervised dataset was subsequently used to enhance the GPT3 model, resulting in GPT-3.5, also known as the SFT model.

Figure 2. Supervised Fine Tuning [17]

97%

MATCHING BLOCK 3/4

W

Step 2: Reward Model After the SFT model is trained in step 1, the model generates better-aligned responses to user prompts. The next refinement comes in the form of training a reward model in which a model input is a series of prompts and responses, and the output is a scaler value, called a reward. The reward model is required in order to leverage Reinforcement Learning in which a model learns to produce outputs to maximize its reward (see step 3).
Figure 3. Reward model [19]

Step 3: Reinforcement Learning Model The model receives a random cue in the last stage and responds with an answer. The 'policy' the model learns in step 2 is used to construct the response. The policy is an approach

50%

MATCHING BLOCK 4/4

W

the computer has learned to adopt to accomplish its objective, in this example, maximizing its reward. A scaler reward value is then established for the prompt and answer pair

using the reward model created in step 2 as a basis. The model then incorporates the reward to modify the policy. While employing the same techniques as InstructGPT, this model employs Reinforcement Learning from Human Feedback (RLHF), with a few minor variations in the data-gathering configuration. An early model was trained through supervised fine-tuning, in which human AI instructors gave interactions in which they took on the roles of both the user and the AI assistant. Trainers have access to sample written recommendations to use as a guide when crafting their answers. The InstructGPT dataset, which was converted into a conversation format, is combined with this new dialogue dataset. It was necessary to gather comparison data, which consisted of two or more model replies graded by quality, in order to develop a reward model for reinforcement learning.

Figure 4. Reinforcement Learning Model [19]

We used the chatbot interactions that AI trainers conducted with it to get this data. We chose a model-written statement at random, sampled a number of potential conclusions, and asked AI trainers to rate them. The model may be improved via Proximal Policy Optimization utilizing these reward models.

IV. GENERATIVE PRE-TRAINED TRANSFORMER MODELS: A class of machine learning models called Generative pre-trained transformer (GPT) is used to solve issues with natural language processing. These models are pre-trained on enormous amounts of data, including books and web pages, in order to create a language that is both contextually relevant and semantically coherent. Simply defined, GPTs are computer programs having the capacity to generate language that mimics that produced by humans without the use of explicit programming. As a result, they may be customized for a range of NLP tasks, including text summarization, language translation, and question-answering. So why are GPTs important? With the help of GPTs, which represent a significant leap in natural language processing, robots can now grasp and generate language with fluency and accuracy that was previously unheard of. In the sections below, we investigate the four GPT models, from the original to the most current GPT-4, and analyze their capabilities and drawbacks.

GPT1: In 2018, OpenAI launched GPT-1, their inaugural Transformer-based Transformer language model iteration[2]. It performed far better than older, 117 million parameter state-of-the-art language models. One of the features of the GPT-1 was its ability to create clear, understandable English when given a prompt or context. The model was trained using a combination of the Common Crawl dataset, a big dataset of web pages with billions of words, and the BookCorpus dataset, a collection of over 11,000 books on a variety of themes. With the use of these numerous datasets, GPT-1 developed strong language modeling capabilities. Although GPT-1 marked a significant improvement in natural language processing (NLP), it had certain limitations. For instance, when given instructions that weren't part of its training set, the model tended to produce repetitive text. Additionally, it was unable to reason through several debate rounds and monitor long-term dependencies in text. Larger portions would also lack continuity, and only shorter text sequences could have cohesion and fluency.

GPT2: As a follow-up to GPT-1, OpenAI launched GPT-2 in 2019[16]. Its astonishing 1.5 billion parameter count made it far bigger than GPT-1[2]. By merging data from Common Crawl and WebText, a significantly bigger and more varied dataset was used to train the model. The capacity of GPT-2 to produce cohesive and lifelike text sequences was one of its strengths. It could also produce replies that resembled those of a person, making it a useful tool for a range of natural language processing activities, including translating and creating content. GPT-2, however, was not without its drawbacks. It had trouble doing activities that needed more intricate reasoning and contextual awareness. While GPT-2 performed exceptionally well with little sections and chunks of text, it struggled to keep context and coherence across longer stretches. The next generation of GPT models was made possible by these restrictions. One of the strengths of GPT-2 was its ability to generate coherent and realistic sequences of text. In addition, it could generate human-like responses, making it a valuable tool for various natural language processing tasks, such as content creation and translation. However, GPT-2 was not without its limitations. It struggled with tasks that required more complex reasoning and understanding of context. While GPT-2 excelled at short paragraphs and snippets of text, it failed to maintain context and coherence over longer passages. These limitations paved the way for the development of the next iteration of GPT models.

Figure 5. Training and Pre-training of GPT2

GPT3: With the introduction of GPT-3 in 2020, natural language processing models advanced exponentially. GPT-3 is more than 100 times larger than GPT-1 and more than ten times larger than GPT-2 with 175 billion parameters. A wide variety of data sources, including BookCorpus, Common Crawl, and Wikipedia, are used to train the GPT-3 algorithm. The datasets include about a trillion words, enabling GPT-3 to provide complex replies on a variety of NLP tasks even in the absence of any example data. One of GPT-3's key advantages over its earlier generations is its capacity to produce coherent text, write computer code, and even produce original artwork. In contrast to earlier models, GPT-3 is capable of producing replies that are suitable given the context of a particular text. For applications like chatbots, content generation, and language translation, the capacity to create natural-sounding text has enormous ramifications. One such instance is ChatGPT, a conversational AI bot that rose to prominence very immediately.

Figure 6. Architecture Of GPT3[17] Even if GPT-3 is capable of doing amazing things, it still has drawbacks. The model may, for instance, provide responses that are biased, erroneous, or unsuitable. Because GPT-3 is trained on vast quantities of text that might contain biased and erroneous information, this problem exists. Additionally, there are times when the model responds with completely unrelated text to a request, demonstrating that it is still having trouble comprehending context and prior knowledge.

Concerns regarding the moral ramifications and potential abuse of such strong language models were also sparked by the capabilities of GPT-3. Experts are concerned that the model may be employed maliciously to produce malware, phishing emails, and false news. Indeed, ChatGPT has already been used by criminals to produce malware. Additionally, OpenAI published an updated version of GPT-3[17].

GPT4: The newest model in the GPT series, the GPT-4, debuted on March 14, 2023[18]. It represents a major improvement over the GPT-3, the previous model, which was already excellent. Even while the model's architecture and training set's characteristics haven't been made public, it definitely improves upon GPT-3's advantages while also overcoming some of its drawbacks. GPT-4 is only available to ChatGPT Plus subscribers, although there is a used cap. It may also be accessed by adding your name to the GPT-4 API queue, albeit this may take some time given the number of applicants. However, utilizing Microsoft Bing Chat is the simplest approach to obtaining GPT-4. There is no need to get on a waitlist, and it is totally free. The multimodal functionality of GPT-4 is a notable feature. As a result, the model can now receive input in the form of images and interpret them similarly to text. For instance, an OpenAI engineer supplied the model with a picture of a hand-drawn website prototype during the GPT-4 launch live broadcast and the model shockingly generated functioning code for the website. Additionally, the model performs at a human level on a number of conventional and professional standards and is better able to comprehend complicated requests. It also includes a bigger context window and context size, which relate to the amount of information the model can hold in memory throughout a chat session. GPT-4 is redefining what is now feasible with AI technologies, and it is expected to find use in a variety of different sectors. But just like with any effective technology, there are worries about possible abuse and moral ramifications of such a great instrument.

V. GPT vs BERT The original plan for GPT was to pre-train a network on language modeling tasks across a substantial amount of text before fine-tuning the network for various language activities[23]. In this manner, we would be able to train the model unsupervised—that is, without any labeled data created by humans—while yet benefiting from supervised training over smaller labeled data for particular tasks. Therefore, pretraining. Even more unexpected was the fact that language modeling itself evolved into a very potent instrument. The model would respond to simple questions about how to carry out tasks by surprising you with a reasonably sophisticated response.

The network was just taught to anticipate the next words in a text given as input, but depending on the content given, the model displayed intelligence that was unthinkable just a few years ago. Therefore, we began referring to the input text as a prompt. In addition to GPT, several other large language models (LLMs) have shown that once their size exceeds a certain threshold (often between 50 billion and 100 billion parameters), they display interesting features in their ability to answer queries. As a result, ChatGPT was further educated to make use of these newly discovered LLM characteristics.

VI. WORKING OF CHATGPT ChatGPT evaluates textual content with the use of a so-called neural network to enhance word choice[17]. Information encoding and decoding are all that is necessary, despite the fact that it could appear tough. To replicate how neurons communicate with one another in the human brain, "neural" networks of algorithms are created. ChatGPT was trained using genuine human interactions to learn how to predict outcomes and identify language patterns, much like our brains do when interpreting our surroundings. The real selling point of ChatGPT is its capacity to react to our woefully insufficient inquiries. The first step entails examining as much text that is available to the general public as is reasonably possible—basically, whatever it can locate online. According to Singh, the program "takes a sequence of words, conceals what the next word should be, and attempts to guess it." [19] If it guesses incorrectly, it sort of modifies itself to guess correctly. The language model[1] employs a reward model to distinguish between right and bad in order to properly construct sentences. This intuition was developed using human AI trainers who engaged directly with the language model, according to a recent Open AI blog article. Following the compilation of their answers to a particular topic and comparison with the AI-generated response, additional human trainers intervened to score the solutions according to accuracy. Through Proximal Policy Optimisation, a type of reinforcement learning that is a subset of machine learning, ChatGPT may adjust its language model using the data. ChatGPT began to move, but it wasn't yet capable of walking by itself. The fact that the internet isn't flawless is a major obstacle; even when the model constructs a whole phrase, it still needs assistance in separating reality from fiction. This is when the procedure's second stage enters the picture. There has to be more fine-tuning in order to steer the language model in the proper direction since it's not a good idea to take some of the opinions found on the internet at face value. It's crucial to keep in mind that ChatGPT would have only been trained using these papers. It wouldn't just copy and paste a web article if I asked it to write me an article about Barack Obama. It doesn't truly search for the answer when you ask it a question. Simply attempting to estimate what appears to be the right response. This is a dilemma because ChatGPT's own explanation explicitly states that there isn't any source of truth at the moment. They point out that if the language model is very conservative, it will just deny queries it can't, which eliminates the novelty.

VII. FEATURES OF CHATGPT We are all very aware of the groundbreaking developments in artificial intelligence that have led to the creation and introduction of chat GPT, a cutting-edge technology that produces and delivers output in a human-like manner using a combination of supervised and unsupervised learning. The following are only a few of its many benefits. 1. Self-Informing and Self-Rewarding – By using a self-informing and self-rewarding reinforcement learning system, ChatGPT can improve its language comprehension and response generation capabilities by transferring knowledge from one scenario to another, which causes it to continuously evolve. 2. Language Generation Capabilities- Language models, an artificial intelligence model that can produce and comprehend human-like answers and languages, are mostly used in chatgpt to create output. This model is trained on a lot of data using statistical methods in order to find trends and predict the word order in a particular sentence or context [23][24]. 3. The ability to solve problems consists of imaginative and intuitive talents, such as the ability to develop stories for various scenarios, come up with solutions to any issue, etc. 5. Customizability - ChatGPT can easily tailor replies to each user's needs and questions because of its prior experience learning feature, which helps it understand its users. 6. Scalability- The widespread use of technology in organizations makes it a perfect option for factories and industries, and it also helps in content search on massive data-containing websites, where scalability is essential. 7. Efficiency- ChatGPT is advantageous in generating responses quickly and efficiently due to which its efficiency increases and is efficient in generating valuable customer service in a short amount of time. 8. Future Prediction- ChatGPT can provide specific predictions about trends seen in historical data using statistical and pattern recognition techniques. It is also known as forecasting since it attempts to create a model and provide results based on it by analyzing historical data and data patterns. However, it should be highlighted that its forecasts are not always correct. 9. Security Provider- An article on ChatGPT that attempted to use human hiring to get around security utilizing image or text captchas and its image processing, word processing, and machine learning technologies was recently released. 10. Cost-Effective- Although GPT-3 is free for consumers, GPT-4 is a premium version that offers a more individualized solution. Whether a user has to purchase the premium version or not depends on their needs. 11. Emotionless- Despite the fact that ChatGPT lacks emotions, it can recognize them through language and react accordingly. Additionally, it might offer them psychological help. On the one hand, it offers a lot of benefits, but it also has certain drawbacks. VIII. LIMITATIONS OF CHATGPT 1. Ambiguous – Due to the pre-training modeling technique's knowledge limitations, the ChatGPT's data is out of date and causes confusion when producing answers [24]. 2. Unconfident – ChatGPT occasionally provides replies that make it challenging for the user to fully and accurately comprehend the context, which causes bewilderment for geeks.

Figure 7. Answer by ChatGPT

4. Lack of emotional comprehension - ChatGPT lacks emotional comprehension. The recommendations made by ChatGPT can occasionally be repetitious and irrelevant. ChatGPT creates interactions that lack empathy and personal connection.

Figure 8. conversation with ChatGPT

4. Vision Impairment - Since chatgpt is unable to view or interact with any website's websites directly, it is unable to resolve the captcha, which is used to prevent the entry of unauthorized or automated bots into the system. With the aid of a third-party provider, it can be resolved. 5. Biased response - Because ChatGPT is programmed to produce replies based on patterns and trends in the data it has been trained on, the accuracy of its output and the quality of its responses are both dependent on the learned data. If the trained data is biased, the output of ChatGPT will similarly be biased.

IX. FUTURE WORK Since Chat-GPT is an artificial intelligence model that uses language to learn, it should concentrate on enhancing user engagement. No linguistic barrier should exist while interacting with the user. ChatGPT ought to be able to speak with users in any language they choose, and it ought to provide more human-like replies. Some of the improvements or future work that can be done on chatgpt are: 1. Enhanced Natural Language Processing (NLP)- Improving ChatGPT's Natural Language Processing (NLP) capabilities would enable it to comprehend and interpret languages more like a person would, including idioms, slang, and colloquialisms. 3. Multilanguage Capabilities- It is important to train ChatGPT to communicate with individuals in a variety of languages. It will aid in extending ChatGPT's reach and usage. 3. Image recognition: This is a feature that can be added. It will make it possible for ChatGPT to read user-uploaded photos. Additionally, this will enable us to propose solutions based on photos and patterns, which will increase their effectiveness. [15] 4. Voice aid - Voice assistance may be added to ChatGPT to improve usability. People who are unable to express themselves verbally can do so via voice help. 5. Integration with Other Platforms - We can combine it with other platforms, such as chatbots powered by AI or social media platforms, to offer smooth and effective communication. 6. Text-to-speech conversion: We may provide the user functionality that will speak back to the response. 7. Image providence - The responses should also be provided as graphs, flowcharts, pie charts, etc. 8. Improving generation capabilities- Language models, which are artificial intelligence models that can produce and comprehend human-like replies and languages, are mostly used in chatgpt to create output. This model is trained on a lot of data using statistical methods in order to find trends. It can also predict the word order in a phrase or context. X. CONCLUSION The conclusion of this research paper summarizes the comprehensive analysis of ChatGPT and its advantages and disadvantages. ChatGPT was developed by Open AI that utilizes Generative Pre-Trained transformer (GPT) technology to fulfill a range of text-based requests. It highlights the strength of the ChatGPT model, its limitations, performance, and usability. ChatGPT is a pre-trained model that is not able to provide some answers correctly. ChatGPT is capable of understanding and interpreting user requests, generating appropriate responses quickly and efficiently in nearly natural human-like language, and completing advanced tasks such as Writing Assistance, Language Translation, General Knowledge, Personal Development, addressing productivity issues, etc. Overall, the paper offers valuable insights into the current state of NLP research and the potential for language models to advance the field in the future.

XI. REFERENCES

[1] Ashish Vaswani, Nikki Parmar, Llion Jones, Noam Shazeer, Jakob Uszkoreit, Aidan N. Gomez, Łukasz Kaiser. Attention is All You Need. arXiv:1706.03762v5 [cs.CL] 6 Dec 2017. [2] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, Improving Language Understanding by Generative Pre-Training [3] Jun-Jie Zhu et al. – A conversational AI that will fundamentally change the way we find information. 21st March 2023. [4] Sakib Shahriar, Kadhim Hayawi, Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations [5] Suzanne Fergus, Michelle Botha, and Mehrnoosh Ostovar, Evaluating Academic Answers Generated Using ChatGPT [6] Justin Paul, Akiko Ueno, Charles, Evaluating Academic Answers Generated Using ChatGPT [7] Ali Iskender, Holy or Unholy? Interview with Open AI's ChatGPT. 2023 [8] Grant Cooper, Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. [9] Gunawan, J. (2023). Exploring the future of nursing: Insights from ChatGPT model. Belitung Nursing Journal, 9(1), 1-5. <https://doi.org/10.33546/bnj.2551> [10] Md. Asrafu Haque-, A Brief Analysis of "ChatGPT" – A Revolutionary Tool Designed by OpenAI., (29 March 2023) [11] Renqian Luo 1, Liai Sun2, Yingce Xia 1, Tao Qin 1, Sheng Zhang 3, Hoifung Poon 3 and Tie-Yan Liu1, BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining. [12] Dinesh Kalla Nathan B. Smith, Study, and analysis of ChatGPT, explore its origins, how it works, and its impact on different fields of study, March 2023. [13] Jun-Jie Zhu et al , ChatGPT and Environmental Research, 21st March 2023 [14] Viriya Taecharungroj et al, "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter, 16th February 2023 [15] Tyna Eloundou1, Sam Manning1,2, Pamela Mishkin*1, and Daniel Rock, OpenAI, University of Pennsylvania, GPTs are GPTs: An Early Look at the Labor Market Impact, Potential of Large Language Models, arXiv:2303.10130v4 [econ.GN] 23 Mar 2023 , March 27, 2023, [16] Alec Radford * 1 Jeffrey Wu * 1 Rewon Child , David Luan 1 Dario Amodei ** Ilya Sutskever , Language Models are Unsupervised Multitask Learners [17] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Sandhini Agarwal, Pranav Shyam, Arvind Neelakantan, Language Models are Few-Shot Learners [18] Steven Basart, Sophie Duba, Cèsar Ferri, Heather Frase, Gavin Hartnett, Jake J. Hecla, Dan Hendrycks, Jose Hernandez-Orallo, Alice Hunsberger, Rajiv W. Jain, OpenAI, GPT-4 Technical Report, arXiv: submit/4812508 [cs.CL] 27 Mar 2023 [19] Molly Ruby, How ChatGPT Works: The Model Behind The Bot, 31 January 2023, <https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286> [20] Long Ouyang* Jeff Wu* Xu Jiang* Diogo Almeida* Carroll L. Wainwright* Pamela Mishkin* Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Training language models to follow instructions with human feedback [21] John Schulman, Barret Zoph, Christina Kim, Jacob Hilton, OpenAI 2022 - ChatGPT: Optimizing Language Models for Dialogue, <https://openai.com/blog/chatgpt>. [22] Marco Ramponi, How ChatGPT actually works, limitations and features, Dec 23, 2022 [23] . Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 conference of the north american chapter of the Association for computational linguistics: Human language technologies, volume 1 (long and short papers), 2019, pp. 4171–4186 [24] OpenAI, "How should AI systems behave, and who should decide?." <https://openai.com/blog/how-should-ai-systems-behave>, Feb. 2023.

2

Hit and source - focused comparison, Side by Side

Submitted text As student entered the text in the submitted document.

Matching text As the text appears in the source.

1/4	SUBMITTED TEXT	56 WORDS	100% MATCHING TEXT	56 WORDS
	ChatGPT is a spinoff of InstructGPT, which introduced a novel approach to incorporating human feedback into the training process to better align the model outputs with user intent. Reinforcement Learning from Human Feedback (RLHF) is described in depth in OpenAI's 2022 paper Training language models to follow instructions with human feedback and is simplified below.		ChatGPT is a spinoff of InstructGPT, which introduced a novel approach to incorporating human feedback into the training process to better align the model outputs with user intent. Reinforcement Learning from Human Feedback (RLHF) is described in depth in openAI's 2022 paper Training language models to follow instructions with human feedback and is simplified below.	
	<div>W</div> https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286			

2/4	SUBMITTED TEXT	21 WORDS	70% MATCHING TEXT	21 WORDS
	fine-tune the GPT-3 model by employing 40 contractors to provide a supervised training dataset, in which each input has a		fine-tuning the GPT-3 model by hiring 40 contractors to create a supervised training dataset, in which the input has a	
	W https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286			
3/4	SUBMITTED TEXT	86 WORDS	97% MATCHING TEXT	86 WORDS
	Step 2: Reward Model After the SFT model is trained in step 1, the model generates better-aligned responses to user prompts. The next refinement comes in the form of training a reward model in which a model input is a series of prompts and responses, and the output is a scaler value, called a reward. The reward model is required in order to leverage Reinforcement Learning in which a model learns to produce outputs to maximize its reward (see step 3). Figure 3. Reward model [19]		Step 2: Reward Model After the SFT model is trained in step 1, the model generates better aligned responses to user prompts. The next refinement comes in the form of training a reward model in which a model input is a series of prompts and responses, and the output is a scaler value, called a reward. The reward model is required in order to leverage Reinforcement Learning in which a model learns to produce outputs to maximize its reward (see step 3). To train the model,	
	W https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286			
4/4	SUBMITTED TEXT	30 WORDS	50% MATCHING TEXT	30 WORDS
	the computer has learned to adopt to accomplish its objective, in this example, maximizing its reward. A scaler reward value is then established for the prompt and answer pair		the machine has learned to use to achieve its goal; in this case, maximizing its reward. Based on the reward model developed in step 2, a scaler reward value is then determined for the prompt and response pair.	
	W https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286			