ChatGPT: Bright or Dark Side of AI?

Sugam Sehgal
[1]Third year Law Student, BBA LL.B.
[1]Bharati Vidyapeeth's New Law College, Pune, India

Abstract - Artificial Intelligence (AI) is expeditiously developing our world and has the latent ability to bring notable benefits to society. However, the introduction and implementation of AI systems also create risks. One such AI system is ChatGPT, the latest iteration of the Generative Pre- trained Transformer language models which is developed by OpenAI. While ChatGPT has received significant attention in the past few months for its advanced language generation capabilities but can be misused in many unknown ways. This paper aims to provide an overview of the AI system (ChatGPT), and its potential misuse, including the spread of misinformation, bias amplification, and cybersecurity risks. It highlights the importance of considering these potential risks and developing responsible AI practices to ensure that AI systems are used ethically and for the benefit of society.

Index Terms - ChatGPT, AI system, Cybersecurity, Cyberspace, Human Intelligence, Artificial Intelligence

I. INTRODUCTION

In this 21st century, humans have arrived at a stage where artificial intelligence has made such progress that a question of knowledge has arisen between artificial and human intelligence [1]. But as we know that human intelligence is the one that bought us into the era of Artificial Intelligence. Similarly, bots have been developed, which are a combination of machine and software and act as a computer program for conversation through voice or text commands. Nowadays, the most alluring field of study for scientists is machine learning, a sub-field of AI. Machine Learning is being used by scientists for many decades, to learn the pattern and relationships in the data and how can it be applied to generate new text, recognize images and speech, sum it all up, to analyze data.

In 1964, researchers at the Massachusetts Institute of Technology (MIT) were working on a computer program so ahead of its time that it would allow seamless communication between humans and machines.

In the next two years, they built ELIZA, an app that would set the foundation for all future chatbots. The introduction of keyboard-enabled responses made ELIZA so impressive. It was the first time, that the program was able to understand human input, which was a big success. By 1995, another language processing bot, ALICE, came out and was followed by Smartchild in 2001, setting the stage for all the current generation chatbots. However, a decade later, the AI chatbots like Amazon, Alexa, Google Now, and Siri took the world by storm [2].

An American AI research laboratory called 'OpenAI' introduced ChatGPT, an advanced conversational AI chatbot that made Google declare a code red![3]

In 2015, Sam Altman, the Current CEO of OpenAI, along with others founded the firm originally as a non-profit, but gradually due to its research & development, it has converted to a for-profit.

ChatGPT is a transformer architecture trained on a massive amount of text data available on the internet, i.e., 570 GB, to generate the responses fundamentally.

It is a type of neural network that was introduced in 2017 in a paper titled 'Attention Is All You Need.' During its initial stages of training, this allows the model to better understand the context and meaning of the input and to generate conversational responses. The capacity of ChatGPT to produce responses like a human conversation is one of its main features.

II. HOW IS CHATGPT DIFFERENT FROM OTHER AI CHATBOT?

We've seen a lot of movies on science fiction which shows how AI is rising and how the world will be look in future. For e.g., in the movie 'Iron Man,' the character, Tony Stark, uses AI and his own intelligence to create an armour suit and uses it to save the world!

ChatGPT is called generative AI, it not only analyses the data but also generates new content with the help of the data inputted. For e.g., there's a novel *A- 'Game of Thrones'* and *novel B- 'Harry Potter'* and if we prompt ChatGPT that we want a book confluence of A&B, it has the capability to generate this kind of new data, i.e., *novel AB* from the data existing in its software.

Elon Musk says, *"ChatGPT is scary good but before we come on to ChatGPT, it's important to note that it's not the first go at an AI chatbot."*

In 2016, Microsoft launched *Tay* but within the first 24 hours, Twitter users taught it rude, racist, and misogynistic language which led to its downfall.

In August 2022, Meta also launched *BlenderBot* 3 but again it fell down for similar reasons as well as providing false information even to questions like who became president in 2020.

By observing the other chatbots, OpenAI is trying to avoid these downfalls by using API- Application Programming Interface, a moderation system that is one of the most important factors in developing a chatbot. API assists the AI developers to take action when things go against their content policy like illegal or unsafe information. Although it's not a hundred percent accurate, it certainly helps alongside.

GPT-3 is the third-generation large language model in the GPT series, introduced in May 2020, succeeding GPT-2 created by OpenAI, which does not come into much limelight. GPT-3 is considered state-of-the-art technology.

On March 15, 2022, OpenAI launched new versions of GPT-3 and Codex in its API with edit and insert capabilities under the names "text-DaVinci-003" and "code-DaVinci-002." It was more successful than the earlier generations.
On November 30, 2022, OpenAI began referring to these models as belonging to the generation-"GPT-3.5" series and released ChatGPT, which again is not the same as GPT-3, but a fine-tuned from a model in the GPT-3.5 series.[4]

ChatGPT [5] was launched on the 30th of November 2022 and surpassed over 1 million users in just five days, which is phenomenal! It can take simple text descriptions like 'two girls partying" and convert the text into photorealistic images that do not exist in the real world.
A statement was made by OpenAI, which says- *"We've trained a model called ChatGPT which interacts in a conversational way. The dialog format makes it possible for ChatGPT to answer follow-up questions admit its mistakes, challenge incorrect premises and reject inappropriate requests."*
ChatGPT was first opened up for beta testing in November 2021 and its launch in 2022 [6] shows improvement and avoids giving harmful outputs. Its software has been trained by AI and ML with text data available on the internet and overall it learns from human feedback, i.e., reinforcement learning from human feedback (RLHF)
The first immediate consequence of ChatGPT was that its website had frozen due to the demand of over 1 million users, which forces OpenAI to swiftly secure its AI system.

CHATGPT: IF USED IN GOOD FAITH

- It allows users to communicate with AI which seems to be natural dialogue. While using, it also helps in correcting the grammar.
- It helps coders in writing any *xyz* code and also fixes bugs in the code.
- It helps in converting-
  - Text to image
  - Movie to emoji
  - Difficult text to simpler text
  - Any Mathematical Equation
- From an educational point of view, it helps the students in solving their doubts or question.
- It helps in answering any kind of medical question, which can help medical students in their thesis or experiments.
- To develop and improve cybersecurity tools, such as natural language-based security systems and chatbots for threat detection and mitigation.
- ChatGPT can potentially be used, for enhancing search engine functionality, automating processes like disputing traffic tickets or bills, and creating code, among many other applications.
- It can be used in making CV, cover letter and even one can practice to prepare for interview.
- Students can use it as a reading tool to convert difficult text into simpler one.
- It has the ability to write unique poems and songs, although not able to win Grammy but
- with right tune.
- It can generate text in N number of languages.
- One can use it to improvise his/hers conversational skills.
- It can be used for customer service, virtual assistance, or other tasks.
- It can be used to create unique content by social media influencers.
- It can be used to know what are the best personality traits to have for stock investing

CHATGPT: IF USED IN BAD FAITH

As Stephen Hawking, Theoretical Physicist said: "*The rise of powerful AI will be either the best or the worst thing ever to happen to humanity. We do not yet know which."*
Although ChatGPT uses moderation API and rejects the prompt which violates its content policy, the data scientists and engineers find a way to alter the coding by jailbreaking ChatGPT and asking questions like, how to make a Molotov Cocktail or nuclear bomb, generating arguments in the style of a Neo-Nazi, to make inflammatory statements and many other ways. Also, it has been officially declared [7] by the US that it can pass the Bar exam, US medical exam or any other exam. This is a threat to the economy as the students are the youth of the country and if they will be using such tools for cheating, such professions will no longer be trusted.
Since the use of ChatGPT is now made available to the public for free, OpenAI will get to know in what other way ChatGPT can be misused (which they couldn't think of) and then they will find a way to curb it. As we know, after in-house testing of tech, it is then released for beta testing, which is most crucial in the launch of an AI. Because, after beta testing only, ChatGPT's potential privacy policy was improved when they got to know that the engineers are hampering its authenticity by jailbreaking its coding to do malicious activities. Hence, it can be said that although ChatGPT is now posing a potential threat, gradually it can be improvised and used for many good causes.

GPTZERO: TO DETECT CHATGPT

In the academic world, teachers, scholars and academicians are not happy about it as students are using ChatGPT as a way of cheating in their examination, writing plagiarised essays and many more, which will in turn affect the academics of students in all. New York City Department of Education banned the use of ChatGPT in the computer network at public schools.[8]
In India, ChatGPT is restricted from using in the campus network of IITs as the students were copying the answers of coding, which is a crucial subject of engineering.
In order to retain the originality and authenticity of the assignments, essays and papers submitted by the students, Edward Tian, a senior undergraduate student at Princeton University, created a program, named *"GPTZero,"* which determines how much of a text is AI- generated, lending itself to being used to detect if an essay is human written to combat academic plagiarism.

CHATGPT: THREAT TO CYBER SECURITY

Artificial intelligence can be used for many things, but due to the mentality of the hackers, its true power lies in its ability to be used for every possible evil ways in the cyberspace. With AI tools, the possibilities, for manipulating and exploitering human, infringing its privacy and accessing its personal data, are infinite.

However, the same capabilities that make ChatGPT useful for these applications also make it a potential tool for cyber criminals. ChatGPT's ability to generate text can be used in a bad way by the cyber criminals in doing malicious cyber activities like,

- Accessing Personal Data
- Phishing Attacks and Malware
- Misinformation Campaigns
- Spreading false information: Its ability to generate realistic-sounding text can be used to spread misinformation and false news via cyber space.
- Impersonation: It can be used to create fake profiles or impersonate others online.
- Bias amplification: As it is trained on internet text data, it can amplify existing biases and perpetuate harmful stereotypes.
- Privacy violations: Its ability to generate text based on input data raises concerns about privacy violations and the potential use of the model for malicious purposes such as phishing attacks.
- Job displacement: Its advanced capabilities in natural language processing have the potential to displace human workers in certain industries.

CHATGPT'S EFFECT ON PRACTICE OF LAW [9]

AI is already widely used in the practice of law and has been for several years now. ChatGPT is more likely to be used in the field of law, given its demonstrated ability to write and generate texts like humans do (ironically, to fool humans into believing its output was produced by a fellow human). It is fair to say that AI has already thoroughly suffused the practice of law, like:

- To "translate" legal jargon into "plain English."
- It is useful in the provide legal services to those who cannot afford them.
- Lawyers can use AI program as they serve their clients to perform a variety of legal tasks, such as: produce relevant documents in discovery through the use of predictive coding; draft, review, and manage contracts; perform legal data analytics; predict judicial decisions;
- It can be used to review briefs for "strengths, weaknesses, patterns, and connections, and analyze the vulnerability of certain arguments.
- Judges also use AIML throughout the criminal justice system, for tasks such as helping to make bail determinations and also predicting the likelihood of recidivism as part of setting a carceral sentence.

III. CONCLUSION

ChatGPT is a powerful and advanced language model that has the potential to revolutionize various industries and fields, including cybersecurity. However, its capabilities also raise concerns about potential misuse and negative consequences. It is important for individuals, organizations, and governments to be aware of these risks and to implement measures to prevent misuse, such as responsible development and deployment of AI powered tools and increased education and awareness about the dangers of AI. Additionally, ongoing research and monitoring are needed to understand the full implications of ChatGPT and other advanced AI models and to ensure their use for the betterment of society and must not let them overpower our society. [10]

IV. REFERENCES

[1] Shivangi Sinha, Artificial Intelligence Vs Natural (Human) Intelligence- Global Challenge for Human Rights, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 7, 2019 (Special Issue), Research India Publications, https://www.ripublication.com/ijaerspl2019/ijaerv14n7spl_05.pdf

[2] Floatbot, https://floatbot.ai/blog/the-evolution-of-chatbots-from-origin-to-conversational-ai,

[3] Business Insider India, https://www.businessinsider.in/tech/news/googles-management-has-reportedly-issued- a-code-red-amid-the-rising-popularity-of-the-chatgpt-ai/articleshow/96407949.cms

[4] OpenAI, https://platform.openai.com/docs/model-index-for-researchers

[5] ChatGPT, https://chat.openai.com/

[6] The Times, UK, "ChatGPT is a stunningly lifelike conversational language system and the world's first truly useful chatbot. This all seems a bit too good to be true, we have to think about not only the limitations of the technology but also the challenges and threats that it poses to society."

[7] The Economic Times, https://economictimes.indiatimes.com/news/international/business/chatgpt-bot-clears- us-law-school-exam/articleshow/97328106.cms?from=mdr

[8] Vice, https://www.vice.com/en/article/y3p9jx/nyc-bans-students-and-teachers-from-using-chatgpt

[9] Amy B. Cyphert, University of California, Davis [Vol. 55:401] , "A Human Being Wrote This Law Review Article" ; https://lawreview.law.ucdavis.edu/issues/55/1/articles/files/55-1_Cyphert.pdf

[10] ibid, page 1 , reference [1]