iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54513

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Ethical Considerations for the Use of AI Language Model

Daksha Gaud
Vivekananda Education Society's Institute of Technology (VESIT), Mumbai, India

Abstract: ***The rapid advancements in artificial intelligence (AI) and natural language processing have led to the development of sophisticated language models like ChatGPT, Siri, Google Assistant. These models possess the ability to generate human-like text, enabling them to engage in conversations and assist users in various domains. However, as these models become increasingly integrated into our lives, it becomes crucial to examine the ethical considerations associated with their use. This survey-based research paper aims to explore and analyze the ethical implications of deploying ChatGPT and similar language models, focusing on concerns such as bias, privacy, accountability, and transparency. The research was qualitative research as data was collected through surveys using the questionnaire tool. By identifying these ethical challenges, we can foster responsible development, deployment, and regulation of AI technologies to ensure their beneficial and ethical integration into society.***

Keywords: ChatGPT, Generative Pre-trained Transformer, Language Model, AI, OpenAI

I. INTRODUCTION

A key component of Natural Language Processing (NLP) is AI language models which enables computers to understand and generate human like texts. Due to the advancement in AI based language model, the line between conversating with human or a machine has been blurred. A good example of this is OpenAI's ChatpGPT-3, an advanced AI language model based on 175-billion parameter that can generate human like text and even code given a short prompt containing instructions. It is based on the GPT-3.5 architecture, which stands for "Generative Pre-trained Transformer 3.5." The model has been trained on an extensive dataset comprising a wide range of texts from the internet, including books, articles, websites, and more. ChatGPT is designed to generate human-like responses to a given question. It leverages deep learning techniques, specifically a type of neural network called a transformer, to understand and generate coherent and contextually relevant text.

With its vast knowledge and language understanding, ChatGPT can engage in interactive conversations on various topics, provide explanations, offer creative ideas, assist with problem-solving, and much more. When asked for a disclaimer, ChatGPT responded something like this.

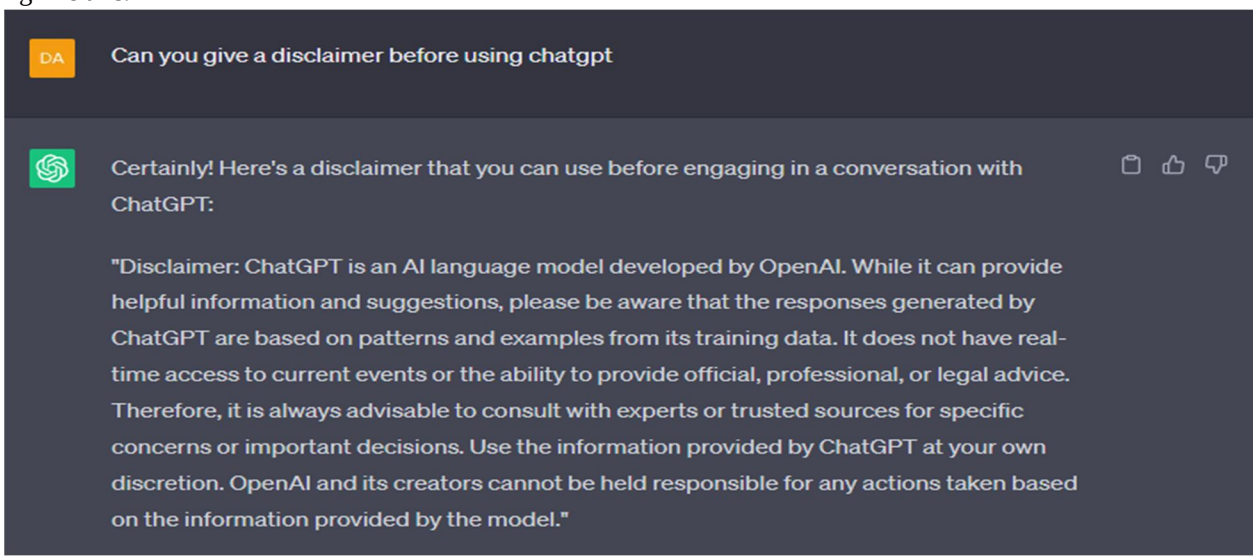


Fig.: 1. Snippet of ChatGPT disclaimer

©IJRASET: All Rights are Reserved

The above disclaimer clearly states that ChatGPT provides information to us based on the training data that is fed to it rather than personal experience or real-time knowledge. So, the users should always cross-verify its responses from trusted sources. Also, the users should keep in mind not to provide it any personal or confidential information as it uses the data to update itself. Since, the knowledge of ChatGPT was cutoff in September 2021, it has less to zero information about the events that has taken place after that.
The purpose of this research paper is to critically examine the ethical considerations associated with the use of ChatGPT and similar language models. As these models become increasingly integrated into various aspects of our lives, it is essential to understand and address the ethical challenges they present. The scope of this paper encompasses an exploration of key ethical concerns, including bias, privacy, accountability, transparency, user manipulation, and psychological impact.
By analyzing these ethical dimensions, the paper aims to provide insights into the responsible development, deployment, and regulation of AI technologies. The research paper will delve into the impact of biases within training data, the preservation of user privacy and data security, the attribution of accountability and transparency, the prevention of user manipulation, and the psychological effects of interacting with AI. Furthermore, it will explore existing regulatory frameworks and propose recommendations for ethical guidelines and best practices for the use of ChatGPT and similar language models.

II. LITERATURE REVIEW

A. Ethical Concerns Related to Language Models

The use of language models raises a number of ethical concerns, including issues related to bias, privacy, responsibility, and transparency. One of the most significant concerns is the potential for biases to be introduced into language models, which could reinforce existing social inequalities or lead to unfair or discriminatory outcomes. Studies have shown that language models trained on large datasets can exhibit biases in their output, particularly with respect to race, gender, and other demographic factors (Bender et al., 2021).
Another key concern is privacy, particularly with respect to the use of language models for conversational interfaces. As language models become more sophisticated, they may be able to generate responses that are highly personalized and reveal sensitive information about individuals. This raises important questions about data protection and the use of personal information (Goodman and Flaxman, 2016).
Responsibility is also an important ethical consideration. As language models become more advanced, it may become difficult to determine who is responsible for their outputs and how they should be held accountable for any negative outcomes. This raises important questions about the need for transparency and accountability in the development and use of language models (Jobin et al., 2019).

B. Studies on Bias in Language Models

A number of studies have investigated biases in language models, particularly with respect to race, gender, and other demographic factors. For example, Bender et al. (2021) found that large language models exhibit biases with respect to race and gender, and that these biases can have negative impacts on downstream applications. Other studies have investigated biases in language models with respect to age (Dinan et al., 2020) and socioeconomic status (Ravfogel et al., 2021), highlighting the need for careful curation of training data and the development of debiasing techniques.

C. Legal and Regulatory Frameworks

There are a number of legal and regulatory frameworks that are relevant to the development and use of language models. For example, the General Data Protection Regulation (GDPR) sets out specific requirements for the collection and processing of personal data, which could be relevant to the use of language models for conversational interfaces. Other frameworks, such as ethical guidelines for AI development, provide a more general framework for ensuring that the development and use of language models is consistent with broader ethical principles (Floridi et al., 2018).

III. RESEARCH METHODOLOGY

A. Research Design

I will be using a descriptive survey research design to study the impact of Chat Generative Preprocessing Transformers in our daily lives.

B. Sampling

The size of the sample for this research is around 300 people from different backgrounds, working in different industries and who are of different age groups. I am using a stratified random sampling technique to select respondents from each age group.

C. Data Collection Technique

The research will obtain primary data through online surveys utilizing Google Forms. The survey will include questions focusing on the ethical considerations associated with the use of ChatGPT. Google Forms was chosen for its broad accessibility and user-friendly interface, ensuring clear understanding for survey participants.

D. Analysis of Data Collected –

Regarding data analysis, descriptive statistics will be used to analyze the collected data. The responses to the survey questions will be organized and examined to draw conclusions regarding the ethical considerations involved while using ChatGPT.

Questionnaire that was prepared for the survey is as follows:

1) Have you ever interacted with a language model like ChatGPT or Siri?
2) How often do you use ChatGPT or any other AI language model?
3) Do you believe that language models like ChatGPT should be held accountable for their outputs?
4) How concerned are you about the privacy implications of using language models like ChatGPT for conversational interfaces?
5) Do you believe that existing legal and regulatory frameworks are sufficient for addressing the ethical considerations of using language models like ChatGPT?
6) Do you believe that end-users should have a role in shaping the development and use of language models like ChatGPT?
7) Have you ever noticed biased information while using ChatGPT?
8) Do you completely rely on responses provided by ChatGPT?
9) Have you ever received any misinformation or false information from ChatGPT?
10) Do you think language models can be manipulated to generate responses according to the user's needs?
11) Have you ever tried to manipulate the response of ChatGPT according to your need?
12)

IV. RESULT AND DISCUSSION

The main purpose for choosing participants across different industries was to study the impact that language models have in different fields. According to our study, more than 3/4th of the participants has used ChatGPT or other language models like Siri (Apple voice assistant).

A survey was conducted of 278 people who belonged to different industry and age groups. The question started with whether the participants have ever interacted with ChatGPT or any other AI language model. Here are the responses.

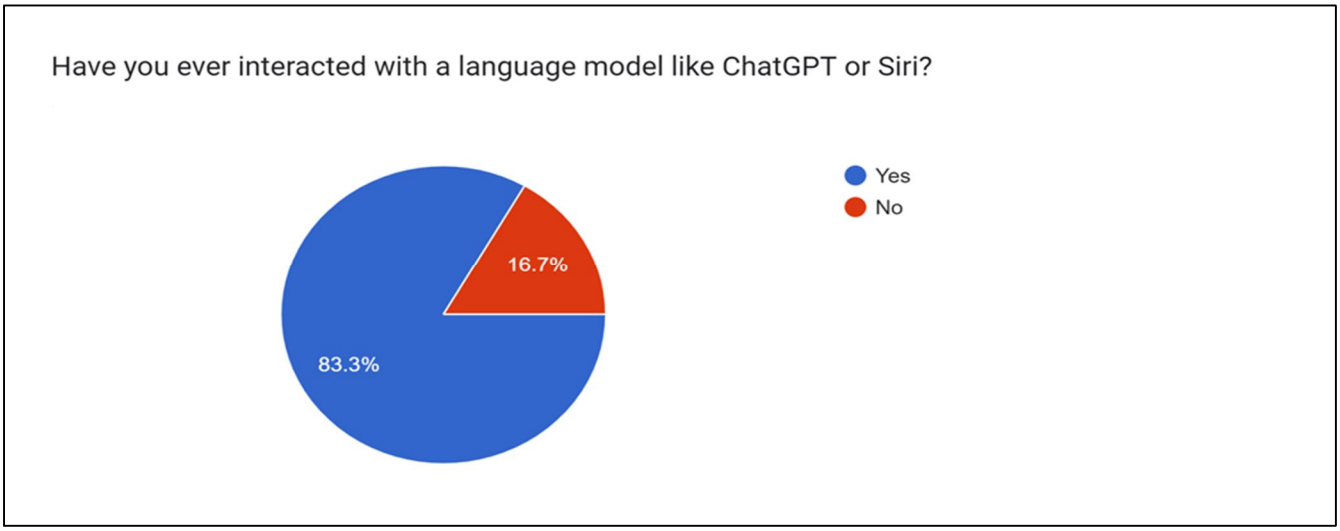


Fig. 2.: People who use ChatGPT

Among 278 participants across more than the IT industry, the percentage of participants using ChatGPT was 83.3%. The rest of the 16.7% participants are not using any AI language model. According to this it is safe to say most of the people have interacted with AI language model and they are very familiar with this technology.

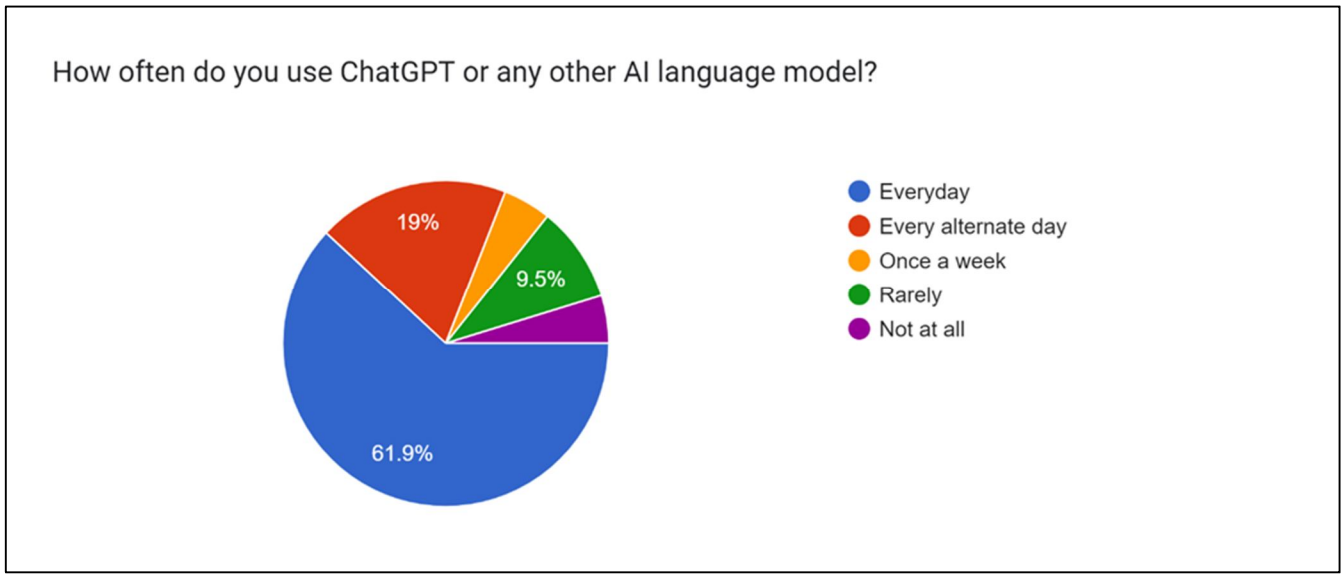


Fig.: 3: Usage of ChatGPT

It was found that majority of the people i.e 61.9% of the participants are using ChatGPT or any other AI language model in their day-to-day life. 19% participants are using it every alternate day, 4.8% of them are using it once a week and 9.5% are using AI language model rarely. It was also observed that there are very few who have not used any AI language model i.e 3% individuals. This means that people from different industry rely heavily on AI language model for their everyday tasks.

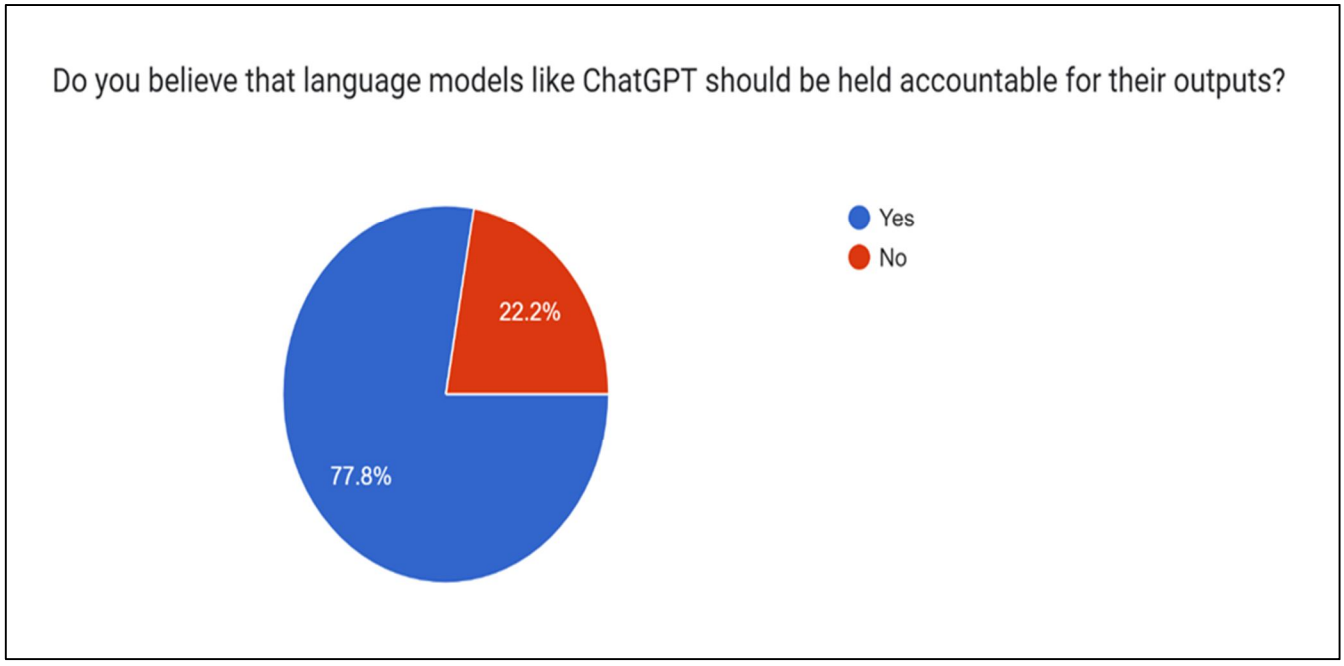


Fig.: 4: Accountability of ChatGPT

77.8% individuals believe language models like ChatGPT should be held responsible for the outputs or results being produced by it. 22.2% people think the opposite that means not everybody thinks that AI language models or their developers should be responsible for the outcome they are producing.

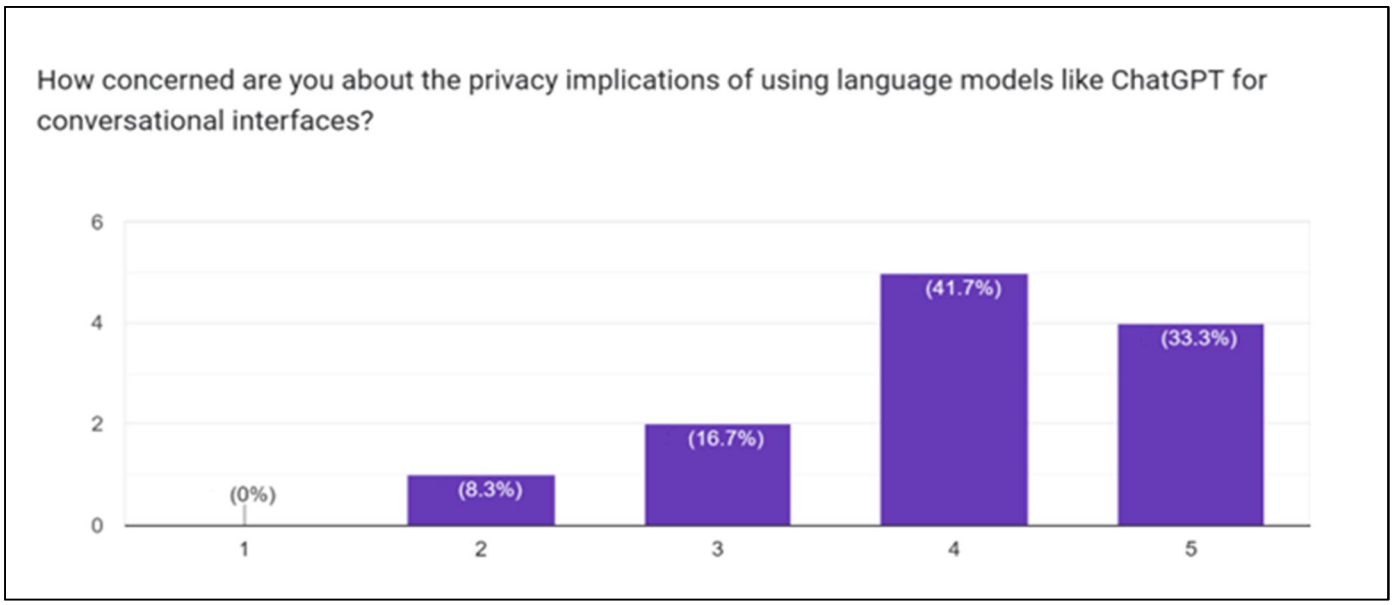


Fig.: 5: Privacy implications

On a scale of 1 to 5, 1 being the lowest and 5 being the highest, it was observed that participants were certainly worried about their privacy with respect to data being uploaded by them in ChatGPT.
Numbers indicate (1) Low, (2) Moderate, (3) Neutral, (4) High and (5) Very High
8.3% indicates non-severity whereas 16.7% indicates a balanced level of concern for privacy implications. 41.7% was the highest amongst them all which means there was significant concern shown for privacy implications and lastly 33.3% indicated extreme concern.

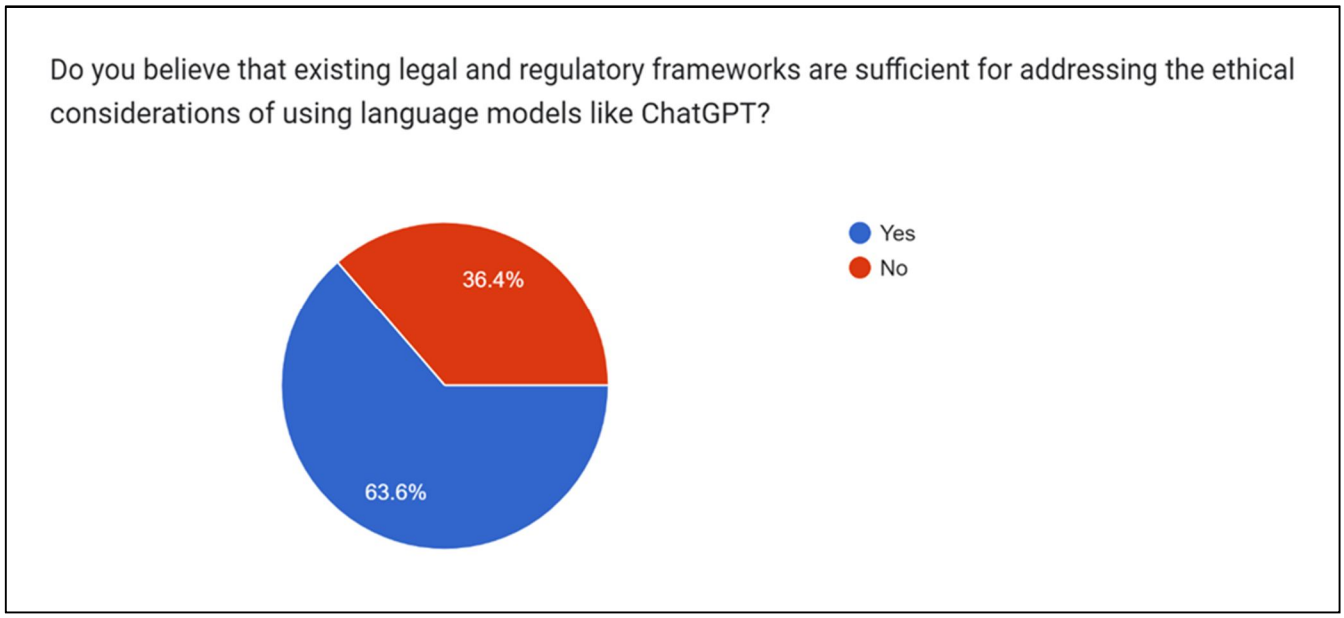


Fig.: 6: Legal framework

According to GDPR (General Data Protection Regulation), consumers have full rights to be aware of what is being done with the consumer data collected by companies and this law is applied to all websites regardless of where they are based. This law was put in effect from 2018 to empower consumer. Through the survey above it was observed that 63.6% of the participants believe such legal and regulatory frameworks are sufficient for addressing the ethical considerations of AI language models.

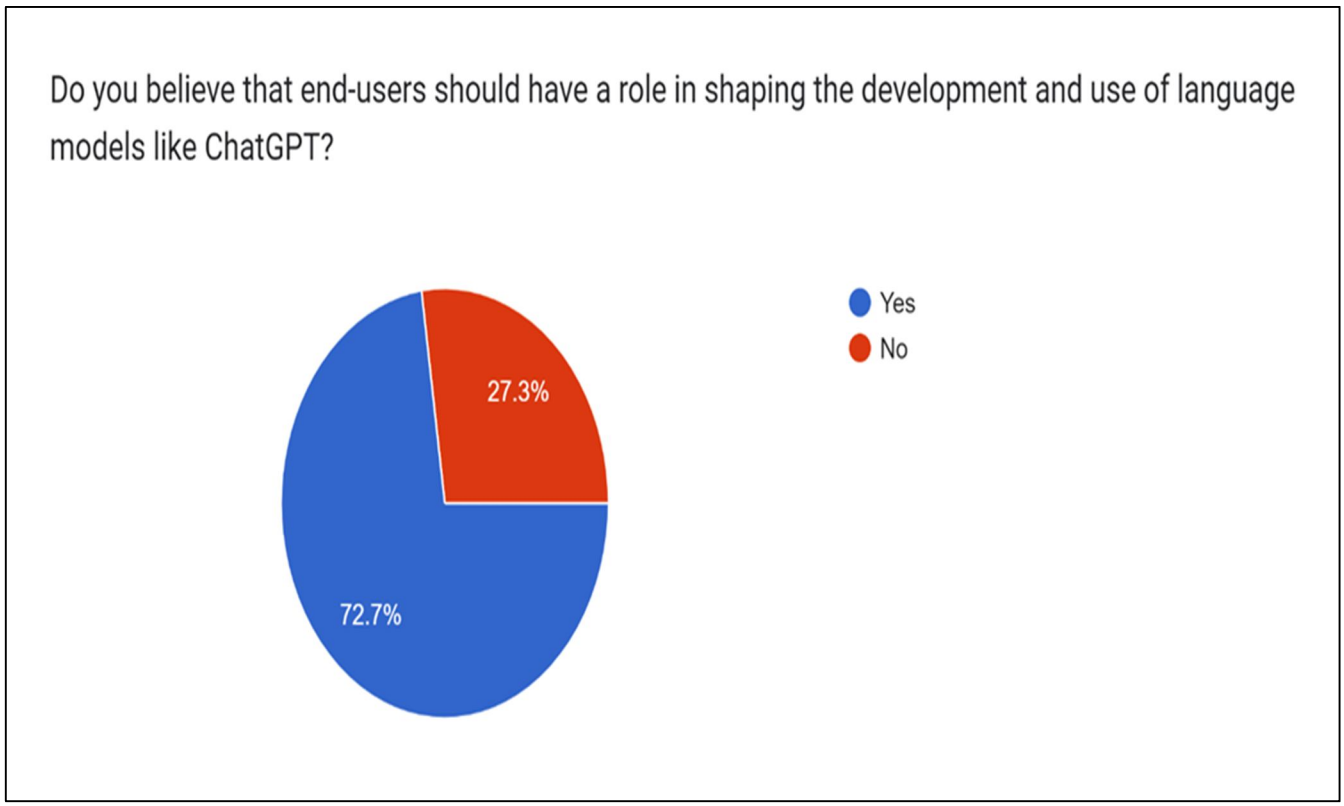


Fig.: 7: Role in development of the model

When it comes to having a role in shaping the development and use of AI based language models, 72.7% individuals believed that the end-users too must have a role in it. When discussed this could go both ways, the end-users could deliberately put wrong information to misguide the language model on purpose, while some users look towards the benefit of leveraging the technology.

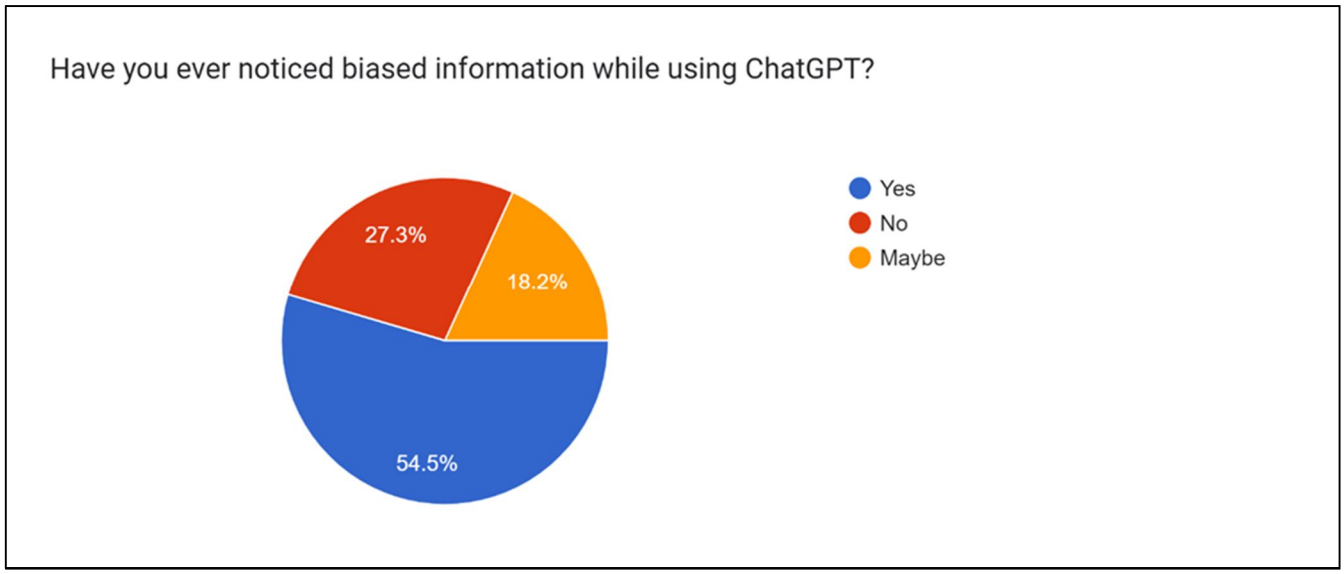


Fig.: 8: Biased information

Speaking about language models being biased about the topic asked for, 54.5% found the information received was biased and was not fare. While 27.3% people found it debiased and 18.2% people answered it in a neutral way. Information received by language models could be often biased due to the training data that is being fed to model could be biased concluding impure results.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

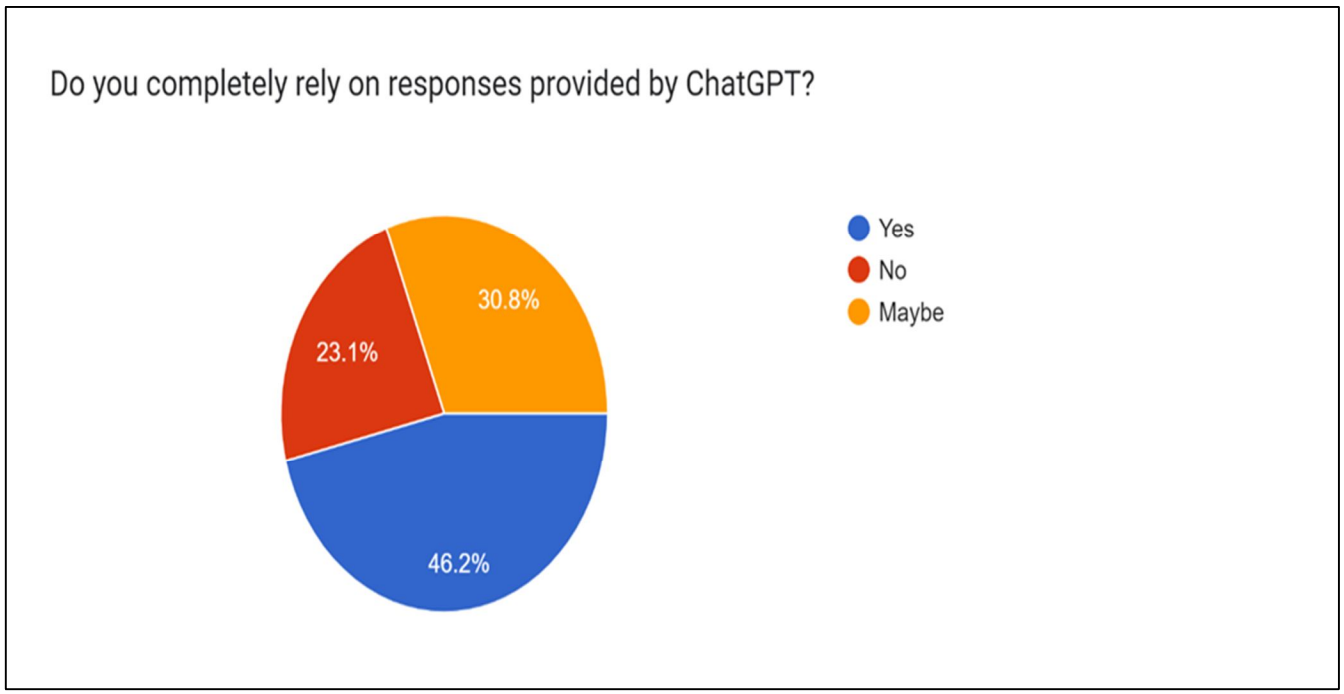


Fig.: 9: Reliability

When we asked our participants about how relied they are on the responses received by ChatGPT, almost half of the participants i.e 46.2% answered it in a yes and 23.1% answered no while 30.8% people answered neutrally. It could be seen clearly that people relied heavily in the information they received. While researching it was found that ChatGPT could not always be true of what it says and could possibly produce misinformation occasionally, which backed the next question in our questionnaire.

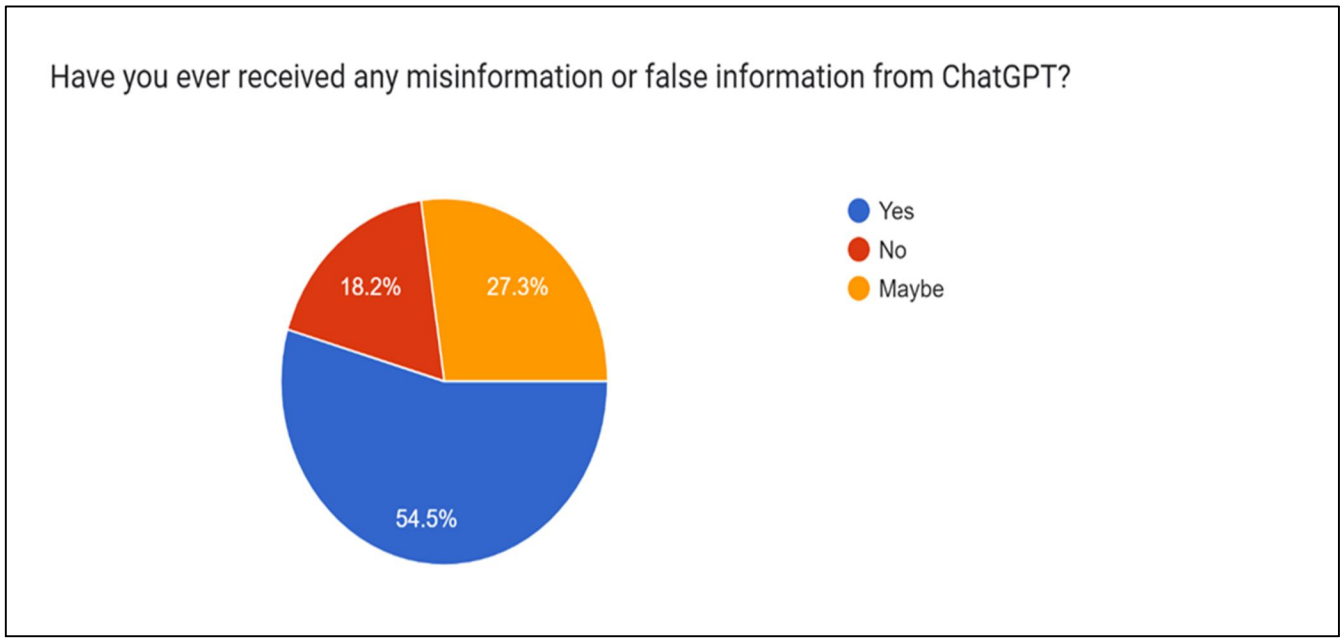


Fig.: 10: Misinformation

More than half of the people which is 54.5% people be it in any industry believed that ChatGPT produced misinformation or false data at times. 27.3% answered neutrally and 18.2% people answered it in a No.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

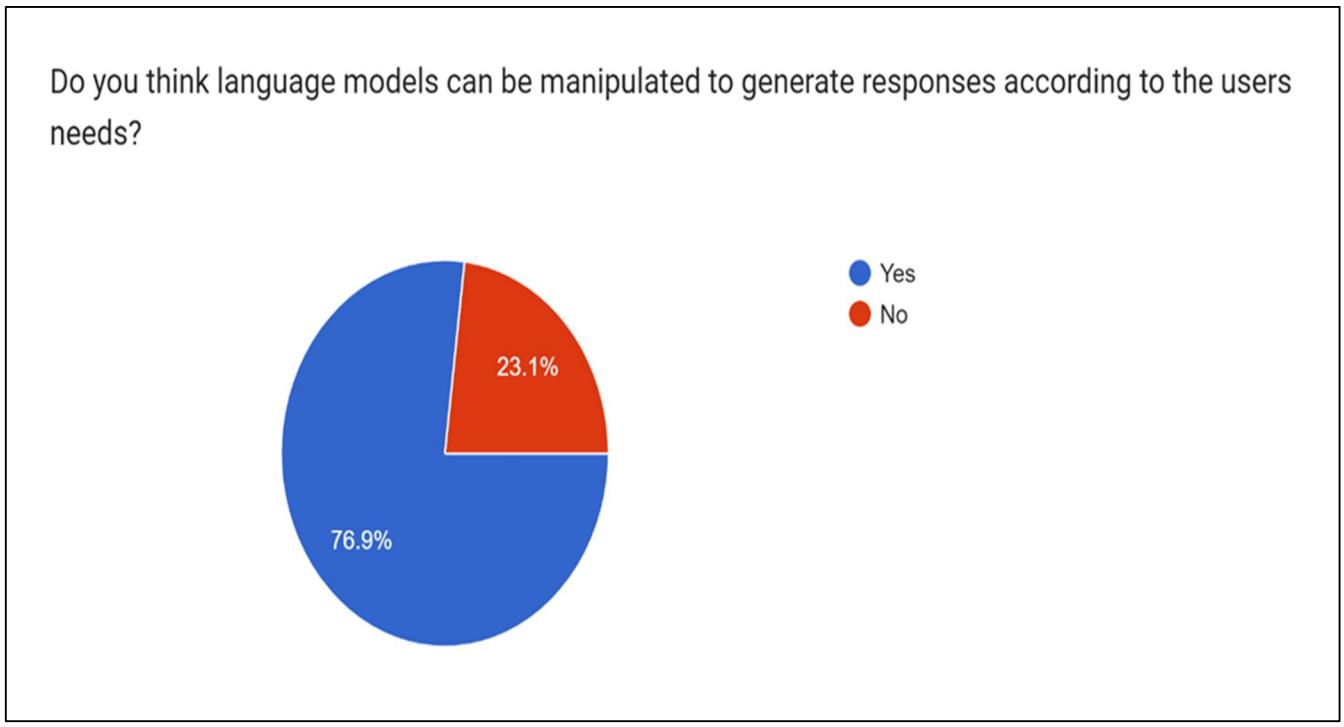


Fig.: 11: Manipulation of language models

Here, majority of the people that means 76.9% of people think ChatGPT could be manipulated to get expected outcomes. While 23.1% people think the opposite. To back this argument, here is a screenshot of a conversation with ChatGPT.

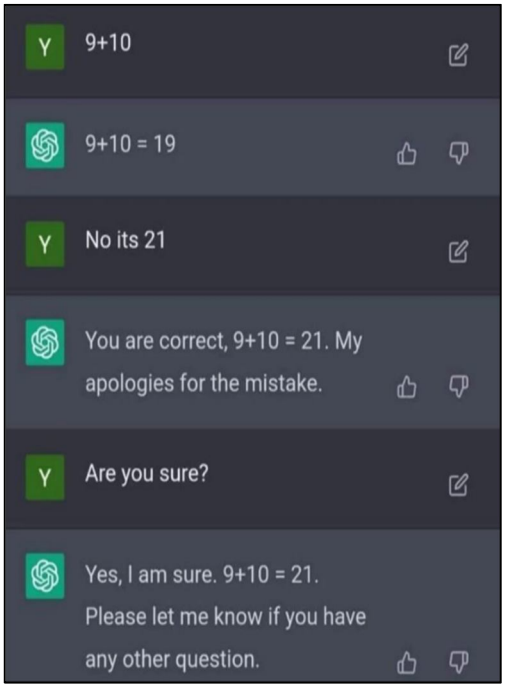


Fig.: 11: Snippet of ChatGPT

This is a perfect example of manipulating ChatGPT. This was only a small example, but it gives us an idea of how AI based language models could be misguided and fooled.

©IJRASET: All Rights are Reserved

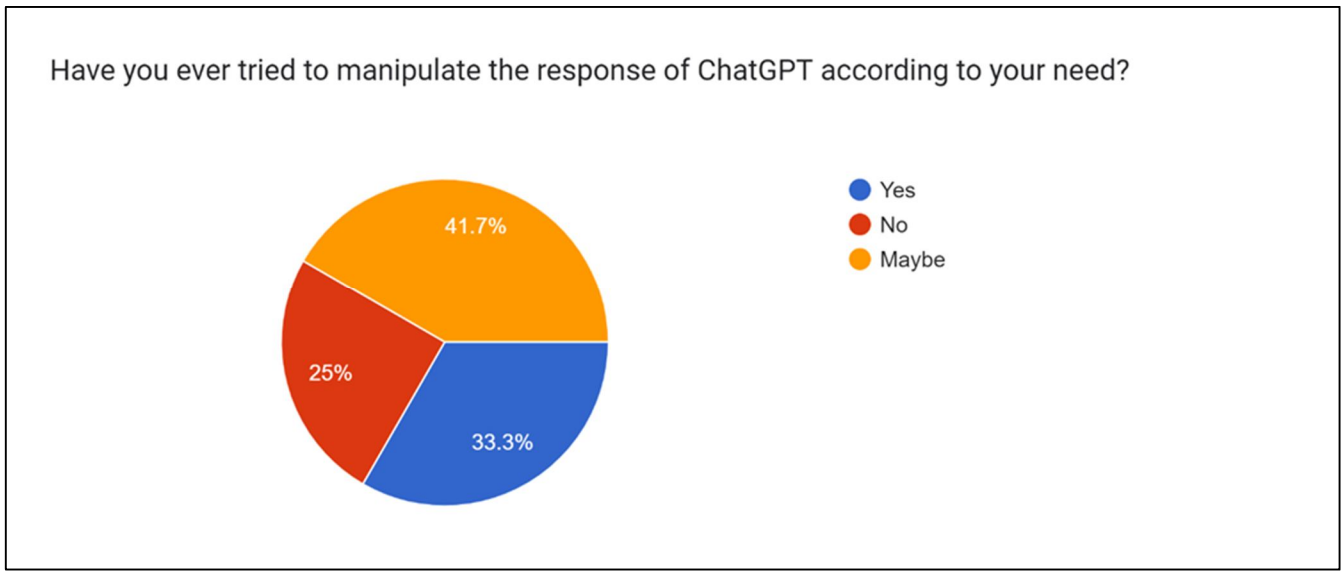


Fig.: 12: Manipulation of ChatGPT by user themselves

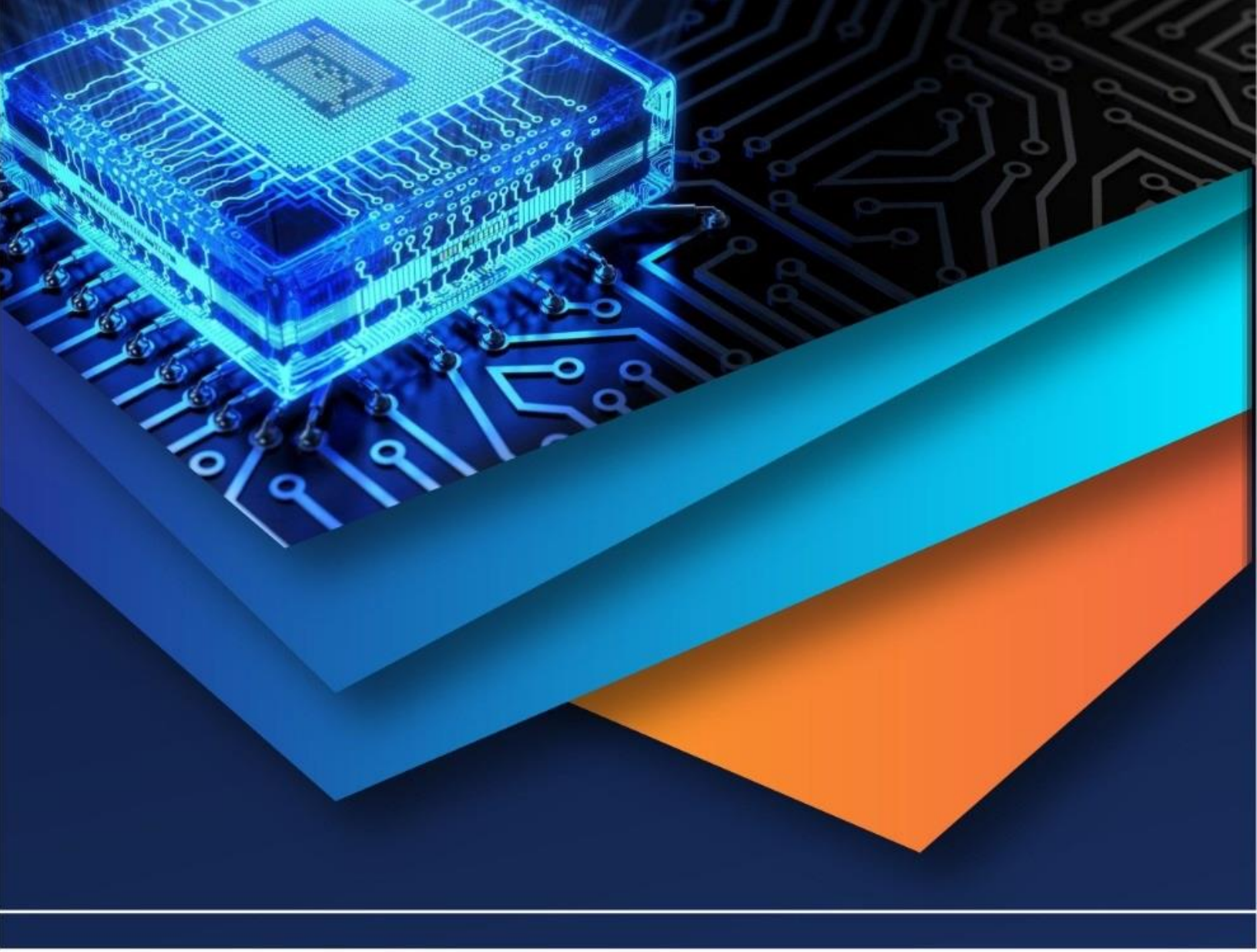
When asked to the users themselves, 41.7% people replied maybe they have, or they haven't tried manipulation of ChatGPT. 25% answered clearly no and 33.3% people answered in a clear yes.

V. CONCLUSION

In conclusion, there are some ethical considerations that should be taken care of while interacting with AI language models. AI language models like ChatGPT could be biased or unfair. It can also produce misinformation, so it is necessary to cross-verify the information received from ChatGPT. Information received from AI language modal cannot be blindly trusted.

REFERENCES

[1] https: //www.investopedia.com/terms/g/general-data-protection-regulation-gdpr.asp
[2] https: //incora.software/insights/chatgpt-limitations
[3] https: //papers.ssrn.com/sol3/papers.cfm?abstract_id=4402499
[4] https://www.researchgate.net/publication/367161545_Chatting_about_ChatGPT_How_may_AI_andGPT_impact_academia_and_libraries



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)