Human Perception of the Sentience of ChatGPT and Artificial Intelligence

Aarushi Gupta

1) Introduction

In the 21st century, Artificial Intelligence (AI) is increasingly known for its role in enhancing social media, such as in the form of chatbots such as My AI in SnapChat, and in the form of voice assistants on different platforms. However, AI has other applications such as in the fields of finance, customer service, and healthcare, among many others. AI can mimic human intellect and replicate the work humans do more precisely and quickly. Although AI has demonstrated its usefulness in various contexts, there has been some fear expressed by the general public that AI may take over certain jobs in the future, leaving humans unemployable (Stahl, 2022). Nevertheless, according to previous studies, while most people seem to trust AI, they do not think of it as a tool to displace the work of humankind just yet (Randieri, 2023). Outside of the trust factor, the consciousness of AI is a factor to be considered as well. Since AI is becoming more humanlike in its functions and responses during conversations with humans, the question of AI having a mind of its own arises. Studies have tried to examine the ability of `sentience` in AI and establish boundaries on where it 'thinks' and can drive actions (Shapiro, 2019). AI operates by utilizing and reproducing the data already fed in its system and its answers are pre-programmed by developers in such a way that it appears that people are conversing with a human. although it shows no signs of sentience or consciousness yet and is also completely pre-programmed, AI's latest large language model, a chatbot called ChatGPT, is creating a stir by its puzzling ability to replicate immensely humanlike and advanced answers.

ChatGPT was launched by OpenAI, an American AI research company, on November 30, 2022. It is a large language model, an algorithm which works by deep learning, and a process that allows artificial intelligence to ingest and process data as a human would. It utilizes deep learning through transformer neural networks, algorithms which focus on context and proceeding to process data through it (Hetler, 2023). ChatGPT can process questions from users and answer them. It is used as a conversational chatbot and is based on Generative Pre-Trained Transformer (GPT-3.5), another language model (Taecharungroj, 2023). ChatGPT operates by the process of web scraping, extracting content from the internet and other databases it has access to, and then providing users with the appropriate responses to the questions posed. ChatGPT has an immensely fast reaction time and its answers are even designed to be friendly and welcoming and to emulate a live conversation amongst humans. Its answers are only targeted and specific towards the question asked unlike other platforms such as Google, Yahoo, or Bing, which also use web scraping, but do not appear to be as intuitive. I have chosen to focus on ChatGPT for this study because of its distinctive features of this eminent chatbot. Further, these features prompt questions about whether it is sentient, actually thinking for itself, and conscious almost as if it perceives human like emotions on its own in the background. There is a need to explore if the users of ChatGPT also have this affiliated opinion, believing that using ChatGPT is similar to just talking to another human behind a computer screen, or in real life.

Some of the broader risks of AI misuse include algorithmic or deliberate deception, deep fakes and videos to spread misinformation, and promoting plagiarism, among others (Hick and Ziefle, 2022). This mix of positives and negatives again highlights the question on how risky yet adept of a tool AI agents can be, and how they should be wielded. Also, with such strong viewpoints towards ChatGPT, it is important to examine if people's impressions of it also influence their impressions of AI in general and how exactly people perceiving it as sentient could affect future usage.

The aim of this research project was to explore whether people view ChatGPT and AI as sentient or conscious, both, or neither. This study also aimed to address how perceptions of sentience of ChatGPT and AI may influence views of the future of AI. The project was exploratory in nature. Conclusions were drawn from the analysis through OLS regression and ANOVA.

2) Methodology

The study was approved by an independent Institutional Review Board (IRB). There were three members in the IRB including two eminent medical professionals and one educator who approved the survey questions and abstract of the project to make sure it was physically and psychologically safe for participants to answer. Minors asked their parents to fill an informed consent form to allow them to take the survey. However, there was no required form for legal adults: people above 18 years, who took the survey. After obtaining IRB approval, I created a survey with questions that would probe participants to think about how exactly they perceived ChatGPT and AI in general, and their perceptions of how risky and adept of a tool these language models can be. The survey was dispersed to a diverse age group, from the ages of 12 years old to 65, and 44 subjects answered that they were above 18. There were 26 females, 26 males, and 2 people who chose not to share their gender identity. Participants were recruited through mediums such as Instagram and WhatsApp; however, the questionnaire itself was sent out via private emails. There was a unanimous positive response towards agreeing to answer the survey. A total of 50 responses were recorded after which the survey form was programmed to stop accepting responses.

The survey started off with a demographics section, in which participants shared some personal identifiers such as their age, gender, and the most recent level of education. This was to observe if any of these factors could turn out to be confounding or lurking variables and unduly influence or distort a statistical analysis. The next set of questions aimed to assess the general knowledge of the participant about ChatGPT, specifically asking them to describe it in their own words, allowing us to gather the vast descriptions of the eminent language model. People were further questioned on whether they had used ChatGPT before, and those who answered 'Yes', were asked to rate their experience on a scale of 1-10 (1 being no experience and 10 being expert knowledge).

The third section of the survey measured people's perceptions on the level of sentience of ChatGPT and AI. In this study, sentience was measured not only based on how humanlike their responses were, but also on whether the subjects believed that AI experienced humanlike emotions. They were also provided with 9 futuristic scenarios, both positive and negative, about

the different effects and outcomes of the usage of ChatGPT and AI. Participants would rate these on a scale of 1 to 10 (1 being the weakest sentiment towards the belief that the situation would occur and 10 being the strongest). These scenarios were provided with the intention to observe if people in general were inclined towards thinking these tools were beneficial or harmful or neither/neutral in the future. The fourth section questioned the participants if they would be interested in using ChatGPT again (or for the first time as applicable), and if they would recommend it to others.

3) Results

To view the survey participants completed, please see the link below.:
https://docs.google.com/forms/d/1gw91z76EbhNZWWDRsA8OXIyAWBAAIlNAxO_7zvHnhIw/edit#responses

Screenshots of the survey results are also included in the analysis section to aid in visual representation and understanding.

People were surveyed about their education level in an open-ended question, which thus generated a variety of responses; however, these were later classified into Master's degree, Bachelor's Degree, and a High School GED, Diploma or less. About 34 of these subjects had used ChatGPT before. Figure 2 displays that nearly all the subjects, who had used ChatGPT before, rated their experience as good (a 5 or above), with only 1 participant rating it a 1 and 2 others rating it as a 3 or 4 respectively.

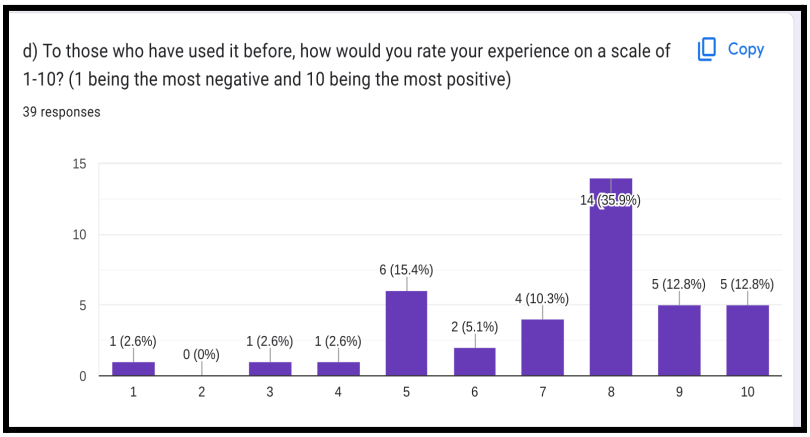


Figure 1: This bar graph represents the numerical rating of people's experience of ChatGPT.

The main focus of the analysis was on respondents' perceptions of sentience. All of the survey questions and their parts were individually numbered and put down in a Python program (Figure 2) to process better and clearer statistics since the sample size was sizably large. A few Python modules for statistics such as linear regression and ordinary least squares (OLS) tests were imported into the program. The order of the questions is provided below.

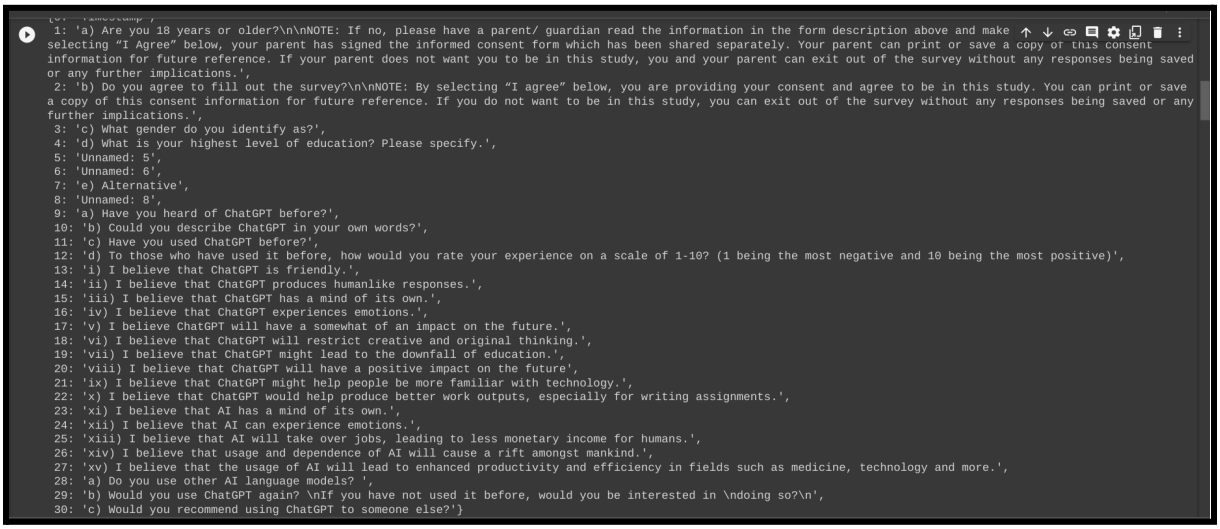


Figure 2: The image displays all the questions of the form numbered appropriately in the Python program.

As the project was not hypothesis driven, statistical significance was tested by examining correlations between certain statements. The main correlation examined if there was a relation between how strongly people believed in ChatGPT's sentience and how strongly they felt negative scenarios would occur upon its usage. Statements 13, 14, and 15 were grouped together:

13: 'I believe that ChatGPT is friendly.'
14: 'I believe that ChatGPT produces humanlike responses.'
15: 'I believe that ChatGPT has a mind of its own.'

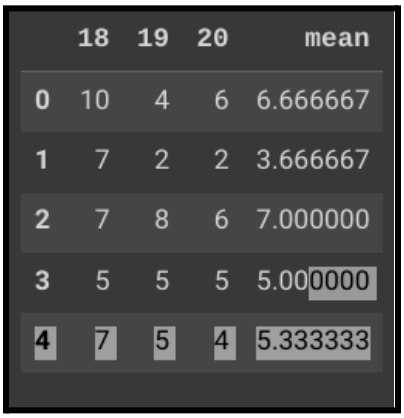
An average of how these statements were rated for each person was calculated, so each person's overall sentience rating for ChatGPT could be measured.

	13	14	15	mean
0	1	9	6	5.333333
1	8	8	4	6.666667
2	6	5	3	4.666667
3	7	7	3	5.666667
4	5	5	5	5.000000

Figure 3(a): The image displays a screenshot of averages of the ratings of the first five users for statements 13, 14, and 15, to measure the overall sentience for each of the users.

Similarly, we grouped statements 18, 19, and 20 together to take the average of how strongly disaster scenarios are thought to occur for each person:

18: 'I believe that ChatGPT will restrict creative and original thinking.'
19: 'I believe that ChatGPT might lead to the downfall of education.'
20: 'I believe that ChatGPT will have a positive impact on the future.'



	18	19	20	mean
0	10	4	6	6.666667
1	7	2	2	3.666667
2	7	8	6	7.000000
3	5	5	5	5.000000
4	7	5	4	5.333333

Figure 3(b): The image displays a screenshot of the averages of the ratings of the first five users for statements 18, 19, and 20 to measure their overall beliefs of AI leading to disaster.

The averages of each person's sentience rating were measured against how strongly they felt a disaster scenario could occur, resulting in the following graph (Figure 3(c)). The sentience rating was put on the x-axis versus the disaster scenarios on the y-axis. The means are scattered about the line of best fit and do not show a linear pattern. However, they do show a slightly negative linear correlation according to the regression plot in Figure 3(c).

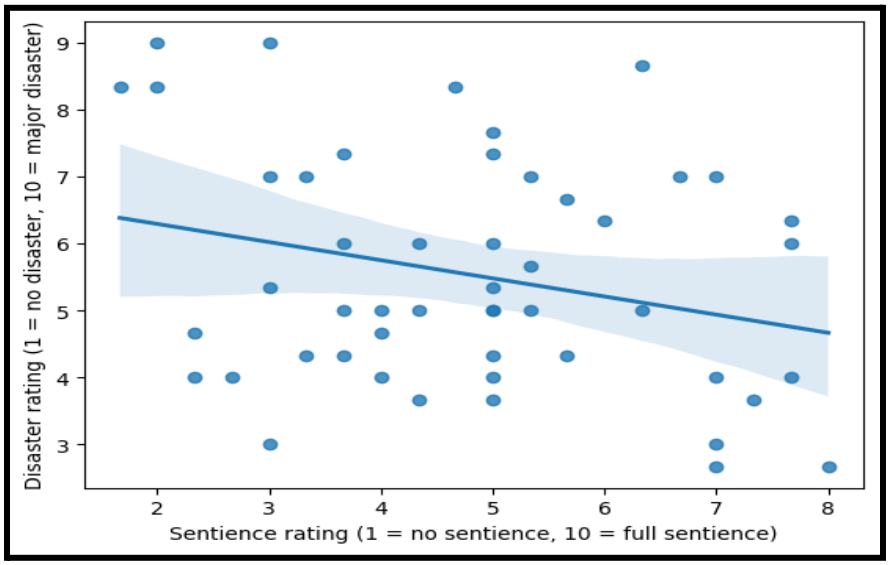


Figure 3(c): The regression plot displays the averages for statements 13, 14, and 15 against the averages for statements 18, 19, and 20.

OLS Regression Results

==========

Dep. Variable: mean	R-squared (uncentered): 0.815
Model: OLS	Adj. R-squared (uncentered): 0.811
Method: Least Squares	F-statistic: 216.0
Date: Fri, 08 Sep 2023	Prob (F-statistic): 1.37e-19
Time: 21:43:36	Log-Likelihood: -121.93
No. Observations: 50	AIC: 245.9
Df Residuals: 49	BIC: 247.8
Df Model: 1	Covariance Type: nonrobust

==========

	coef	std.err	t	P>|t|	[0.025	0.975]
mean	1.0052	0.068	14.697	0.000	0.868	1.143

Omnibus: 2.648	Durbin-Watson: 1.736
Prob(Omnibus): 0.266	Jarque-Bera (JB): 1.975
Skew: -0.482	Prob(JB): 0.373
Kurtosis: 3.187	Cond.No: 1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 3(d): The OLS regression test measures the averages of the two pairs of statements against each other to check if the regression plot data is statistically significant.

Ultimately, an OLS linear regression test (Figure 3(d)) was configured which would help us further solidify whether the correlation between the two statements is statistically significant. Upon observing the results, the major takeaways are $R^2 = 0.763$, $F = 162.3$, $p < 0.001$. The coefficient of determination (R^2) can be interpreted as follows: 76.3% of variance in the disaster scenario rating can be accounted for by the linear model with the sentience rating. The p-value: <0.001 in this case, is below the threshold p-value: 0.05; therefore, as per the rule of statistical analysis for regression tests, we can conclude that this correlation and regression fit is statistically significant. We can further interpret this as the lower a person rates the sentience of ChatGPT, the higher they seem to be inclined towards negative scenarios, and vice versa. This was the main deduction of the research paper and a direct link between perception of sentience and the future of ChatGPT and AI as such. It also aligns with the findings and visual representations of Figure 3(c).

After the main conclusion was obtained, a few more statements were compared to observe if there were any more correlations. One example includes statements 15 and 16 being grouped

together. This was done so we could analyze if there was a difference in people perceiving sentience versus consciousness. For example, we considered if a person giving a high rating to ChatGPT having a mind of its own would assign the same rating to the statement of it experiencing emotions?

15: ‘I believe that ChatGPT has a mind of its own.'
16: ‘I believe that ChatGPT experiences emotions.'

Much like the previous analysis, averages were once again calculated for the sentience statement (15) and the consciousness statement (16) for each participant to measure the strength of their belief that ChatGPT has a mind of its own.

	15	16	mean
0	6	2	4.0
1	4	1	2.5
2	3	1	2.0
3	3	1	2.0
4	5	5	5.0

Figure 4(a): This image displays the averages of the ratings of each user for statements 15 and 16.

After taking a linear regression model again (Figure 4(b)), we get an upward trend in the data, suggesting a positive correlation between the two statements. It appears as if people mostly assign the same rating to both the statement of sentience and the statement of consciousness, offering the conclusion that they are directly related. Since this conclusion was expected, given how the words sentience and consciousness are often used interchangeably, an OLS regression test was not conducted.

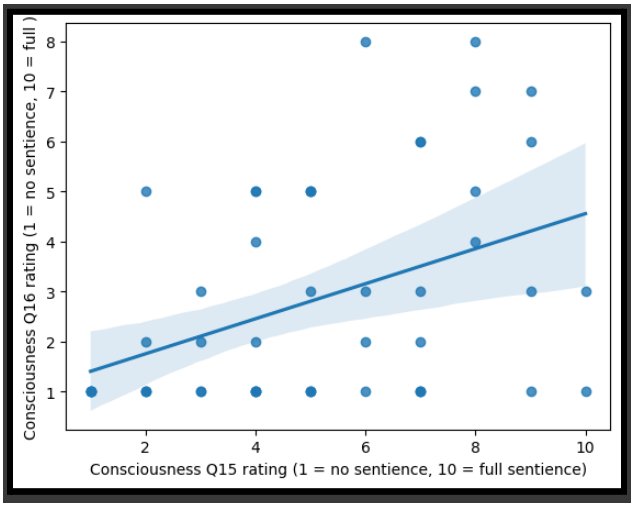


Figure 4(b): The regression plots the averages of statement 15 against the averages of statement 16 to determine if there is a correlation and trend in the pattern.

This time, statements 15 and 16, which most directly measure the sentience of ChatGPT, were plotted against statements 23 and 24, which most directly measure the sentience of AI.

15: 'I believe that ChatGPT has a mind of its own.'
16: 'I believe that ChatGPT experiences emotions.'
23: 'I believe that AI has a mind of its own.'
24: 'I believe that AI can experience emotions.'

The averages of the statements 15 and 16 were calculated for each participant to measure their overall rating of sentience. The same was repeated for questions 23 and 24 to measure each participant's overall rating of sentience as well.

	15	16	mean
0	6	2	4.0
1	4	1	2.5
2	3	1	2.0
3	3	1	2.0
4	5	5	5.0

Figure 5(a): This image displays a screenshot of the averages of the ratings of the first five users for statements 15 and 16 to measure the sentience of ChatGPT.

	23	24	mean
0	4	2	3.0
1	5	2	3.5
2	7	4	5.5
3	8	3	5.5
4	6	5	5.5

Figure 5(b): This image displays a screenshot of the averages of the ratings of the first five users for statements 23 and 24 to measure the sentience of AI.

After the means were computed and inputted into a linear regression model (Figure 5(c)), it was observed there is a strong positive correlation. This indicates people who rate ChatGPT high in terms of sentience, also rate AI high in this area as well.

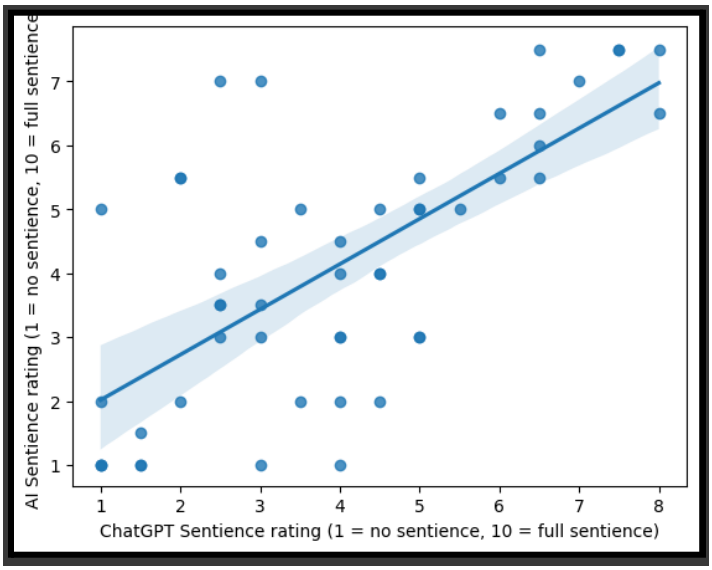


Figure 5(c): The regression plot is used when the averages of statements 15 and 16 are plotted against the averages of statements 23 and 24 to check if there is a correlation between how people perceive the sentience of AI and how they perceive the sentience of ChatGPT.

The final set of conditions tested was between the education levels provided in Statement 12 and how participants rated ChatGPT's sentience in statement 15, and if they viewed AI

positively in Statement 27. Statement 12 was an open-ended question so it elicited a variety of responses, which would have been difficult to analyze on a graph. Therefore, the education levels were individually sorted out into one of the three categories: Bachelor's degree, Master's degree, and High School Diploma/GED or less. The means were taken for each participant across the questions and respondents with Bachelor's degrees rate sentience averaging about 6.2000, higher than those with Master's degree averaging about 4.905, and high school Diploma/GED or less which had an average about 4.215. An Analysis of Variance (ANOVA) test was used since this set involves categorical and quantitative variables, unlike earlier, when only quantitative variables were statistically examined. According to the test, the F statistic is 0.874, and the p-value: 0.42 is greater than the threshold value of 0.05; therefore, according to statistical rules, education level and belief of sentience do not have a statistically significant correlation.

4) Discussion

This project was meant to be exploratory in nature. Still, after examining the results in Python, aided with visual representations, a few conclusions can be drawn about sentience and the implications of ChatGPT and AI in the future.

The main deduction from the analysis is that people who generally seem to view AI as less sentient or humanlike place a higher rating on the possibility of AI leading to disaster. This suggests that people who have a lower belief of sentience view AI in a more negative light. A possible interpretation of this result could be that humans perceive technology that is most aligned with their line of thought or the most humanlike as something that would benefit mankind more. This is contrary to the popular belief that AI agents having a mind of their own and thinking like humans would cause more harm in the future to humans.

The media has played a large role in portraying AI as the root of all evil in the dystopian world. For example, major box office hit horror 'Megan', which was released in 2022, showed a humanoid AI-powered doll robot initially created to counsel children and teenagers. Once the doll, Megan started becoming self-aware of herself and her purpose, she tried overriding her assigned duties and became almost fatal to the protagonists. This raised awareness on social media platforms about how AI can go wrong. This is similar to the movie 'I, Robot', wherein robots operating on AI, retain higher positions than humans, in a futuristic world. A murder took place in the film and the primary protagonist, who was meant to solve the mystery, immediately assigned the blame to one of the robots as he feared their sentience. These movies have fueled the fears and suspicion of the general population towards AI.

Other results of this study include that people rated sentience and consciousness about the same, almost blurring the difference between the two terms. Since AI having human emotions and thoughts is equated to having a mind of its own, the two terms sentience and consciousness are often used interchangeably.

In addition, participants rated ChatGPT and AI as almost equal in terms of disaster scenarios and the possibility of having a mind of their own. While ChatGPT is based on AI, it is also an independent language model, so initially there was some difference expected, but the results

from this study do not suggest this. ChatGPT is also partly the first major AI agent that the world has been introduced to, so individuals may equate AI and ChatGPT as interchangeable although this would need to be explored further.

Another result as stated in the analysis was that people with Bachelor's degrees in various fields treated sentience higher than people with Master's degrees, and the equivalent of a high school diploma, GED, or less. This was not statistically significant to the study.

Further, to rule out gender as a lurking variable, it was confirmed that people identifying as male and female mostly had the same sentience rating.

5) Limitations and Scope for Further Study

From the statistical analysis, some limitations of this project were first that the difference between ChatGPT and AI was perhaps not explained thoroughly, leading to confusion for participants when choosing answers. This is evident in the finding that people seem to put a rating of the disaster scenarios and sentience of ChatGPT that is almost identical to what they put for AI. Future research could include definitions of each that highlight similarities and differences. In addition, the terms sentience and consciousness could also have been more well defined.

An experimental manipulation could also have been implemented. This process would include some subjects who are easy to reach for an in-person experiment and who answered 'No' to using ChatGPT being asked to use it for the first time while being monitored. They could then have taken the survey again and one could analyze their results before and after the experiment to see if they viewed ChatGPT any differently after.

6) Conclusion

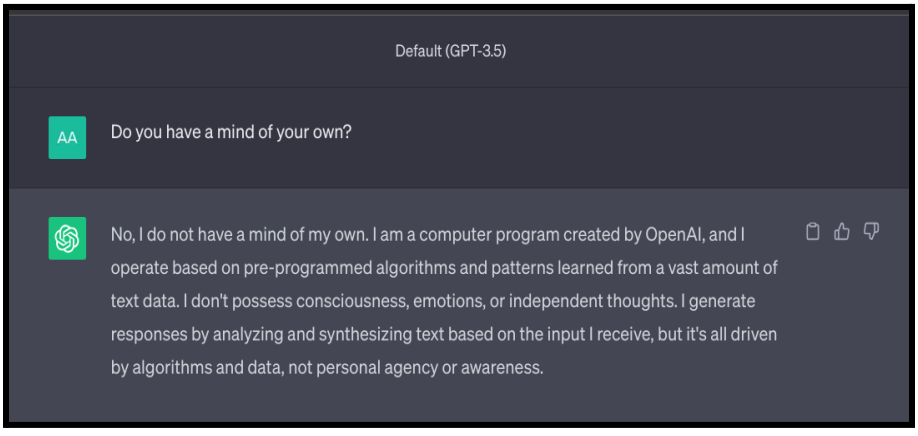


Figure 8: The image displays ChatGPT's answer when it is questioned, "Do you have a mind of your own?"

As shown in Figure 8, an interesting observation to be noted is when ChatGPT itself is asked the question, "Do you have a mind of your own?" it answers that it does not and it is a program running on servers (Introducing ChatGPT). It adds that it does not possess consciousness, self-awareness, or independent thought. Although not confirmed, it may be that the developers themselves were trying to mitigate beliefs right from the start about its supposed sentience. Additionally, while a majority of people from this study's survey have answered that they believe ChatGPT and AI cannot perceive emotions, a significant number of people are also inclined towards the belief that AI can think on its own. A few decades, actually even a few years ago, the fact that a chatbot like ChatGPT would become the fastest growing computer application in history or that computer-generated AI could give us such specific narrowed down responses unlike other web scraping platforms, would not have been unexpected. Yet, developers have done so, and while the algorithm certainly has its limitations such as not grasping the full tone and severity of the language and situation at times, it has still marked a major milestone in the journey of AI.

In regard to the question of whether AI will become a menace to society, it is important to remember that AI is only as good as the people who use it. While there are concerns about the potential misuse or abuse of AI, it is up to us as a society to ensure that AI is developed and used in a responsible and ethical manner.

This requires a multi-disciplinary approach that involves not just technology experts, but also ethicists, policymakers, and other stakeholders. It is important to establish ethical frameworks and guidelines for the development and deployment of AI to ensure that it serves the greater good and does not cause harm.

Moreover, it is important to ensure that AI is transparent and accountable. This means that AI systems should be able to explain their decision-making processes and be subject to external audits and oversight. By doing so, we can build trust in AI systems and ensure that they are used for the benefit of society.

AI has helped automate repetitive tasks, as well as increase efficiency and productivity in several industries, yet the field is still seen with skepticism owing to the dangers and risks it has a potential to create. Amidst all of this, experts in AI are cognizant of the lurking danger and probable misuse of this technology. As we move forward, governments, public policy experts, and academia in general, will have to come together to shape legislation and create guardrails so that we continue to reap the benefits of AI, while risks are minimized.

It is also important to consider that the sentience and even consciousness of AI could still develop, if it continues to grow by human feedback and development to the point where it evolves on its own. The use of AI powered technology has rapidly increased throughout the years, with AI voice changing and 'deep faking/ face swap' softwares taking the lead. These lead to tarnishing of images, of public figures or rather anyone. There have been several cases of identity theft under these AI softwares, where images and voices have been superimposed on people for immoral purposes. There have been incidents wherein cyber attackers have used AI to cause severe harm to the environment and the world. These include the Colonial Pipeline hack where the AI powered system to provide fuel and gasoline was forcibly shut down by an

anonymous hacker and this caused extreme fuel shortage and a rise of gas prices in Florida. Another similar incident is at Metro-Goldwyn-Mayer (MGM) casinos wherein hackers created a cybersecurity data breach and millions of dollars in revenue were lost, affecting thousands of customers and the stocks of the company. These events go on to show that the responsibility of using AI for appropriate functions does depend on the AI designer.

Therefore, there should be a legal moral code of conduct placed on the usage of AI, to make sure no user oversteps and causes harm to others: physically, emotionally and socially. There should be penalties and negative consequences assigned to those who choose to violate it. The designers, who refine AI should also be subject to this code and its consequences as they are the ones ultimately shaping one of the world's most utilized mediums. When they are recruited for this job, it must be made sure that they are absolutely trustworthy and in turn, the citizens themselves must be made aware of the ways AI should be used. Henceforth in the future, we must learn how, when, and where to wield AI, as it can be a double-edged sword.

References

[1] Brauner, P., Hick, A., Philipsen, R., & Ziefle, M. (2023). "What Does the Public Think About Artificial Intelligence?—A Criticality Map to Understand Bias in the Public Perception of AI." *Frontiers in Computer Science*, vol. 5, Frontiers Media, Mar. 2023, https://www.frontiersin.org/articles/10.3389/fcomp.2023.1113903/full

[2] Hetler, Amanda. "ChatGPT." *WhatIs.com*, Sept. 2023, www.techtarget.com/whatis/definition/ChatGPT#:~:text=ChatGPT%20is%20an%20artificial%20intelligence,%2C%20essays%2C%20code%20and%20emails.

[3] Taecharungroj, Viriya. "'What Can ChatGPT Do?' Analyzing Early Reactions to the Innovative AI Chatbot on Twitter." *Big Data and Cognitive Computing*, vol. 7, no. 1, Multidisciplinary Digital Publishing Institute, Feb. 2023, p. 35. https://www.mdpi.com/2504-2289/7/1/35

[4] Stahl, Ashley. "The Rise of Artificial Intelligence: Will Robots Actually Replace People?" *Forbes*, 3 May 2022, www.forbes.com/sites/ashleystahl/2022/05/03/the-rise-of-artificial-intelligence-will-robots-actually-replace-people/?sh=70f6c8fb3299.

[5] Hick, Alexander, and Martina Ziefle. "A Qualitative Approach to the Public Perception of AI." *International Journal on Cybernetics & Informatics*, vol. 11, no. 4, Aug. 2022, pp. 1–17. https://www.researchgate.net/publication/363225909_A_Qualitative_Approach_to_the_Public_Perception_of_AI

[6] Randieri, Christian. "Can AI replace human curiosity?" Forbes, 22 March 2023, https://www.forbes.com/sites/forbestechcouncil/2023/03/22/can-ai-replace-humancuriosity/?sh=7fc767831991

[7] Shapiro, Daniel. "Can Artificial Intelligence 'Think'?" Forbes, 23 October 2019, https://www.forbes.com/sites/danielshapiro1/2019/10/23/can-artificial-intelligencethink/?sh=5b9a00c42d7c