Summer 2023

Sentiment Analysis of Tweets about ChatGPT

Hanna Dodd

Sentiment Analysis of Tweets about ChatGPT

Hanna Dodd

LIS9704: Librarianship & Evolving Technologies

August 6, 2023

Introduction

There is no surprise that ChatGPT has exploded in popularity since it's release in November 2022. There have been many other systems that aim to use large-language models for conversational bots, but none that have been as successful. The user base of ChatGPT has grown significantly since launch, reaching 100 million monthly active users by January 2023, (Hu, 2023). With this large spike in popularity, there has been a lot of chatter online regarding the implications of this model and what it means for the future. People are trying to make sense of this innovation, some worrying about the future, some excited for the opportunities this may bring. This project aims to quantify public perception of ChatGPT. To do this, I use Python to complete an analysis of 500,000 tweets from January to March 2023. I aim to discover the sentiment and popularity of ChatGPT on Twitter and gain some insight into the future perceptions of these models. As a note, throughout this paper I refer to Twitter and tweets, despite the recent rebranding of the company to X. This choice was made for consistency and identification of the platform.

Methods

The analysis for this project was completed using the Python programming language in Google Colab. I used the pandas python library (McKinney, 2010) for the data analysis and the VADER sentiment analysis tool (Hutto & Gilbert, 2014). VADER is geared towards social media sentiment and can easily identify slang, punctuation, or emojis as sentiment factors. This made it the ideal tool for the project described above. When preprocessing the data, 62 tweets were removed because they contained a null value. The content of each tweet was processed to remove links, new line characters, mentions, hashtags, and additional spaces. This was done to ensure there was no abnormal skew towards neutral sentiment due to inconsequential content

such as a link. I chose to leave emojis and punctuation in the content of the tweet, as VADER is able to factor these features into the analysis.

Prior to the sentiment analysis, I looked at some key features of the dataset to understand the data. I separated the most liked and retweeted tweets in order to understand what the most popular conversations were about ChatGPT. I also segmented the dataset to chart the number of tweets per day, week, and month. Finally, I identified at the most mentioned users and the most used hashtags. This general analysis gave an overview of ChatGPT and familiarized me with the dataset before continuing to the sentiment analysis. For the sentiment analysis, I used VADER to analyze the processed content of each tweet and assign them a compound polarity score. The compound score is a sum of the positive, negative, and neutral sentiment which is then adjusted to be between -1 and 1. By using this score, I have a consistent way to see if each tweet is read as positive, negative or neutral. I then plotted these scores to show the overall sentiment, and the sentiment during each month that the dataset contains.

Results

As mentioned previously, this analysis was completed on a set of 499974 tweets. To get a sense of what is being talked about online, I looked at the most liked and retweeted tweets in the set. Table A1 shows the five most liked tweets, their like counts, and the user that posted them. From these top five liked tweets, we can see some common themes. The tweets by users @rgay and @kevinschawinski both express concern for the use of ChatGPT. Twitter user Roxane Gay states confidently that “you’ll know it when you see it” (Gay, 2023) when receiving student work written with assistance from ChatGPT. Kevin Schawinski is more concerned with the capabilities of AI, as the response he received from his prompt “tell me a subtle lie” was “I am a human being,” (Schawinski, 2023). The other most liked tweets were interested in finding new

ways to market or use AI tools, including ChatGPT. This is similar to the most retweeted tweets, as seen in Table A2. They revolve around using the tools to sell, market, or influence people to buy their product or interact with their content.

Second, I charted the number of tweets per day, week, and month throughout the duration of the dataset. In these figures we can see a few key spikes in the conversation, especially during pivotal moments. As seen in Figure B1, there was a large increase in discussion during early February and mid-March. The spike in early February appears to be on February 7th, meaning we can assume this may be due to the release of Google Bard on that day. Similarly, the spike on March 15th aligns with the announcement of GPT-4. When looking at Figure B3, we can see that there is an overall increase in conversation surrounding ChatGPT throughout the duration of the dataset, and we can assume that this trend continued after March 2023.

Third, I made note of the users that were mentioned the most, as well as the hashtags used most often. The users that were mentioned most often can be seen in Table C1. They include many large companies such as OpenAI, YouTube, Microsoft, and Google, as well as main players such as Sam Altman and Elon Musk. As seen in Table C2, there were many duplicates within the hashtags, with variations of #ChatGPT, #AI, and #artificialintelligence.

Finally, for the sentiment analysis I looked at the sentiment of the entire dataset, as well as in segments by month. Figure D1 shows the sentiment of the entire dataset, and we can see that it skews to a positive view. The average compound sentiment score for the dataset is 0.198, which is positive. Throughout each month, the sentiment stays consistent: January has an average sentiment of 0.20, February has an average sentiment of 0.186, and March has an average sentiment of 0.20. Each of the figures showing the sentiment of each segment can be seen in Appendix D. The dip in February could be explained by a lack of news surrounding

ChatGPT during this month, or the release of competing models that caused public opinion of ChatGPT to decrease.

Discussion

From these results, we can see that ChatGPT is incredibly popular on Twitter and is only increasing in popularity. The number of tweets per month increases throughout the dataset, and can be assumed to increase in the future, especially after the release of GPT-4 in March 2023. The sentiment on Twitter is skewing towards positive, meaning that despite much of the discussion relating to the dangers of models like ChatGPT, people are still responding well to the software. This is important for two reasons. The first is that OpenAI and other software companies who are looking to create AI models can see this positive reception and continue making more advanced systems. The second is that despite personal fears, this confirms that artificial intelligence is not going to disappear anytime soon. This will affect all of us in the coming years, as AI systems like ChatGPT become more normalized in our society.

Conclusion

In conclusion, the data provided by Twitter showed a snapshot of public perception and opinion. It is clear that Twitter users believe that artificial intelligence is a positive technology, and it is increasing in popularity over time. This shows us that we cannot ignore the development of these models and instead need to find ways to embrace their potential and mitigate any damage that they may do. The development of ChatGPT is continuing at lightning speed, and therefore the dataset is already out of date as it only includes up until March 2023. In the future, it could be informative to look at more recent datasets to view sentiment after events such as the launch of the ChatGPT mobile app or other future updates. Overall, ChatGPT is here to stay, and we can only hope to prepare for what is to come.

References

Ansari, K. (2023). *500k ChatGPT-related Tweets Jan-Mar 2023* [Data set]. https://www.kaggle.com/datasets/khalidryder777/500k-chatgpt-tweets-jan-mar-2023

Ansari, K. (2023, April 19). Cracking the CHATGPT code: A deep dive into 500,000 tweets using advanced NLP techniques. Medium. https://medium.com/@ka2612/the-chatgpt-phenomenon-unraveling-insights-from-500-000-tweets-using-nlp-8ec0ad8ffd37

Gay, R. [@rgay]. (2023, March 6). *I am pretty sure I am reading my first CHAT GPT student essay and like, teachers, don't panic. You'll know it when you see it.* [Tweet].

Hu, K. (2023, February 2). *CHATGPT sets record for fastest-growing user base - analyst note*. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).

Schawinski, K. [@kevinschawinski]. (2023, January 12). *First #ChatGPT answer which made me pause. https://t.co/nbc9uRoWlA*. [Tweet]. Twitter. https://twitter.com/kevinschawinski/status/1613446745328963584

Appendix A

Table A1.

Most liked tweets about ChatGPT from January to March 2023.

Like Count	Content	User
64094	Writing erotic prompts for Chat GPT hoping to be the first person to make AI cum. It's like the Turing Test	MoistCr1TiKaL
63835	Best AI Tools You Need To Know #chatgpt #chatgpt3 #ArtificialIntelligence #ai https://t.co/0jfr8cOMoo	johnvianny
44940	I am pretty sure I am reading my first CHAT GPT student essay and like, teachers, don't panic. You'll know it when you see it.	rgay
42125	ultra-modern generative ai🗨: \n\n• AI2\n• AI21\n• mdm\n• gpt-J\n• gpt-3\n• x-clip\n• bloom\n• cohere\n• gopher\n• dall•e 2\n• craiyon\n• tabnine\n• jukebox\n• chatGPT ***\n• anthropic\n• codegeex\n• nvidia get3d\n• dreamfusion\n• stable diffusion\n• meta make-a-video https://t.co/ON5eIGvnEQ	aaronsiim
38278	First #ChatGPT answer which made me pause. https://t.co/nbc9uRoWlA	kevinschawinski

Table A2.

Most retweeted tweets about ChatGPT from January to March 2023.

Retweet Count	Content	User
16080	Best AI Tools You Need To Know #chatgpt #chatgpt3 #ArtificialIntelligence #ai https://t.co/0jfr8cOMoo	johnvianny
11501	ultra-modern generative ai🗨: \n\n• AI2\n• AI21\n• mdm\n• gpt-J\n• gpt-3\n• x-clip\n• bloom\n• cohere\n• gopher\n• dall•e 2\n• craiyon\n• tabnine\n• jukebox\n• chatGPT ***\n• anthropic\n• codegeex\n• nvidia get3d\n• dreamfusion\n• stable diffusion\n• meta make-a-video https://t.co/ON5eIGvnEQ	aaronsiim
10498	AI copywriting tools to check out:\n\n1. Chat GPT - Research\n2. QuillBot - Paraphrasing\n3. StoryLab - Hooks and outlines\n4. Grammarly - Grammar/spelling\n5. Hemingway -	writingtoriches

	Conciseness/clarity\n6. Power Thesaurus - Thesaurus\n7. Tweet Hunter - Content creation\n\nWhat else?	
8096	Creators and writers are making $1,000,000s using ChatGPT.\n\nBut 99% of people don\'t know how to build a Business and make money using it.\n\nI just built ChatGPT Business Crash Course.\n\nFree for the next 24 hours!\n\nLike, RT & comment "GPT" and I\'ll DM it to you\n\n(Must be following) https://t.co/L3qVoUudOk	garryflix
8060	$150 || 7250 PHP || 2.1 IDR\n\n$100\n✅ RT & Follow @gptaiinu\n\n+$50\n✅ Join Telegram (post proof)\nhttps://t.co/sUAaxEo8D0\n\nends in 48 hours\n\n----------\nFirst ChatGPT lead Crypto Project. 🤖🚀\nDEX- https://t.co/CzcEyt8xy8 #ChatGPT https://t.co/rH00LZM3oK	crypto_bearr

Appendix B

Figure B1.

Number of tweets per day throughout the duration of the dataset.

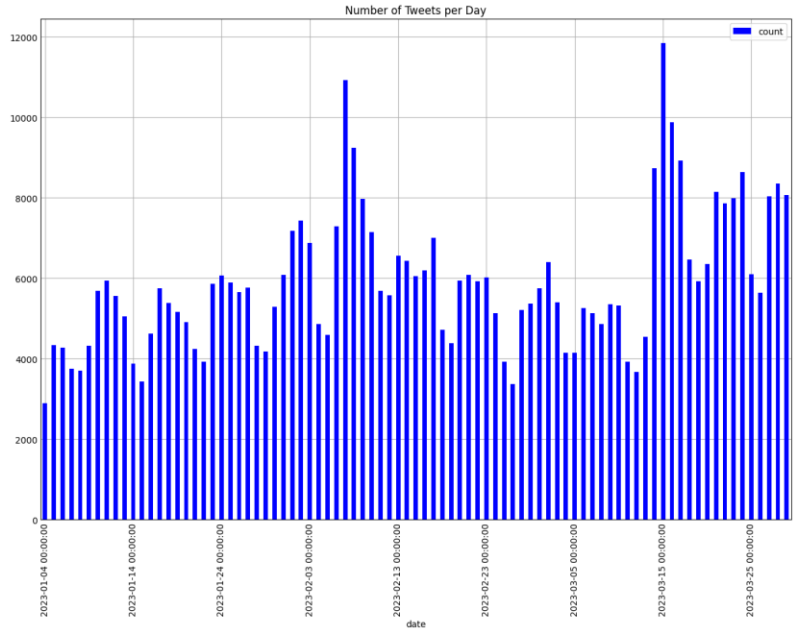


Figure B2.

Number of tweets per week throughout the duration of the dataset.

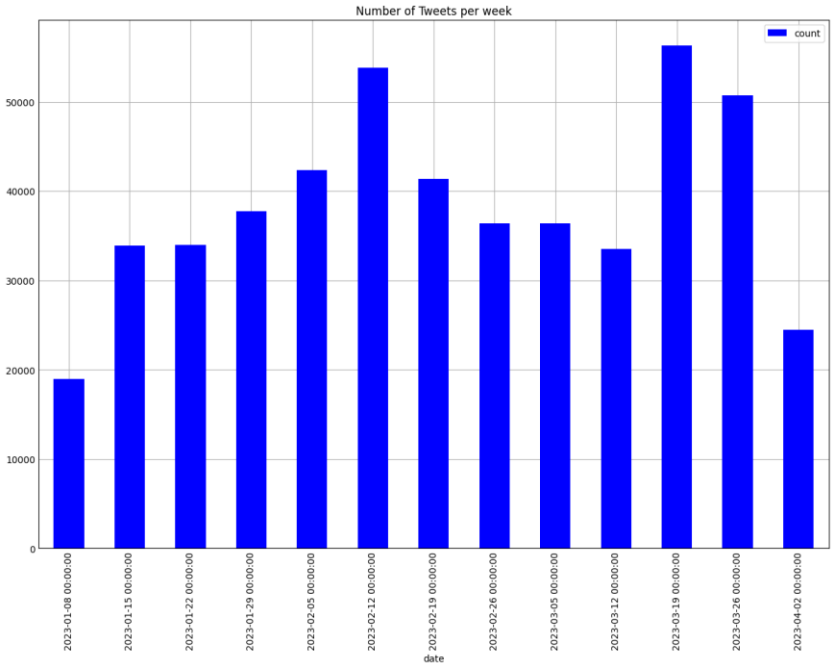
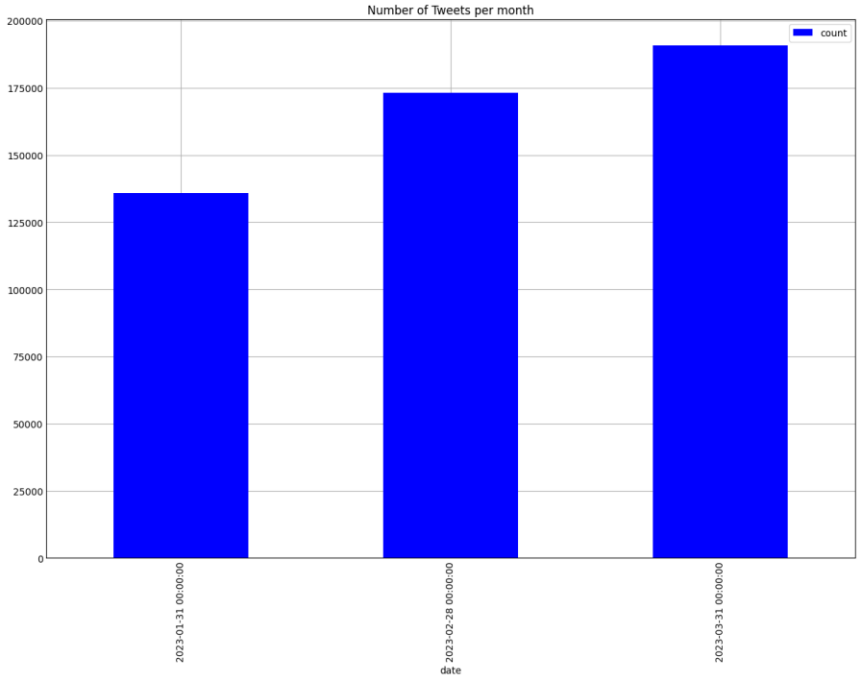


Figure B3.

Number of tweets per month throughout the duration of the dataset.



Appendix C

Table C1.

Users mentioned most often within the dataset.

Username	User	Times Mentioned
@OpenAI	OpenAI	15394
@elonmusk	Elon Musk	8727
@ChatGPT	ChatGPT	4075
@YouTube	YouTube	3107
@Microsoft	Microsoft	2838
@Google	Google	2543
@sama	Sam Altman	2010
@bing	Bing	1348
@BetaMoroney	Tony Moroney	885
@SpirosMargaris	Spiros Margaris	882

Table C2.

Hashtags used most often within the dataset.

Hashtag	Times Used
#ChatGPT	283901
#AI	80181
#chatgpt	54299
#ai	24776
#OpenAI	19996
#ArtificialIntelligence	17545
#chatGPT	14295
#openai	13031
#artificialintelligence	8213
#technology	7802

Appendix D

Figure D1.

Sentiment of the entire dataset.

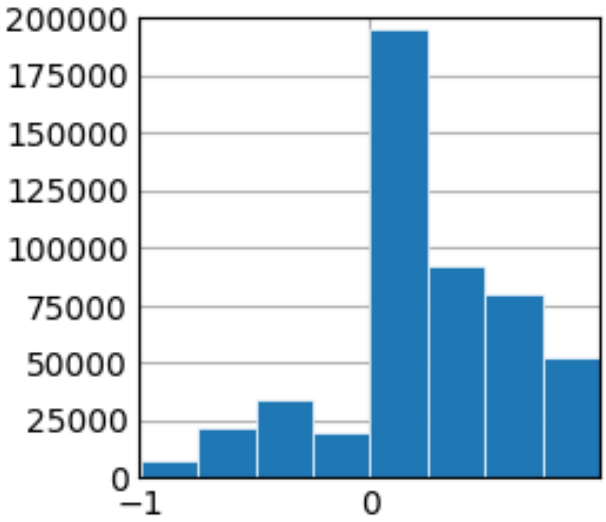


Figure D2.

Sentiment during the month of January.

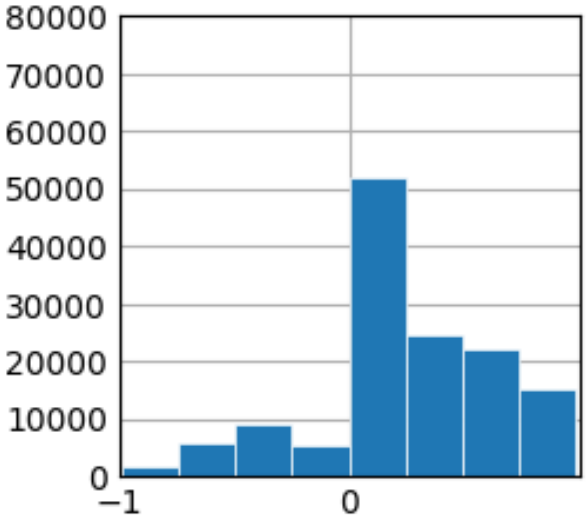


Figure D3.

Sentiment during the month of February.

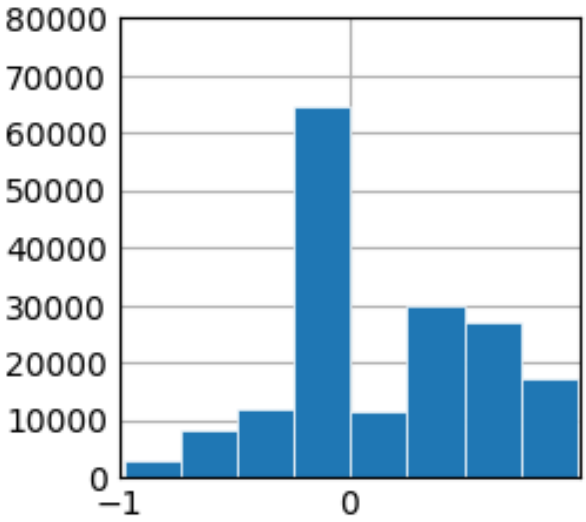


Figure D4.

Sentiment during the month of March.