A Sentiment Analysis On ChatGPT Tweets

Barbara Ruvolo
EPFL
Lausanne, Vaud, Switzerland
barbara.ruvolo@epfl.ch

Ajkuna Seipi
EPFL
Lausanne, Vaud, Switzerland
ajkuna.seipi@epfl.ch

Hongyi Shi
EPFL
Lausanne, Vaud, Switzerland
hongyi.shi@epfl.ch

ABSTRACT

This paper presents a comprehensive analysis of the release of ChatGPT, an AI-based conversational model, on Twitter. The study focuses on the first month (December 2022) and the last month (April 2023), aiming to understand the evolution of the public sentiment through time. In addition, a classification of users based on their scientific background is performed to identify any differences of sentiment between different users groups. Leveraging sentiment analysis techniques, we used the multi-lingual sentiment classifier from Barbier et al. (2022) [3] to predict the sentiment labels. Furthermore, we employed topic analysis techniques to identify the prevalent themes and discussions surrounding ChatGPT. Additionally, we employed a Roberta model to conduct emotion analysis, providing insights into the emotional responses evoked by ChatGPT.

Our findings show that ChatGPT is generally viewed as of high quality, with a majority of positive sentiment when excluding the neutral ones, and that the emotions of approval and *admiration* are predominant. Overall we noticed a slight increase of neutral sentiment during the period analysed, that could indicate that public view is becoming more rational about chatGPT after an initial hype. Concerning the topics, the most prevalent topics remain *Science & technology* and *Business & entrepreneurs*, indicating general concerns about technical aspects from the public view.

CCS CONCEPTS

• **Human-centered computing → Social media**.

KEYWORDS

datasets, neural networks, text embedding, language processing, tweets, analysis

ACM Reference Format:
Barbara Ruvolo, Ajkuna Seipi, and Hongyi Shi. 2023. A Sentiment Analysis On ChatGPT Tweets. In *Computational Social Media.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Artificial Intelligence (AI) has rapidly emerged as a transformative force in our society, reshaping various aspects of our lives. As AI advances, it becomes increasingly important to understand its impact on the population. Our purposes focus on exploring the implications of AI, specifically through sentiment analysis of tweets about ChatGPT. By analysing public sentiment towards ChatGPT, we gain valuable insights into how individuals perceive and interact with this AI technology. Furthermore, we conducted a deeper analysis of the release of ChatGPT on Twitter, focusing on the first month (December 2022) and the last month (April 2023) to examine the evolution of sentiments over time. Additionally, we classified users based on their scientific background to understand the engagement and opinions of experts in the field.

The emergence of AI, including ChatGPT, brings a range of implications for the population. Firstly, it enhances accessibility by providing instant access to information, services and support. Furthermore, AI's automation capabilities can streamline repetitive and mundane tasks, freeing up human workers to focus on more complex and creative endeavours. However, this automation may also raise concerns about job displacement and the need for individuals to adapt their skills to new roles in an evolving job market.

Ethical considerations are another critical aspect. Privacy, data security, algorithmic bias and transparency must be addressed to ensure responsible deployment.

To understand public sentiment towards ChatGPT and AI, sentiment analysis plays a crucial role. Sentiment analysis can be defined as the process of computationally identifying and categorizing opinions expressed in a piece of text, documents, or tweets through Natural Language Processing techniques. In our study, we focused on classifying sentiments using a multi-class approach, so organizing tweets in positive, negative and neutral.

Section 2 introduces the dataset used in our study. Section 3 investigates the methodology and the approaches applied. Finally, the results are described and analysed inside Section 4.

During the whole project, we worked and met each other to discuss about the work done individually, the challenges faced and the future work to do. We took in consideration each member's ideas and suggestions, and provided help to each other when it was necessary. Barbara did researches about AI and ChatGPT, while also analysing the first month of the dataset entirely. Ajkuna did state-of-art researches about sentiment analysis on Twitter dataset. She also classified and analysed the scientific user group. Hongyi did in depth analysis on the users of the dataset, while working on the analysis of the last month and comparing the differences between the first and last month. Each member wrote about their respective work in the report.

2 DATASET

The dataset used in our project is the Kaggle dataset containing tweets about ChatGPT [2]. The tweets were extracted through tweets extraction by selecting a sample of hashtags. The hashtags chosen were respectively: $\#ChatGPT$, $\#chatgpt$, $\#gpt$, $\#gpt2$, $\#gpt3$

DH-500, Spring 23, Lausanne, CH

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

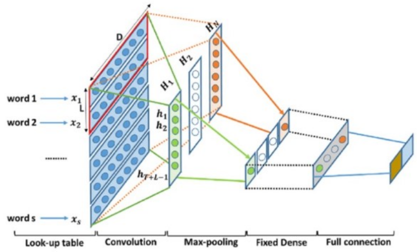

Figure 1: Standard CNN model

and #*gpt*4. The dataset was updated every day with the corresponding tweets from December 5th 2022 to April 26th 2023, resulting to a dataset containing 400,775 English tweets.

In total, the dataset is composed of 12 columns, but only 4 were used through the entire analysis. The columns selected are respectively the user name, the user description, the text of the tweet and the date of creation of the tweet.

3 METHODOLOGY AND APPROACHES

3.1 Model selection

The first step of our analysis remains to select the model to perform the sentiment analysis over our dataset.

Convolutional Neural Network (CNN)

As a first step, a cleaning process of the tweets was performed. We removed emojis and special symbols, as well as the punctuation and the hyperlinks. We also applied tokenization and stemming with the help of the NLTK library.

We used the standard CNN for sentiment analysis, seen in Figure 1, taking as input a four dimensions vector and giving as output a three dimensions vector.

To train the model, we used a labelled Kaggle dataset of ChatGPT tweets [5]. We obtained a 0.81 test accuracy, thus making the model appropriate for our analysis. Unfortunately, after predicting the labels for our dataset and further analysis, we noticed incoherence in the label predictions. As an example, in a sample of twenty random tweets, only eight out of twenty were correctly classified. Also, many of them were misclassified between positive and negative sentiments, making the results false and misleading.

XML-Roberta

After some state-of-art research, we decided to use another model, which showed to work well with sentiment analysis on ChatGPT tweets. We used the multi-lingual sentiment classifier from Barbier et al. (2022) [3]. to predict the sentiment labels. The language model is trained on 198 million tweets and it was finetuned on Twitter sentiment dataset in eight different languages. The model performance varies among languages but the model yields a good result for English with an F1-score of 71%. In addition, we cleaned the text of the tweets by removing special symbols and hyperlinks before predicting the labels.

To ensure the sentiment results are consistent, we selected random tweets and observed the sentiment prediction. Out of twenty tweets, fourteen were correctly classified, giving us a rough accuracy of 0.7. We noticed, through the misclassified tweets, that the model tends to classify more on the extreme sentiments rather than the neutral sentiment. Therefore, we concluded that the results obtained are consistent for further analysis.

3.2 User classification

To classify the users into scientists and non-scientists, we decided to approach the problem by using the user name and the user description fields. After some cleaning process of the two fields and some analysis, we decided to categorize users as scientists when the word *phd* is included inside the user name, and the words *science* or *scientist* are found inside the user description.

The latter condition was selected so that the sample of users contains exclusively scientists or researchers and avoids having undesirable profiles. We obtained 152 users in our sample, who generated 534 tweets inside the dataset.

The main profiles found are related to medical fields, such as neuroscientists. We also observe physicians, college professors and a majority of data scientists.

We are aware that this approach may be quite restrictive, resulting in missing some desirable profiles. But, as we prefer to obtain more precise and coherent results, we decided to favour the quality of the sample over the quantity.

3.3 Topic modeling

To delve more into the analysis of the sentiments, we decided to perform a topic modelling with a monolingual topic classification model developed by Antipas et al. (2022) [1]. This Roberta-based model was trained on 124 million tweets and finetuned for multi-label topic classification on a corpus of over 11k tweets. The model has 19 classes of topics and we focused on the five main topics that appear in our sample datasets.

To briefly evaluate the model with our dataset, we selected some random tweets and observed the topics predicted. Out of twenty tweets, 15 were associated to a coherent topic. In addition, for the remaining tweets, the model was not able to classify them in any of the categories, resulting to very rare misclassification between the categories.

3.4 Emotion analysis

To have a more fine-grained analysis of the sentiments, we performed an emotion analysis with a BERT based model on the tweets. We used the emotion classifier finetuned on the GoEmotions dataset [4], which contains 58k English comments from Reddit labelled for 27 emotion categories or neutral, with a respective score. We categorized the tweets into the corresponding emotion by choosing the highest scored one.

Out of twenty randomly selected tweets, we got 15 categorized in a meaningful way, resulting to a rough accuracy of 0.75 of the model. Therefore, we conclude that the model is expected to perform the correct categorization of the tweets into emotions.

4 RESULTS

This section introduces the results of our sentiment, topic and emotion analysis. We first give an overview of the entire dataset,

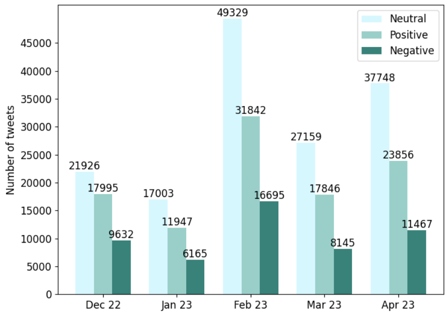

Figure 2: Monthly sentiment analysis

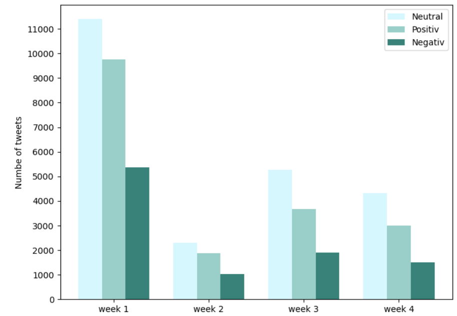

Figure 3: December 2022, weekly sentiment analysis

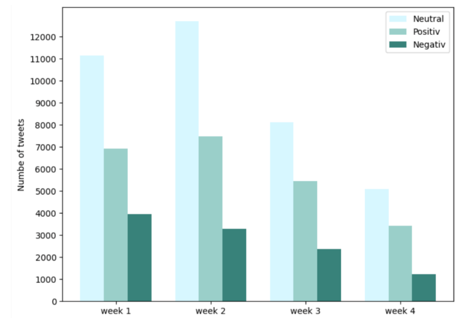

Figure 4: April 2022, weekly sentiment analysis

then focus on the periods of December 2022 and April 2023. Finally, we explore the tweets generated exclusively by scientific users.

4.1 Whole dataset

To examine the sentiment change over time, we plot the amount of neutral, positive and negative tweets monthly. Overall, each month, the proportion of the number of generated tweets remains the same. As observed in the Figure 2, February and April are the months with the greatest amount of tweets about ChatGPT. Those results could be related to the introduction of ChatGPT Plus and GPT-4. In fact, on February 9th and 13th, the introduction of ChatGPT Plus sparks significant discussions and engagements. ChatGPT Plus, a subscription-based service, offers users additional benefits and enhanced features, further increasing its appeal and attracting attention from users on Twitter. Moreover, on March 14th and 23rd, the announcement of the release of GPT-4, the successor to ChatGPT, generates reactions and discussions among the public.

4.2 First month analysis

This time, a sentiment analysis was conducted exclusively on the month of December, which contains 51,422 tweets generated between the 5th and the 31st December. Figure 3 depicts the weekly distribution of positive, negative and neutral opinions. The first week was the most active one, probably because of the initial hype resulting from the news of the release. Between December 5th to 11th, neutral and positive sentiments are predominant, and they remain the majority of feelings during the whole month. On the other hand, negative sentiment remains consistently lower among the three classes of opinion, although the number is very high during the first week.

4.3 Last month analysis

The month of April 2023 consists of 91,174 tweets generated between the 1st and the 26th April. Figure 4 shows the sentiment through the weeks. The number of positive tweets remains about the same throughout the weeks, and so does the number of negatives, which tends to decrease. The amount of neutrals changes between the second week and the third, decreasing by about 4,000 posted tweets. Initially, the first week (3rd-9th April) has the most generated tweets, which can be linked to the introduction of GPT-4 creating discussions around it. The last week has the least tweets as it contains tweets generated on only three days (24th-26th April).

4.4 Comparison of first and last month

As we were aiming to understand the evolution of the public sentiment on the release of ChatGPT, we focused exclusively on comparing the first month after the release and the last month of our dataset in order to have a large time span between the two samples.

Sentiment analysis

We observe in Figure 5 a significant increase of neutral tweets, an increase of positive tweets as well and that the number of negative tweets stayed more or so the same. Indeed, the number of neutral tweets increased by +109%, positive tweets increased by +56%, while negative tweets only increased by +25%. We can explain the increase of neutral tweets through the rising popularity of ChatGPT, which lead to a huge quantity of advertisements related to ChatGPT powered tools, often generated by bots, or the release of GPT-4, which also raised a lot of reactions and made several media owned accounts post news on the topic.

In addition, we perform a Wilcoxon Signed-Rank test to evaluate quantitatively if there is any significant difference in sentiment between the first and last month. By selecting a sample of each dataset, we obtain a p-value of approximately 0.59, which is greater than the significance level, i.e. $\alpha = 0.05$, thus failing to reject the null hypothesis. This suggests that there is no significant difference in sentiment analysis between the two months.

Topic analysis

The top topics between the two months are more or so the same. Figure 6 shows negative tweets main topics, we observe a significant

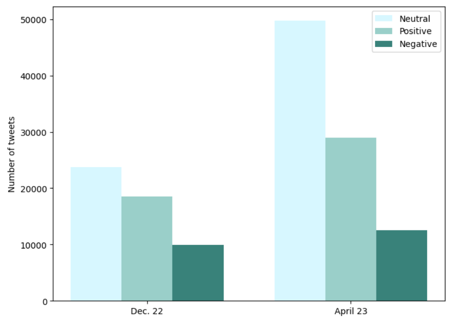

Figure 5: Sentiment comparison between first and last month

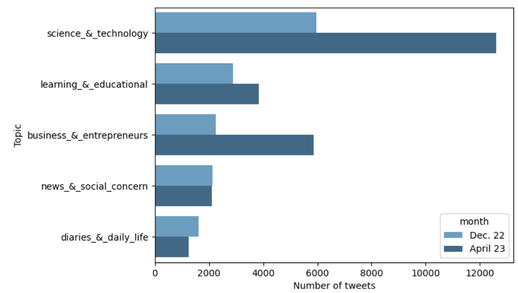

Figure 7: Top 5 positive tweets topics comparison

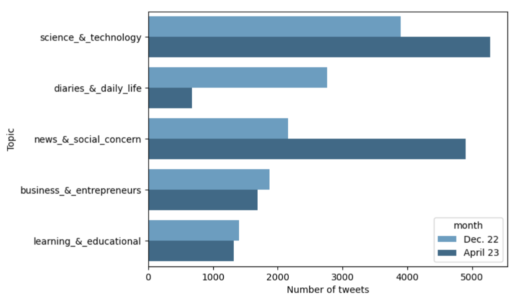

Figure 6: Top 5 negative tweets topics comparison

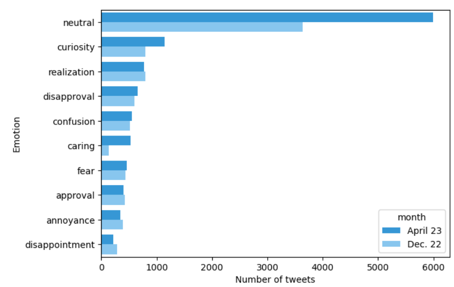

Figure 8: Top 10 negative tweets emotion comparison

increase of +126% in the number of tweets related to the topic *News & social concern* which contains various news brought by the democratization of ChatGPT and the release of GPT-4, and a decrease of -75% in the number of tweets related to the topic *Diaries & daily life* which contains mainly tweets talking about ChatGPT users experience and were generated in large amount during the first month due to the original hype. The increase of +35% in the number of *Science & technology* tweets are inherent to the increase of the global number of April tweets compared to December as it is the topic to which ChatGPT relates to the most.

The positive tweets topics are depicted in Figure 7. We observe a similar *Science & technology* topic increase of +111% and explain it the same way as for the negative tweets. The increase of +160% in the topic *Business & entrepreneurs* can be explained by the increase of advertisements for business related tools powered by GPT which were in the clear increase during the month of April. Those accounts were often powered by bots that generate tweets in aggressive quantities.

Emotion analysis

Overall, we observe a predominance of *neutral* emotion labeled tweets. Those often contains strictly factual information, like news related to the use of GPT and tips helping the democratization of GPT tools.

For the negative tweets, we observe in Figure 8 that the most frequent emotions are *curiosity* and *realization* which according to the GoEmotions dataset, are both classified as ambiguous emotions [4]. We have to wait the third emotion, *disapproval*, to get an actual negative emotion. Then the fourth one, *confusion*, is again an ambiguous emotion. This fades down the relevance of the negative tweets set. Indeed, looking back at the sentiment analysis, putting aside the neutral tweets, we already had a majority of positive tweets. Now, knowing that in the few negative tweets left, most of them are actually not fully negative but are ambiguous, thus close to neutral, this unbalances even more the sentiment analysis and confirm the global positive sentiment toward GPT within our dataset. The increase of +65% of neutral emotion is inherent to the increase of the total number of tweets during the month of April.

Notice the presence and the increase of +304% of the emotion *caring* in the negative tweets. At first, it seems like it is a positive emotion; GoEmotions dataset also classify it as positive. However, digging into their content, those are actually mostly prevention tweets against ChatGPT, people warning others about danger that could be introduced by the use of ChatGPT, which explain why they are classified as negative tweets. As more and more people start using the tool by the months, more concerns about its uses are also raised, which is not surprising. This shows us the nuance between the sentiment analysis and the emotion analysis: positive, negative and neutral emotions classified by the emotion classifier are related to the author emotion which does not necessarily match with sentiment of the tweet toward our subject ChatGPT; one could express a positive emotion through a tweet while the content brings down ChatGPT.

Figure 9 depicts emotions of positive tweets. Unlike in the negative ones where the number of tweets per top emotions was increasing gradually, here we have a polarization toward the emotions *approval* and *admiration*, if excluding *neutral*. It is however still worth to notice that the *neutral* emotion increased by +118% which can be explained as previously, by the news related to GPT-4, the

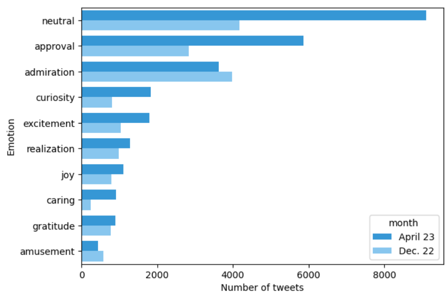

Figure 9: Top 10 positive tweets emotion comparison

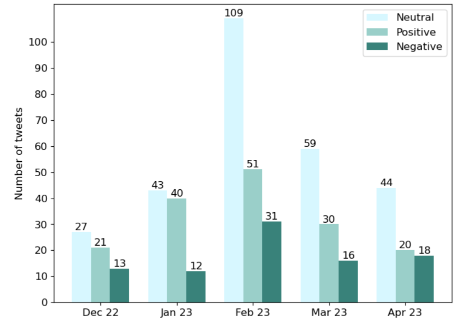

Figure 10: Monthly sentiment of scientist users

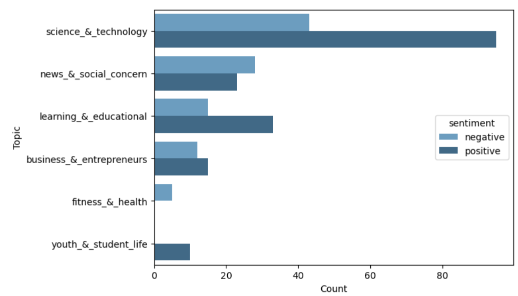

Figure 11: Top 5 topics of scientific user group

advertisements or the bots, or more generally by the global increase in the number of tweets during the month of April. Similarly as in the positive tweets, there is a significant increase of +279% in the *caring* tweets, however the content is highly different to the negative ones. The positive caring tweets are directed toward advises on how to get the most from ChatGPT and ChatGPT powered tools. Finally, we observe a significant increase of the emotions *approval*, *curiosity* and *excitement*, which increased by +106%, +126% and +75% respectively. Remember the total number of April tweets is higher than in December. As the other emotions did not increase as much, we can infer that most of the new reactions are classified in one of the emotions just mentioned (including the *neutral* emotion). GPT-4 release could have brought a lot of curiosity and excitement and since it has been a few months since the release of ChatGPT, its democratization might have lead to a better approval of the app by the global public.

4.5 Scientists analysis

Sentiment analysis

As previously seen, the neutral sentiment is overall predominant in the dataset with 52.8%. Aside the neutral sentiment, we observe that the majority of tweets are categorized as positive (30.3%), and the negative ones are a clear minority (16.9%).

In Figure 10, we remark the ratio of positive and negative sentiments remains steady through the months. Nevertheless, we notice an increase of positive sentiments in January and, in the other hand, an increase of negative sentiments compared to the positive ones in April. A possible interpretation of the increase of negative sentiments during the last month could be that the scientific community realized the limits of ChatGPT, the regulations and ethical concerns about the chatbot.

In addition, we performed a Kruskal-Wallis test to evaluate quantitatively if there is any significant difference in sentiment between the scientist users and overall users. By selecting a sample of the two groups, we obtain a p-value of approximately 0.467, which is greater than the significance level, i.e. $\alpha = 0.05$, thus failing to reject the null hypothesis. This suggests that there is no significant difference in sentiment between the scientific and non-scientific user groups.

Topic analysis

Modelling the five top topics of the scientist users, we observe (Figure 11) that *Science & technology* remains the top one topic in the negative and the positive tweets. *Learning & education* is the second one in the positive tweets, while it is placed in the third position in the negative ones. Interestingly, we observe for the first time the *Fitness & health* topic showing up in the top five.

To get a more meaningful analysis of the topics, we take a closer look on their corresponding tweets. In the negative tweets, the topic of *News & social concern* and *Science & technology* regroups mostly concerns about the efficiency of the chatbot and its limits. Concerning the topic of *Fitness & health*, we observe some tweets indicating that ChatGPT spreads false medical information. In the positive tweets, the topic of *Science & technology* includes tweets concerning the effectiveness and speed of response of ChatGPT. In addition, the topic of *Learning & education* is about how efficient the chatbot could be as an academic tool for students as well as for professors.

Emotion analysis

In Figure 12, we observe that, for the negative tweets, the predominant emotion, excluding *neutral*, is *realization*. We also notice that some tweets are also categorized with the emotion of *curiosity* and *fear*. For the positive tweets, the predominant emotions remain *admiration* and *approval*.

Through a sample of tweets, we observe that users *admire* and *approve* of the chatbot because it has the ability to answer complex medical questions as well as to generate responses quickly. For positive tweets, the *realization* emotion is more related to the user considering the improvements that can bring ChatGPT in the medical field. On the other hand, negative tweets categorized into the emotion of *realization* are related to tweets about ChatGPT

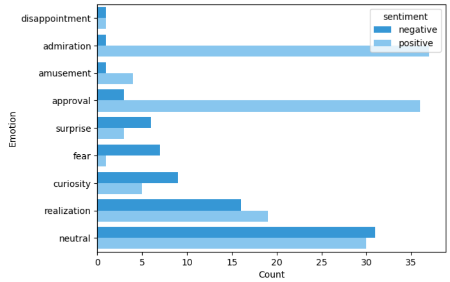

Figure 12: Top 9 emotions of scientific user group

answering incorrectly to some questions and/or requests. Some users also wrote about their concerns over the replacement of the actual search engines such as Google. Finally, the tweets emotionally categorized to *curiosity* and *fear* state concerns about the lack of concrete academic and ethical regulations.

5 DISCUSSION

There are three main concerns that are worth to be pointed out in our research, as they could have lead the results displayed to be biased.

First, we note that we have 91,174 unique user names who tweeted during the month of April, which we state as unique users (this is not necessarily true). After delving in our dataset and analysing some samples of random tweets, we have remarked that a lot of tweets were generated by bot accounts or advertisement accounts, which mostly generate neutral and positive sentiment tweets. After some investigation, we observe that we have 57 users who generated at least 60 or more tweets during the month of April. For instance, this results in 0.2% of the total unique users generating 17,038 of tweets in the last month, which corresponds to 18.7% of our April dataset. As said above, when we analyse the top users, most of them are bots and ad accounts, the only exception being the official account of ChatGPT. Therefore, we decided to remove the users that generate 60 tweets or more per month from the entire dataset, which corresponds to users that generate at least two tweets per days containing the hashtag #ChatGPT or similar. Even though, this approach removes the most active bot accounts, we could have probably missed other bot accounts or advertisement accounts, which increase the percentage of positive sentiments.

Secondly, we concluded that the scientific user group has clearly a positive opinion on AI, and more precisely on ChatGPT. But, the selection of the scientific users might have introduce bias, leading to inaccurate results. In fact, by an analysis of the users profiles in the scientific group, we observed mostly technical profiles, such as data scientists or neuro scientists, that tend to have a more positive opinion on AI. They will mostly raise comments about the technical improvements that leads the chatbot, rather than the ethical concerns that result from the integration of ChatGPT in our daily life. Selecting the scientific users with other words which might select more social scientists, such as *sociology* or *psychology*, may give us different results. Therefore, a possible future improvement of the analysis would be to have a different selection of scientists and evaluate the difference of sentiments between them.

However, it is worth to note that it might be difficult to select the sample of scientists we want to focus on because of the noisiness of the tweets text. In fact, users could use different words to mean the same thing, leading to difficult categorization through the whole dataset.

Finally, the tweet extraction itself might have introduced bias in our results. The tweets in our dataset were extracted by a hashtag selection of different versions of words defining ChatGPT. As stated in the research paper of Tufekci Z. (2014) [6], users that include the hashtag related to the subject have generally a more positive opinion about it. On the other hand, users with negative opinion may not include the hashtag to avoid giving much importance to the subject. In addition, a lot of them may react negatively to the subject in the comment sections without generating a proper tweet. Therefore, we might have missed a non-negligible amount of negative tweets which lead us to less accurate results.

6 CONCLUSION

Through our analysis, we found out that the global public sentiment and emotion toward ChatGPT are predominantly neutral; when excluding neutral opinions, reactions are mostly positive. More specifically, when comparing the first and the last month, the number of positive tweets increased and the number of negative tweets remains the same, but the number of neutral tweets doubled: this significant increase could indicate a rationalization of the public view on ChatGPT after the initial hype. The main topic is *Science & technology*, both for the positive and the negative sub-dataset. Notice the decrease of *Diaries & daily life* and the increase of *News & social concern* and *Business & entrepreneurs* motivated by the democratization of ChatGPT. Emotions that increased the most are *neutral*, *approval*, *curiosity*, *excitement* and *caring*, thus mostly positive emotions. We observe a predominance of the emotions *neutral*, *approval* and *admiration*.

Focusing on the scientific user base, we found that January has a greater ratio of positive sentiment, and April shows an increase of negative sentiments: scientific community might be more conscious of the limit and regulation concerns of the chatbot. In the positive sub-dataset, we found a larger amount of tweets about technology compared to other topics, the discussion is focused on the efficiency of the model. As opposite, in the negative sub-dataset we found a majority of *News & social concern* topic, as well as technology concerns about the efficiency and the limits. Similarly, the most commons emotions are also *admiration* and *approval*. Notice that proportionally, the quantity of *neutral* emotion is much lower than in the overall user base datasets.

Overall, Twitter users have a positive opinion of ChatGPT that keeps increasing, which shows the quick democratization of GPT tools within a few months. This is confirmed by ChatGPT frequent mentions in various news despite a decrease of the global public reactions, which, on Twitter, translates to an increase of neutral sentiment tweets and a global increase in the number of ChatGPT related tweets.

Nevertheless, we notice that some of our results could have been biased, leading to non-optimal results. Therefore, to enhance our work, further analysis should be done and mitigation methods should be applied.

REFERENCES

[1] Dimosthenis Antypas, Asahi Ushio, Jose Camacho-Collados, Vitor Silva, Leonardo Neves, and Francesco Barbieri. 2022. Twitter topic classification. *Proceedings of the 29th International Conference on Computational Linguistics*, (July 2022), 3386–3400. https://aclanthology.org/2022.coling-1.299.

[2] Konrad Banachewicz. 2023. Large language models: the tweets. Retrieved June 11, 2023 from https://www.kaggle.com/datasets/konradb/chatgpt-the-tweets.

[3] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. 2022. Xlm-t: multilingual language models in twitter for sentiment analysis and beyond. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, (June 2022), 258–266. https://aclanthology.org/2022.lrec-1.27.

[4] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: a dataset of fine-grained emotions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Oct. 2020), 4040–4054. DOI: 10.18653/v1/2020.acl-main.372.

[5] Charuni SA. 2023. Chatgpt sentiment analysis. Retrieved June 11, 2023 from https://www.kaggle.com/datasets/charunisa/chatgpt-sentiment-analysis.

[6] Zeynep Tufekci. 2014. Big questions for social media big data: representativeness, validity and other methodological pitfalls. *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*. https://doi.org/10.48550/arXiv.1403.7400.

Received 14 June 2023