Analysis and Research of Vaccine Discussion Based on Twitter

Zhaolin Fang, Qiuyi Bao, Xin Yang, and Yongqing Liang

Vrije Universiteit Amsterdam, Amsterdam, Netherlands
`z2.fang/q.bao/x7.yang/y4.liang@student.vu.nl`

Abstract. COVID-19 vaccines play an extremely important role in preventing the spread of the coronavirus. However, there are different views on vaccination in the current society. The social web represented by Twitter is a platform for the masses to express their views on the COVID-19 vaccination. Excavating these tweets can help us understand people's different attitudes towards COVID-19 vaccines and their respective reasons. This research paper uses Twitter's API to obtain all vaccine-related tweets from January 9, 2020 to December 11, 2021, and performs sentiment analysis and keyword extraction on these tweets to study people's attitudes towards vaccination. The results indicate that the discussion of the COVID-19 vaccine on Twitter peaked from January to April 2021. And even as time changes, the tweets showing a positive attitude towards the COVID-19 vaccine have always been the majority. Most of this group of tweets praised the government's free vaccination measures, the good protective effect of the COVID-19 vaccine and its sufficient supply. The anti-vaccine campaigners on Twitter expressed their concerns about the safety of the COVID-19 vaccine, especially on the side effects.

Keywords: COVID-19 · Vaccine· Twitter· Sentiment Analysis · Social Web.

1   Introduction

The COVID-19 pandemic spread rapidly around the world from the end of 2019. According to the official website of the World Health Organization, as of 10 December 2021, the pandemic has caused a total of 5.28 million deaths globally[1]. The International Monetary Fund estimates that the pandemic will cost the global economy 21 trillion GBP [2].

Because the COVID-19 virus is highly infectious and highly mutable, it is difficult to eradicate the spread of the COVID-19 virus solely through prevention and control measures in a globalised economy. In addition, studies show that even after recovery, patients with COVID-19 still suffer some irreversible damage. Vaccination against COVID-19 is therefore a matter of urgency from the perspective of national development and personal health.

Many people support vaccination to control the spread of the pandemic, but there are also people who oppose vaccination, with oppopnents subscribing to

an alternative, non-scientific worldview [3]. Twitter, which represents part of the Social Web, has become a major forum for citizens to express their opinions, and these comments contain a wealth of valuable information.

This research paper explores two questions using Twitter as a method of entry:

(1) How have attitudes towards vaccines changed over time?

(2) What are the most commonly used keywords and what do they represent?

With the collection and analysis of comments on COVID-19 vaccination, we can identify people's emotional tendencies toward vaccination at different points in time and the proportion of different emotions they express.

In exploring the meaning behind the high-frequency words in vaccine-related discussions, we can understand why people express different emotions regarding the vaccine.

In this paper, sentiment analysis and keyword extraction are conducted on Twitter's tweet data in order to address the two questions above. In all probability, it is of significant importance to the government in order to carry out targeted vaccination propaganda and guidance more efficiently.

2   Research Overview

Vaccination is the primary tool for governments to fight against COVID-19. Since the end of 2020, vaccines have been distributed worldwide in an effort to contain the virus and prevent its further spread. Since the beginning of the pandemic, a lot of different views on the fight against COVID-19 have been expressed on different social media platforms. In the beginning, the vaccine was brought to the people as the only solution. However, as time passed by and new mutations arose, the effectiveness of the vaccine decreased. This led to new restrictions and the call for a booster vaccination in many countries. Our research focuses on the sentiment surrounding the vaccine since the start of the pandemic. We will look into if and how the sentiment towards the vaccine changed in the past two years.

In literature, there are already some research works on the study of social sentiment towards COVID-19 vaccination via tweet analysis. Lyu et al. [4] did a sentiment analysis on COVID-19 related discussions on Twitter. They used R to clean the data and retain tweets that contained certain keywords. Their research covered the period from the start of the pandemic in early 2020 until February 2021. They concluded that there is an increasingly positive sentiment regarding COVID-19 vaccines.

In their research on COVID-19 hesitancy, Thelwall et al. (2021) [5] used the Twitter API and keyword query to collect their data, analysing tweets from between March 10 and December 5,2020. They categorised different types of vaccine hesitations, finding the main groups of vaccine hesitaters were right-wing or African Americans.

H. Lyu, J. Wang, W. Wu et al. (2021) [6] conducted a multinomial regression and counterfactual analysis on public opinions on potential COVID-19 vaccines during the period of September to November, 2020. They used a human-guided

machine learning framework to analyse over six million tweets. They classified the opinions into three groups: pro-vaccine, vaccine hesitant and anti-vaccine. They found that socioeconomically disadvantaged groups are more likely to have polarized opinions, either pro-vaccine or anti-vaccine. Furthermore, regarding the COVID-19 vaccines, the public in the U.S. is most concerned over the safety, effectiveness, and political issues regarding the COVID-19 vaccines.

Despite their inspiring findings, the above-mentioned research works still have some limitations. The research of Lyu et al. was ended before some of the major side-effects became apparent, which was the cause for a lot of discussion. Also, their research was concluded before it became apparent that the effectiveness of the vaccine decreases rather quickly over time, with the need for a booster vaccine already within one year after the first vaccines were distributed. As for Thellwall et al. and H. Lyu, J. Wang, W. Wu et al., both of their works were conducted in 2020, which concluded their research before the first vaccines were approved and administered.

While a considerable amount of research works have been carried out on the public discussions and opinions on COVID-19 vaccines in the period from March 2020 until early 2021, much less is known about the public discussions and opinions up to the end of 2021. As stated earlier, many events relating to the COVID-19 vaccinations occurred in 2021, which are very likely to influence relevant discussions. Therefore, the primary aim of our research is to provide a updated result of how the sentiment regarding COVID-19 vaccines has changed over time. Furthermore, we aim to summarize the most frequently used keywords and classify them into three categories, i.e. positive, neutral, or negative in regard to the vaccine.

3   Methods and Pipeline

We divided our analysis into two major tasks: A) Content sentiment analysis, B) Keyword extraction. The main objective of our project is to perform sentiment analysis on the content of Twitter tweets. Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, and computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information[1], such as people's views, emotions, evaluations and attitudes towards products, services, organizations, individuals, problems, events, topics and their attributes.

A wide range of techniques have been developed that can be applied to multiple tasks in sentiment analysis, including both supervised and unsupervised methods. For supervised methods, early papers utilized methods such as support vector machines, maximum entropy, naive Bayes, and feature combinations. On the other hand, unsupervised methods use affective dictionaries, grammatical analysis and syntactic patterns [7]. There are three types of granularity associated with sentiment analysis: document granularity, sentence granularity,

---

[1] https://en.wikipedia.org/wiki/Sentiment_analysis

and aspect granularity. Document level sentiment analysis is concerned with the analysis of tweets from the social networking service Twitter. As part of this section, we will introduce the method we use in accordance with our pipeline.

3.1   Data Extraction and Preprocessing

There are four parts in our pipeline as shown in Fig. 1. To start, using the Twitter API we obtained all tweets with the hashtag " #vaccine "from 9 January 2020 to 11 December 2021, totaling 338,572 tweets. Afterwards, the data was preprocessed by removing the duplex tweets, punctuation marks, single characters and double spaces, and URLs, etc. Following the data cleansing, a sentiment analysis was carried out on the cleaned data. Finally, we performed the Keyword Extraction phrase.
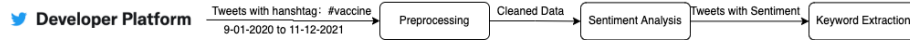

Fig. 1. The pipeline of our work.

3.2   Sentiment Analyses

Using sentiment analysis, it is possible to determine the polarity of a given document; the document can be assigned a score indicating whether the expressed opinion is positive, negative, or neutral. In addition to polarity, emotion analysis is able to provide scores for different emotions, such as fear, anger, sadness, happiness, and love [8].

In order to perform sentiment analysis on the tweets that form our dataset, we used the TextBlob library. In the context of natural language processing (NLP), TextBlob provides an easy-to-use API to perform common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, and translation. TextBlod().sentiment.polarity is used to calculate the polarity value for each tweet,defining negative polarity values as [-1, -0.01), neutral polarity values as [-0.01, 0.01), and positive polarity values as [0.01, 1].In addition, WORDCLOUD was used to extract keywords from all tweets, as well as for negative, neutral, and positive tweets separately.

As a final step to ease the process of analysis, PLOTLY EXPRESS was used for the visualisation of all the results.

4   Results and Discussion

We first counted the total number of vaccine-related tweets from Twitter users per day (see Fig. 2) and discovered that the peak of vaccine discussions occurred between January and April 2021. This corresponds to the start of widespread
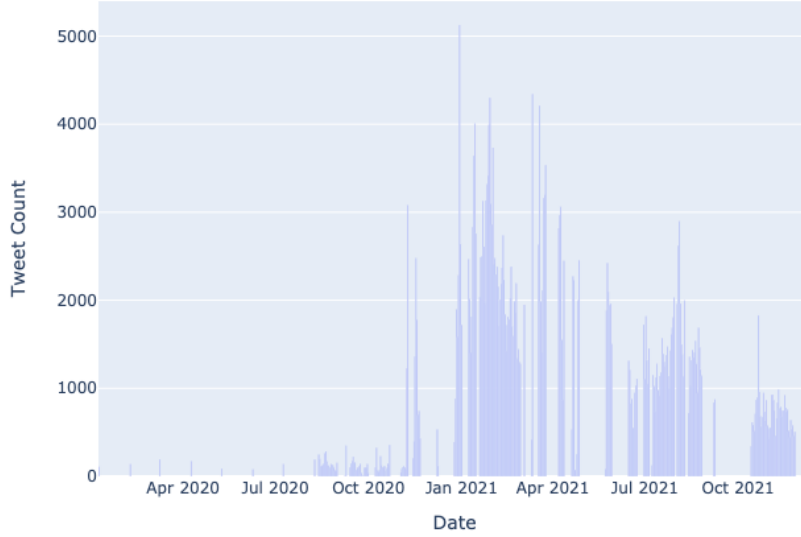
Fig. 2. The number tweet over time.

vaccination throughout a number of countries. Concerns regarding the effectiveness and safety of vaccines were causing a great deal of discussion.

Since the vaccine was in clinical trials and not widely available before January 2021, there have been fewer discussions among users then. As of April 2021, the number of tweets remains at a constant level, but is lower than the time when the vaccine was first introduced. Additionally, the sentiment analysis of people's attitudes towards vaccines revealed that more than 180,000 tweets had polarity values[2] in the range of (0, 0.25), with over half of the tweets having polarity values above zero (see Fig. 3).

According to the sentiment label distribution chart (see Fig. 4), attitudes towards vaccinations are concentrated in the neutral and positive categories. The results indicate that people's attitudes towards vaccines remain generally positive, although there are still a sizeable proportion of people who hold negative attitudes towards vaccines. As a result, we present the sentiment of each tweet in chronological order (see Fig. 5) and find that, similar to the trend in the number of vaccine- related tweets discussed earlier, the positive and negative sentiment are increasing simultaneously, and the gap of number between the positive and negative tweets are smaller after July 2020, but it is still larger than negative sentiment overall.

---

[2] Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement.
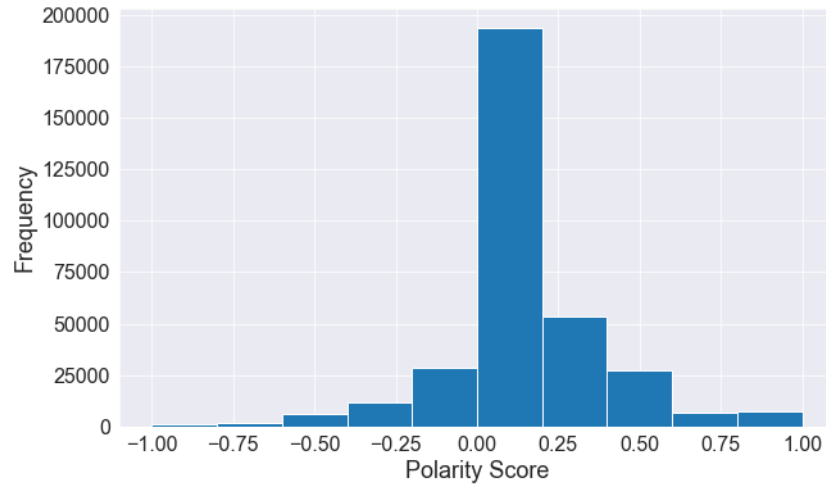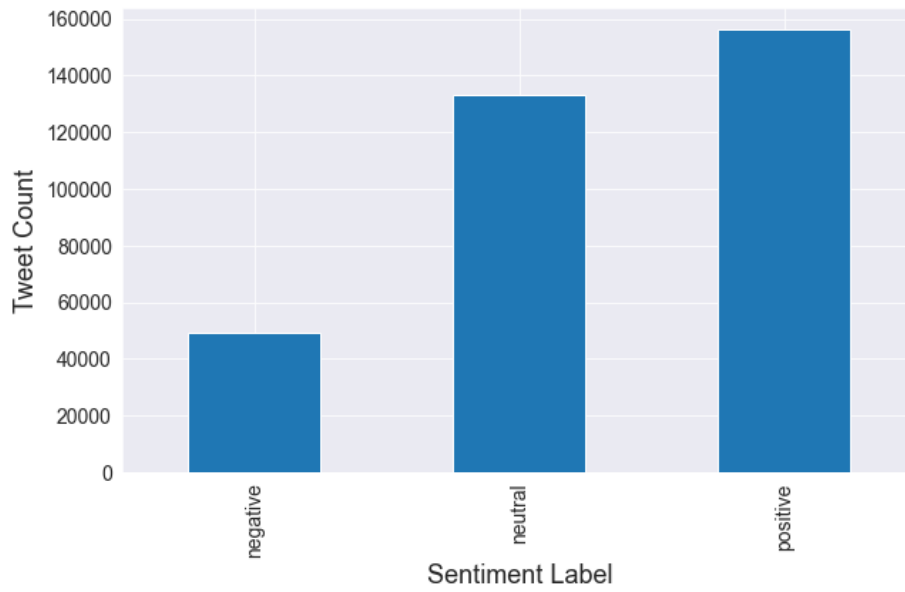
Fig. 3. Distribution of polarity values.

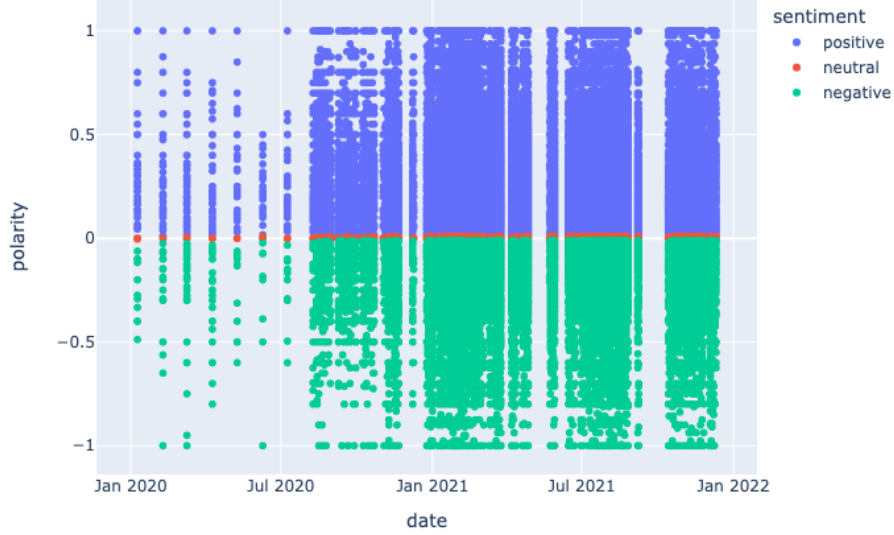**Fig. 4.** The sentiment value counts.

Fig. 5. The sentiment and polarity of each tweet in chronological order.

From the word cloud for all tweets relating to the vaccine (see word could in Fig. 6), it was found that besides the topics related to "COVID vaccine" and "COVID-19", there are also other keywords such as "side effects", "slots", "convisheild", "availability", etc. In order to better analyse the reasons behind them we categorised these keywords (see Fig. 7) into positive, neutral and negative attitudes.

Based on the findings of the study, among the keywords associated with negative sentiment, "sick", "worse", "illness" and "wrong" were among the most frequently used words, suggesting that people's negative attitudes toward vaccines were primarily due to concerns regarding vaccine safety and side effects. The most frequently used keywords that convey a positive attitude are: "available", "free", "great", "good", "effective", etc. For neutral sentiment, keywords with a high frequency of use were significantly greater than those related to the topic, but were not significantly associated with vaccine effectiveness or satisfaction.

5   Privacy Risks and Mitigation measures

5.1   Discussion on Privacy and Ethics

The topic of vaccination has many different arguments. It relates to personal lives, including religious beliefs, morals, and values. On Twitter, people were

Fig. 6. Word cloud of all tweets in the dataset.

Fig. 7. Word cloud based on sentiment tag classification.

allowed to have and create discussions and debates regarding the topic of vaccination. They post their questions, concerns, and doubts that they had about vaccination, as well as positive examples they had seen of vaccine use. There are also many different people that are against the idea of vaccination, such as those who are not vaccinated or have been vaccinated but do not believe in the value of vaccination and those who have been misinformed about their vaccinations. There is also a group of people who believe that vaccines can aid in causing cancer and autism.

But there is a group of people who do not know what side they fall on and why they believe what they do regarding vaccines. The main issue that should be taken into consideration is whether it is ethical to ignore the arguments and perceptions of others when discussing vaccines. For example, is it unethical to ignore the belief of someone who believes that vaccines can lead to cancer because medical science has proven that they do not?

When conducting research, certain measures need to be taken to ensure the safety and privacy of the individuals who are studied. This is because the freedoms of these people are limited, which also means they can be more easily exploited. These measures include obtaining consent, anonymising data where possible, and getting ethical approval. They are essential for avoiding ethical problems that might occur during the research.

Researching the surrounding vaccination on Twitter also means that researchers need to analyse all the posts on this platform.This means that there will be a lot of affected individuals whose privacy can be violated as their personal information, such as names and personal data, might be shared in the analysis report. Furthermore, these individuals might not be aware that their posts are being researched and might not agree with it.

Moreover, the discussions on the Twitter platform can lead to heated arguments regarding the vaccination, which can bring a lot of positive results or negative consequences. Therefore, it is important to know the ethical aspects of this research topic and the possible consequences that can be faced if there are privacy issues with this study. We should be able to recognise whether there are privacy issues with the research and how they could be solved without violating the rights of these individuals.

5.2   Privacy Risks

Collecting data always comes with risks. Data processing raises several privacy concerns. The use of the internet and extensive data collection is a reality in everyone's life. Data in the wrong hands can cause harm. For our research, we have to collect a large amount of data. We have to think of ways to do this in a safe and responsible manner. We have to discuss how much and what type of data we would require for our research. Our first research question focuses on the aspect of the sentiment changing over time. To answer this question, we had to collect a large number of posts, order them chronologically and assess their content. For this purpose, it did not matter to us what user posted what. Our second research question focuses on what keywords are used most frequently.

Just as for the first question, we had to collect large amounts of data and assess their content. Again, it did not matter to us what user posted the content

Our research focuses on the vaccine discussions on Twitter. In our opinion, the vaccine discussion can be marked as a sensitive topic. Not only because there are many different opinions about this topic, but also because this can be related to the medical integrity of a person. The medical integrity is interesting in our discussion about our research's privacy and ethical concerns.

What if, for instance, the user posts or replies that they will not get vaccinated? To what extent can this be considered the medical integrity of a person and thus personal data? Can this content still be used for our research? We believe that a person always has the right to medical integrity. A person never has to disclose any medical condition, treatment or anything related to their medical history to an unauthorised third party. However, we believe that once a person chooses to disclose such information openly on the internet, it can be used for our research. However, we still have a responsibility to process such information carefully. That is why we choose not to quote any posts in our paper. This is to protect the privacy of the users we collected data from for our research.

The main privacy risk during our research lays in the data collection phase. There is a risk that we will gather the personal data of users, which we are unauthorised to do. Collecting and storing personal data is subject to many laws and regulations. The most prominent law related to privacy regulations in the EU is the General Data Protection Regulation (GDPR). Article five of the GDPR states personal data shall be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation'); adequate, relevant and limited to what is necessary for relation to the purposes for which they are processed ('data minimisation')." [9] Following Article five of the GDPR, we are not allowed to collect any personal data since it is not necessary for the purpose of our research. The definition of personal data is as follows:'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" [10]. To meet our research goal the only information we collected was content of public posts and the time at which the content was posted. No other information will be gathered. Since this data is not classified as personal data it is not subject to the strict GDPR regulations.

5.3   Mitigation Measures

As a measure to mitigate the privacy risk, we used a Twitter API to collect our data. We created a developer account on Twitter and collected our data from

there. Since Twitter also has to comply with the GDPR, they were not allowed to share any personal data with us. All the data we received was anonymous to us. We did not know which user posted the contents. This means we would not be collecting any personal data unauthorised. However, this method would entirely rely on the compliance of third parties. In order to assure our compliance also, we checked all the results that came up to make sure we did not collect any unauthorised data.

6   Conclusion

This research paper identified the main mood changes and high frequency topics in COVID-19 vaccine-related discussions on Twitter between January 9, 2020, and December 11, 2021.

The peak in COVID-19 discussions on Twitter was between January and April 2021, when mass vaccination policies were introduced in multiple countries, illustrating how many discussions about COVID-19 on Twitter were sparked by major national public health events. Additionally, although the amount of discussions about vaccines has decreased since April 2021, the number of tweets expressing positive feelings toward COVID-19 vaccines has been higher than those expressing neutral and negative feelings, so we can conclude that people are generally accepting of COVID-19 vaccines. As well, by examining the keywords of tweets with positive views on COVID-19 vaccines, we are able to observe that these tweeters are positive about free COVID-19 vaccinations in their country, the protective effects of COVID-19 vaccines, and their sufficient enough supply. Note that the majority of tweets expressing negative feelings about the COVID-19 vaccines expressed concerns about the safety and side effects of the vaccines. We believe this may be helpful to governments in designing corresponding interventions, such as providing credible explanations for vaccine safety and side effects to correct anti-vaccine misperceptions and encourage voluntary acceptance of the COVID-19 vaccination.

7   Limitations of the Research

There are three limitiations to this research that stand out. It should be noted that the collected data used in this study comes solely from Twitter and it does not represent other social networks, nor can it reflect everyone's opinion of the vaccine. Furthermore, the data collected is only derived from the hashtag "vaccine", whereas there may be many tweets referring to vaccines without such hashtags or using other hashtags. Another limitation is that the overall classification of users' vaccine-related tweets (Negative, Neutral, Positive) reveals only the overall trend and does not investigate the reasons for why users have such feelings about vaccines.

Due to time constraints and API restrictions, which make it impossible to obtain larger amounts of data, the study is unable to overcome these limitation. To process more data using multiple hashtags would have taken us significantly

longer to process. This is because we would have to filter out all duplicate tweets before processing. Tweets usually have multiple hashtags, so one can imagine there will be a lot of duplicate tweets to be processed.Further, the cause cannot be inferred from the way in which the user responds in their tweets to vaccines. There are a number of ways in which social media data can be used to determine causal relationships between vaccine-related tweets. Yet, it is not the scope of this study to analyze these data because they may contain sensitive information regarding the privacy of the users, such as their nationality and religion etc.

The source of our research was limited to Twitter only and we used only the hashtag "vaccine". This gives us a limitation in the amount of data we processed for our research. However, the amount of data (338,572 tweets) was still large enough to process and conduct our research. Also, as mentioned before, tweets usually have multiple hashtags. Although we would have had duplicate data to filter, if we have included more hashtags in our research. It is not without a doubt to say that this would have resulted in better analysis. Even though the tweet data was collected, it is not possible to explore their attitudes towards vaccines in a large space due to the limitations mentioned above. In addition, it appears difficult to infer the causes of people's attitudes towards vaccines from their tweets and how such causes are the result of such attitudes. However, our research questions did not focus on the cause of such attitudes, but merely on how the attitudes have changed over time and what the most frequently used keywords are. Hence, this limitation does not affect our research to an unacceptable extent.

Research should be conducted in the future to gain more insights into the usage of other hashtags or tweets without the hashtag "vaccine" on Twitter or other social media platforms. In this way, we will be able to investigate whether there is any correlation between the individual's Twitter identity and their tweets concerning vaccines. Secondly, we would like to examine how negative emotions are portrayed in tweets about vaccines. Lastly, we will look at how the vaccine generates data from users. It is widely accepted that social media is an open and transparent source for gathering public opinion and providing valuable information that can enhance the quality of decision-making on various issues, including public health issues.

References

1. https://covid19.who.int/
2. (2020). World Economic Outlook, October 2020. World Economic Outlook.
3. Browne, M. (2018). Epistemic divides and ontological confusions: The psychology of vaccine scepticism. Human Vaccines & Immunotherapeutics, 14, 2540 - 2542
4. Joanne Chen Lyu,"COVID-19 Vaccine–Related Discussion on Twitter: Topic Modelingand Sentiment Analysis"(J Med Internet Res 2021;23(6):e24435) doi: 10.2196/24435
5. Thelwall, Mike; Kousha, Kayvan; Thelwall, Saheeda (2021). "Covid-19 vaccine hesitancy on English-languageTwitter". Profesional de la información, v. 30, n. 2, e300212.
6. Hanjia Lyu, Junda Wang, Wei Wu, Viet Duong, Xiyang Zhang, Timothy D. Dye, Jiebo Luo, Social media study of public opinions on potential COVID-19 vaccines: Informing dissent, disparities, and dissemination, Intelligent Medicine, 2021,ISSN 2667-1026, https://doi.org/10.1016/j.imed.2021.08.001.
7. Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words:Liwc and computerized text analysis methods.Journal of language and social psychology, 29(1):24–54, 2
8. Plutchik R. A general psychoevolutionary theory of emotion. In: Robert P, Henry K, editors. Theories of Emotion. Cambridge, MA: Academic Press; 1980:3-33
9. https://gdpr-info.eu/art-5-gdpr/
10. https://gdpr-info.eu/art-4-gdpr/

Appendix

A Individual Contribution

– **Zhaolin Fang**:

For the Jupyter notebook assignment A3, I finished the whole coding and question answering tasks according to the steps in the notebook, then I showed and explained to my teammates how it works. And for the Jupyter notebook assignment A1, I helped my teammates to solve the problems of configuring the python environment.

For the rubrics part "research question and motivation", Yongqing LIANG and I proposed the research direction - Sentiment Analysis, and discussed with other teammates to finalize our research questions. I also reviewed one paper related to our topic and extracted the key parts of the review results of other teammates. After discussing the feedback, I provided the outline of the literature review part of our assignment to Qiuyi BAO, then she finished this part.

For the rubrics part "Methods & Data", I first obtained the data set from the Twitter API, consulted the relevant materials according to the data set, and finally determined the methods and libraries to be used. Then I used these methods and tools to complete the whole notebook of Sentiment Analysis and Keywords Extraction of our project.

For the rubrics part "Visualisation and conclusion", I used the Plotly library to visualize the results of data analysis, and wrote the initial version of the result analysis to Qiuyi BAO for further optimization.

For the rubrics part "Privacy and Ethical considerations", I provided a brief introduction on which privacy risk we will face during the process of data analysis, and gave an outline and some ideas to Qiuyi BAO according to my experience during the process of data analysis for the rubrics part "Limitations of your research".

For the rubrics part "Layout", I copied all the contents of our paper into the template provided by the professor and formatted it to make the paper clearer and easier to read.

– **Qiuyi Bao**:

For the rubrics part "research question and motivation" we discussed with all four members of our group what research questions would be suitable for our research topic. After discussing with all four group members, we came up with two research questions. I worked on the literature review together with Xin YANG and Yongqing LIANG. Xin YANG, Yongqing LIANG and I all reviewed one paper related to our topic and provided each other with feedback. After discussing the feedback, I wrote the literature review part for our assignment.

For the rubrics part "method and data", Zhaolin FANG wrote a draft for the "Method" chapter and the "Data extraction and preprocessing" subchapter.

He sent the draft to me. I finalized the draft by checking the grammar and spelling and to make the style consistent with the rest of the paper.

For the rubrics part "Visualisation and conclusion", Zhaolin FANG provided me with a draft version of the chapter "Result and Visualisation". I then finalized the draft by checking the grammar and spelling and to make the style consistent with the rest of the paper. Xin YANG provided me with a draft of the chapter "Conclusion". I then again finalized the draft by checking the grammar and spelling and to make the style consistent with the rest of the paper.

For the rubrics part "Privacy and Ethical considerations", I wrote the chapter 5 "Privacy Risks and mitigation measures" and added this to the pape.

For the rubrics part "Limitations of your research", I wrote chapter 7 "Limitation of the research" and added this to the paper.

Finally, I did the final proofreading work for the whole research paper after we finished.

For the assignment of Jupyter Notebook 1 on Twitter data collection, I finished this assignment. In the beginning, I was struggling with installing twitter PrettyTable matplotlib, which was successfully installed in the end after the help from my teammate Zhaolin FANG.

For the assignment of Jupyter Notebook 2 on data representation, I have helped the team to answer the Task 1 question of "Is KML a microformat, why (not)?"and Task 3.1 question of "Compare the schema.org information about a band on last.fm to the Facebook Open Graph information about the same band from Facebook. What are the differences? Which format do you think supports better interoperability?".

– **Xin Yang**:
Regarding the research data and research questions, the four members of our group asked the professor and determined that the research data must come from the social web represented by Facebook and Twitter. And now vaccines are a hot topic, so the topic of the research paper was finally determined to be the analysis and research of vaccine discussion based on twitter through online communication and offline session discussions. After determining the topic, the four members of our group discussed and determined four research questions together, and then TA suggested that we need to reduce them to two.

Regarding the abstract part, Zoe suggested that I should write according to the structure of background, methods, and results. I accepted her suggestion because it does make the structure clear and the content easy to understand. I wrote the method part in detail at first, because I think this part is very important. Zhaolin FANG recommended deleting some after reading it, because the abstract needs to be concise. Finally, I added the modified abstract to the paper.

Regarding the introduction section, after reading several related documents, I decided to explain the importance of Covid 19 vaccination in terms of national development and personal health. Furthermore, Twitter is one of

the most popular social media platforms. These factors are the research motivation of this paper. Then I first explained the two research questions of this paper by narrating the meaning of the research questions. After reading it, Zoe suggested that I write down the two research questions directly to make it clearer. Then I sent the modified introduction to Zoe to check the grammar and spelling to make it consistent with the style of other parts of the paper, and finally added it to the paper.

Regarding the literature section, that is, the content in the theoretical background section of the paper was completed by Zoe, Yongqing LIANG and me. I chose three articles with similar research topics and sent them to the discussion group, and then we each chose one to write a short overview, and after giving feedback and discussion to each other, Zoe finally combined the three parts to write the literature review part.

Regarding the conclusion section, I was responsible for summarizing the main findings of the paper. Then Zoe checked the grammar and spelling to make it consistent with the other parts of the paper, and finally added it to the paper. For the assignment of Jupyter Notebook 2 on data representation, I completed the code filling. During the period, Zhaolin FANG helped me solve the problem of incomplete access to the map information of Amsterdam.

– **Yongqing Liang:**
  For the assignment of Jupyter Notebook 4, I have done all the relevant work, including answering theoretical questions and reviewing relevant materials for programming the code.

  For the rubrics part "research question and motivation" , I read the relevant literature "Social media study of public opinions on potential COVID-19 vaccines: Informing dissent, disparities, and dissemination" and extracted from it some research perspectives and technical approaches that are relevant to our project, such as the background knowledge and motivation for the research and the perspectives from which the conclusions were analysed with, and applied them to our project.

  For the rubrics part "Visualisation and conclusion", I used some of the data generated by the code, worked on the plots and wrote through the visualisation of the results, analysing the corresponding conclusions.

  For the rubrics part "research question and motivation", I reviewed the information to provide part of the idea for our project, as well as the research question, motivation and research topic.

  For the rubrics part "Privacy and Ethical considerations", I have found the relevant slides from the web data processing system to support the information in this section.

  While working on the four notebooks, I provided some support and guidance to my teammates in configuring their environments. Finally I checked and corrected some grammatical and spelling errors in the paper.