Unveiling Sentiments and Topics in COVID-19 Vaccine Comments on YouTube Over Time: from the First Vaccine Approval to the Post-Pandemic Era

s1122141
Radboud University, Nijmegen, Netherlands

ABSTRACT

While extensive research has examined emotional responses, public sentiments, and topic modeling related to COVID-19 vaccines, the majority of them have primarily utilized data from the early breakout of coronavirus or vaccine rollout until 2021 or 2022. However, a deeper insight into the public's opinions can be gained by taking into account discourses on YouTube videos related to the coronavirus vaccines from the first vaccine approval until the post-pandemic era. The paper analyses a dataset that consists of YouTube comments related to COVID-19 vaccine with a focus on comments published from August 11, 2020, when it was the world's first COVID-19 vaccine approval from Russia to December 2023, which is seven months beyond the official declaration that COVID-19 is no longer considered a global threat (May 2023). Through BERT language models, the paper presents a monthly sentiment analysis and topic modeling, offering insights into how discussions evolved as vaccine development extended and the pandemic progressed. The analysis highlights persistent negative sentiments, especially around vaccine efficacy and safety, while also capturing moments of positive sentiment reflecting support for healthcare efforts and vaccine developments. This extended timeframe allows for a subtle insight into the public discourse through the various stages of the pandemic's impact.

1 INTRODUCTION

1.1 Background

The recent global spread of the SARS-CoV-2 coronavirus began in December 2019 in Wuhan City, China, causing infections primarily in the respiratory system of affected individuals. Within around two months, this virus outbreak escalated into a global pandemic, as declared by the World Health Organization (WHO) on March 11, 2020[10]. By March 2020, the number of confirmed COVID-19 cases reached 96,000 worldwide [14] and increased rapidly thereafter.

The need for emergency approval of vaccines was very high due to the asymptomatic nature of COVID-19 cases which led to widespread virus transmission. Various initiatives have been established to develop vaccines against COVID-19, with the first approval granted by Russia of the "Sputnik V" on August 11, 2020[1]. Subsequently, numerous vaccines have been approved globally, with the most prominent ones that received an Emergency Use Authorization from the WHO being: Pfizer, Moderna, AstraZeneca, Janssen, and SinoVac, with different approaches being applied to each vaccine[16]. Following December 2020 and the COVID-19 vaccine roll-out, vaccination campaigns have been launched worldwide, with the acceptance rate standing at 64.9%[12].

TxMM'23-'24, Radboud University, Nijmegen, Netherlands

Meanwhile, more and more people have begun expressing their opinions about the effectiveness and other important concerns of the recently approved COVID-19 vaccines on different social media platforms[16]. For this reason, previous research papers have highlighted the important role that social media plays in disseminating information and in the case of COVID-19, shaping attitudes toward vaccines. The systematic review by Cascini F. et al[2] establishes a link associated between vaccine hesitancy and social media, highlighting the dominance of anti-vaccination discourses on different social media. Furthermore, according to this study, these social media discourses can offer beneficial insights not only into the reasons for people's vaccine acceptance or refusal but also into the influential role of social media figures, particularly doctors, in shaping individual attitudes toward vaccines. Additionally, the study underlines the potential role of social media platforms in comprehending the public's views on the COVID-19 vaccination and making targeted and on-time interventions to encourage the public's vaccine acceptance rate.

The pandemic spanned about three years until the WHO declared it a non-emergency global threat on May 11, 2023. Since then, health policies about the COVID-19 vaccination have become less strict in many countries worldwide. In the post-pandemic era, there have been various studies that investigate the post-effects of the COVID-19 vaccines for different scopes, for example, on human fertility and neurological complications[3, 15].

Following the official end of the pandemic, there has been little work on analyzing the discussions on social media about the COVID-19 vaccines in the post-pandemic period. The majority of research has primarily relied on data from the early breakout of coronavirus or vaccine roll-out until 2021 or 2022, often drawing from the Twitter platform. A more meaningful exploration would involve investigating how the post-vaccination effects in the period when COVID-19 is not a global threat influence the discussions across different social media. In that way, based on the review research of Cascini F. et al. [2], analyzing the connection between social media and COVID-19 vaccination could support establishing effective policies to manage vaccine reluctance and promoting vaccination worldwide.

1.2 Research Objectives and Questions

While understanding how the sentiments and topics discussed over social media evolved during the pandemic period is crucial, it is equally important to uncover how these were particularly expressed in YouTube comments and evolved over time in response to these significant milestones, from the first approval of Russia's vaccine until the post-pandemic era.

Recognizing the need to unveil these sentiments and topics, this research project aims to comprehensively examine the changing

dynamics by investigating the following essential research questions:

- RQ1: What are the public sentiments in discussions on COVID-19 vaccine-related YouTube videos from the first COVID-19 vaccine approval to the post-pandemic period?
- RQ2: What are the primary topics discussed over that period?
- RQ3: How do the sentiments of the major topics evolve over that period?

2 RELATED WORK

As this study seeks to analyze the changing sentiments and topics of COVID-19 vaccine discussions over time, it is essential to contextualize this research project within the existing body of research.

In a recent study conducted by Huangfu L. et al[5], COVID-19 vaccine-related tweets spanning from December 14, 2020, to April 30, 2021, were analyzed using a sentiment-based topic modeling approach, highlighting the significance of this method in identifying topics and trends based on sentiment groups. The study reveals that negative sentiment patterns remained stable over the examined period, whereas the positive ones were more susceptible to fluctuations. The findings show how sentiment distributes among the various topics, explaining exhaustively the sentiment scores in the extracted comments over time for different vaccines and the evolving negative topics on a weekly basis.

Another important study by Ljajić A. et al.[9], similar to the research by Huangfu L. et al[5], tries to unveil the reasons behind the COVID-19 vaccine hesitancy in Serbia, extracting tweets from January 2021 to June 2022 and using BERT-based classifiers. The research aimed to identify prevailing topics with a negative sentiment score regarding vaccination. The results highlighted significant reasons for vaccine hesitancy, including "concerns over vaccine side effects, effectiveness, insufficiently tested vaccines, mistrust of authorities, and conspiracy theories".

Additionally, a study conducted by Luy C. L. et al.[11] analyzed the major sentiments and topics concerned on Twitter discussions and discovered that they were largely influenced by significant events and progress regarding COVID-19 vaccines over the selected period, from March 11, 2020, to January 31, 2021. The study further indicates a general overall positive sentiment trend over time, which may imply a higher acceptance rate of this vaccine than other traditional ones.

In these three studies, it is essential to acknowledge certain limitations, especially the need for analyzing users' comments on different social media platforms to identify alternations in the sentiments and topics discussed. However, it is equally crucial to examine the evolving public attitudes and topics discussed over the post-pandemic period, starting with the first vaccine approval by Russia, and understand their dynamics during and after the pandemic. To address these considerations, this project examines comments extracted from YouTube videos related to vaccines over the selected era. Moreover, getting inspiration from the insights gained in the analysis of the first and second studies cited earlier[5, 9], this project aims to explore how the sentiment score evolves for each dominant topic by using the BERT training approaches to achieve high performance at the targeted tasks.

3 APPROACH

3.1 Extraction and Cleaning the dataset

The dataset collection is based on the comments on YouTube videos related to the COVID-19 vaccines. To achieve this, 1195 videos were scraped by searching with the query "Covid-19 vaccines" through the YouTube API. The relevant language was set to English to retrieve only English videos. A timeline restriction was applied to maximize the number of videos retrieved, spanning from November 1, 2019, to November 16, 2023. Subsequently, these video IDs were employed to extract their comments, focusing on a specific timeframe from August 11, 2020, to December 13, 2023. Furthermore, the duplicate entries have been removed, and the English comments have been distinguished to be used by utilizing the fastText library to detect the English language in the extracted comments. FastText uses trained models for language detection for over 176 languages[6, 7] and has been selected due to its time efficiency. After that, cleaning techniques were applied to lowercase the text, remove URLs, special characters, emojis, punctuation characters, and empty values, and filter out comments created by the same author in the same video, by spammers, and by the video creator. Lastly, English stop words were dropped to enhance the quality of the extracted comments only for the BERT topic modeling approach.

After applying these data-cleaning approaches, the final dataset consists of 907,380 comments from 1,195 videos and 592,780 different users.

3.2 Stratified Sampling

Due to the large-scale extracted dataset and the limited resources available, a stratified random sampling approach was applied. This approach was chosen to retrieve a smaller and more manageable set of comments that fit in the available memory to perform the analysis using the BERT language model approach and for its ability to maintain homogeneity in the produced sample dataset. The stratified method works by partitioning the dataset based on certain features into groups [13] and in this case, the strata were the video IDs and the months each comment was published every year. This technique enables the acquisition of a random sample that represents each selected stratum, thereby eliminating bias in the sampling and ensuring a manageable dataset.

With this approach, the final dataset is composed of 591,016 comments from 1,195 videos and 415,298 different users.

3.3 Sentiment Analysis

For the sentiment analysis, the BERTSent pre-trained sentiment analysis model[8][1] was used due to its fine-tuning in COVID-19-related tweets. This model distributes the given text into 3 sentiment categories: positive, neutral, or negative, using probability scores. During the cleaning of the dataset, the stopwords haven't been removed because the model has been trained on a dataset that includes them because they contribute to the meaningfulness of the text. An example of comments from each sentiment category is illustrated in Table 1.

[1] https://huggingface.co/rabindralamsal/BERTsent

It is assumed for the purposes of this analysis that the target of the sentiment in the comments is the COVID-19 vaccine entity, and the sentiment experiencer is the author of the comments. These assumptions are justified on the basis of the fact that comments under videos usually reflect the feelings of the authors themselves, and the videos are related to COVID-19 vaccines, so it is expected that this is the main subject of comment.

In order to ensure the external validity of the sentiment analysis process, a small dataset subset (N=30) has been manually labeled in each sentiment category, and the F1-score has been computed. The achieved score on the negative sentiment was 90.9%, the neutral 80%, and the positive 85.7%. Moreover, the related research studies and relevant events have been used to validate and confirm the output from both the sentiment analysis and topic modeling tasks.

Sentiment Category	Comment Text
Positive	"got the sputnik vector vaccine it s a very good vaccine"
Neutral	"vaccine is to new we don t know the long term effects"
Negative	"which is worse catching covid virus and ending up on a ventilator life support spreading the virus or getting the vaccine and suffering serious side effects of such"

Table 1: Examples of comments distributed across positive, neutral, and negative sentiment categories.

3.4 Topic Modeling

For the topic modeling, the BERTopic framework[4] has been applied to this analysis because of the coherent topics that it can generate and its capabilities over the dynamic topic modeling task of this study. By adopting a class-based version of the TF-IDF approach, this technique leverages language models for topic generation allowing to represent topics as distributions of words, enhancing the precision in representation modeling[4]. One main assumption that BERTopic has is that each document contains only one topic, which sometimes does not reflect reality. However, its flexibility, performance, and capability of fine-tuning can address this limitation. During the embedding creation step, the default sentence pre-trained transformer model "all-MiniLM-L6-v2"[2] was utilized to identify the similarity semantically across the comments. As for the dimensionality reduction, the UMAP solution was applied by altering the parameters to improve the representation of the embeddings. Moreover, for fine-tuning the topics, three representation models, the default KeyBert, the maximal marginal relevance model, the zero-shot classification model assigning some candidate labels to the generated topics, and the OpenAI with "gpt-3.5-turbo" representation model, were combined, and compared to retrieve the most coherent topics using the multi-aspect topic modeling.

4 RESULTS AND ANALYSIS

The results are organized into three distinct sections: initially presenting the Sentiment Analysis, followed by Topic Modeling, concluding with an exploration of the top topics combined with the sentiment trends.

[2]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

4.1 Sentiment Analysis

The results to answer the RQ1 revealed that the negative sentiment was stronger than the positive on this selected social media platform compared to the results of Huangfu et al.(2022)[5] on Twitter. More specifically, as illustrated in Figure 4, they were found 307,931 negative, 221,129 neutral, and 61,956 positive comments in the total extracted dataset.

The two line charts illustrate the distribution of sentiment of comments over time in occurrences(Figure 1) and in percentages(Figure 3 in Appendix). These figures show that the negative trend increased from November 2020, when the COVID-19 vaccine discussions increased due to Pfizer stating of 90%effectiveness vaccine[5] and, in general, vaccine roll-out worldwide, with some fluctuations in the meantime, until March 2022, when there was a sudden drop. Before November 2020, the negative sentiment seemed to show an increase from August 2020 with the highlight in October 2020. This can be justified because during that period AstraZeneca vaccine clinical trials paused due to safety concerns. Furthermore, there was an increase in negative sentiment in March 2021 and April 2021 which can be explained since the Astra Zeneca and Johnson & Johnson vaccinations were halted for the side effects reasons.

The neutral sentiment, overall, follows the same pattern as the negative sentiment. The positive sentiment seems to be, from the beginning till the end, very weak with a slight increase after the period where the negative sentiment shows a drop until approximately September 2022 where it shows a roughly stable percentage until December 2023.

After September 2022, the negative sentiment shows a sudden rise which may be explained because of the wave of concerns about the necessity and safety of a booster shot approved by the FDA. The fluctuations in the negative sentiment continued with May 2023 showing a slight decrease probably due to the official ending of COVID-19, which caused this drop, as people finally noticed an ending to this pandemic and the vaccines. However, the opposite trend follows for the next months, when after October 2023, there is a slight proportional increase in the positive sentiment possibly due to the initiations of autumn vaccine campaigns in different countries, and the Novavax vaccine approval made with a different type of vaccine technology than the mRNA technology which is linked with conspiracy theories as stated by Ljajić A. et al.,[9].

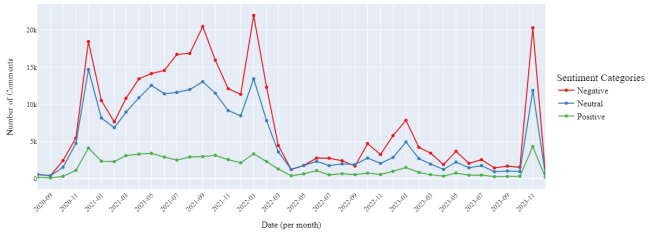

Figure 1: Monthly Sentiment Trend line chart for the extracted YouTube comments about COVID-19 vaccines. The chart shows the trend in public sentiment, categorized into negative(red line), neutral(blue line), and positive(green line) comments from August 2020 till December 2023

4.2 Topic Modeling

In addressing RQ2, the analysis revealed 10 dominant themes in discussions over time. Figures 2 and 5(Appendix) show the frequency of each topic over time, with the most frequent topic being the controversy surrounding the COVID-19 vaccines efficacy. This topic shows a rapid increase at the end of 2020, with some fluctuations in 2021 and 2022, when there is a significant drop around May 2022. However, the topic indicates a resurgence around January 2023 and at the end of 2023, indicating that it continues to be a major and debated issue after the official end of the pandemic. Other major themes discussed in these years include the safety and effectiveness of vaccines, appreciation, and gratitude for healthcare workers, vaccination efforts, or positive vaccine experiences, similar to the theme covered by Ljajić A. et al. study,[9], following the same trends as the first topic, indicating the influence that similar events had on these themes. Topics about vaccines as a global issue(similar to Luy C. L. et al. [11] theme), discussing about different countries (e.g. China and India) and other impacts, such as climate change, began to exhibit a trend over November 2020 until May 2021, which started to show a decreasing trend until January 2023 and September 2023 and later showing increase.

Notably, conspiracy theories, such as the 5G network, the magnetic vaccine theory, and the depopulation theory, and side effects and vaccination hesitancy, particularly for specific groups(e.g. kids and pregnant women), displayed an initial surge until March 2022, followed by a steady decrease and a renewed increase after September 2023. Discussions around vaccine information, certification, and celebrity opinions peaked between December 2020 and May 2022, aligning with wider vaccine availability. Two consistent but less prominent topics involved different vaccine discourse and vaccine development, with a notable emphasis on the USA's operation warp speed in late 2020.

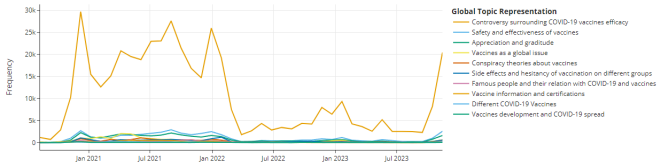

Figure 2: Monthly Topic Trend line chart for the extracted YouTube comments about COVID-19 vaccines. The line chart shows the trends of each topic representation.

4.3 Sentiment-based Topic Modeling on main topics

In addressing RQ3, the analysis of the top 5 topics revealed varying sentiment trends. Figure 6 and 7(Appendix) illustrate that for topics around the efficacy, effectiveness, and safety of vaccines, the negative sentiment leads over the whole period with the neutral sentiment following and the positive sentiment being very low. The topic of showing gratitude and appreciation indicates a strong sentiment of support and gratitude(Figure 8), while at the beginning of the examined period and in August 2023, the topic seemed to score positive low levels as the neutral and negative sentiments.

The topic related to vaccines as a global issue generally maintains a neutral sentiment with a peak around May 2021 followed by a drop until November 2023, when showing a slight increase in both negative and neutral sentiment(Figure 9 in Appendix). Positive sentiment also increases over the same period with the neutral remaining the dominant one.

Lastly, Figure 10(Appendix) presents that in the topic of conspiracy theories, both negative and neutral sentiments demonstrate similar patterns as the previous topic with some fluctuations, but the positive one remains relatively low throughout the entire period. The topic continues to have a negative sentiment even after the end of the pandemic, with discussions focusing on urging people to "wake up" to the perceived lyings surrounding the vaccines' effectiveness and technology.

5 DISCUSSION AND OUTLOOK

This paper presents how the sentiment and topics of discussions distributed about the COVID-19 vaccines over a timeframe of 3 years from the pandemic to the post-pandemic era. The results provide valuable insights into the trends of the discussions over the selected period. The sentiment analysis displays a prevailing negative sentiment, punctuated by spikes corresponding to key vaccine-related events. Topic modeling reveals the efficacy, effectiveness, and safety of vaccines as a dominant theme, with significant discourse fluctuations reflecting major events. Sentiment trends within major topics show a persistent negative sentiment towards vaccine efficacy and safety, while gratitude towards healthcare efforts holds a positive view.

Before November 2020, it seems the influence of the first vaccine approval and the clinical trials development of other vaccines influences steadily the public's discussions and sentiment. Following the pandemic's official end, there's a notable shift: negative sentiment drops, suggesting a collective relief, while positive sentiment increases slightly. However, after September 2023, possibly influenced by the new vaccination campaigns and vaccine approvals, the topics and negative sentiments follow an increasing trend. The analysis highlights the public's complex and evolving engagement with vaccine information and its impact on sentiment and topics of discussion over time.

6 LIMITATIONS AND FUTURE WORK

In this study, the analysis was constrained by the available comments of videos extracted from the YouTube platform using a specific query 'COVID-19 vaccines'. Future research is needed to focus on each COVID-19 vaccine and examine the sentiment and topic discussion. Additionally, future studies might extend these findings and compare them with datasets from different social media platforms such as TikTok and Facebook, exploring the discussion trends. Additionally, a bias in the topic modeling task might occur due to the selection and interpretation of the major topics. Furthermore, due to the time limit, the dataset was not cleaned from comments that may not have been directly related to COVID-19 vaccines and have different targets expressing negative or positive sentiments. The scope of the research was limited by the nature of the dataset containing only English and not targeted at specific geographical areas limiting the representation of global perspectives. A research with multilingual dataset could aim to include a diverse sample and generalize the sentiments and topics of this discourse over the selected timeframe.

ACKNOWLEDGMENTS

I would like to thank the Teaching Assistants and the Professor for the help and feedback to conduct this research.

REFERENCES

[1] Nikolay I Briko, Vladimir A Korshunov, Alla Ya Mindlina, Roman V Polibin, Maksim O Antipov, Alexey I Brazhnikov, Yurii E Vyazovichenko, Ekaterina V Glushkova, Kirill S Lomonosov, Alena V Lomonosova, Platon D Lopukhov, Artem A Pozdnyakov, Tatiana S Saltykova, Nikolay V Torchinsky, Natalia N Tsapkova, Olga P Chernyavskaya, and Arseny V Shamis. 2022. Healthcare Workers' Acceptance of COVID-19 Vaccination in Russia. *Int J Environ Res Public Health* 19, 7 (March 2022).

[2] Fidelia Cascini, Ana Pantovic, Yazan A Al-Ajlouni, Giovanna Failla, Valeria Puleo, Andriy Melnyk, Alberto Lontano, and Walter Ricciardi. 2022. Social media and attitudes towards a COVID-19 vaccination: A systematic review of the literature. *EClinicalMedicine* 48 (May 2022), 101454.

[3] Ravindra Kumar Garg and Vimal Kumar Paliwal. 2021. Spectrum of neurological complications following COVID-19 vaccination. *Neurol Sci* 43, 1 (Oct. 2021), 3–40.

[4] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

[5] Luwen Huangfu, Yiwen Mo, Peijie Zhang, Daniel Dajun Zeng, and Saike He. 2022. COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment-Based Topic Modeling. *J Med Internet Res* 24, 2 (Feb. 2022), e31726.

[6] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. 2016. FastText.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651* (2016).

[7] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification. *arXiv preprint arXiv:1607.01759* (2016).

[8] Rabindra Lamsal, Aaron Harwood, and Maria Rodriguez Read. 2022. Twitter conversations predict the daily confirmed COVID-19 cases. *Applied Soft Computing* 129 (2022), 109603.

[9] Adela Ljajić, Nikola Prodanović, Darija Medvecki, Bojana Bašaragin, and Jelena Mitrović. 2022. Uncovering the Reasons Behind COVID-19 Vaccine Hesitancy in Serbia: Sentiment-Based Topic Modeling. *J Med Internet Res* 24, 11 (17 Nov 2022), e42261. https://doi.org/10.2196/42261

[10] Meng Lv, Xufei Luo, Janne Estill, Yunlan Liu, Mengjuan Ren, Jianjian Wang, Qi Wang, Siya Zhao, Xiaohui Wang, Shu Yang, Xixi Feng, Weiguo Li, Enmei Liu, Xianzhuo Zhang, Ling Wang, Qi Zhou, Wenbo Meng, Xiaolong Qi, Yangqin Xun, Xuan Yu, Yaolong Chen, and COVID-19 evidence and recommendations working group. 2020. Coronavirus disease (COVID-19): a scoping review. *Euro Surveill* 25, 15 (April 2020).

[11] Joanne Chen Lyu, Eileen Le Han, and Garving K Luli. 2021. COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *J Med Internet Res* 23, 6 (29 Jun 2021), e24435. https://doi.org/10.2196/24435

[12] Dechasa Adare Mengistu, Yohannes Mulugeta Demmu, and Yohanis Alemeshet Asefa. 2022. Global COVID-19 vaccine acceptance rate: Systematic review and meta-analysis. *Front Public Health* 10 (Dec. 2022), 1044193.

[13] J. Qian. 2010. Sampling. In *International Encyclopedia of Education (Third Edition)* (third edition ed.), Penelope Peterson, Eva Baker, and Barry McGaw (Eds.). Elsevier, Oxford, 390–395. https://doi.org/10.1016/B978-0-08-044894-7.01361-0

[14] Tanu Singhal. 2020. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr* 87, 4 (March 2020), 281–286.

[15] Chao Wang, Min Wang, Guanjian Li, Bing Song, Qiong Xing, and Yunxia Cao. 2023. Effects of COVID-19 vaccination on human fertility: a post-pandemic literature review. *Ann Med* 55, 2 (Sept. 2023), 2261964.

[16] Charles Yap, Abulhassan Ali, Amogh Prabhakar, Akul Prabhakar, Aman Pal, Ying Yi Lim, and Pramath Kakodkar. 2021. Comprehensive literature review on COVID-19 vaccines and role of SARS-CoV-2 variants in the pandemic. *Therapeutic Advances in Vaccines and Immunotherapy* 9 (2021), 25151355211059791. https://doi.org/10.1177/25151355211059791 arXiv:https://doi.org/10.1177/25151355211059791 PMID: 34870090.

A WORK REPORT

The research project was composed of different steps including gathering the dataset, cleaning and filtering the dataset, performing the stratified sampling, preparing the environment for the BERT models, performing the sentiment analysis, topic modeling, and plotting the necessary charts. The most time-intensive task was extracting comments due to YouTube API's daily limits. Subsequent steps involved cleaning the dataset, selecting English-language comments, and preprocessing the data for sentiment analysis and topic modeling. In the meantime, the report was drafted up to the Research Objectives and dataset extraction sections. Handling the large dataset posed a challenge, especially fitting it into memory for BERT-based methods. After spending some days finding the best number of comments fitting into the 30GB of RAM for the BERT approaches, stratified sampling was implemented. Then, considerable time was spent on sentiment analysis and especially on topic modeling to discover the most effective representation of the topics, merging some topics and labeling them. Lastly, the final steps were composed of producing the plots, interpreting them, and writing about the results, and the discussion in the report. For refining the report's language, Chat-GPT was used in accordance with the course guidelines.

B FIGURES

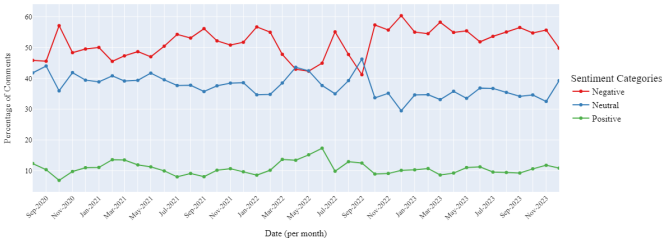

Figure 3: Monthly Sentiment Trend line chart for the extracted YouTube comments about COVID-19 vaccines using percentages. The line chart shows the trends of each sentiment category using the percentages of comments in each category.

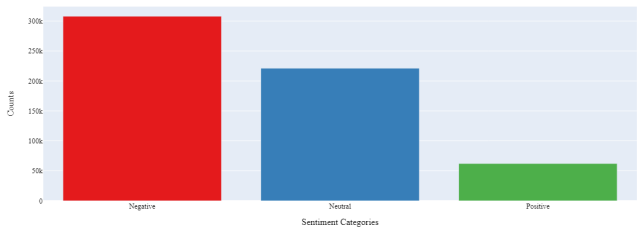

Figure 4: Overall Sentiment Distribution for the extracted YouTube comments about COVID-19 vaccines. The bar chart shows the categorization of the comments in negative, neutral, and positive sentiment, with a predominant occurrence of negative comments.

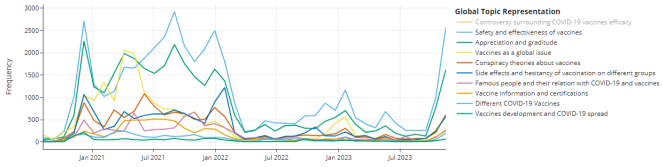

Figure 5: Monthly Topic Trend line chart for the extracted YouTube comments about COVID-19 vaccines. The line chart shows the trends of each topic representation excluding the main one: "Controversy surrounding COVID-19 vaccine efficacy" for a clear view of the other topics.

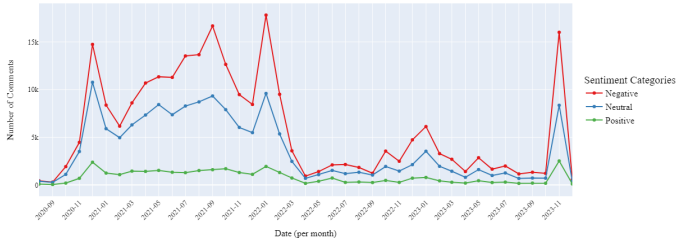

Figure 6: Monthly Sentiment Trend line chart for the topic "Controversy surrounding COVID-19 vaccine efficacy". The line chart shows the trends of the topic in each sentiment category.

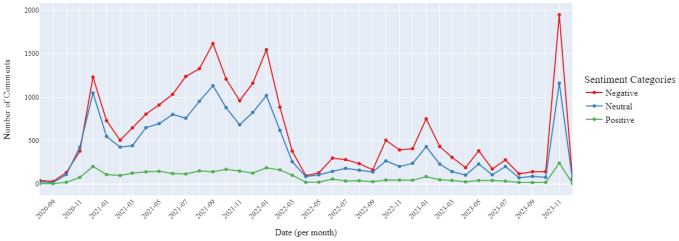

Figure 7: Monthly Sentiment Trend line chart for the topic "Safety and effectiveness of vaccines". The line chart shows the trends of the topic in each sentiment category.

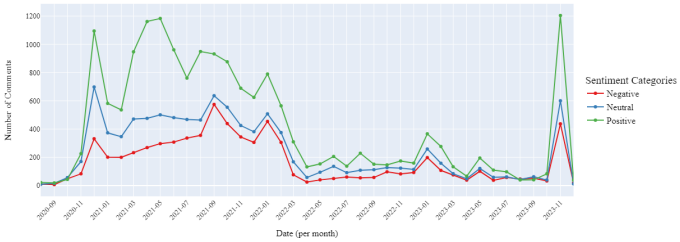

Figure 8: Monthly Sentiment Trend line chart for the topic "Appreciation and gratitude". The line chart shows the trends of the topic in each sentiment category.

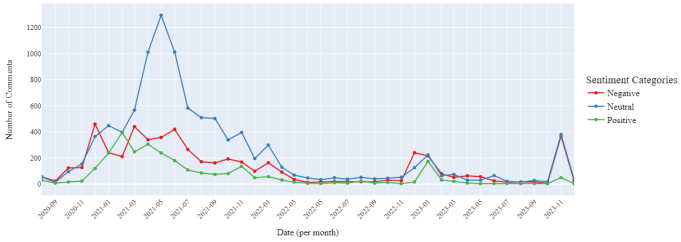

Figure 9: Monthly Sentiment Trend line chart for the topic "Vaccines as a global issue". The line chart shows the trends of the topic in each sentiment category.

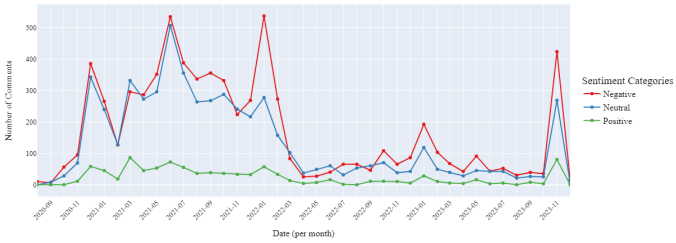

Figure 10: Monthly Sentiment Trend line chart for the topic "Conspiracy theories about vaccines". The line chart shows the trends of the topic in each sentiment category.