

Ensuring Transparency in Using ChatGPT for Public Sentiment Analysis

CHUN-HUA TSAI, University of Nebraska at Omaha, USA

GARGI NANDY, University of Nebraska at Omaha, USA

DEANNA HOUSE, University of Nebraska at Omaha, USA

JOHN M. CARROLL, The Pennsylvania State University, USA

The advancement of generative AI, involving the utilization of large language models (LLMs) like ChatGPT to assess public opinion and sentiment, has become increasingly prevalent. However, this upsurge in usage raises significant questions about the transparency and interpretability of the predictions made by these LLM Models. Hence, this paper explores the imperative of ensuring transparency in the application of ChatGPT for public sentiment analysis. To tackle these challenges, we propose using a lexicon-based model as a surrogate to approximate both global and local predictions. Through case studies, we demonstrate how transparency mechanisms, bolstered by the lexicon-based model, can be seamlessly integrated into ChatGPT's deployment for sentiment analysis. Drawing on the results of our study, we further discuss the implications for future research involving the utilization of LLMs in governmental functions, policymaking, and public engagement.

CCS Concepts: • **Information systems** → **Wrappers (data mining); Sentiment analysis; Language models.**

Additional Key Words and Phrases: AI Ethics and Governance, Civic Engagement, COVID, CDC

ACM Reference Format:

Chun-Hua Tsai, Gargi Nandy, Deanna House, and John M. Carroll. 2024. Ensuring Transparency in Using ChatGPT for Public Sentiment Analysis. In *25th Annual International Conference on Digital Government Research (DGO 2024)*, June 11–14, 2024, Taipei, Taiwan. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3657054.3657128>

1 INTRODUCTION

Analyzing public sentiment is a crucial process that aids in comprehending the emotions, attitudes, and opinions of the general population [31]. These analyses hold broad applications, from influencing policy decisions, refining product development, managing brand reputation, and handling crisis communication [19]. Businesses derive value from sentiment analysis for market research, understanding user perception, and customer satisfaction [26]. Furthermore, political science researchers and analysts employ it to gauge voter preferences and tailor messages effectively [35]. Public sentiment analysis offers valuable insights that inform decision-making, enhance public engagement, and contribute to a deeper understanding of societal trends and behaviors [33]. Sentiment analysis employs natural language processing (NLP), statistical methods, and machine learning algorithms to identify trends and patterns among attitudes, emotions, and opinions expressed in text, subsequently classifying them into categories or sentiment scores [22].

The advent of ChatGPT marks a significant breakthrough in the general public's utilization of large language models (LLMs). This groundbreaking tool allows users to tailor conversations according to their preferred language styles and structures, opening up new possibilities for human-AI interaction. One of its most promising applications is its ability to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

gauge public sentiment and perception effectively [29]. ChatGPT has undergone evaluation on various test sets and in different scenarios, comparing its performance to fine-tuned language models like BERT [36]. The results have shown that ChatGPT’s efficacy in public sentiment analysis is superior, demonstrating its potential in this domain. Due to ChatGPT being a closed-source model, it can be perceived as a black-box model, raising concerns about the transparency and accountability of its outputs [38]. A crucial aspect that remains inadequately explored pertains to the reliability of the results produced by ChatGPT and the methods for interpreting the predictions, enabling decision-makers in the public sector to grasp the reliability of the prediction outcomes. Additionally, there is a potential threat of the widespread adoption of algorithms in governmental operations, including sentiment analysis, aimed at comprehending and leveraging public sentiment for community involvement. This issue has been identified as the ‘challenge of the technical black box [21].

This paper seeks to answer the research question of *how can we ensure transparency in using ChatGPT for public sentiment analysis?* To answer this question, we employed ChatGPT to analyze public sentiment data collected from the United States Centers for Disease Control and Prevention (CDC) during the global COVID-19 pandemic. Our analysis focused on user comments concerning the CDC’s health preventive measures and policies. We presented a case study on the public sentiment toward the stay-at-home guidance published by the CDC to demonstrate the effectiveness of using ChatGPT to analyze public policy sentiment. To ensure transparency when interpreting these results, we used a lexicon-based approach as a surrogate model to investigate ways to explain the results to *approximate* ChatGPT’s predictions. The surrogate model was deemed an engineering approach when directly measuring or computing a desired outcome is challenging, i.e., we used an approximate mathematical representation of the predicted outcome [20]. We anticipate the surrogate model’s performance may not outperform ChatGPT but could provide transparency for understanding the prediction. We adopted quantitative and qualitative methods to assess the surrogate model’s effectiveness. Specifically, we conducted correlation analysis to gauge the relationship between lexicon-based and ChatGPT-based sentiment scores. Furthermore, we analyzed specific instances to enhance our understanding of the localized predictions made by our proposed models.

Through our investigations, we affirmed ChatGPT’s effectiveness in the domain of public sentiment analysis. Employing the lexicon-based surrogate model, we demonstrated their capacity to elucidate the predictions generated by ChatGPT. Our experimental results suggested that the lexicon-based surrogate model could provide a moderate approximation of the predictions. These experimental outcomes support the concept of employing explainable and transparent models to approximate the opaque predictions of AI, offering decision-makers enhanced confidence in understanding and trusting ChatGPT’s or other large language models’ predictions. Nevertheless, our findings also underscore that while transparent surrogate models offer value, they might not entirely unravel the intricacies of AI models, highlighting the ongoing necessity for exploring more robust explanatory models. By analyzing the outcomes of our study, we delve into the potential consequences for future research that entails employing LLMs in government functions, policies, and public engagement.

Through our investigation, we gained valuable insights into using a surrogate model to improve the transparency of complex generative AI methods in public sentiment analysis. Our findings are essential for understanding the adaptability of LLMs, especially in analyzing public sentiment amidst a global pandemic. This study enhances our understanding of the potential of LLMs and sets the stage for further exploration and improvement in this critical area. Building on our research, we highlight the implications for future studies in public sentiment analysis. Additionally, we shed light on the adoption and impact of generative AI in governmental functions, policies, and public engagement.

2 RELATED WORKS

2.1 Sentiment Analysis in Government Social Media

Sentiment analysis can be considered an important technique that uses natural language processing (NLP), statistical methods, and different machine learning algorithms to identify different trends, patterns, attitudes, emotions, and opinions associated with a particular text into three categories, which include positive, negative, and neutral [23]. Studies by multiple authors have discussed the growing importance of social media platforms more than ever before and considered social media as a significant proportion of people get their news from social media [4, 6]. Some of the most popular social media applications include Facebook, Twitter, Instagram, and Reddit [4]. Statistical studies show that these social media platforms heavily influence people [4]. Because of the growing importance of social media in the lives of the people, various local and federal government agencies use various social media platforms to establish a relationship between the government and the citizens [9]. Hence, considering the importance of social media in influencing public trust, several authors have examined the role of government in its social media use to establish a positive relationship with the citizens. [25] in his research used sentiment analysis to study the attitude of American citizens toward the adoption of social media by the US government and found that citizens responded positively towards government use of social media. Sentiment analysis, also known as opinion mining, is one such technique that is widely used in the field of text mining and can be used to extract valuable insights from the vast amount of user-generated data and help in understanding human emotions as the user engages themselves in the various social media platforms [3, 22]. During the COVID-19 pandemic, the government used various social media platforms to disseminate information and keep the citizens vigilant of their surroundings. [39].

Government agencies can use social media platforms to communicate with the citizens and gather their feedback effectively. As mentioned by Graham & Avery, social media has the ability to help the government establish an effective, transparent relationship with citizens [13]. Increasing people's engagement on government social media pages allows them to express their opinions on government decisions. Scholars and practitioners have looked at the effectiveness of sentiment analysis in various domains, including tourism, education, politics, and health, for various beneficial purposes like analyzing public opinions and gathering real-time data trends [3]. [41] in their study used sentiment analysis of US local government tweets to determine the overall connection between the tone and the citizen involvement, i.e., whether positive tweets encourage more citizen participation than negative tweets. Their study found that the positive tone in government tweets can influence citizens to participate and engage more than neutral or negative tweets. [26] in their study established a strong correlation between the rise/fall in stock prices of a company with that of the public sentiments and opinions expressed in Twitter [26]. Recent study by [5] used aspect based sentiment analysis to gain insights on customer's expectations and needs when using government mobile apps. They used hybrid approach integrating lexicon and rule based approach to extract different aspects from the customer reviews and categorize the associated sentiments. Hence, sentiment analysis of government-citizen interaction in social media platforms can be considered a powerful tool to help policymakers better understand public concerns and improve their service decision-making process [10].

2.2 Sentiment Analysis and Techniques

Sentiment analysis is broadly classified into three distinct levels: document level, sentence level, and aspect level [34]. At the document level, sentiment analysis is conducted on the entire text to gauge its overall sentiment, making it essential for the document to be centered around a single subject. This level assesses the general sentiment polarity of the text,

identifying it as either positive, negative, or neutral. On the other hand, sentence-level analysis focuses on evaluating the sentiment of individual sentences within a document. Each sentence is analyzed to determine its sentiment polarity, classifying it into positive, negative, or neutral categories. The method is usually adopted lexicon-based method [32]. The aspect-level analysis, also referred to as feature-level analysis, delves into the finer details by examining specific aspects or features mentioned within the text. For example, in the sentence "The camera of the iPhone 11 is awesome," the analysis would focus on the "camera" of the "iPhone 11," identifying the sentiment as positive toward this specific feature. This level of analysis is particularly useful for pinpointing precise likes and dislikes in user feedback. Aspect-level sentiment analysis often requires more nuanced techniques, such as machine learning, to accurately identify and evaluate sentiments related to specific features or aspects within the text, which could more accurately estimate the public sentiment [40].

The lexicon-based approach has been extensively utilized in social media sentiment analysis, as highlighted by [32]. This method involves the initial segmentation and tokenization of words, which are subsequently assigned polarity values (e.g., positive = 1, negative = -1, neutral = 0) based on a predefined dictionary [24]. The cumulative sentiment score is derived by totaling these values, thereby establishing the text's overall sentiment as positive, negative, or neutral. For example, in analyzing the phrase "The food is awesome," tokens such as "the," "food," "is," and "awesome" are identified. Upon removing stop words and any punctuation, the focus shifts to "food" and "awesome," with "food" likely being deemed neutral (0) and "awesome" positive (1). The lexicon-based methodology bifurcates into dictionary-based and corpus-based categories. The dictionary-based strategy involves the creation of a sentiment dictionary from an initial set of sentiment-laden words. This dictionary is expanded by identifying synonyms and antonyms within various lexicons, thereby enriching the dictionary with alternative expressions. This expansion continues until no further additions are possible. On the contrary, the corpus-based strategy starts with an existing set of words, using this as a basis for further analysis [23]. Despite its wide application and explainable nature, the lexicon-based approach may face limitations in accurately capturing the nuanced meanings of sentiments expressed, particularly due to the context-specific usage of words, which may lead to lower accuracy in estimating public sentiment [27].

The machine learning approach to sentiment analysis fundamentally depends on analyzing historical data to improve the prediction of outcomes. This method incorporates various machine learning algorithms broadly classified into supervised and unsupervised learning categories. The principal difference between these categories is that supervised learning algorithms require pre-defined output labels for training, facilitating a direct comparison between predicted and actual outcomes. In contrast, unsupervised learning algorithms derive insights and patterns directly from the data without predefined labels, enhancing prediction accuracy by analyzing past data [23]. An illustrative study by [14] compared the efficacy of various sentiment analyzers against two machine learning algorithms, Naive Bayes (NB) and Support Vector Machine (SVM), in the context of analyzing sentiments related to election discussions on Twitter. The findings indicated that the Word Sense Disambiguation (W-WSD) method yielded higher accuracy when used in conjunction with the NB classifier. Similarly, TextBlob and W-WSD showed improved accuracy when evaluated using the SVM algorithm. While machine learning approaches can achieve high accuracy in sentiment analysis, they may encounter challenges related to explainability, which refers to the ability to interpret and understand the decision-making process of the algorithms [34].

2.3 Utilization of Large Language Models (LLMs) for Sentiment Analysis

Models such as GPT-3 (Generative Pre-trained Transformer) have gained prominence in the field of Natural Language Processing (NLP) due to their remarkable ability to engage in human-like interactions and provide accurate answers to

questions [11]. Large Language Models (LLMs) have emerged as potent tools in sentiment analysis, capitalizing on their capacity to grasp context semantics and generate human-like text. Notably, OpenAI’s ChatGPT, an LLM introduced on November 30, 2022, has demonstrated exceptional performance across various NLP tasks, including question answering, translation, essay writing, computer programming, and sentiment analysis [18].

By harnessing extensive pre-training data, LLMs capture subtle sentiment nuances and contextual cues, enhancing sentiment classification accuracy, especially on government social media data. ChatGPT, developed by OpenAI and released in November 2022, signifies a significant advancement in AI capabilities, particularly in natural language understanding and reasoning [18]. This technology’s accessibility offers comprehensive insights on diverse subjects and enables complex conversations. It generates accurate and coherent responses to challenging queries requiring comprehensive information synthesis, analysis, and application. Leaning on OpenAI’s GPT-3 large language model, ChatGPT employs deep learning to comprehend and generate natural language content [18]. These models have acquired an understanding of human language intricacies and nuances through extensive training on vast textual datasets. ChatGPT is designed to produce text responses closely resembling human writing, capable of multilingual interactions, and finds applications in customer service, content creation, healthcare, and language translation [29].

The work of [29] compared the effectiveness of the BERT model with ChatGPT in sentiment analysis, analyzing 5,643 abstracts related to Lyme Disease from scientific journals. The results suggested that ChatGPT’s interpretations of text subjectivity and sentiment scores were closely aligned with the BERT model’s. Another preliminary study posited that ChatGPT could act as a "universal and well-behaved sentiment analyzer" across various domains [36]. These technological advancements highlight the potential of leveraging LLMs in public sentiment analysis. However, further research is needed to address concerns regarding the reliability of these outcomes in critical areas such as public sentiment analysis to avoid the 'challenge of the technical black box' [21].

3 METHOD

In this paper, we propose to utilize a lexicon-based approach to elucidate the sentiment analysis predictions produced by ChatGPT, inspired by the work of [16]. Our objective is to investigate the efficacy of employing an explainable, albeit potentially less accurate, lexicon-based method as a means to approximate the "black box" nature of language models in sentiment analysis tasks.

3.1 Data Collection

To answer the research questions, we chose to focus on the *Centers for Disease Control and Prevention (CDC) Facebook page*¹. According to the group’s description, the page aimed to fulfill CDC’s goal of “*dedicated to protecting health & promoting quality of life through prevention and control of disease, injury, and disability*”. The Facebook page was used to share public health information and a channel through which CDC communicates with the public [15]. We selected the CDC’s Facebook page as our study site because 1) Facebook is one of the primary social media platforms that the CDC uses to perform public health communication. Millions of active users would read and interact with the posts from the CDC during the pandemic. 2) Based on the CDC’s guideline of *comment policy* [12], the comments from the viewer are encouraged in a respectful manner in the Facebook page. It provides a unique lens to observe the public perception and engagement of the public health authority agency’s response to the COVID-19 pandemic.

¹<https://www.facebook.com/CDC>

Table 1. Descriptive statistics of the collected dataset

	All (N=553)			Pre-Pandemic (N=203)			Post-Pandemic (N=350)		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
Number of Likes	5,299.3	1050	19,713.5	478.5	311	502.9	8095.3	1900	24354.9
Number of Shares	3,328.8	626	20,301.3	515	154	946.7	4960.9	1200	25378.6
Number of Comments (Shown)	885.4	421	1,763.4	318.6	122	430.8	1215.0	629	2125.9
Number of Comments (Archived)	519.3	236	977.1	241.7	94	336.1	608.8	349	1172.6
Post Characters	188.9	211	104.9	176.5	209	103.5	196.1	215.5	105.2
Comment Characters	31.1	30.42	14.8	31.7	29.38	21.6	30.8	30.8	8.8

Based on the timeline of COVID-19 development in 2020, the first local case was identified in late January [7]. The US declared COVID-19 as a national emergency on March 13, 2020, following the World Health Organization (WHO) declared COVID-19 as a global pandemic on March 11, 2020 [1]. We used a PHP script on June 26, 2020, to parse and collect threads from the CDC Facebook page through Facebook’s application programming interface (API). Our preliminary dataset included 662 posts created by CDC associates and all the user-generated comments from November 26, 2019, to June 26, 2020. This time period covers the initial pre-to-post-outbreak transition in the United States. After the data collection, we first analyzed all posts to gain an initial understanding of the preliminary dataset. We noticed the first time the CDC Facebook page first mentioned the keyword “COVID-19” on Feb 12, 2020. On Feb 25, 2020, the CDC announced more “*aggressive measures*” against COVID-19. Based on the COVID-19 development timeline and the preliminary dataset, we then defined March 1st, 2020, as the cut point between the pre and post-pandemic periods. To compare the disparity in public engagement before and after the pandemic, we selected a span of three months before and three months following the designated cut-off date. This encompassed a total duration of six months. The final dataset included 554 posts in six months, which combined 203 pre-pandemic posts from December 2019 to February 2020 and 350 post-pandemic posts from March to May 2020.

Table 1 provides an overview of the collected dataset’s descriptive statistics². Upon initial examination, a noteworthy surge in post-pandemic engagement (likes, comments, etc.) was observed relative to the pre-pandemic period. This observation is congruent with findings from a similar study conducted on the same Facebook page [30]. To ensure a comprehensive range of public sentiment dynamics in our analysis, we opted for the top eight posts with the highest comment counts during the post-pandemic period. Furthermore, comments containing fewer than 50 characters were excluded from consideration. This filtration process was necessary due to the prevalence of short phrases like *Thank you* or solitary emojis, which offer minimal substance for sentiment identification. After the filtration, the final dataset encompassed 13,547 comments, each possessing the desired character count (mean = 111.86, SD = 82.08) for sentiment analysis. Detailed statistics for this refined dataset are presented in Table 2.

3.2 Sentiment Analysis

To obtain the sentiment score of each comment, we utilized manual prompts to request ChatGPT sentiment analysis scores for individual comments. The data analysis was performed using ChatGPT model 3.5³, and all requests were

²Please note that although all CDC posts data were archived, certain user-generated data were inaccessible due to privacy settings or data removal. The number of displayed and downloaded comments is outlined in Table 1.

³<https://platform.openai.com/docs/models/gpt-3-5>

Table 2. Descriptive statistics of the final dataset

Post ID	Date	Number of		Number of Comments			Subject
		Likes	Shares	Shown	Archived	Filtered	
1	Apr 3, 2020	192,000	460,000	24,000	9,003	2,830	Stay at home guidance.
2	May 1, 2020	50,000	22,000	10,000	6,625	1,262	Hospitalization rates.
3	May 5, 2020	83,000	30,000	11,000	7,098	1,701	Social distancing.
4	May 7, 2020	104,000	51,000	9,200	5,104	1,191	Virus spreading.
5	May 12, 2020	152,000	50,000	9,200	5,116	1,217	Protective measures.
6	May 19, 2020	12,000	22,000	9,700	8,872	1,801	Virus spreading.
7	May 22, 2020	45,000	17,000	10,000	5,486	1,980	Protective measures.
8	May 29, 2020	58,000	16,000	9,500	5,575	1,565	Social distancing.

made and completed in March 2023. The prompt used during the study was as follows: **Categorize each comment into negative, neutral, or positive. Show only the ID and your rating on a 1-9 scale (negative to positive).** We appended this prompt to the corresponding archived comment data (ID + text). We sent approximately 10-50 comments in each request to accommodate the token limitation. Below is a sample request that we used in this experiment.

Prompt: *Categorize each comment into negative, neutral, or positive. Show only the ID along with your rating on a 1-9 scale (from negative to positive)*

73829 Thank you for the valuable information on staying safe during flu season!

50214 I'm curious about the latest COVID-19 variants. Any updates on their impact?

92675 Your vaccination efforts are truly commendable. Let's beat this pandemic together!

30587 Could you share some tips on mental health during these challenging times?

64890 How effective are masks in preventing the spread of respiratory illnesses?

81123 Kudos to your team for the insightful travel guidelines. Essential information for travelers!

47956 As a teacher, I'm concerned about classroom safety. Any advice for schools reopening?

20784 Your weekly updates have become my go-to source for reliable health info. Keep up the great work!

Here is the example reply from the ChatGPT API (Application Programming Interface). It's important to mention that the API might provide sentiment scores in various formats. We manually standardized the format and extracted the sentiment score to prepare the dataset for this experiment.

Response: Based on the provided comments and IDs:

73829 - "Thank you for the valuable information on staying safe during flu season!" Positive (Rating: 8/9)

50214 - "I'm curious about the latest COVID-19 variants. Any updates on their impact?" Neutral (Rating: 5/9)

92675 - "Your vaccination efforts are truly commendable. Let's beat this pandemic together!" Positive (Rating: 9/9)

30587 - "Could you share some tips on mental health during these challenging times?" Neutral (Rating: 5/9)

64890 - "How effective are masks in preventing the spread of respiratory illnesses?" Neutral (Rating: 5/9)

81123 - "Kudos to your team for the insightful travel guidelines. Essential information for travelers!" Positive (Rating: 8/9)

47956 - "As a teacher, I'm concerned about classroom safety. Any advice for schools reopening?" Neutral

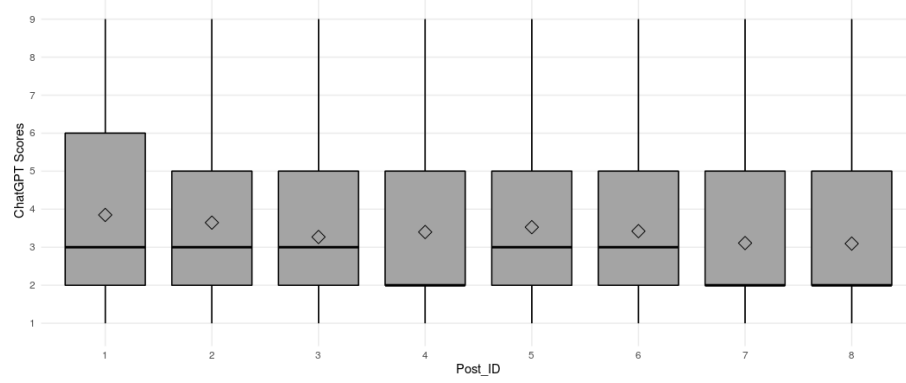


Fig. 1. The sentiment analysis scores acquired from each post using ChatGPT.

(Rating: 5/9)

20784 - “Your weekly updates have become my go-to source for reliable health info. Keep up the great work!” Positive (Rating: 9/9)

The surrogate model utilized in this study employed the functions of the *Syuzhet* package in R [17]. We selected this model for its easily comprehensible features; all words relied on a human-annotated dictionary, enabling straightforward sentiment analysis at the word level, specifically in unigrams. The package encompasses four sentiment dictionaries and offers access to a robust sentiment extraction tool. We utilized the *get_nrc_sentiment* function, which is built upon Saif Mohammad’s NRC Emotion lexicon. This function generates a list of words and associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). To render the sentiment scores comparable to the ChatGPT scores, we merged the two sentiment scores by deducting the negative score from the positive score. For instance, if a comment includes five positive and two negative words, the lexicon-based score is calculated as 5 minus 2, resulting in a score of 3.

To explore the surrogate model’s effectiveness in approximating the predictions from ChatGPT scores, we conducted a correlation analysis to examine the relationship between the scores obtained from the surrogate model and those generated by ChatGPT. This correlation provided a mean to approximate the sentiment score using an interpretable surrogate model [8]. By interpreting the surrogate model, we can gain insights into the black box model’s behavior, making it useful as a global surrogate model. Additionally, we conducted a qualitative content analysis [42] to explore local explanations of specific instances, providing insights at a more detailed and instance-specific level. We categorized the data based on the predicted sentiment score and reported the summary analysis as a case study.

4 FINDINGS

4.1 Public Sentiment During COVID-19: a Case Study

Based on the analysis, our findings indicate that sentiment scores generated by ChatGPT reflect predominantly negative public sentiment ($M=3.43$, $SD=2.02$). The distribution and skewness of numerical data are presented in Figure 1, illustrating quartile groups for each post. Although Post 1 appears to have garnered a somewhat neutral public sentiment, all other posts consistently exhibit a negative perception pattern. These sentiment trends align with other

studies on social media during the pandemic [2]. To delve deeper, we selected Post 1 for a case study to explore how ChatGPT can aid in comprehending public sentiment toward public policy announcements.

On April 3rd, Post 1 was published on the CDC’s official Facebook page, conveying the message: “*While the White House 30 Day Stay At Home guidance remains our most powerful weapon against COVID-19, CDC’s recommendation to wear a cloth face covering may help protect the most vulnerable. Watch U.S. Surgeon General Jerome Adams make a face covering in a few easy steps.*” This post garnered significant public engagement, amassing over 192,000 likes and 460,000 re-posts. Among the 24,000 comments received, we archived 9,003, with a final analysis including 2,830 comments. We randomly sampled ten positive comments (score > 7), ten negative comments (score < 3), and ten neutral comments (score = 5). Comments containing usernames were removed from the analysis to safeguard user privacy. All selected comments have been provided in Appendix 1.

Positive Sentiment: The public sentiment towards the certain policy is varied, encompassing understanding of scientists’ ongoing learning about the virus, appreciation for proactive measures such as the creation of emergency masks in India, and support for easy DIY solutions like the one presented by Surgeon General Adams. Gratitude is expressed for showing effective protective measures and invoking a sense of divine protection, while confusion arises from skepticism about community protection and concerns about people dismissing the effectiveness of such measures. Some comments reflect faith in overcoming the pandemic, and personal stories highlight the positive impact of shared efforts within communities. Overall, the sentiment varies between positive support, frustration towards skeptics, and appeals for collective responsibility in preventing the virus’s spread.

Negative Sentiment: The public sentiment towards the policy of recommending cloth masks for protection against the virus appears mixed and skeptical. Some express disbelief that such basic materials like rubber bands and cloth can effectively safeguard against the virus, while others criticize the changing guidelines of the CDC, questioning the credibility of the measures and lamenting the perceived misinformation. There’s frustration about the evolving stance on masks, with some feeling that the situation is not clearly understood, and calls to prioritize healthcare workers. Concerns are raised about the limited efficacy of cloth masks, their inability to protect against viruses at a microscopic level, and doubts about their practicality. Critics highlight cloth masks’ potential flaws and limitations as protective measures, advocating for more advanced solutions like N95 masks, comprehensive coverage, longer quarantine periods, and increased testing. Amidst these comments, there’s a sense of disappointment and disillusionment with the perceived decline of authoritative institutions like the CDC.

Neutral Sentiment: The public sentiment regarding the mentioned policy of using cloth masks for protection against the virus is diverse. While some acknowledge the value of any protection but suggest enhancements like including eye protection and tightly woven fabrics, others propose alternative materials such as cloth napkins. Concerns about supply shortages, doubts about the efficacy of cloth masks, and practical limitations are voiced. People share personal experiences, including challenges with mask fit and alternative material uses. Criticisms of authority figures’ contradictory statements and calls to acknowledge earlier contributors are present. Amidst discussions, humor emerges comparing the Surgeon General to a young Mike Tyson, turning his focus from fighting individuals to viruses.

4.2 Global Surrogate Model

While ChatGPT generated remarkable results by assigning sentiment scores to each comment and facilitating a pandemic-related case study on public sentiment toward guidance, the crucial question remains: How do we ensure the reliability and trustworthiness of these outcomes? To address this, we employed a lexicon-based model as a surrogate, approximating sentiment predictions from ChatGPT. In our sentiment analysis. The ChatGPT model ($M=3.43$, $SD=2.02$)

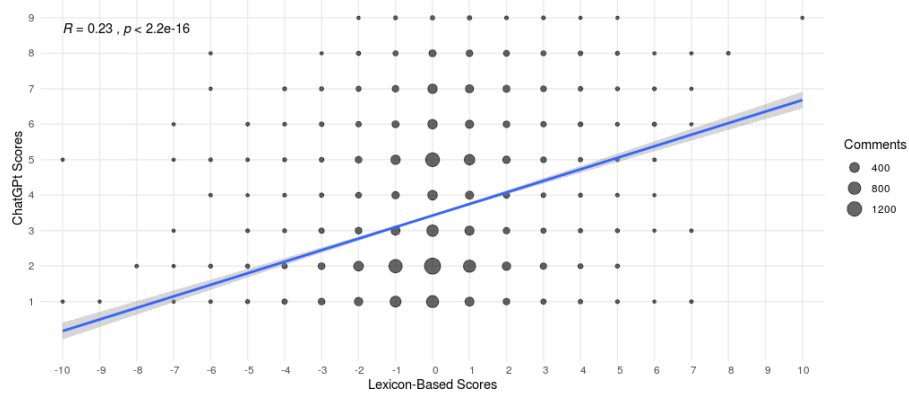


Fig. 2. Spearman's rank correlation Analysis between ChatGPT and the surrogate models.

yielded more negative sentiment scores than the Lexicon method ($M=0.03$, $SD=1.42$). These scores suggest a right-skewed distribution for ChatGPT, indicating a stronger leaning toward negative sentiment. Conversely, the lexicon method produces scores closer to a normal distribution, implying neutrality or a slight positive. Notably, the lexicon method frequently yields a zero (neutral sentiment) score, potentially due to a lack of identified positive or negative words in the dictionary or an equal balance between them. It's essential to recognize that the two scores stem from distinct methodologies, preventing direct comparisons.

To better understand the relationship between the two scores, we utilized Spearman's rank correlation to assess the connection between ChatGPT's sentiment scores and those derived from the lexicon-based approach. The analysis revealed a positive correlation between these variables, yielding a correlation coefficient of $r(13,545) = 0.23$ and a statistically significant p-value of $< 2.2 \times 10^{-16}$. While the correlation coefficient ($r=0.23$) indicates a weak positive link, the statistical significance ($p < 2.2 \times 10^{-16}$) verifies the existence of a relationship between the scores. For a visual representation, please consult Figure 2. Our analysis suggests that the lexicon-based method can function as an approximate measure of sentiment, albeit with limited explanatory capacity. Given ChatGPT's intricate nature [37], achieved through extensive text data training for heightened accuracy, it's unsurprising that the explanatory scope of the lexicon-based method remains confined. Despite providing only a modest overall correlation for the ChatGPT model, the lexicon-based method maintains its value as a transparent model, aiding users in assessing prediction reliability. Furthermore, this result also implies that LLMs consider multiple factors beyond singular unigrams, underscoring the need for awareness when employing such models for public sentiment analysis.

4.3 Local Surrogate Models

We extended our exploration of explanation methods by employing a surrogate model to approximate local predictions, focusing specifically on explanations for individual instances. To illustrate how a lexicon-based model could emulate ChatGPT predictions, we handpicked a few comments as examples. The inherent understanding of the lexicon model allowed us to surface the words used in the comments and subsequently compare the outcomes with ChatGPT predictions. Here's an example comment from Post 8: while the Lexicon method assigns a neutral sentiment (Score=0; 2 positive and two negative words), ChatGPT predicts a notably negative score (Score=1). In the quote, we've differentiated emotional words using distinct colors: positive words are highlighted in blue, and negative words are in red.

wonderful government wants us sick! They are telling us to do things that will make us sick! Please know this that they make a lot of money off sick people :-(don't comply with their BS please

From this comment, we observe that there are two positive words and two negative words mentioned, leading the lexicon model to consider the sentiment as neutral. However, upon reading the comment, it becomes evident that the underlying emotion is quite negative, and the ChatGPT model accurately captures this sentiment. Here's an additional illustration demonstrating differing ratings. The comment originates from Post 1 and is analyzed using the Lexicon method, which assigns a neutral sentiment (Score=0; Positive 0 and negative 0). However, ChatGPT predicts a significantly positive score (score=9).

finallyyyyy....CDC recommended about this! Last time WHO warned us about to do not to use the face mask lol

The lexicon model failed to detect any emotional words in this comment. Nevertheless, it's apparent that the poster expressed gratitude towards the new policy adopted by the CDC, which offers official guidance to stay at home and avoid social contact. The lexicon model could also over-estimate the emotion due to the word picking. Below is an example comment from Post 8: The lexicon method assigns a very negative sentiment (Score=-5; Positive 0 and negative 5), whereas ChatGPT predicts the score as neutral (score=5).

We need to open the economy though. Keeping people at home forever is not the answer. You lower their resistance and it becomes worse. If you want to stay home knock yourself out. 87,000 died from the swine flu. No lockdown!

In this comment, the poster used several words to describe their opinion on the CDC policy, with many of them being classified as negative emotions in the lexicon model. However, upon careful examination of this comment, it becomes evident that the poster intended to express a neutral opinion and emotion on this policy. The opposite example could also happen, just like the comment of Post 8 that the Lexicon method assigns a very positive sentiment (Score=5; Positive five and negative 0), whereas ChatGPT predicts the score as neutral (score=5).

I would love to see the science behind this. Every study I have read states it is not beneficial. Can you provide the science in which you are basing this suggestion to the public??

This is another example where the poster conveyed their opinion with a neutral emotion, even though they used numerous words that could be interpreted as positive. The lexicon method assigns a very positive sentiment (Score=7; Positive 11 and negative 4), whereas ChatGPT predicts the score as very negative (score=1). The ChatGPT model may lead to overestimation based on the method used, which could be misleading as it does not consider the context.

I believe this is the most helpless solution. The United States seems to be very poor in this COVID-19 response. No wonder the president will transfer the contradiction to other countries like a clown, to blame others, not to try to make good precautions! Very disappointed! May all medical staff and people be healthy! God bless freedom and democracy!

Although the author utilized numerous positive word choices in this comment, the underlying sentiment seems to lean towards more negative emotions. A similar example can be observed in the comment for Post 4, where the poster employed several negative words in their sentence but managed to convey a more positive emotion. For instance, the lexicon method assigns a very negative sentiment (Score=-6; Positive 2 and negative 8), whereas ChatGPT predicts the score as very positive (score=8).

*Having just completed six months of down and **dirty** chemotherapy for **leukemia** and approaching 60, I have an extremely suppressed immune system and am at high **risk** of **infection** and complications (mainly **Death**) should I come in **contact** with this COVID 19, or even the common **influenza virus**. My only **hope** is **isolation**, **PPE**, and meticulous hygiene practices. I limit outings to once weekly for fresh produce, dairy and meat.*

As evident from the text, despite the poster’s use of numerous negative emotion words to describe their situation and opinion on the policy, we can still sense a positive emotion in the sentence reflecting their agreement with the CDC’s measurement policy for COVID-19.

5 DISCUSSION

This paper explores strategies to ensure transparency in analyzing public sentiment using ChatGPT. By examining a Facebook CDC dataset, we demonstrate ChatGPT’s effectiveness in accurately reflecting public opinions and sentiments [36]. We qualitatively analyzed specific comments and their correlations with both methodologies. We employed a lexicon-based surrogate model designed to approximate and clarify ChatGPT’s outputs, aiming for a more transparent and explainable approach to interpreting the predictions. Our findings reveal a moderate correlation between the outcomes generated by the lexicon-based approach and those derived from ChatGPT, supporting the effectiveness of surrogate models in improving transparency in public sentiment analysis with ChatGPT. While our results show that the lexicon-based method does not surpass ChatGPT in the performance of sentiment analysis, it does offer a higher level of transparency. Our efforts shed light on integrating interpretability from the lexicon-based method with the superiority of using ChatGPT for public sentiment analysis.

Based on our investigations, we affirmed the efficacy of ChatGPT within the realm of public sentiment analysis. It provides an accessible interface or API that allows us to submit public opinions conventionally and receive associated sentiment labels. Previous research has already examined the model’s accuracy compared to diverse baseline models [29, 36], and our aim is not to replicate those cross-validation experiments. Instead, we conducted a case study to showcase how ChatGPT could be utilized to analyze real-world public sentiment. The primary focus of this paper lies in employing a surrogate model to assess its potential to explain the model’s effectiveness and implications for decision-makers in the public sector. Our experimental outcomes support the notion of employing a transparent model to approximate the opaque predictions of AI, aiding decision-makers in trusting and comprehending predictions from ChatGPT or other large language models with greater confidence. Nevertheless, our findings also indicate that while transparent and explainable models hold value, they may not completely convey the nuances of AI models, underscoring the need for continued exploration in the quest for more potent explanatory models.

Our approach can potentially enhance public government transparency when utilizing LLMs for public sentiment analysis [21]. The public sector can leverage generative AI techniques to analyze and understand public sentiment, thereby improving public policy-making and enhancing public communication and engagement [33]. Public sentiment analysis can empower the government to make data-driven decisions, thus helping to proactively manage the crises. Our proposed approach aims to maintain practical transparency while harnessing the powerful analytical features of these methods, a crucial factor when communicating with the public. This increased transparency can build greater public trust in government decision-making [28]. Our examination of local prediction cases reveals instances of disagreement between the original and surrogate models, offering insights into the mechanisms that require attention to align predictions with the surrogate model approximations. Our proposed work provides implications for establishing such mechanisms when the public sector employs generative AI in high-stakes use cases.

We acknowledge certain limitations in our research. Firstly, the performance of the surrogate model is not on par with generative AI or machine learning models. Secondly, our analysis and case study are based solely on one social media dataset, limiting the generalizability of our findings to other contexts of public sentiment analysis. Thirdly, we did not possess ground truth data to demonstrate the performance of the ChatGPT model. Further quantitative research would be required to assess the model's performance more comprehensively.

6 CONCLUSION

Public sentiment analysis is pivotal in understanding prevalent emotions, attitudes, and viewpoints, with far-reaching implications spanning policy formulation and beyond. ChatGPT, an innovative large language model, facilitates versatile human-AI interactions and holds promise in assessing public sentiment. Nevertheless, persistent concerns exist regarding its transparency and interpretability. This study presents our investigation into the application of ChatGPT to analyze public sentiment within the context of a global pandemic. We introduce a lexicon-based surrogate model to ensure transparency and interpretability. Employing a mixed-method approach, we validate the surrogate model's effectiveness, shedding light on the utility of ChatGPT's sentiment analysis in public domains. This research advances our understanding of the use of large language models in public sentiment analysis, providing insights into the application of generative AI techniques in the public sector and governmental functions.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2153509.

REFERENCES

- [1] 2021. A Timeline of COVID-19 Developments in 2020. <https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020>
- [2] Folashade B Augusto, Eric Numfor, Karthik Srinivasan, Enahoro A Iboi, Alexander Fulk, Jarron M Saint Onge, and A Townsend Peterson. 2023. Impact of public sentiments on the transmission of COVID-19 across a geographical gradient. *PeerJ* 11 (2023), e14736.
- [3] Abdullah Hussein Alamoodi, Bilal Bahaa Zaidan, Aws Alaa Zaidan, Osamah Shihab Albahri, Khalid Ibrahim Mohammed, Rami Qays Malik, Esam Motashar Almahdi, Mohammed A Chyad, Ziadoon Tareq, Ahmed Shihab Albahri, et al. 2021. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert systems with applications* 167 (2021), 114155.
- [4] Kashif Ali, Hai Dong, Athman Bouguettaya, Abdelkarim Erradi, and Rachid Hadjidj. 2017. Sentiment analysis as a service: a social media based sentiment analysis framework. In *2017 IEEE international conference on web services (ICWS)*. IEEE, 660–667.
- [5] Omar Alqaryouti, Nur Siyam, Azza Abdel Monem, and Khaled Shaalan. 2024. Aspect-based sentiment analysis using smart government review data. *Applied Computing and Informatics* 20, 1/2 (2024), 142–161.
- [6] Gil Appel, Lauren Grewal, Rhonda Hadi, and Andrew T Stephen. 2020. The future of social media in marketing. *Journal of the Academy of Marketing science* 48, 1 (2020), 79–95.
- [7] Allison Aubrey. 2020. Trump Declares Coronavirus A Public Health Emergency And Restricts Travel From China. <https://www.npr.org/sections/health-shots/2020/01/31/801686524/trump-declares-coronavirus-a-public-health-emergency-and-restricts-travel-from-c>
- [8] Jorge A Chan-Lau, Ruofei Hu, Maksym Ivanyna, Ritong Qu, and Cheng Zhong. 2023. Surrogate Data Models: Interpreting Large-scale Machine Learning Crisis Prediction Models. (2023).
- [9] Hsuanwei Michelle Chen and Patricia C Franks. 2016. Exploring Government Uses of Social Media through Twitter Sentiment Analysis. *Journal of Digital Information Management* 14, 5 (2016).
- [10] Olfa Belkahla Driss, Sehl Mellouli, and Zeineb Trabelsi. 2019. From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly* 36, 3 (2019), 560–570.
- [11] Luciano Floridi and Massimo Chiriatti. 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines* 30 (2020), 681–694.
- [12] Centers for Disease Control and Prevention. 2021. Facebook Guidelines and Best Practices. <https://www.cdc.gov/socialmedia/tools/guidelines/facebook-guidelines.html>
- [13] Missy Graham and Elizabeth Avery. 2013. Government public relations and social media: An analysis of the perceptions and trends of social media use at the local government level. *Public relations journal* 7, 4 (2013), 1–21.
- [14] Ali Hasan, Sana Moin, Ahmad Karim, and Shahaboddin Shamshirband. 2018. Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Mathematical and Computational Applications* 23, 1 (2018). <https://doi.org/10.3390/mca23010011>

- [15] Amy Burnett Heldman, Jessica Schindelar, and James B Weaver. 2013. Social media engagement and public health communication: implications for public health organizations being truly “social”. *Public health reviews* 35, 1 (2013), 13.
- [16] Andreas Holzinger, Anna Saranti, Christoph Molnar, Przemyslaw Biecek, and Wojciech Samek. 2022. Explainable AI methods-a brief overview. In *International workshop on extending explainable AI beyond deep models and classifiers*. Springer, 13–38.
- [17] Matthew Jockers. 2017. Package ‘syuzhet’. URL: <https://cran.r-project.org/web/packages/syuzhet> (2017).
- [18] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023), 102274.
- [19] Manpreet Kaur, Rajesh Verma, and Frank Nana Kweku Otoo. 2021. Emotions in leader’s crisis communication: Twitter sentiment analysis during COVID-19 outbreak. *Journal of Human Behavior in the Social Environment* 31, 1-4 (2021), 362–372.
- [20] Slawomir Koziel, David Echeverría Ciaurri, and Leifur Leifsson. 2011. Surrogate-based methods. *Computational optimization, methods and algorithms* (2011), 33–59.
- [21] Han-Wei Liu, Ching-Fu Lin, and Yu-Jie Chen. 2019. Beyond State v Loomis: artificial intelligence, government algorithmization and accountability. *International journal of law and information technology* 27, 2 (2019), 122–141.
- [22] Wala Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal* 5, 4 (2014), 1093–1113.
- [23] Pooja Mehta and Sharnil Pandya. 2020. A review on sentiment analysis methodologies, practices and applications. *International Journal of Scientific and Technology Research* 9, 2 (2020), 601–609.
- [24] Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *National Research Council, Canada* 2 (2013), 234.
- [25] Taewoo Nam. 2012. Citizens’ attitudes toward open government and government 2.0. *International review of administrative sciences* 78, 2 (2012), 346–368.
- [26] Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. 2016. Sentiment analysis of Twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPEs)*. IEEE, 1345–1350.
- [27] Azeema Sadia, Fariha Khan, and Fatima Bashir. 2018. An overview of lexicon-based approach for sentiment analysis. In *2018 3rd International Electrical Engineering Conference (IEEC 2018)*. 1–6.
- [28] Changsoo Song and Joocho Lee. 2016. Citizens’ use of social media in government, perceived transparency, and trust in government. *Public Performance & Management Review* 39, 2 (2016), 430–453.
- [29] Teo Susnjak. 2023. Applying bert and chatgpt for sentiment analysis of lyme disease in scientific literature. *arXiv preprint arXiv:2302.06474* (2023).
- [30] Sue Yeon Syn. 2021. Health information communication during a pandemic crisis: Analysis of CDC Facebook Page during COVID-19. *Online Information Review* 45, 4 (2021), 672–686.
- [31] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen, and Xiaofei He. 2013. Interpreting the public sentiment variations on twitter. *IEEE transactions on knowledge and data engineering* 26, 5 (2013), 1158–1170.
- [32] David Valle-Cruz, Vanessa Fernandez-Cortez, Asdrúbal López-Chau, and Rodrigo Sandoval-Almazán. 2022. Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. *Cognitive computation* (2022), 1–16.
- [33] Sanjeev Verma. 2022. Sentiment analysis of public services for smart society: Literature review and future research directions. *Government Information Quarterly* 39, 3 (2022), 101708.
- [34] Bin Wang, Pengfei Guo, Xing Wang, Yongzhong He, and Wei Wang. 2021. Transparent Aspect-Level Sentiment Analysis Based on Dependency Syntax Analysis and Its Application on COVID-19. *ACM Journal of Data and Information Quality (JDIQ)* 14, 2 (2021), 1–24.
- [35] Hao Wang, Doğan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 system demonstrations*. 115–120.
- [36] Zengzhi Wang, Qiming Xie, Zixiang Ding, Yi Feng, and Rui Xia. 2023. Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv preprint arXiv:2304.04339* (2023).
- [37] Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang. 2023. A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica* 10, 5 (2023), 1122–1136.
- [38] Kailai Yang, Shaoxiong Ji, Tianlin Zhang, Qianqian Xie, and Sophia Ananiadou. 2023. On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis. *arXiv preprint arXiv:2304.03347* (2023).
- [39] Yi Yang, Wen Deng, Yi Zhang, and Zijun Mao. 2021. Promoting public engagement during the COVID-19 crisis: how effective is the Wuhan local Government’s information release? *International journal of environmental research and public health* 18, 1 (2021), 118.
- [40] Lin Yue, Weitong Chen, Xue Li, Wanli Zuo, and Minghao Yin. 2019. A survey of sentiment analysis in social media. *Knowledge and Information Systems* 60 (2019), 617–663.
- [41] Staci M Zavattaro, P Edward French, and Somya D Mohanty. 2015. A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. *Government information quarterly* 32, 3 (2015), 333–341.
- [42] Yan Zhang and Barbara M Wildemuth. 2009. Qualitative analysis of content. *Applications of social research methods to questions in information and library science* 308, 319 (2009), 1–12.

A SAMPLE COMMENTS IN POST 1

Positive Comments: 1) Everyone needs to understand that the scientists and Doctors are learning things about this virus everyday. Just because they are suggesting it now does not mean they kept information from the public, they are learning about it too. 2) In India video creating this type of emergency masks are circulated 2 weeks back... God bless all :) 3) I love our Surgeon General Adams showing a quick and easy solution for everyone. Home craft project that can change daily at little to no expense. Plus, hair ties can be used instead of rubber bands! We're all in this together! 4) Thank you for showing us how easily this is done!!! 5) Thank you, creator, Lord, for this creative way to protect us and protect all human beings from this epidemic 6) It baffles me that so many people dont even consider this could help protect people in the community. Instead Im reading people write over and over that this wont help protect me. 7) Fantastic one. God will stands with us to wipe out corona virus in the country, in Jesus Christ name Amen 8) The great woman at our church made them for us. I where wear mine when I go outside. God bless them. Love all of you. 9) Great job! Its easy and everyone can make it. 10) I guess things are not bad enough for some of you. So you decide to carry on with your hatefulness. The rest of as usual will follow this and help not to spread the virus. Stop blaming those who are trying to stop the spread and do something for yourself for a change. If not you are to blame for causing another persons death. Time to pull up your socks and do the right thing.

Negative Comments: 1) So weve come to this? Rubber bands and a piece of cloth as protective gear? 2) with how many times the CDC has changed the precautions for this virus, it's hard to believe anything anymore. A piece of cloth isn't going to help.... But if wearing it makes you feel better have at it. I just think it's sad to spread this kind of misinformation. 3) This is so sad. First no masks...now masks. Ya'll don't even know what's going on except you're letting people die. Unass the supplies and give them to our healthcare workers!!!! 4) This is so sad. First no masks...now masks. Ya'll don't even know what's going on except you're letting people die. Unass the supplies and give them to our healthcare workers!!!! 5) But regular cloth does not protect against a virus....only actual droplets which means you can see it.....not microscopic. 6) But regular cloth does not protect against a virus....only actual droplets which means you can see it.....not microscopic. 7) And just like that, America ran out of rubber bands. 8) Whats the point it allows 95% percent of the bacteria and virus through, plus its the perfect Petri dish for other bacteria and viruses to live in. Just another gimmick. 9) A cloth mask is barely help. You need an N95. You beed to cover eyes and ears as well. Two week quarantine is not conservative enough when the majority of people are contagious up to 42 days. 6 weeks. Test test test. 10) I always thought CDC is the leader in rhe world and best in this field! But now under Trump CDC Ilis a joke, just like NASA, which can be easily taken over by Space X....

Neutral Comments: 1) I agree that something is better than nothing but they need to mention that we need eye protection too and tightly woven fabric works better. Masks made in England for example are made of material similar to swimsuits. Stay home if you possibly can... 2) Seems like cloth napkins might be perfect for this type of mask . 3) Now rubber bands will be off the shelves! 4) Does tha really work? Its cloth material not a sponge ... 5) but i use my hand towel to whipe myself after sex and i dont have scarf and m shirts cost too much to be ripping... 6) Fit to long or big ears; I tested; I have small ears. 7) Please show us how to make toilet paper. Since all these MFs want to hoard it 8) I knew we needed to wear masks when we were out of our home running essential errands. I used my common sense ignored the IDIOT surgeon general who has endangered many lives by not encouraging ppl to wear masks. He said not to wear masks.The main thing is hand washing repeatedly TRY to avoid touching your face. 9) They should at least acknowledging the nurse that posted this 2 weeks ago. Side note! I see a lot of people complaining. Before you

complain learn to use spell check!!! 10) The Surgeon General is a young Mike Tyson who instead of fighting people decided to fight viruses.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009