



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Research Article

# Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination

Hanjia Lyu<sup>1</sup>, Junda Wang<sup>2</sup>, Wei Wu<sup>1</sup>, Viet Duong<sup>2</sup>, Xiyang Zhang<sup>3</sup>, Timothy D. Dye<sup>4</sup>, Jiebo Luo<sup>2,\*</sup>

<sup>1</sup> Goergen Institute for Data Science, University of Rochester, Rochester, New York 14627, United States

<sup>2</sup> Department of Computer Science, University of Rochester, Rochester, New York 14627, United States

<sup>3</sup> Department of Psychology, University of Akron, Akron, Ohio 44325, United States

<sup>4</sup> Department of Obstetrics and Gynecology, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, United States



## ARTICLE INFO

## Keywords:

Vaccine hesitancy  
 COVID-19  
 Public opinion  
 Social Media  
 Twitter

## ABSTRACT

**Background** The current development of vaccines for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is unprecedented. Little is known, however, about the nuanced public opinions on the vaccines on social media.

**Methods** We adopted a human-guided machine learning framework using more than six million tweets from almost two million unique Twitter users to capture public opinions on the vaccines for SARS-CoV-2, classifying them into three groups: pro-vaccine, vaccine-hesitant, and anti-vaccine. After feature inference and opinion mining, 10,945 unique Twitter users were included in the study population. Multinomial logistic regression and counterfactual analysis were conducted.

**Results** Socioeconomically disadvantaged groups were more likely to hold polarized opinions on coronavirus disease 2019 (COVID-19) vaccines, either pro-vaccine ( $B = 0.40$ ,  $SE = 0.08$ ,  $P < 0.001$ ,  $OR = 1.49$ ; 95%  $CI = 1.26$ – $1.75$ ) or anti-vaccine ( $B = 0.52$ ,  $SE = 0.06$ ,  $P < 0.001$ ,  $OR = 1.69$ ; 95%  $CI = 1.49$ – $1.91$ ). People who have the worst personal pandemic experience were more likely to hold the anti-vaccine opinion ( $B = -0.18$ ,  $SE = 0.04$ ,  $P < 0.001$ ,  $OR = 0.84$ ; 95%  $CI = 0.77$ – $0.90$ ). The United States public is most concerned about the safety, effectiveness, and political issues regarding vaccines for COVID-19, and improving personal pandemic experience increases the vaccine acceptance level.

**Conclusion** Opinion on COVID-19 vaccine uptake varies across people of different characteristics.

## 1. Introduction

Researchers suggest that the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) will quickly rebound if interventions (e.g., quarantine and social distancing) are relaxed [1]. Vaccination has greatly reduced the burden of many infectious diseases throughout history, and developing SARS-CoV-2 vaccines that can be used globally is, therefore, a priority for ending the pandemic [2]. Nevertheless, as scientists and medical experts around the world are developing and testing coronavirus disease 2019 (COVID-19) vaccines, the United States public is now divided over whether or not to obtain COVID-19 vaccines. According to a recent Pew Research Center study (<https://www.pewresearch.org/science/2020/09/17/u-s-public-now-divided-over-whether-to-get-covid-19-vaccine/>), in May,

2020, 71% of United States adults indicated that they would definitely or probably obtain a vaccine to prevent COVID-19 if it were available. The percentage dropped sharply, however, to 51% in September, 2020. The survey shows that the United States public is concerned about the safety and effectiveness of possible vaccines, and the rapid pace of the approval process.

Previous studies show that the sharing of public concerns about vaccines might lead to delaying or not getting vaccination [3], which could compromise global COVID-19 vaccine distribution strategies. This phenomenon is termed “vaccine hesitancy” [4] which is a complex issue driven by a variety of context-specific factors [5]. Researchers have investigated public opinions on existing vaccines for vaccine-preventable diseases like MMR [6–7], HPV [8–9] and H1N1 [10]. Hesitancy and opinions can vary, however, according to the vaccine involved [11].

\* Corresponding author: Jiebo Luo, Department of Computer Science, University of Rochester, 3101 Wegmans Hall, Rochester, NY 14627, USA (Email: [jluo@cs.rochester.edu](mailto:jluo@cs.rochester.edu)).

<https://doi.org/10.1016/j.imed.2021.08.001>

Received 8 June 2021; Received in revised form 25 July 2021; Accepted 6 August 2021

2667-1026/© 2021 The Authors. Published by Elsevier B.V. on behalf of Chinese Medical Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Lazarus et al. [12] and Feleszko et al. [13] have investigated the potential acceptance of a COVID-19 vaccine using survey methods, yet little is known about the scope and causes of public opinions on COVID-19 vaccines on social media platforms. Although the survey data of a traditional design can lead to detecting causality, it is labor-intensive and expensive [14], thus, being difficult to deploy surveys at a large scale without introducing social desirability biases [15] and in a timely manner, compared to social media data [14]. In addition, due to the passive nature of collecting social media data, observing social media data can potentially capture a different (and unperturbed) view of human behaviors [16]. To the best of our knowledge, there is no other study that has tracked and understood the public opinion regarding COVID-19 vaccines using social media data.

Meanwhile, the development and testing of COVID-19 vaccines has drawn great attention and response on social media platforms like Twitter and Reddit that allow fast sharing of health information [17–19] and are found to play a major role in disseminating information about vaccinations [20–24]. Public attitudes towards the vaccines, therefore, can be reflected by analyzing comments and posts in social media [25–26].

In the current study, we adopt a human-guided machine learning framework based on state-of-the-art transformer language models to capture individual opinions on COVID-19 vaccines, and categorize these opinions into three groups: pro-vaccine, vaccine-hesitant, anti-vaccine. We use more than 40,000 rigorously selected tweets (out of over six million tweets collected using keywords) posted by over 20,000 distinct Twitter users ranging from September to November of 2020. We aggregate the tweets to reflect the state-level and the national attitudes towards COVID-19 vaccines. To characterize the opinion groups, we extract and infer individual-level features such as demographics, social capital, income, religious status, family status, political affiliations, and geo-locations. Lazarus et al. [12] suggested that personal experience such as COVID-19 sickness in the people and their family, and the external perception such as cases and mortality per million of a nation's population are associated with the vaccine acceptance level. To quantitatively measure and confirm these two effects, we extract the sentiment of personal pandemic experience and non-pandemic experience for each Twitter user. We collect the number of COVID-19 daily confirmed cases from the data repository maintained by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) to measure the county-level pandemic severity perception. In our study, we hypothesize that:

- **Hypothesis 1:** There will be differences in demographics, social capital, income, religious status, family status, political affiliations and geo-locations among opinion groups.
- **Hypothesis 2:** The personal pandemic experience will have an impact on shaping the attitude towards potential COVID-19 vaccines.
- **Hypothesis 3:** The county-level pandemic severity perception will have an impact on shaping the attitude towards potential COVID-19 vaccines.

We conduct multinomial logistic regression and find that there are differences in demographics, social capital, income, religious status, political affiliations and geo-locations among the opinion groups. People who have the worst personal pandemic experience are more likely to hold anti-vaccine opinion. In addition, people who have the worst pandemic severity perception are more likely to be vaccine-hesitant. We further show that the individual-level features can be used to anticipate whether this person is in favor of the potential COVID-19 vaccines - or not - over time. By incorporating the individual-level features and additional factor indicators, and by conducting counterfactual analyses, we find that the United States public is most concerned about the safety, effectiveness, and political issues with regard to potential vaccines for COVID-19 and improving personal pandemic experience increases the vaccine acceptance level.

## 2. Materials and methods

The Methods section is structured as follows. We describe the datasets we use in Methods M1 and how we infer or extract features in Methods M2. We describe our strategy for opinion mining and the standard of labelling in Methods M3. In Methods M4, we discuss the experimental procedures.

### 2.1. M1 DataSets

#### 2.1.1. Twitter

We use the Tweepy API [27] to collect the related tweets which are publicly available. The search keywords and hashtags are COVID-19 vaccine-related or vaccine-related, including “vaccine”, “COVID-19 vaccine”, “COVID vaccine”, “COVID19 vaccine”, “vaccinated”, “immunization”, “covidvaccine”, “#vaccine” and “covid19 vaccine”. It is noteworthy that the capitalization of non-hashtag keywords does not matter in the Tweepy query. Slang and misspellings of the related keywords are also included which are composed of “vacinne”, “vacine”, “anti-vax” and “anti vax”. In the end, 6,314,327 tweets (including retweets) from September 28 to November 4, 2020 posted by 1,874,468 unique Twitter users are collected. To collect as many related text as possible, both COVID-19 vaccine-related and vaccine-related search keywords are used. However, the tweets collected using the vaccine-related search keywords are not necessarily related to COVID-19 vaccines. For example, MMR vaccine-related or HPV vaccine-related tweets might be crawled as well. In addition, the data collection is carried out during the flu shot season, resulting in collecting many influenza shot-related tweets. We apply a keyword-based search in tweets to remove all the tweets containing MMR, autism, HPV, tuberculosis, tetanus, hepatitis B, flu shot or flu vaccine (4.0% removed).

The tweet content and other Twitter profile information are used to extract or predict demographics, user-level features like the number of followers, income, religious status, family status, political affiliations, geo-locations, sentiment about the COVID-19-related experience and non-COVID-related experience. To infer the sentiment, we use Tweepy API to collect the publicly available tweets posted by each user for the last three months. For example, if the tweet containing the search keywords or hashtags was posted on October 1, 2020, then all the publicly available tweets posted by this Twitter user from July 1 to October 1, 2020 are collected as well. It should be noted that only the last 3,200 tweets can be collected per the Tweepy API limitations.

The preprocessing pipeline is shown in Figure 1. First, the features of the Twitter users are inferred or extracted. To better understand the relationships between all characteristics, we choose to only keep the users of which we can infer all the features except for sentiment. Next, we achieve the mining of opinions via a human-guided machine learning framework. 25,407 unique users with all the features except for the sentiment scores are used to study the temporal and spatial patterns of the opinions. 10,945 of them with sentiment scores are further included in the characterization study and counterfactual analyses.

#### 2.1.2. Johns Hopkins University CSSE

We extract the number of COVID-19 daily confirmed cases from the data repository maintained by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [28]. The median relative change of the number of daily confirmed cases of the last three months at the county level is calculated to measure the county-level pandemic severity perception.

### 2.2. M2 feature inference

#### 2.2.1. Demographics

Following the methods of Lyu et al. [29], we use the Face++ API [30] to infer the gender and age information of the users using their profile images. The invalid image urls and images with multiple or zero



Figure 1. The diagram of data preprocessing procedures.

faces are excluded. The gender and age information of the remaining users (i.e., there is only one intelligible face in the profile image) is inferred. Since our study focuses on the opinions of United States adults, the users who are younger than 18 are removed. Face++ can achieve a good accuracy of the gender and age inference of Twitter data [31].

### 2.2.2. User-level features

Seven user-level features are crawled by Tweepy API as well which include the number of followers, friends, listed memberships, favourites, statuses, the number of months since the user account was created, and the verified status. Moreover, we normalize the number of followers, friends, listed memberships, favourites, and statuses by the number of months since the user account was created.

### 2.2.3. Geo-locations

For Twitter, we choose to resolve the geo-locations using users' profiles. Similar to Lyu et al. [29], the locations with noise are excluded, and the rest are classified into urban, suburban, or rural.

### 2.2.4. Income

Following the method of Preoțiuc-Pietro et al. [32], we use a supervised ensemble model to predict the income of Twitter users. The ensemble model includes Gradient Boost Decision Tree (GBDT), Random Forest, Logistic Regression, and XGboost. We use the income datasets of Twitter users [32] to train our model(s). The features include age, days of Twitter history, the number of followers, friends, listed memberships, favourites, and sentiment score calculated by Vader [33]. We categorize income into three classes (low, medium, high) based on the income levels of Kochhar [34] and turn regression problems into classification problems. The accuracy of is 70.02%.

### 2.2.5. Religious status

We assign each user a Boolean value for whether she/he is religious based on the tweets and the description in the profile [35].

### 2.2.6. Family status

By applying regular expression search, we identify users who show evidence that they are either fathers or mothers [35].

### 2.2.7. Political affiliations

The political attribute is labelled based on whether this Twitter user followed the Twitter accounts of the top political leaders. The incumbent president (Joe Biden. Joe Biden was the presidential candidate when the data were collected.) and the former president (Donald Trump) are included in the analysis. Due to limitation of Twitter API, only about half of Donald Trump's follower ID was crawled.

### 2.2.8. Sentiment

In our study, we intend to infer the sentiment of personal pandemic experience and non-pandemic experience. First, we use keyword search methods to classify the three-month historical tweets into COVID-related and non-COVID-related. If a tweet does not contain any of the keywords: "corona", "covid", "covid19", "coronavirus", "chinese virus", "china virus", "wuhan virus", "wfh", "work from home", "pandemic", "epidemic", "herd immunity", "quarantine", "lockdown", "mortality", "morbidity", "social distancing", "mask", "social distance", "respirator", "state of emergency", "ventilator", "isolation", "fatality", "community spread", "vaccine", "vaccinated", "vaccination", "panic buying", "hoard", it is categorized as non-COVID-related. The example tweets are

"< user > I can not wait to take the last name of my husband! I feel so good to solidify our union by taking his name. I also cringe a little bit at the whole keep the maiden name on social media thing some girls do...I'm more leave-and-leave type." and

"< user > what a distinguished day that was."

The remaining tweets are categorized as COVID-19-related. The example tweets are

"I am the type of person who does half an hour of meditation and yoga from my peloton app before going to bed to read some chapters of my book and be fast asleep before 11pm. Quarantine changed me." and

"< user > Oooorr...I can wear a mask, get on an plane, in a limited space, with NO social distancing, with people from hundreds of different households, ALL going to various destinations, and then take my mask OFF to eat/drink once I'm in my seat < hashtag > < hashtag >".

For each Twitter user, the tweets of the two categories are concatenated, respectively. Next, a normalized, weighted composite score is calculated to measure the sentiment of the tweet content using Vader [33]. The score is between -1 (most extreme negative) and +1 (most extreme positive). Vader outperforms individual human raters when assessing the sentiment of tweets [33].

## 2.3. M3 opinion mining

To capture the opinions expressed through text by Twitter users, we adopt a human-guided machine learning framework inspired by Sadilek et al. [36]. The text are classified into four categories: (1) pro-vaccine, (2) vaccine-hesitant, (3) anti-vaccine, and (4) irrelevant.

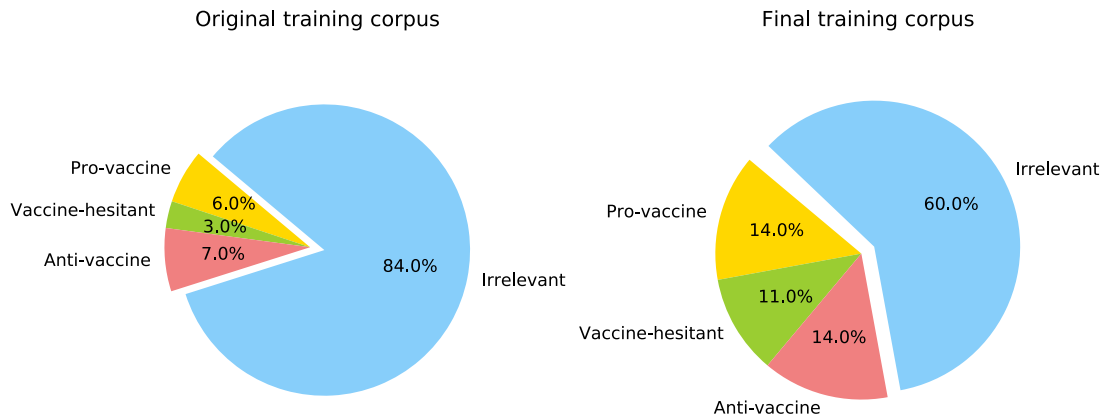
Tweets might be retweeted for multiple times. We observe that there are 6,703 non-unique tweets in the initial batch of over 90,000 tweets. These non-unique tweets, combined with their retweets constitute 62.9% of all tweets. As a result, the tweets are divided into two groups - the unique-tweet group and the non-unique-tweet group. 430 non-unique tweets which have been retweeted for at least 20 times are included in the non-unique-tweet group. These tweets and their retweets constitute 41.5% of all tweets. The rest are included in the unique-tweet group. All the tweets of the non-unique-tweet group are manually annotated. However, only a subgroup of the unique-tweet group are manually annotated. The state-of-the-art transformer-based language model [37], trained with the subgroup, is used to make estimates of the rest of the unique-tweet group.

### 2.3.1. Human-guided machine learning framework

We annotate the opinions of the tweets as pro-vaccine, vaccine-hesitant, or anti-vaccine using a human-guided machine learning framework to strike the best balance between automation and accuracy. In total, we stream over six million publicly available tweets from Twitter using Tweepy API between September 28 to November 4, 2020 with search keywords that are vaccine-related or COVID-19 vaccine-related. Unlike Tomeny, Vargo, and El-Toukhy [26], a majority of the tweets crawled with the search keywords in our study is irrelevant to the actual individual opinions about the vaccines for COVID-19, which causes a challenging class imbalance problem that may not only slow down the annotation process but also hinder the performance of automated classifiers [38]. To address this problem, we adopt a human-guided machine learning framework [36] based on the state-of-the-art transformer language model to label the opinions of the tweets. After extracting or inferring the features of these tweets and their authors, we only keep the ones with all the required informative features available.

**Table 1** Labeling scheme for Tweets

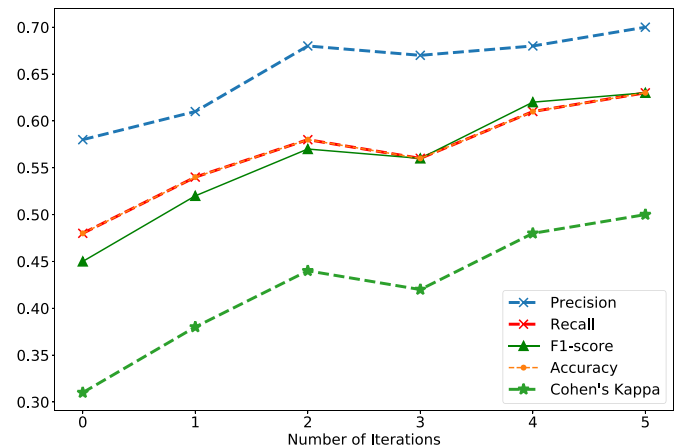
Category	Description
Pro-vaccine	i. Claiming that they would take the vaccine once it is available ii. Advocating and supporting vaccine/vaccine-associated entities like vaccine experiment trials iii. Believing that the vaccine will be the solution to the pandemic
Vaccine-hesitant	i. Claiming that they would like to take the vaccine after the vaccine is proven safe/effective ii. Claiming that they would wait for a while and see whether a vaccine is truly safe/effective if there is one iii. Showing worries about the effectiveness of a rushed vaccine
Anti-vaccine	i. Promoting/arguing in favor of conspiracy theory about vaccine/vaccine-associated entities ii. Believing that an effective vaccine would not be invented quickly and help overcome the pandemic iii. Believing that a covid-19 vaccine is dangerous for whatever reasons and would not take it even though the commenters claim that they are not anti-vaccine
Irrelevant	i. Vaccine News. No written opinion from the commenters ii. Including vaccine and the commenters opinions, but the focus is something else (i.e., insurance, politics, personal life experience, economics, emotional complaints, etc.) iii. Comments/questions on vaccines/vaccine-associated entities but with unclear meanings



**Figure 2.** Distributions of different categories of the original and final training corpora.

We initialize the human-guided machine learning framework by sampling 2,000 unique tweets from the corpus  $C$  with 244,049 tweets. Three researchers independently read each tweet and make a judgement whether this tweet is irrelevant, pro-vaccine, vaccine-hesitant, or anti-vaccine. Table 1 describes the labeling scheme for each opinion category. We label each tweet as one of the categories as long as it matches one of the descriptions of that category. The label of the tweet is assigned with the consensus votes from three researchers. If three researchers vote entirely differently, the senior researcher determines the label of this tweet after discussing with the other two researchers. The Fleiss’ Kappa score of the three researchers is 0.52. The corpus  $C_{train}$  of the initial 2,000 labelled tweets is fed to the XLNet model [37]. The four-class classification model  $H_1$  is trained and validated on an external validation set  $D_{validation}$  with 400 annotated tweets. The distribution of the four categories is balanced. We then construct another binary classification model  $H_2$  that is trained with only two classes of data. The data for  $H_1$  and  $H_2$  are almost the same except for the label of the output variable. For  $H_2$ , one class includes all the irrelevant tweets of the data for  $H_1$  and the other includes all the relevant tweets that are pro-vaccine, vaccine-hesitant, or anti-vaccine in the data for  $H_1$ . After training,  $H_2$  is used to make estimates for a corpus of 4,500 unlabeled tweets sampled from  $C$  regarding whether they are irrelevant or relevant. 90% of a new batch of corpus is composed of the top 10% of the most likely relevant tweets. The other 10% of the new batch is sampled uniformly at random to increase diversity. This new batch of corpus of 500 tweets is annotated by the three researchers as aforementioned and is added to the corpus  $C_{train}$ .  $H_1$  is trained with the updated  $C_{train}$  and validated again. This whole process is considered as one iteration. For each iteration, the three researchers annotate a new batch of corpus of 500 tweets.

This framework actively searches for relevant tweets to increase the sizes of the relevant datasets. Figure 2 shows the percentages of the



**Figure 3.** Performance of  $H_1$  of each iteration.

different opinion groups of the original  $C_{train}$  and the final  $C_{train}$  after five iterations. In each iteration, humans guide the machine to learn the irrelevant, pro-vaccine, vaccine-hesitant, and anti-vaccine tweets by updating the training set. Figure 3 shows the performance of  $H_1$  of each iteration. As a result, the framework allows us to label the opinions of the tweets and build the model more efficiently.

### 2.3.2. Tweets preprocessing

We adopt a tweet preprocessing pipeline from Baziotis et al. [39] which can transform the specific text often used in Twitter to special tokens. For example, if the original tweet is

**Table 2** Performance of the four-class XLNet model  $H_1$ 

Class	Precision	Recall	F1-score
Irrelevant	0.45	0.84	0.59
Pro-vaccine	0.78	0.52	0.62
Vaccine-hesitant	0.77	0.54	0.64
Anti-vaccine	0.79	0.61	0.69
Overall	0.70	0.63	0.63

“Scientists develop a COVID vaccine that could initiate a 10-times stronger immune response < url >”

After preprocessing, the tweet becomes

“scientists develop a < allcaps > covid < /allcaps > vaccine that could initiate a < number > - times stronger immune response < url >”.

### 2.3.3. Performance of the XLNet model

Table 2 summarizes the performance of the final four-class XLNet model  $H_1$  on the external validation set with 400 samples. The final accuracy is 0.63 and the Cohen’s Kappa score is 0.5, which indicates a good agreement.

## 2.4. M4 analysis details

### 2.4.1. Statistical analysis

To understand what opinion (i.e., pro-vaccine, vaccine-hesitant, and anti-vaccine) the people ( $n = 10,945$ ) would hold based on their demographics, social capital, income, religious status, family status, political affiliations, geo-location, sentiment about COVID-19-related experience and non-COVID-related experience, and relative change of the number of daily confirmed cases at the county level, we conduct multinomial logistic regression, selecting vaccine-hesitant group as the reference category.

### 2.4.2. Counterfactual analyses

Following Chang et al. [40], we intend to estimate the impact of communication-related strategies by constructing a hypothetical machine learning model that reflects the expected effect. To assess the potential outcomes of the communication-related strategies, we build the machine learning model using the real data, and apply the constructed model to the hypothetical data.

The data ranging from September 28 to October 21, 2020 are used to train a support vector machine (SVM)  $H_3$  which makes predictions about the opinion group of the data of the latest two weeks (October 22 - November 4, 2020). The real percentage of pro-vaccine users and the prediction percentage are plotted in Figure 4. The real percentage falls within in one standard deviation of the predicted percentage, indicating a good simulation performance.

We further analyze the relationship between the opinions and the topics of the tweets using the Latent Dirichlet Allocation (LDA) topic modelling [41] with 10 topics as shown in Figure 5. The coherence score is 0.31. In the word cloud of each topic, top 30 keywords are plotted. As we can see from the figure, people are most concerned about the safety and effectiveness of the vaccine which is consistent with the Pew Research Center survey (<https://www.pewresearch.org/science/2020/09/17/u-s-public-now-divided-over-whether-to-get-covid-19-vaccine/>). Some politics-related keywords like “administration”, “white house”, and the names of political figures like “Trump” and “Kamala” are presented as well. To label the factor indicators, we narrow down the 10 topics to two major ones: “safety and effectiveness” and “politics”, and use keyword search methods. The keywords for the safety and effectiveness include “safe”, “effective”, and “efficacy”. The keywords for the politics include “administration”, “politics”, “politician”, “political” and the names of Donald Trump, Mike Pence, Joe Biden and Kamala Harris. Each tweet is labelled 1 if it contains the related keywords, and 0 if it does not. Table 3 shows the descriptive statistics of

**Table 3** Descriptive statistics of the factor indicators

Variables	Number	Mean	SD	Min	Max
Politics	10,945	0.2512	0.4337	0	1
Safety and effectiveness	10,945	0.1801	0.3843	0	1

these two variables. The basic settings for the counterfactual classifiers are the same as  $H_3$ . We analyze one factor at a time. We train the classifier with the basic variables and the factor indicator with the real value. The basic variables include user demographics, Twitter usage patterns, sentiment of the pandemic and non-pandemic experience, income, religious status, family status, political affiliation, as well as the population density. The prediction is plotted in orange in Figure 4. Then we change the value of the factor indicator which was originally 1 into 0, keeping other variables constant. The trained classifier is applied to the hypothetical data, and the prediction is plotted in green.

## 3. Results

### 3.1. Characterization of different opinion groups

The proportions of the different opinion groups of the United States public change over time as shown in Figure 6a, which roughly correspond to the major pandemic-related events. Figure 6b shows the number of Twitter users. Overall, 57.65% (6,218 of 25,407) are pro-vaccine, 19.30% (2,469 of 25,407) are vaccine-hesitant, and the rest are anti-vaccine. By aggregating people at the state level, we estimate the opinions about the potential COVID-19 vaccines of each state as shown in Figure 7. The Southeast of the United States shows a relatively lower acceptance level, so does the cluster of Ohio, Indiana and Kentucky.

After performing the Granger Causality Test with a one-day lag, we find that (Figure 8), in Nevada, Tennessee and Washington, the percentage of the pro-vaccine people deviates the most from the national average ( $P > 0.05$ ). The percentage of the pro-vaccine group of Washington is above the national average during the most of the time, while the acceptance level of Nevada is relatively lower than the national average. More drastic changes are observed for the acceptance level of Tennessee.

Descriptive statistics and bi-variate correlations of the variables of the multinomial logistic regression are shown in Table 4. Table 5 summarizes the results of the multinomial logistic regression. The Chi-square test shows that the variables significantly predict the opinion on potential COVID-19 vaccines:  $\chi^2(40, N = 10,945) = 1,340.94, P < 0.001$ , McFadden’s pseudo  $R^2 = 0.06$ , which supports our hypotheses. Next, we show the predictive effects of these variables with paired comparisons.

#### 3.1.1. Women are more likely to hold hesitant opinions

Gender is statistically significant ( $\chi^2 = 91.83, P < 0.001$ ). Women are likely to hold hesitant opinions rather than polarized opinions (i.e., pro-vaccine, anti-vaccine). Specifically, comparing the anti-vaccine group and vaccine-hesitant group, we find that women are less likely to be anti-vaccine ( $B = -0.25, SE = 0.06, P < 0.001, OR = 0.78; 95\% CI = 0.69-0.88$ ). Comparing the pro-vaccine group and vaccine-hesitant group, we find that women are also less likely to be pro-vaccine ( $B = -0.47, SE = 0.05, P < 0.001, OR = 0.63; 95\% CI = 0.57-0.69$ ).

#### 3.1.2. Older people tend to be pro-vaccine

Age is statistically significant ( $\chi^2 = 72.47, P < 0.001$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we do not find significant evidence that older people are more anti-vaccine. However, comparing the pro-vaccine group and vaccine-hesitant group, we find that people who are one year older are 1.01 ( $B = 0.01, SE = 0.00, P < 0.001, OR = 1.01; 95\% CI = 1.01-1.02$ ) times more likely to be pro-vaccine instead of vaccine-hesitant, which echoes the findings

**Table 4** Descriptive statistics and the bi-variate correlations

Variables	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Gender (0 = male, 1 = female)	0.46	0.50																			
2. Age (years)	39.89	14.69	-.08**																		
3. Verified (0 = no, 1 = yes)	0.04	0.20	-.00	.03**																	
4. Twitter history (months)	91.30	43.47	-.02	.03**	.09**																
5. # Followers	1.60	1.63	.04**	-.01	.37**	-.08**															
6. # Friends	1.95	1.25	.04**	-.00	.09**	-.29**	.68**														
7. # Listed memberships	-1.62	0.93	-.00	.08**	.49**	.22**	.69**	.31**													
8. # Favorites	4.17	1.95	.12**	-.14**	-.03**	-.21**	.38**	.47**	.05**												
9. # Statuses	4.09	1.43	.02*	-.11**	.07**	-.09**	.53**	.45**	.29**	.58**											
10. Higher-Income (0 = no, 1 = yes)	0.00	0.05	-.00	-.03**	.03**	-.01	.05**	.02*	.04**	.02**	.01										
11. Lower-Income (0 = no, 1 = yes)	0.76	0.43	-.00	-.43**	-.04**	.05**	-.15**	-.18**	-.12**	-.13**	-.09**	-.10**									
12. Religious (0 = no, 1 = yes)	0.04	0.19	.01	.07**	-.03**	-.03**	.03**	.07**	-.03**	.02	.02	-.00	-.06**								
13. Having kids (0 = no, 1 = yes)	0.12	0.32	.09**	.09**	.03**	.02*	.04**	.05**	.04**	.01	-.05**	-.01	-.05**	.09**							
14. Following Trump (0 = no, 1 = yes)	0.11	0.31	-.05**	.06**	-.03**	-.17**	-.05**	.04**	-.12**	-.01	-.05**	-.00	-.05**	.06**	.04**						
15. Following Biden (0 = no, 1 = yes)	0.17	0.38	.09**	.07**	.06**	.10**	.05**	.19**	.06**	.09**	.03**	.00	-.06**	-.03**	.07**	.01					
16. Rural (0 = no, 1 = yes)	0.19	0.40	-.02	.09**	-.05**	-.04**	-.07**	-.02	-.08**	-.01	-.04**	.00	-.05**	.06**	.04**	.05**	-.00				
17. Suburban (0 = no, 1 = yes)	0.14	0.35	-.01	.07**	-.01	-.02*	-.05**	-.03**	-.05**	-.03**	-.02**	-.02	-.01	.03**	.04**	.03**	-.00	-.20**			
18. Pandemic experience (sentiment)	0.06	0.80	.03**	-.04**	.10**	.03**	.09**	-.00	.15**	-.08**	-.08**	.02*	.04**	-.03**	-.00	-.06**	.01	-.03**	.00		
19. Non-pandemic experience (sentiment)	0.62	0.75	.07**	-.06**	.07**	.04**	.14**	.09**	.14**	.06**	.01	.02	.02	.00	.03**	-.06**	.01	-.03**	-.02*	.27**	
20. Pandemic severity perception (relative change of # daily confirmed cases)	0.01	0.00	-.01	.03**	-.03**	-.02*	-.05*	-.01	-.07**	-.00	-.04**	.01	-.03**	.06**	.05**	.04**	-.01	.27**	.12**	-.01	.01

Note: \*  $P < 0.05$ . \*\*  $P < 0.01$ . # The numbers of followers, friends, listed memberships, favorites, statuses are normalized by the months of Twitter history and log-transformed.

**Table 5** Multinomial logistic regression outputs for the opinion on potential COVID-19 vaccines against demographics and other variables of interest

Predictor	Anti-vaccine			Pro-vaccine		
	B	SE	OR (95% CI)	B	SE	OR (95% CI)
Intercept	-1.82***	0.26		0.79***	0.20	
Age (years)	0.00	0.00	1.00 (1.00, 1.01)	0.01***	0.00	1.01 (1.01, 1.02)
Twitter history (months)	0.00	0.00	1.00 (1.00, 1.00)	-0.003***	0.001	0.997 (0.996, 0.999)
# Followers	0.28***	0.04	1.32 (1.22, 1.42)	0.08**	0.03	1.09 (1.02, 1.16)
# Friends	-0.18***	0.04	0.83 (0.77, 0.90)	-0.07*	0.03	0.93 (0.88, 0.99)
# Listed memberships	-0.63***	0.06	0.53 (0.47, 0.60)	0.10*	0.05	1.10 (1.01, 1.20)
# Favorites	0.04*	0.03	1.12 (1.06, 1.19)	0.04*	0.02	1.04 (1.01, 1.08)
# Statuses	0.11***	0.03	1.12 (1.06, 1.19)	-0.06*	0.02	0.94 (0.90, 0.99)
Pandemic experience (sentiment)	-0.18***	0.04	0.84 (0.77, 0.90)	0.21***	0.03	1.24 (1.16, 1.32)
Non-pandemic experience (sentiment)	-0.04	0.04	0.96 (0.89, 1.04)	0.13***	0.03	1.14 (1.07, 1.22)
Pandemic severity perception (relative change of # daily confirmed cases)	-14.99	8.13	0.00 (0.00, 2.58)	-22.68***	6.59	0.00 (0.00, 0.00)
Female	-0.25***	0.06	0.78 (0.69, 0.88)	-0.47***	0.05	0.63 (0.57, 0.69)
Verified user	-0.61*	0.27	0.54 (0.32, 0.91)	-0.16	0.14	0.85 (0.65, 1.12)
Higher-income	-170.67	5.00e+36	0.00 (0.00, Inf)	0.47	0.43	1.60 (0.70, 3.68)
Lower-income	0.40***	0.08	1.49 (1.26, 1.75)	0.52***	0.06	1.69 (1.49, 1.91)
Religious	0.74***	0.17	2.10 (1.52, 2.91)	0.37*	0.15	1.45 (1.07, 1.95)
Having kids	-0.11	0.10	0.90 (0.74, 1.09)	0.00	0.08	1.00 (0.86, 1.15)
Following Trump	0.41***	0.10	1.51 (1.26, 1.83)	0.06	0.08	1.06 (0.90, 1.25)
Following Biden	-1.22***	0.10	0.29 (0.24, 0.36)	-0.34***	0.06	0.71 (0.63, 0.80)
Rural	0.17*	0.08	1.19 (1.01, 1.39)	0.07	0.07	1.07 (0.94, 1.22)
Suburban	0.18*	0.09	1.20 (1.01, 1.43)	0.11	0.07	1.12 (0.97, 1.29)
Chi-square				1,340.94***		
df				40		
-2 log likelihood				20,171.90		
McFadden's pseudo R <sup>2</sup>				0.06		
Sample size				10,945		

Note: \*  $P < 0.05$ . \*\*  $P < 0.01$ . \*\*\*  $P < 0.001$ . The vaccine hesitant group is selected as the reference category. # The numbers of followers, friends, listed memberships, favorites, statuses are normalized by the months of Twitter history and log-transformed.

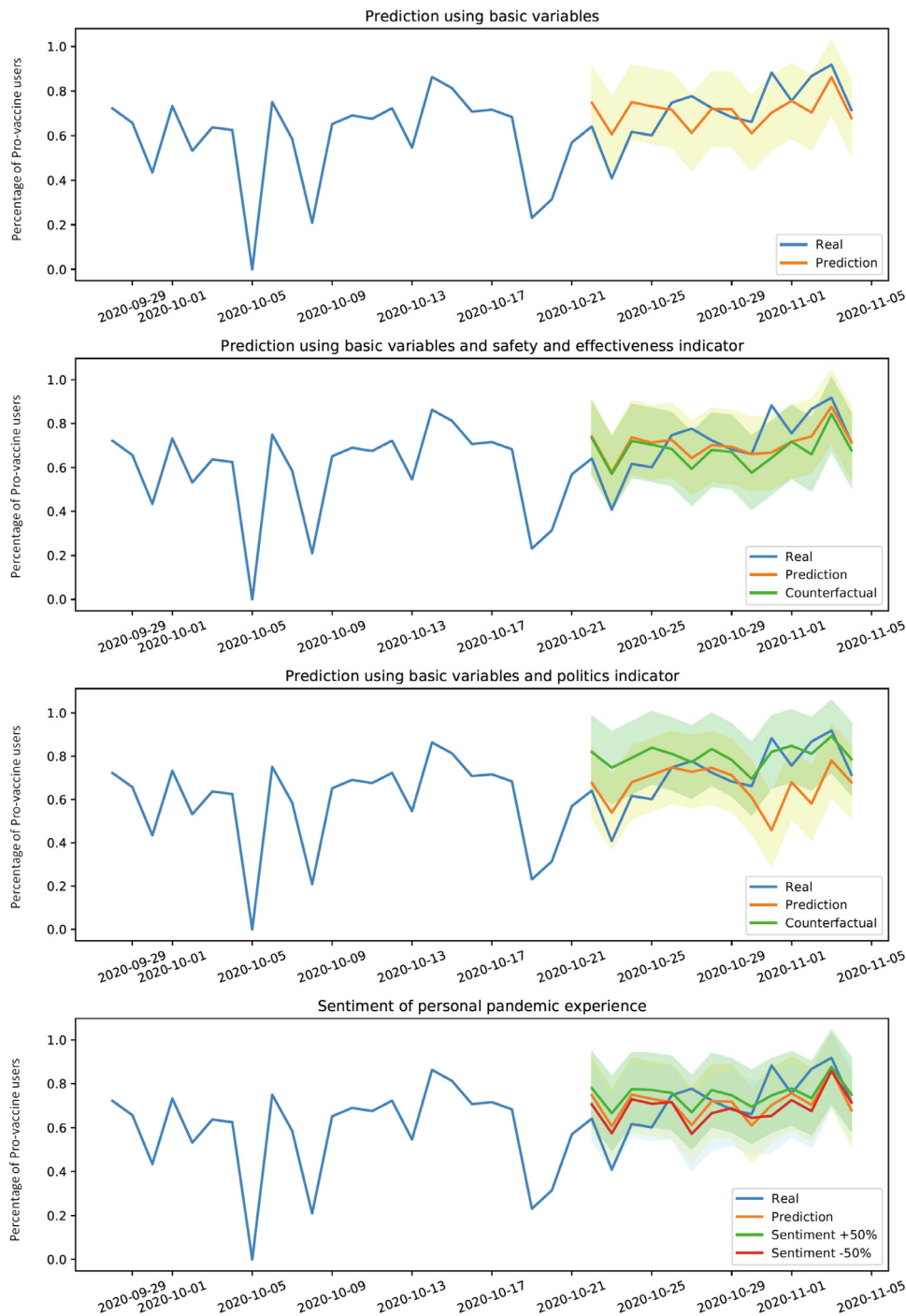


Figure 4. Counterfactual analyses illustrate the importance of politics, safety and effectiveness factor indicators, and personal pandemic experience.

of Lazarus et al. [12]. One potential explanation is that the risk of dying with COVID-19 increases with age [42], and the benefits of not getting infected with COVID-19 outweigh the risk of getting vaccinated.

### 3.1.3. Different patterns of Twitter usage

A verified Twitter account must represent or otherwise be associated with a prominently recognized individual or brand [43]. In our study, verified status is statistically significant ( $\chi^2 = 6.12, P < 0.05$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we find verified users are less likely to be anti-vaccine ( $B = -0.61, SE = 0.27, P < 0.05, OR = 0.54; 95\% CI = 0.32-0.91$ ), however, comparing the

pro-vaccine group and vaccine-hesitant group, we do not find significant differences.

Months of Twitter history is statistically significant ( $\chi^2 = 17.52, P < 0.001$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we do not find significant differences, however, comparing the pro-vaccine group and vaccine-hesitant group, we find if the months of Twitter history were to increase by one month, it is 0.997 ( $B = -0.003, SE = 0.001, P < 0.001, OR = 0.997; 95\% CI = 0.996-0.999$ ) less likely to be pro-vaccine than vaccine-hesitant.

After normalizing the number of followers, friends, listed memberships, favorites, and statuses with the number of months of Twitter history, we still find that the social capital is statistically significant. Specif-



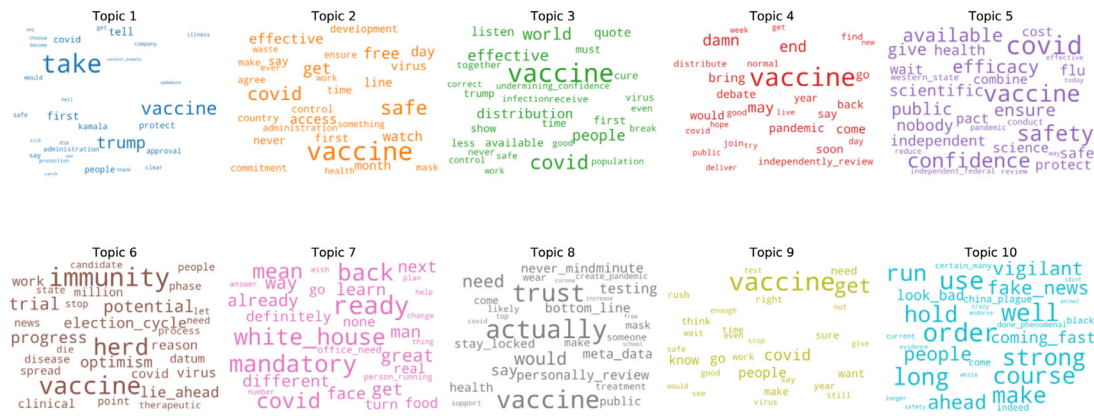


Figure 5. 10 topics extracted from the tweets with the top 30 keywords.

ically, there are significant differences in terms of followers counts ( $\chi^2 = 51.06$ ,  $P < 0.001$ ), friends counts ( $\chi^2 = 21.28$ ,  $P < 0.001$ ), listed memberships counts ( $\chi^2 = 199.51$ ,  $P < 0.001$ ), favorites counts ( $\chi^2 = 6.10$ ,  $P < 0.05$ ), statuses counts ( $\chi^2 = 47.37$ ,  $P < 0.001$ ).

Comparing the anti-vaccine group and vaccine-hesitant group, if the log-scale followers count were to increase by one unit, it is 1.32 ( $B = 0.28$ ,  $SE = 0.04$ ,  $P < 0.001$ ,  $OR = 1.32$ ; 95%  $CI = 1.22$ – $1.42$ ) times more likely to be anti-vaccine. If the log-scale friends count were to increase by one unit, it is less likely to be anti-vaccine ( $B = -0.18$ ,  $SE = 0.04$ ,  $P < 0.001$ ,  $OR = 0.83$ ; 95%  $CI = 0.77$ – $0.90$ ). If the log-scale listed memberships count were to increase by one unit, it is less likely to be anti-vaccine ( $B = -0.63$ ,  $SE = 0.06$ ,  $P < 0.001$ ,  $OR = 0.53$ ; 95%  $CI = 0.47$ – $0.60$ ). If the log-scale favorites count were to increase by one unit, it is 1.04 ( $B = 0.04$ ,  $SE = 0.02$ ,  $P < 0.05$ ,  $OR = 1.04$ ; 95%  $CI = 1.00$ – $1.09$ ) times more likely to be anti-vaccine. If the log-scale statuses count were to increase by one unit, it is 1.12 ( $B = 0.11$ ,  $SE = 0.03$ ,  $P < 0.001$ ,  $OR = 1.12$ ; 95%  $CI = 1.06$ – $1.19$ ) times more likely to be anti-vaccine.

Comparing the pro-vaccine group and vaccine-hesitant group, if the log-scale followers count were to increase by one unit, it is 1.09 ( $B = 0.08$ ,  $SE = 0.03$ ,  $P < 0.01$ ,  $OR = 1.09$ ; 95%  $CI = 1.02$ – $1.16$ ) times more likely to be pro-vaccine. If the log-scale friends count were to increase by one unit, it is less likely to be pro-vaccine ( $B = -0.07$ ,  $SE = 0.03$ ,  $P < 0.05$ ,  $OR = 0.93$ ; 95%  $CI = 0.88$ – $0.99$ ). If the log-scale listed memberships count were to increase by one unit, it is 1.11 ( $B = 0.10$ ,  $SE = 0.05$ ,  $P < 0.05$ ,  $OR = 1.10$ ; 95%  $CI = 1.01$ – $1.20$ ) times more likely to be pro-vaccine. If the log-scale favorites count were to increase by one unit, it is 1.04 ( $B = 0.04$ ,  $SE = 0.02$ ,  $P < 0.05$ ,  $OR = 1.04$ ; 95%  $CI = 1.01$ – $1.08$ ) times more likely to be pro-vaccine. If the log-scale statuses count were to increase by one unit, it is less likely to be pro-vaccine ( $B = -0.06$ ,  $SE = 0.02$ ,  $P < 0.05$ ,  $OR = 0.94$ ; 95%  $CI = 0.90$ – $0.99$ ).

Twitter users who have more followers or fewer friends, or give more favourites are more likely to hold polarized opinion. The larger listed memberships count is, the more likely the Twitter user is pro-vaccine. Twitter users who post more statuses tend to be anti-vaccine.

### 3.1.4. The lower-income group is more likely to hold polarized opinions

Income is statistically significant ( $\chi^2 = 79.09$ ,  $P < 0.001$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we find that the lower-income group is 1.49 ( $B = 0.40$ ,  $SE = 0.08$ ,  $P < 0.001$ ,  $OR = 1.49$ ; 95%  $CI = 1.26$ – $1.75$ ) times more likely to be anti-vaccine than the medium-income group, however, the difference between the higher-income group and medium-income group is not significant. Comparing the pro-vaccine group and vaccine-hesitant group, we find that lower-income group is 1.69 ( $B = 0.52$ ,  $SE = 0.06$ ,  $P < 0.001$ ,  $OR = 1.69$ ; 95%  $CI = 1.49$ – $1.91$ ) times more likely to be pro-vaccine than

medium-income group. The difference between the higher-income group and medium-income group is not significant. Inconsistent with Lazarus et al. [12] that the higher the income is, the more likely people are pro-vaccine, we find the effect of income more nuanced. Lower-income people tend to be polarized.

### 3.1.5. Religious people are more likely to be polarized

Religious status is statistically significant ( $\chi^2 = 21.34$ ,  $P < 0.001$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we find that religious people are more likely to be anti-vaccine than non-religious people ( $B = 0.74$ ,  $SE = 0.17$ ,  $P < 0.001$ ,  $OR = 2.10$ ; 95%  $CI = 1.52$ – $2.91$ ). Comparing the pro-vaccine group and vaccine-hesitant group, we find that religious people are also more likely to be pro-vaccine than non-religious people ( $B = 0.37$ ,  $SE = 0.15$ ,  $P < 0.05$ ,  $OR = 1.45$ ; 95%  $CI = 1.07$ – $1.95$ ). This is in line with Larson et al. [5] that the effect of religious status is complicated.

### 3.1.6. Political diversion indicates a divided opinion about the potential COVID-19 vaccines

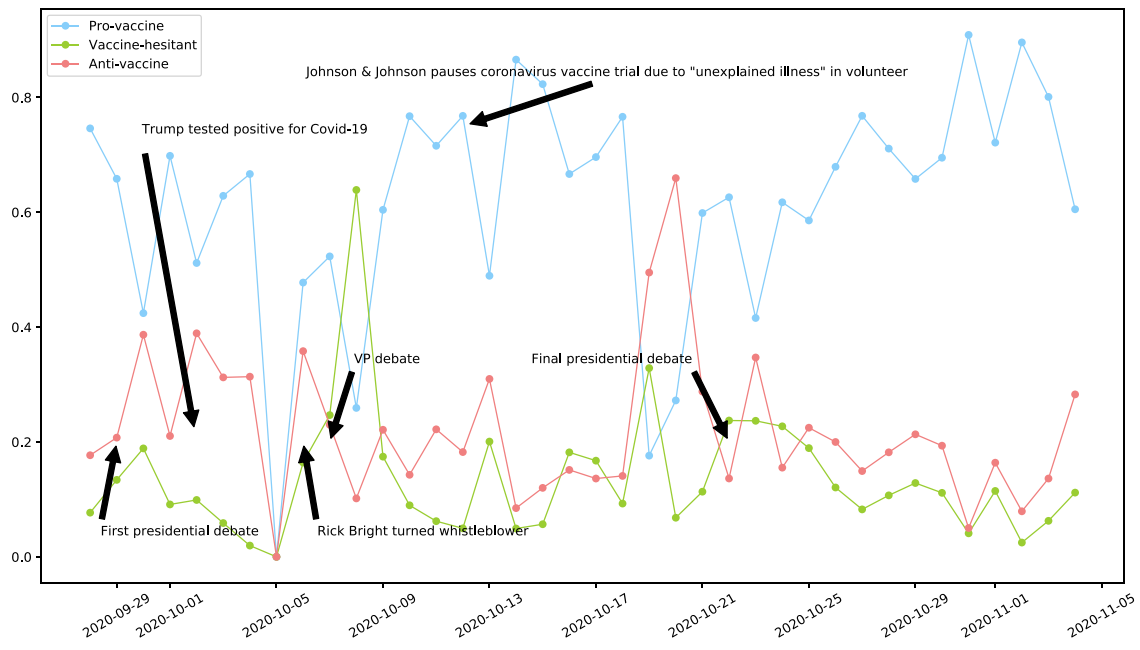
Following Donald Trump is statistically significant ( $\chi^2 = 25.22$ ,  $P < 0.001$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we find that the Twitter users who follow Donald Trump are 1.51 ( $B = 0.41$ ,  $SE = 0.10$ ,  $P < 0.001$ ,  $OR = 1.51$ ; 95%  $CI = 1.26$ – $1.83$ ) times more likely to be anti-vaccine than the Twitter users who do not. Comparing the pro-vaccine group and vaccine-hesitant group, following Donald Trump is not significant.

Following Joe Biden is statistically significant ( $\chi^2 = 177.96$ ,  $P < 0.001$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we find that the Twitter users who follow Joe Biden are less likely to be anti-vaccine than the Twitter users who do not ( $B = -1.22$ ,  $SE = 0.10$ ,  $P < 0.001$ ,  $OR = 0.30$ ; 95%  $CI = 0.24$ – $0.36$ ). Comparing the pro-vaccine group and vaccine-hesitant group, we find that the Twitter users who follow Joe Biden are also less likely to be pro-vaccine than the Twitter users who do not ( $B = -0.34$ ,  $SE = 0.06$ ,  $P < 0.001$ ,  $OR = 0.71$ ; 95%  $CI = 0.63$ – $0.80$ ).

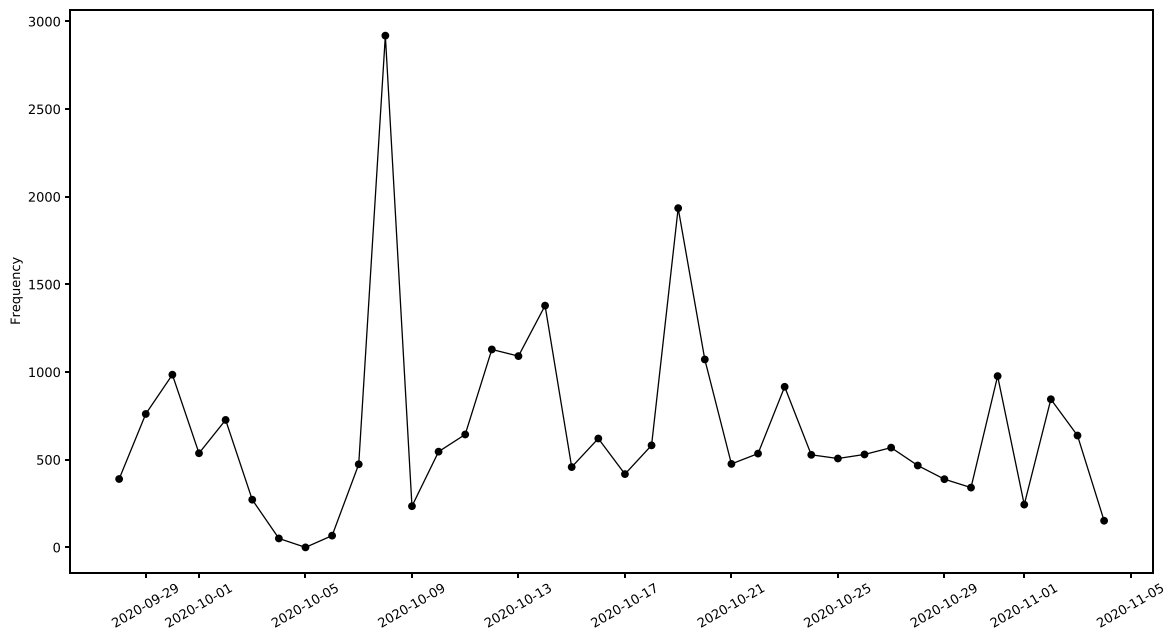
Twitter users who follow Donald Trump tend to be anti-vaccine, while those who follow Joe Biden tend to be vaccine-hesitant.

### 3.1.7. People living in suburban or rural areas are more likely to be anti-vaccine

Although the population density of the area is not statistically significant across three opinion categories, we still find differences between the anti-vaccine group and vaccine-hesitant group. People living in suburban areas are 1.20 ( $B = 0.18$ ,  $SE = 0.09$ ,  $P < 0.05$ ,  $OR = 1.20$ ; 95%  $CI = 1.01$ – $1.43$ ) times more likely to be anti-vaccine than people living in urban areas. People living in rural areas



(a)



(b)

**Figure 6.** (a) The proportions of the opinion groups from September 28 to November 4, 2020. (b) Number of Twitter users from September 28 to November 4, 2020. The data of October 5, 2020 are missing due to a data collection issue.

are 1.20 ( $B = 0.17$ ,  $SE = 0.08$ ,  $P < 0.05$ ,  $OR = 1.18$ ; 95%  $CI = 1.01$ – $1.39$ ) times more likely to be anti-vaccine than people living in urban areas.

Most of the results are consistent with Hypothesis 1. There are significant differences in demographics, social capital, income, religious status, political affiliations and geo-locations among opinion groups, however, we do not find significant difference in family status.

### 3.1.8. Personal experience with COVID-19 and the county-level pandemic severity perception shape the opinion

The sentiment score of personal experience with COVID-19 is statistically significant ( $\chi^2 = 146.50$ ,  $P < 0.001$ ). Comparing the anti-vaccine group and vaccine-hesitant group, we find that if the sentiment score of personal experience with COVID-19 were to increase by

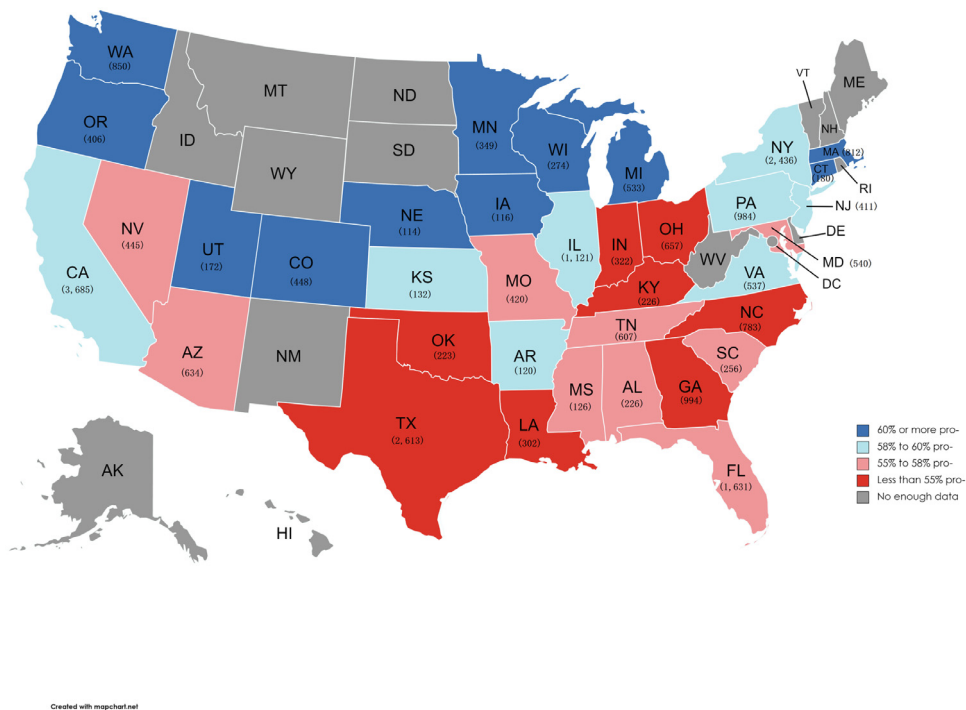


Figure 7. State-level public opinions about potential COVID-19 vaccines. The numbers in parentheses stand for the sizes of the study populations.

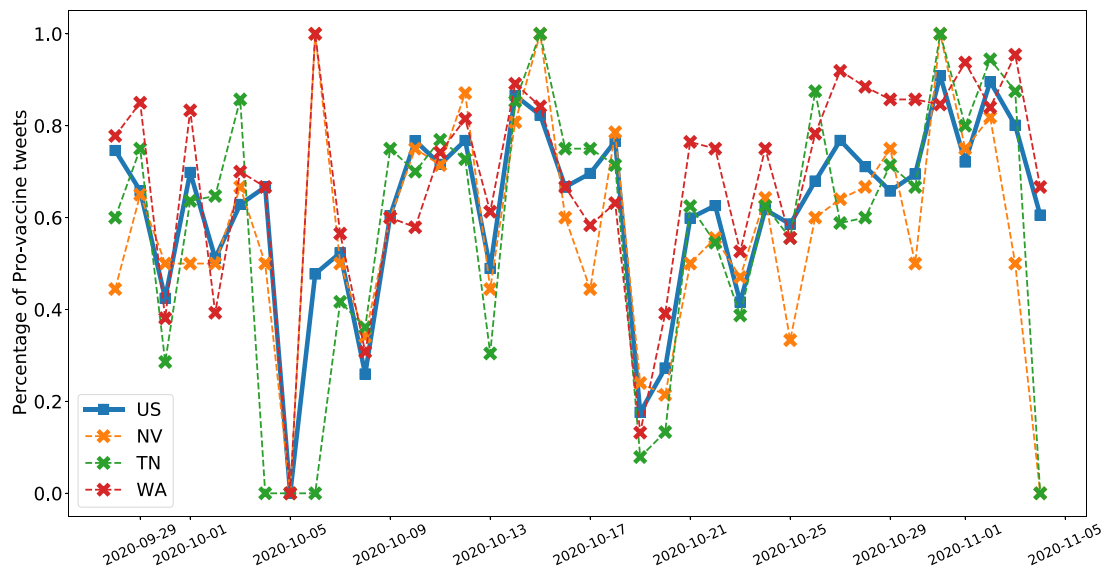


Figure 8. The percentages of the pro-vaccine groups of the national average (US), Nevada (NV), Tennessee (TN), and Washington (WA).

one unit (i.e., the sentiment became more positive), the person would be less likely to hold anti-vaccine opinion ( $B = -0.18, SE = 0.04, P < 0.001, OR = 0.84; 95\% CI = 0.77-0.90$ ). Comparing the pro-vaccine group and vaccine-hesitant group, we find if the sentiment score of personal experience with COVID-19 were to increase by one unit (i.e., the sentiment became more positive), the person would be 1.24 times more likely to hold pro-vaccine opinion ( $B = 0.21, SE = 0.03, P < 0.001, OR = 1.24; 95\% CI = 1.16-1.32$ ), which is consistent with Hypothesis 2.

The sentiment score of non-COVID-related personal experience is overall statistically significant ( $\chi^2 = 29.28, P < 0.001$ ), but comparing the anti-vaccine group and vaccine-hesitant group, we find no significant difference. However, comparing the pro-vaccine group and vaccine-hesitant group, we find if the sentiment score of non-COVID-related personal experience were to increase by one unit (i.e.,

the sentiment became more positive), the person would be more likely to hold pro-vaccine opinion ( $B = 0.13, SE = 0.03, P < 0.001, OR = 1.14; 95\% CI = 1.07-1.22$ ).

The county-level pandemic severity perceptions are overall statistically significant ( $\chi^2 = 11.76, P < 0.01$ ), supporting Hypothesis 3, but we find no significant difference comparing the anti-vaccine group and vaccine-hesitant group. However, comparing the pro-vaccine group and vaccine-hesitant group, if the relative change of the number of daily confirmed cases at the county level were to increase by one unit, the person would be less likely to be pro-vaccine ( $B = -22.68, SE = 6.59, P < 0.001, OR = 0.00; 95\% CI = 0$ ).

At the individual level, the personal pandemic experience is a strong predictor of the opinion about COVID-19 vaccines. People who have the worst personal pandemic experience are more likely to hold anti-

vaccine opinion. However, the non-pandemic experience is not a strong predictor of anti-vaccine opinion. At the county level, people who have the worst pandemic severity perception (i.e., the relative change of the number of daily confirmed cases is the largest) are more likely to be vaccine-hesitant.

### 3.2. Counterfactual analyses

#### 3.2.1. Removing the safety and effectiveness factors reduces the vaccine acceptance level. However, removing the politics factor increases it

Figure 4 shows the results of counterfactual analyses of factor indicators. Using counterfactual analysis by turning the factor indicator of safety and effectiveness into 0, there is a clear decrease (4.42% on average) of the percentage of the pro-vaccine people. However, by turning the factor indicator of politics into 0, there is a clear increase (22.65% on average) of the percentage of the pro-vaccine people. This indicates that people are most concerned about the relationship between the politics and the potential COVID-19 vaccines, which is also mirrored by the news report [44].

#### 3.2.2. Improving personal pandemic experience increases the vaccine acceptance level

Figure 4 shows the results of counterfactual analyses of different sentiment levels of personal pandemic experience. By increasing the sentiment scores with a factor of 50%, the percentage of the pro-vaccine people increases by 6.39%. However, by reducing the sentiment scores of a factor of 50%, the percentage of the pro-vaccine people decreases by 2.82%.

### 3.3. Robustness verification

The multinomial logistic regression and counterfactual analyses are conducted based on the opinion derived from the social media data. By adopting the human-guided machine learning framework, the dataset is composed of human-annotated data and machine-inferred data. Although the Cohen's Kappa score of the machine and human annotators reaches 0.5 after five iterations, the results of the analyses could potentially change given the seemingly limited prediction performance. Therefore, we additionally verify the robustness of the findings of multinomial logistic regression and counterfactual analyses by examining two different combinations of the human-annotated data and machine-inferred data:

- Only human-annotated data ( $n = 2,939$ ) are used to conduct the multinomial logistic regression and counterfactual analyses
- Human-annotated data ( $n = 2,939$ ) and 50% machine-inferred data (randomly sampled,  $n = 4,003$ ) are used to conduct the multinomial logistic regression and counterfactual analyses

The results of the two robustness verification experiments are shown in Appendices. Both of them are basically consistent with the reported results in Table 5 and Figure 4. The findings do not fundamentally change, confirming that the opinion computationally derived from the social media data are reliable for our study.

## 4. Discussion

We conduct multinomial logistic regression to investigate the scope and causes of public opinions on vaccines and test three hypotheses. The current study shows the hypothesized effects of most of the characteristics in predicting the odds of being pro-vaccine or anti-vaccine against vaccine-hesitant. The findings suggest that women are more vaccine-hesitant, which is consistent with the Reuters/Ipsos survey (<https://www.reuters.com/business/healthcare-pharmaceuticals/more-women-than-men-us-nervous-about-fast-rollout-covid-vaccine-thats-problem-2020-12-11/>), and older people tend to be pro-vaccine. With respect to social capital, people who have more followers or fewer friends, or give more favorites, are more likely to hold polarized

opinions. Verified status, months of Twitter history, listed memberships counts and statuses counts are statistically significant as well. We also show that the lower-income group is more likely to hold polarized opinions. This is inconsistent with the finding by Lazarus et al. [12]. Moreover, religious people tend to hold polarized opinions. As for political affiliations, Twitter users who follow Donald Trump are more likely to be anti-vaccine rather than vaccine-hesitant, while those who follow Joe Biden tend to be vaccine-hesitant rather than anti-vaccine or pro-vaccine. In addition, we find people who live in rural or suburban areas tend to be anti-vaccine. However, we do not find the hypothesized predictive effect of family status on the opinion about vaccines.

Furthermore, the current study shows the hypothesized predictive effects of the personal pandemic experience and the county-level pandemic severity perception. In particular, personal experience with COVID-19 is a strong predictor of anti-vaccine opinion. The more negative the experience is, the more negative the opinion on vaccines is. People are more likely to be vaccine-hesitant if their pandemic severity perceptions are worse.

Our current study has limitations. The public opinions of some (less populated) states cannot be reflected due to the inadequate data. The findings could be further validated in other populations. In addition, the performance of the XLNet model of this study could be improved so that the proposed human-guided machine learning framework could identify the public opinions on the COVID-19 vaccines more accurately. Specifically, future work could address the class imbalance issue by annotating more tweets or augmenting the training data via back translation (i.e. translating English tweets to another language and then translating them back to English). Despite the limitations, our study broadly captures the public opinions on the potential vaccines for COVID-19 on Twitter. The final F1-score of our study is 0.63, indicating a good performance, which is achieved via the human-guided machine learning framework that is composed of both the state-of-the-art transformer-based language model and the large annotated dataset. By aggregating the opinions, we find a lower acceptance level in the Southeast part of the United States. The changes of the proportions of different opinion groups correspond roughly to the major pandemic-related events. We show the hypothesized predictive effects of the characteristics of the people in predicting pro-vaccine, vaccine-hesitant, and anti-vaccine group. Using counterfactual analyses, we find that people are most concerned about the safety, effectiveness and politics regarding potential COVID-19 vaccines, and improving personal experience with COVID-19 increases the vaccine acceptance level.

Our results can guide and support policymakers making more effective distribution policies and strategies. First, more efforts of dissemination should be spent on the socioeconomically disadvantaged groups who are exposed to potentially higher risks [40,45–46] and already possess more polarized attitudes towards the vaccines. Second, messaging for the vaccines is extremely important because the vaccine acceptance level can be increased by removing the politics factor. Third, safety and effectiveness issues need to be well addressed because the acceptance level is reduced by removing this factor. Finally, improving personal pandemic experience may increase the vaccine acceptance level as well and thus all helpful measures should be integrated to maximize the vaccine acceptance. In the future, by combining social media data and more traditional survey data, we hope to acquire deeper insights into the public opinions on potential COVID-19 vaccines and thus inform more effective vaccine dissemination policies and strategies.

### Conflicts of interest statement

The authors declare that there are no conflicts of interest.

### Funding

This research was supported in part by a University of Rochester Research Award, and National Institutes of Health (Grant No. RF1AG063811-01S2).

## Author contributions

Hanjia Lyu, Timothy D. Dye and Jiebo Luo conceived and designed the analysis. Hanjia Lyu collected the data. Hanjia Lyu, Junda Wang, Viet Duong contributed to data analysis. All authors wrote the paper. Hanjia Lyu and Jiebo Luo made critical revision.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.imed.2021.08.001.

## References

- [1] Ferguson N, Laydon D, Nedjati Gilani G, et al. Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020. doi:10.25561/77482.
- [2] Yamey G, Schäferhoff M, Hatchett R, et al. Ensuring global access to COVID-19 vaccines. *Lancet* 2020;395(10234):1405–6. doi:10.1016/S0140-6736(20)30763-7.
- [3] Gust DA, Darling N, Kennedy A, et al. Parents with doubts about vaccines: which vaccines and reasons why. *Pediatrics* 2008;122(4):718–25. doi:10.1542/peds.2007-0538.
- [4] Dubé E, Loberge C, Guay M, et al. Vaccine hesitancy: an overview. *Hum Vacc Immunother* 2013;9(8):1763–73. doi:10.4161/hv.24657.
- [5] Larson HJ, Jarrett C, Eckersberger E, et al. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. *Vaccine* 2014;32(19):2150–9. doi:10.1016/j.vaccine.2014.01.081.
- [6] Motta M, Callaghan T, Sylvester S. Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Sci Med* 2018;211:274–81.
- [7] Deiner MS, Fathy C, Kim J, et al. Facebook and twitter vaccine sentiment in response to measles outbreaks. *Health Inf J* 2019;25(3):1116–32. doi:10.1177/1460458217740723.
- [8] Abdelmutti N, Hoffman-Goetz L. Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil in north American news magazines. *J Cancer Educ* 2010;25(3):451–6. doi:10.1007/s13187-010-0087-9.
- [9] Pan S, Zhang D, Zhang J. Caught in the crossfire: how contradictory information and norms on social media influence young women's intentions to receive HPV vaccination in the United States and China. *Front Psychol* 2020;11:3469. doi:10.3389/fpsyg.2020.548365.
- [10] Henrich N, Holmes B. What the public was saying about the H1N1 vaccine: perceptions and issues discussed in on-line comments during the 2009 H1N1 pandemic. *PloS One* 2011;6(4):e18479. doi:10.1371/journal.pone.0018479.
- [11] Bedford H, Lansley M. More vaccines for children? Parents views. *Vaccine* 2007;25(45):7818–23. doi:10.1016/j.vaccine.2007.08.057.
- [12] Lazarus JV, Ratzan SC, Palayew A, et al. A global survey of potential acceptance of a COVID-19 vaccine. *Nat Med* 2020:1–4. doi:10.1038/s41591-020-1124-9.
- [13] Feleszko W, Lewulis P, Czarnecki A, et al. Flattening the curve of COVID-19 vaccine rejection—a global overview Available at SSRN; 2020.
- [14] Mokdad AH, Remington P. Measuring health behaviors in populations. *Prev Chronic Dis* 2010;7(4):A75.
- [15] Krumpal I. Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quantity* 2013;47(4):2025–47. doi:10.1007/s11135-011-9640-9.
- [16] Heikinheimo V, Minin ED, Tenkanen H, et al. User-generated geographic information for visitor monitoring in a national park: a comparison of social media data and visitor survey. *ISPRS Int J Geo-Inf* 2017;6(3):85. doi:10.3390/ijgi6030085.
- [17] Scanzfeld D, Scanzfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control* 2010;38(3):182–8. doi:10.1016/j.ajic.2009.11.004.
- [18] Singh L, Bansal S, Bode L, et al. A first look at COVID-19 information and misinformation sharing on twitter. arXiv:2003.13907v1.
- [19] Yeung N, Lai J, Luo J. Face off: polarized public opinions on personal face mask usage during the COVID-19 pandemic. In: *Proceedings of the IEEE international conference on big data*; 2020.
- [20] Stahl JP, Cohen R, Denis F, et al. The impact of the web and social networks on vaccination. New challenges and opportunities offered to fight against vaccine hesitancy. *Med Mal Infect* 2016;46(3):117–22. doi:10.1016/j.medmal.2016.02.002.
- [21] Wei W, Hanjia L, Jiebo L. Characterizing discourse about COVID-19 vaccines: a Reddit version of the pandemic story. arXiv preprint arXiv:2101.063212021.
- [22] Dunn AG, Surian D, Leask J, et al. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine* 2017;35(23):3033–40. doi:10.1016/j.vaccine.2017.04.060.
- [23] Brainard J, Hunter PR. Misinformation making a disease outbreak worse: outcomes compared for influenza, monkeypox, and norovirus. *Simulation* 2020;96(4):365–74. doi:10.1177/0037549719885021.
- [24] Tangcharoensathien V, Calleja N, Nguyen T, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced who technical consultation. *J Med Internet Res* 2020;22(6):e19659. doi:10.2196/19659.
- [25] Kim H, Han JY, Seo Y. Effects of facebook comments on attitude toward vaccines: the roles of perceived distributions of public opinion and perceived vaccine efficacy. *J Health Commun* 2020;25(2):159–69. doi:10.1080/10810730.2020.1723039.
- [26] Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on twitter, 2009–15. *Social Sci Med* 2017;191:168–75. doi:10.1016/j.socscimed.2017.08.041.
- [27] Tweepy API2021; Accessed July 21, <https://www.tweepy.org/>.
- [28] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis* 2020;20(5):533–4. doi:10.1016/S1473-3099(20)30120-1.
- [29] Lyu H, Chen L, Wang Y, et al. Sense and sensibility: characterizing social media users regarding the use of controversial terms for COVID-19. *IEEE Trans Big Data* 2020.
- [30] Face ++2021; Accessed July 21, <https://www.faceplusplus.com/>.
- [31] Jung S-G, An J, Kwak H, et al. Assessing the accuracy of four popular face recognition tools for inferring gender, age, and race Twelfth international AAAI conference on web and social media; 2018.
- [32] Preoțiu-Pietro D, Volkova S, Lamos V, et al. Studying user income through language, behaviour and affect in social media. *PloS one* 2015;10(9):e0138717. doi:10.1371/journal.pone.0138717.
- [33] Gilbert C, Hutto E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. 2014;81:82. Available from <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>.
- [34] Kochhar R. *The American middle class is stable in size, but losing ground financially to upper-income families*, vol 6. Pew Research Center; 2018.
- [35] Xiyang Z, Yu W, Hanjia L, et al. The influence of COVID-19 on the well-being of people: big data methods for capturing the well-being of working adults and protective factors nationwide. *Front Psychol* 2021;12:2327.
- [36] Sadilek A, Brennan SP, Kautz HA, et al. nemesis: Which restaurants should you avoid today? HCOMP. Citeseer; 2013.
- [37] Yang Z, Dai Z, Yang Y, et al. XLNet: generalized autoregressive pretraining for language understanding. In: *Proceedings of advances in neural information processing systems*; 2019. p. 5753–63.
- [38] Japkowicz N, Stephen S. The class imbalance problem: a systematic study. *Intell Data Anal* 2002;6(5):429–49.
- [39] Baziotis C, Pelekis N, Doukeridis C. DataStories at SemEval-2017 task 4: deep LSTM with attention for message-level and topic-based sentiment analysis. In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Vancouver, Canada: Association for Computational Linguistics; 2017:747–754.
- [40] Chang S, Pierson E, Koh PW, et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 2020:1–8.
- [41] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3(Jan):993–1022.
- [42] Lloyd-Sherlock P, Ebrahim S, Geffen L, et al. Bearing the brunt of COVID-19: older people in low and middle income countries. 2020.
- [43] Twitter. About verified accounts. Accessed July 21, 2021; <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>.
- [44] LaFraniere S, Thomas K, Weiland N, et al. Scientists worry about political influence over coronavirus vaccine project. *The New York Times*; 2020. Accessed from <https://www.nytimes.com/2020/08/02/us/politics/coronavirus-vaccine.html>.
- [45] Hopman J, Allegranzi B, Mehtar S. Managing COVID-19 in low-and middle-income countries. *Jama* 2020;323(16):1549–50.
- [46] Adams-Prassl A, Boneva T, Golin M, et al. Inequality in the impact of the coronavirus shock: new survey evidence for the UK2020.