

Original Paper

# Leveraging Transfer Learning to Analyze Opinions, Attitudes, and Behavioral Intentions Toward COVID-19 Vaccines: Social Media Content and Temporal Analysis

Siru Liu<sup>1</sup>, PhD; Jili Li<sup>2</sup>, BS; Jialin Liu<sup>3</sup>, MD

<sup>1</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>2</sup>West China Medical School, Sichuan University, Chengdu, China

<sup>3</sup>Department of Medical Informatics, West China Hospital, Sichuan University, Chengdu, China

**Corresponding Author:**

Jialin Liu, MD

Department of Medical Informatics

West China Hospital

Sichuan University

No 37 Wainan Guoxuexiang Street

Chengdu, 610041

China

Phone: 86 28 85422306

Email: [djlj18@163.com](mailto:djlj18@163.com)

## Abstract

**Background:** The COVID-19 vaccine is considered to be the most promising approach to alleviate the pandemic. However, in recent surveys, acceptance of the COVID-19 vaccine has been low. To design more effective outreach interventions, there is an urgent need to understand public perceptions of COVID-19 vaccines.

**Objective:** Our objective was to analyze the potential of leveraging transfer learning to detect tweets containing opinions, attitudes, and behavioral intentions toward COVID-19 vaccines, and to explore temporal trends as well as automatically extract topics across a large number of tweets.

**Methods:** We developed machine learning and transfer learning models to classify tweets, followed by temporal analysis and topic modeling on a dataset of COVID-19 vaccine-related tweets posted from November 1, 2020 to January 31, 2021. We used the F1 values as the primary outcome to compare the performance of machine learning and transfer learning models. The statistical values and *P* values from the Augmented Dickey-Fuller test were used to assess whether users' perceptions changed over time. The main topics in tweets were extracted by latent Dirichlet allocation analysis.

**Results:** We collected 2,678,372 tweets related to COVID-19 vaccines from 841,978 unique users and annotated 5000 tweets. The F1 values of transfer learning models were 0.792 (95% CI 0.789-0.795), 0.578 (95% CI 0.572-0.584), and 0.614 (95% CI 0.606-0.622) for these three tasks, which significantly outperformed the machine learning models (logistic regression, random forest, and support vector machine). The prevalence of tweets containing attitudes and behavioral intentions varied significantly over time. Specifically, tweets containing positive behavioral intentions increased significantly in December 2020. In addition, we selected tweets in the following categories: positive attitudes, negative attitudes, positive behavioral intentions, and negative behavioral intentions. We then identified 10 main topics and relevant terms for each category.

**Conclusions:** Overall, we provided a method to automatically analyze the public understanding of COVID-19 vaccines from real-time data in social media, which can be used to tailor educational programs and other interventions to effectively promote the public acceptance of COVID-19 vaccines.

(*J Med Internet Res* 2021;23(8):e30251) doi: [10.2196/30251](https://doi.org/10.2196/30251)

**KEYWORDS**

vaccine; COVID-19; leveraging transfer learning; pandemic; infodemiology; infoveillance; public health; social media; content analysis; machine learning; online health

## Introduction

The outbreak of COVID-19 has affected 219 countries and territories with 102,083,344 confirmed cases causing 2,209,195 deaths as of January 31, 2021, as reported by the World Health Organization (WHO) [1]. As a significant global health threat, long-term control of COVID-19 relies on the development and acceptance of a preventive vaccine [2-4]. Fortunately, in November 2020, Pfizer-BioNTech and Moderna reported more than 95% efficacy of their vaccines [5], which were subsequently authorized by the US Food and Drug Administration (FDA) for emergency use. Since the preventive vaccine has been successfully developed, the current barrier is obtaining a sufficient proportion of the population to accept vaccines to slow the spread of the outbreak [6]. However, according to a recent survey, only 51% of 10,093 adults in the United States indicated that they would be willing to receive the COVID-19 vaccine when it becomes available [7], which would not achieve the recommended threshold of 70% to reach herd immunity [8].

Vaccine hesitancy, defined as “a behavior with delay in acceptance or refusal of vaccines despite available services,” was identified by the WHO as a global threat in 2019 [9]. The SAGA Working Group developed the Vaccine Hesitancy Determinant Matrix, including contextual influences (ie, related to historic, sociocultural, environmental, institutional, economic, or political factors), individual and group influences (ie, factors related to personal perception or social environment), and vaccine/vaccination-specific issues [10]. Unlike other common vaccines, the COVID-19 vaccines are associated with many factors that might amplify vaccine hesitancy [11,12]. Previous studies have reported widespread public concern about the rapid speed of vaccine development, novelty of the development technology (mRNA), unknown long-term side effects, and politicization of vaccines [13,14]. Furthermore, the social environment is polarized, with distrust of science among some groups and a plethora of conspiracies and misinformation about vaccines spreading across social media platforms [15,16]. For these reasons, it might be more difficult to achieve the coverage goal for COVID-19 vaccines. Therefore, it is urgent to efficiently collect information on public perceptions to tailor education materials for public and clinical guidance, which will enable primary care physicians to promote COVID-19 vaccines.

With the increased growth of internet-based applications, more people have begun sharing their opinions on social media platforms. In particular, during the current COVID-19 pandemic, people may increase their use of social media due to social distancing [11]. Social media is awash with virus conspiracies and misinformation [15]. Various social media platforms (eg, Facebook, Instagram, Reddit) are currently providing health information to researchers; among them, the Twitter platform has a more prominent role in gathering public perceptions on health care [17]. Twitter has become a good data source to collect real-time perceptions from a large-scale population for public health research. Over the past decades, researchers have used social media analytics tools to monitor public sentiment and communication patterns in a global pandemic crisis (eg, Ebola and Zika outbreaks) [18-20]. Mavragani [21] performed

a time-series analysis on Google trends data and found a significant correlation between search interests with reported COVID-19 cases. Li et al [22] developed a taxonomy of Weibo posts on COVID-19 topics, and Liao et al [23] analyzed Weibo posts to identify public engagement and government responsiveness. Fadda et al [24] performed a content analysis to examine the extent of vaccine conspiracy theories reflected in tweets. Our study focused on the behavioral intentions related to COVID-19 vaccines, which is different from previous studies that performed a general analysis of COVID-19 tweets or vaccine conspiracy theories. Findings of this study could directly help researchers and policymakers to develop more targeted implementation strategies to improve acceptance rates of COVID-19 vaccines.

Machine learning and deep learning techniques have been used as efficient methods to detect public perceptions on social media platforms. In health care, researchers have developed deep learning models to perform longitudinal and geographic analyses to understand human papillomavirus (HPV) vaccine discussions [25]. These models also achieved good performance in predicting diagnosis or identifying patients in a high-risk group [26-28]. Transfer learning, as an emerging deep learning technique, has been applied to classify computed tomography images and notes. In transfer learning, a pretrained model is first used, which is then fine-tuned based on the specific datasets and tasks. Because the pretrained model already contains large-scale domain knowledge, the classification performance can achieve high values even with fine-tuning on relatively small datasets [29]. In this study, we applied Google's bidirectional encoder representations from transformers (BERT) model as the pretrained model, which has achieved new state-of-the-art results in the natural language processing domain [29].

Although previous studies have explored additional knowledge in the context of other vaccines using machine learning and deep learning methods, several questions related to COVID-19 vaccines remain unknown: What is the prevalence of user opinions on a social media platform? How many tweets express positive/negative attitudes and behavioral intentions to take vaccines? Which topics are mostly associated with these contents? To answer these questions, we developed machine learning models (logistic regression, random forest, support vector machine) and transfer learning models to detect the content expressing user opinions, attitudes, and behavioral intentions toward COVID-19 vaccines. We then performed a temporal analysis to explore trends over time and developed probabilistic topic models to obtain the most important and valuable topics. We believe that this study will be of great benefit to the timely rollout of COVID-19 vaccines by extracting the latest public opinions, attitudes, and behavioral intentions that can help tailor promotion programs to fit different populations.

## Methods

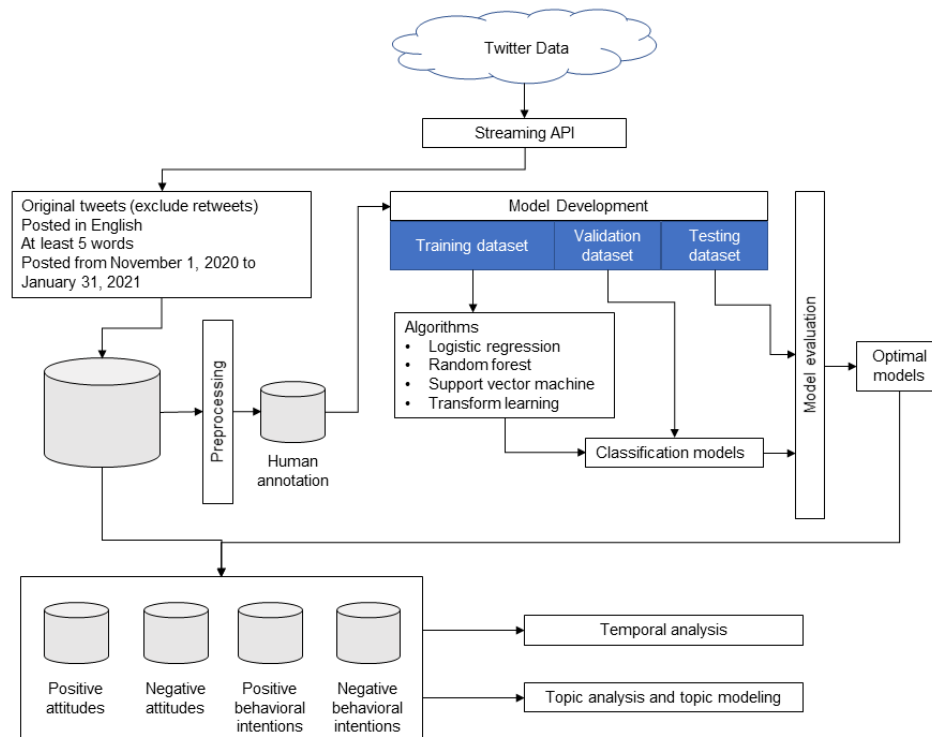
### Study Overview

We collected tweets related to COVID-19 vaccines posted from November 1, 2020 to January 31, 2021, and annotated 5000

tweets as the gold standard. We developed machine learning and transfer learning models to classify tweets for three tasks: (1) opinions (yes, no); (2) attitudes (positive, negative, neutral); and (3) behavioral intentions (positive, negative, unknown). The above tasks all focused on COVID-19 vaccines. We then applied the models to predict unlabeled tweets and performed a temporal analysis to capture trends in the unlabeled tweets.

In addition, we performed a topic analysis using word clouds and a latent Dirichlet allocation (LDA) model to further understand the content of tweets in the following categories: positive attitudes, negative attitudes, positive behavioral intentions, and negative behavioral intentions. The overall framework is shown in Figure 1.

**Figure 1.** Overall study framework. API: application programming interface.



**Data Collection**

We used a combination of keywords and hashtags related to COVID-19 vaccines to collect tweets in English published from November 1, 2020 to January 31, 2021. We intentionally chose November, following the announcement of the first effective vaccine on November 9, 2020, to determine if the announcement of successful vaccine trial results might influence the perceptions of vaccines or vaccination. The search strategy employed the following search terms: “(#covid OR covid OR #covid19 OR covid19) AND (#vaccine OR vaccine OR #vacine OR vacine OR vaccinate OR immunization OR immune OR vax) since:2020-11-01 until:2021-01-31 lang:en.” We used snsrape and tweepy in Python 3 to collect data and to exclude retweets. To clean up the original tweets, we removed nonalphanumeric characters and converted the text to lowercase. We randomly selected 5000 tweets from November 1, 2020 to November 22, 2020, annotated by two independent reviewers (SL and JL) in batches of 200. Any annotation disagreements were discussed and adjudicated by the supervising investigators. For each tweet, we first labeled whether it included a user opinion toward the COVID-19 vaccines (yes or no). We considered a tweet to include an opinion about the COVID-19 vaccines if it met both of the following conditions: (1) targeted at the COVID-19 vaccines and (2) generated by a user. For the tweets that expressed user opinions toward the COVID-19 vaccines, we

labeled the attitude (positive, negative, or neutral) and the behavioral intention (positive, negative, or unknown) toward COVID-19 vaccines. The attitude category used the traditional emotional polarity. The analysis of attitude was performed on the aspect level. If both positive and negative attitudes toward COVID-19 vaccines were present in the same tweet, we labeled it in the unknown category. The coding rules were iteratively developed by our group in which an independent review was performed, disagreements were discussed, and coding rules were revised. This process continued until the interrater agreement reached  $\geq 0.80$ . The annotated corpus was used as a gold standard to train and evaluate the machine learning and transfer learning models.

**Model Development and Evaluation**

For data preprocessing, we used the tweet-preprocessor package in Python 3 to remove URLs, hashtags, mentions, reserved words (eg, RT, FAV), emojis, smileys, and numbers in each tweet. We split the annotated dataset into three parts: training (60%), validation (20%), and testing (20%). The training and validation datasets were used to train models and select optimal hyperparameters through 5-fold cross-validation. We applied transfer learning using text frequency-inverse document frequency to compare traditional machine learning algorithms (logistic regression, random forest, and support vector machine)

to transfer learning models. The machine learning models were developed using the scikit-learn package in Python 3.

For transfer learning, we used the BERT-base-cased as the pretrained language model and the “BERT for sequence classification” model as the pretrained classification model. Because the BERT model requires each sentence to be the same length, we padded each tweet with 64 tokens, as most tweets have lengths in this range. We then fine-tuned this model on the training and validation datasets using the Adam algorithm with weight decay (AdamW) as an optimizer. We performed three text classification tasks. We first developed a binary classifier to determine whether the tweets state an opinion related to the COVID-19 vaccines. We then developed two multiclass classifiers to categorize attitudes and behavioral intentions, respectively. The BERT models were generated using the huggingface package in Python 3. The models were developed with the Google Colab platform using a high-RAM GPU.

We evaluated the models on the testing dataset and report outcomes with 1000 rounds of bootstrapping. The primary outcome was the macro-F1 value and the secondary outcomes were recall, precision, and accuracy. We performed the Nemenyi test to compare the F1 values of traditional machine learning models and transfer learning models [30]. The model with the highest F1 value was considered the optimal model.

### Temporal Analysis

We applied the optimal models to predict the unlabeled data for 3 months starting from November 1, 2020. For the task of extracting opinions, we calculated the proportion of tweets classified as containing opinions to the total number of tweets posted each day about the COVID-19 vaccines. For the tasks of classifying attitudes and behavioral intentions toward the COVID-19 vaccines, we calculated the percentage of tweets predicted to exhibit a particular attitude or behavioral intention

to all tweets indicating attitudes or behavioral intentions, respectively. To assess the statistical significance of variability over time, we performed the Augmented Dickey-Fuller (ADF) test [31] with a significance threshold of  $P < .05$ . The ADF test is a unit root test, which is commonly used to determine the stationarity of a time-series sample.

### Topic Analysis and Topic Modeling

To understand the content of tweets in each category, we used word clouds to illustrate the frequency of words appearing in the content. The more frequently used words have larger sizes, indicating more importance in the category [32]. Furthermore, we performed the LDA analysis to extract the main topics of discussion. LDA is a widely used unsupervised method that automatically clusters text based on content and identifies keywords in each topic through a probabilistic model [33,34]. We performed 5-fold cross-validation to tune hyperparameters in the LDA model (number of components and learning rate). After obtaining the results of the LDA models, we visualized extracted topics using the pyLDAvis library [35] in Python 3, which is an interactive visualization tool for displaying the distribution of topics and the top 30 most relevant terms with their weights in each topic.

## Results

### Performance of Classification Models

We annotated 5000 tweets from 4796 unique users with an average interrater reliability ( $\kappa$ ) of 0.76. The prediction performances of models on the testing dataset using four different algorithms for three tasks are presented in Table 1. The transfer learning model significantly outperformed the machine learning models in identifying tweets that included opinions, attitudes, and behavioral intentions, achieving the highest F1 values of 0.792, 0.578, and 0.614, respectively.

**Table 1.** Metrics of transfer learning models and machine learning models in classifying tweets related to COVID-19 vaccines.

Task	Recall, mean (95% CI)	Precision, mean (95% CI)	F1, mean (95% CI)	Accuracy, mean (95% CI)
<b>Opinions</b>				
BERT <sup>a</sup>	0.762 (0.759-0.766)	0.862 (0.858-0.866)	0.792 <sup>b</sup> (0.789-0.795)	0.854 (0.852-0.856)
Logistic regression	0.774 (0.770-0.779)	0.757 (0.753-0.762)	0.764 (0.761-0.767)	0.807 (0.805-0.810)
Random forest	0.754 (0.750-0.758)	0.732 (0.728-0.735)	0.740 (0.737-0.743)	0.783 (0.781-0.786)
Support vector machine	0.767 (0.764-0.771)	0.752 (0.748-0.755)	0.758 (0.755-0.761)	0.803 (0.801-0.806)
<b>Attitudes</b>				
BERT	0.529 (0.521-0.536)	0.698 (0.686-0.710)	0.578 <sup>b</sup> (0.572-0.584)	0.873 (0.871-0.875)
Logistic regression	0.475 (0.468-0.482)	0.530 (0.520-0.541)	0.495 (0.490-0.500)	0.859 (0.856-0.861)
Random forest	0.518 (0.511-0.526)	0.558 (0.545-0.570)	0.508 (0.502-0.514)	0.830 (0.827-0.833)
Support vector machine	0.506 (0.498-0.514)	0.551 (0.541-0.562)	0.523 (0.517-0.530)	0.863 (0.860-0.865)
<b>Behavioral intentions</b>				
BERT	0.562 (0.549-0.575)	0.734 (0.716-0.752)	0.614 <sup>b</sup> (0.606-0.622)	0.961 (0.960-0.962)
Logistic regression	0.472 (0.461-0.483)	0.725 (0.699-0.752)	0.527 (0.519-0.536)	0.951 (0.949-0.952)
Random forest	0.447 (0.437-0.457)	0.577 (0.543-0.611]	0.466 (0.457-0.476)	0.935 (0.934-0.937)
Support vector machine	0.469 (0.458-0.479)	0.710 (0.684-0.737)	0.523 (0.513-0.533)	0.950 (0.948-0.951)

<sup>a</sup>BERT: Bidirectional encoder representations from transformers.

<sup>b</sup> $P=.001$  in the Nemenyi test.

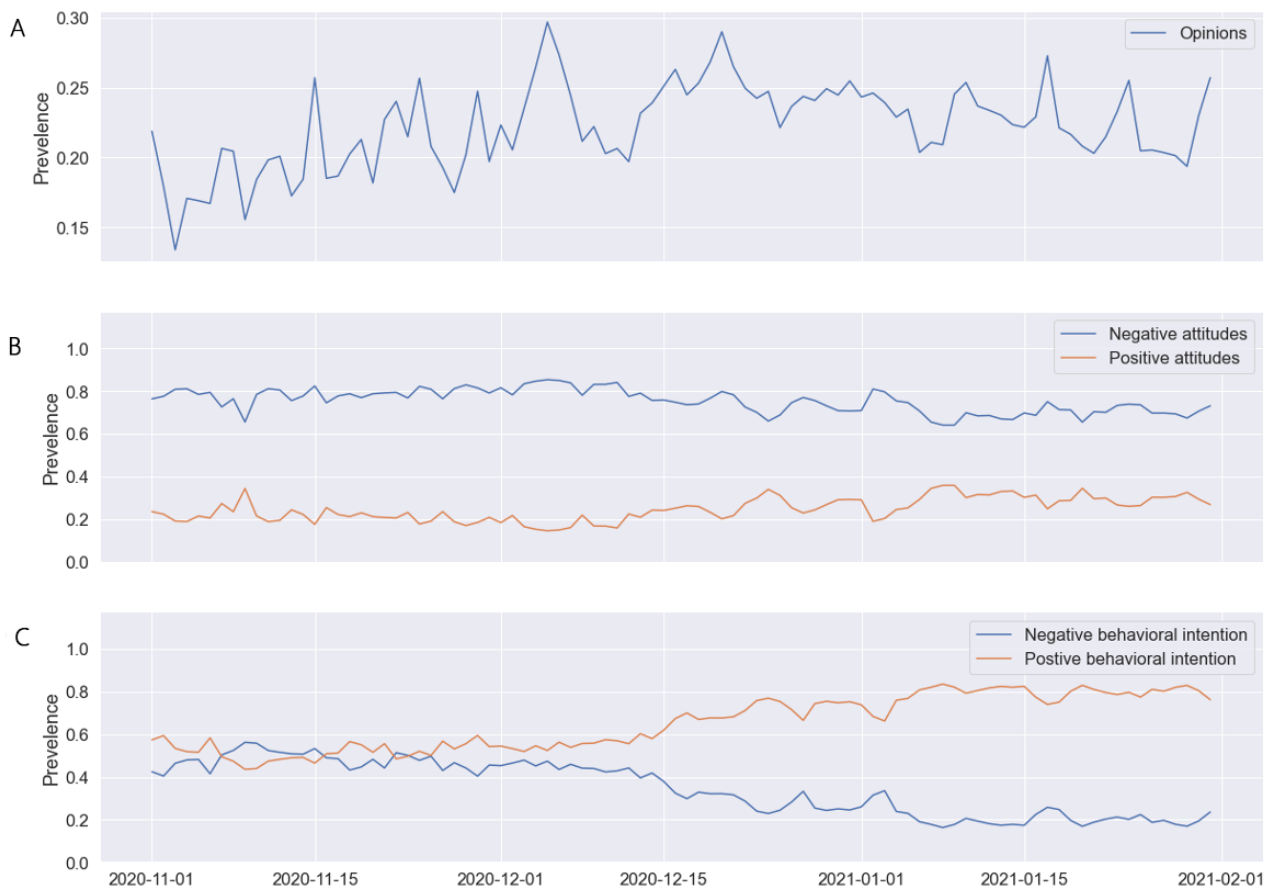
## Temporal Analysis

We collected 2,678,372 tweets related to COVID-19 vaccines posted by 841,978 unique users from November 1, 2020 to January 31, 2021. The daily prevalence distributions of opinions, attitudes, and behavioral intentions are shown in Figure 2. The daily prevalence of tweets expressing users' opinions was 0.222 (95% CI 0.202-0.245). The ADF statistic was  $-4.341$  ( $P<.001$ ), indicating that the time-series data were stationary. This reflects that the prevalence of tweets expressing opinions did not change significantly over time. For tweets containing attitudes toward the COVID-19 vaccines, the rate of negative attitudes was 0.754 (95% CI 0.707-0.795), while the rate of positive attitudes was

only 0.246 (95% CI 0.204-0.293). The daily prevalence of attitudes was nonstationary (ADF  $-1.137$ ,  $P=.70$ ), which indicated a significant change in users' attitudes toward vaccines over time. Among tweets related to behavioral intentions, the rate of tweets indicating that users will not get vaccinated was 0.342 (95% CI 0.229-0.461), whereas the rate of tweets indicating that users will get vaccinated was 0.652 (95% CI 0.539-0.771). The behavioral intention prevalence was also nonstationary (ADF  $-0.980$ ,  $P=.76$ ), indicating that it varied significantly over time. Notably, we observed a substantial increase in the prevalence of tweets expressing positive behavioral intention starting from mid-December 2020.



**Figure 2.** Distribution of the prevalence of the tweets containing opinions (A), attitudes (B), and behavioral intentions (C) about COVID-19 vaccines for each day from November 1, 2020 to January 31, 2021.



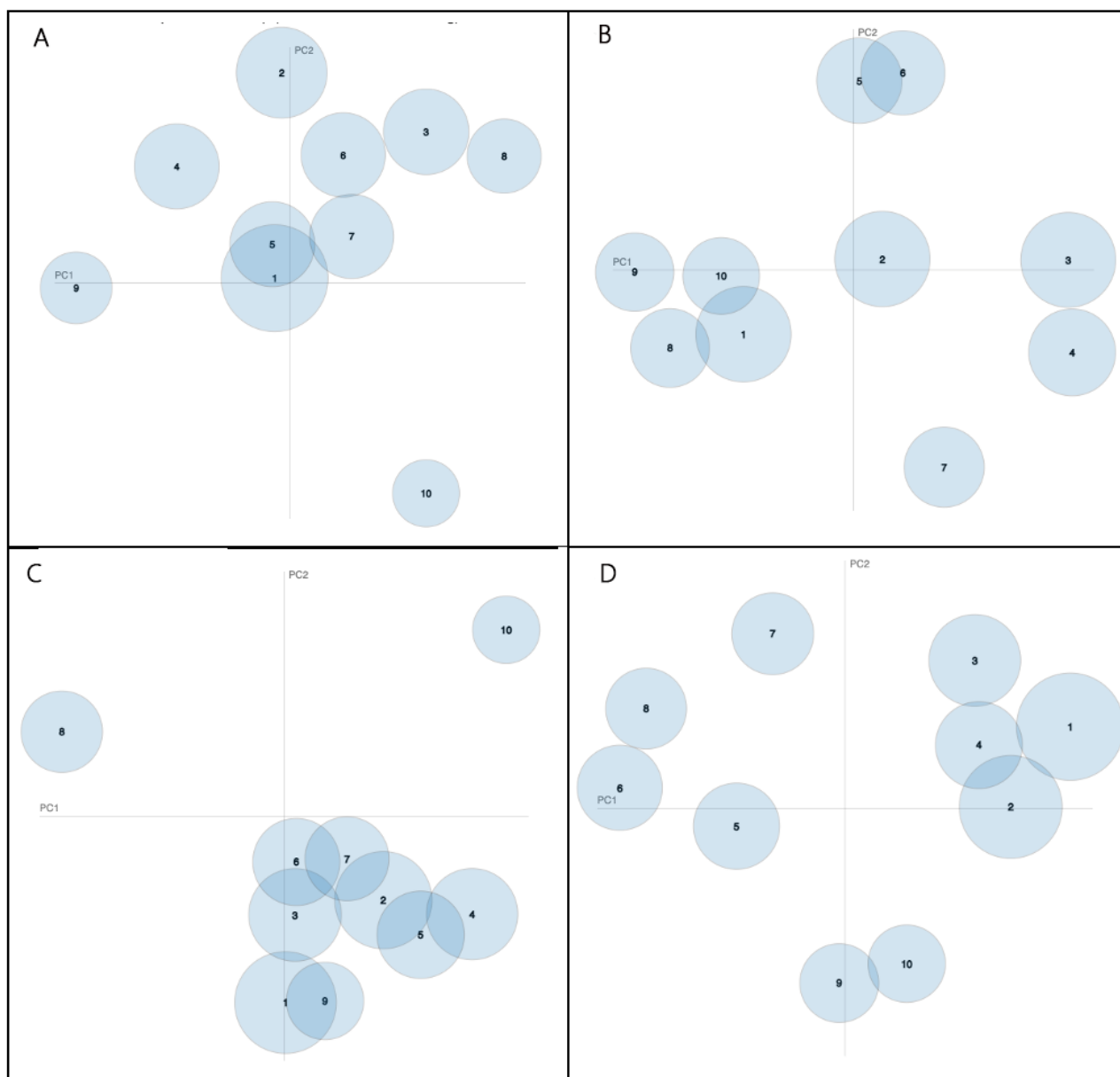
**Topic Modeling and Analysis**

**Primary Domain Topics**

After tuning hyperparameters of the LDA models, each model had 10 components (topics). Figure 3 presents intertopic distance maps generated by tweets containing positive/negative attitudes

and positive/negative behavioral intentions. The size of bubbles represents the ratio of relevant tweets in that topic to the total number of tweets. In the following sections, we selected several domain topics for tweets in each category to describe the potential inferred themes based on identified relevant key terms. The overall top 15 keywords for each topic are listed in Textbox 1.

**Figure 3.** Intertopic distance maps for tweets that contained information in the following categories: negative attitudes (A), positive attitudes (B), negative behavioral intentions (C), and positive behavioral intentions (D).



**Textbox 1.** Top 15 key terms for each topic.

**Negative attitudes**

1. worry, prevent, covid, stop, need, spread, symptom, transmission, catch, people, reduce, infection, virus, eat, doesn
2. death, covid, case, people, rate, die, number, cause, population, test, trial, fear, report, survival, day
3. risk, covid, test, people, health, worker, trial, know, need, woman, work, child, pregnant, safe, age
4. effect, long, term, know, covid, bad, unknown, risk, affect, people, study, concern, damage, potential, impact
5. covid, year, make, anti, month, mask, rush, people, want, safe, need, just, know, sense, wear
6. covid, dose, use, virus, immune, antibody, body, immunity, trial, second, make, protein, cell, test, response
7. virus, new, covid, strain, effective, work, mutate, year, develop, mutation, research, cold, variant, different, make
8. covid, people, just, say, think, make, know, trust, want, cure, government, believe, thing, come
9. covid, die, people, life, chance, treatment, old, kill, effective, want, say, sick, save, safe, family
10. flu, covid, reaction, shot, drug, adverse, expect, people, shoot, allergic, just, high, bad, year, polio

**Positive attitudes**

1. covid, thank, work, great, today, day, make, worker, scientist, happy, mom, care, just, hard, nurse
2. covid, feel, effect, day, long, arm, just, little, work, fine, hour, term, good, excited, sore
3. safe, stay, end, covid, news, pandemic, effective, trial, good, amp, light, continue, home, hope, step
4. covid, hope, soon, look, forward, normal, life, hopefully, come, available, new, world, news, return, year
5. covid, good, year, just, time, wait, thing, hope, think, come, pray, love, wish, news, day
6. people, covid, want, need, know, die, risk, just, really, say, think, make, life, safe, fear
7. covid, dose, receive, today, grateful, second, family, feel, patient, able, thankful, protect, friend, happy, excited
8. flu, virus, covid, make, immune, fight, sure, body, new, immunity, just, strain, world, distribute, cause
9. mask, wear, covid, stop, social, spread, distancing, hand, catch, need, distance, people, virus, stay, help
10. covid, vaccinate, amp, case, symptom, prevent, ready, immunity, just, mean, virus, reduce, life, rate, infection

**Negative behavioral intentions**

1. covid, virus, stop, prevent, symptom, test, dose, immune, spread, mask, antibody, sick, just, catch, body
2. covid, flu, shot, shit, shoot, just, allow, work, win, scare, dead, year, virus, arm, sure
3. risk, covid, say, immune, make, high, virus, people, disease, just, healthy, sense, dangerous, case, good
4. want, covid, vaccinate, child, use, kill, kid, new, wait, way, cure, effective, doctor, just, people
5. covid, body, rate, vaccination, survival, choice, eat, mandatory, know, worry, life, fear, want, hear, need
6. covid, anti, just, tell, say, refuse, vaxxer, afraid, reason, people, stop, right, make, job, stupid
7. covid, year, trust, chance, inject, month, government, test, old, develop, cold, make, research, come
8. effect, know, long, term, covid, dna, affect, change, people, bad, rush, chance, unknown, study, test
9. people, covid, die, need, think, just, kill, family, care, damn, believe, say, real, death, chance
10. covid, force, try, reaction, people, bad, severe, look, allergic, medical, receive, say, fine, pay

**Positive behavioral intentions**

1. covid, people, want, just, think, say, know, mask, wear, make, really, ask, scare, right
2. covid, want, need, look, tomorrow, let, know, life, forward, ready, dose, morning, normal, receive, volunteer
3. covid, wait, long, turn, effect, line, term, finally, eat, excited, worried, afraid, use, drink, polio
4. just, dose, covid, second, got, day, effect, symptom, receive, fever, ache, hour, experience, headache, body
5. flu, covid, shot, year, time, bad, shoot, sick, immune, just, need, month, make, think, doctor
6. covid, arm, sore, sign, just, feel, today, hour, little, hurt, yesterday, injection, far, nervous, appointment
7. work, covid, home, thank, stay, patient, hospital, help, safe, care, protect, family, trial, receive, vaccinate
8. covid, risk, immune, die, people, virus, chance, high, know, need, vaccinate, healthy, live, just, catch
9. covid, today, hope, mom, test, dose, happy, able, soon, dad, positive, receive, good, grateful



10. feel, covid, day, week, fine, great, make, shit, ago, better, worker, body, job, good, healthcare

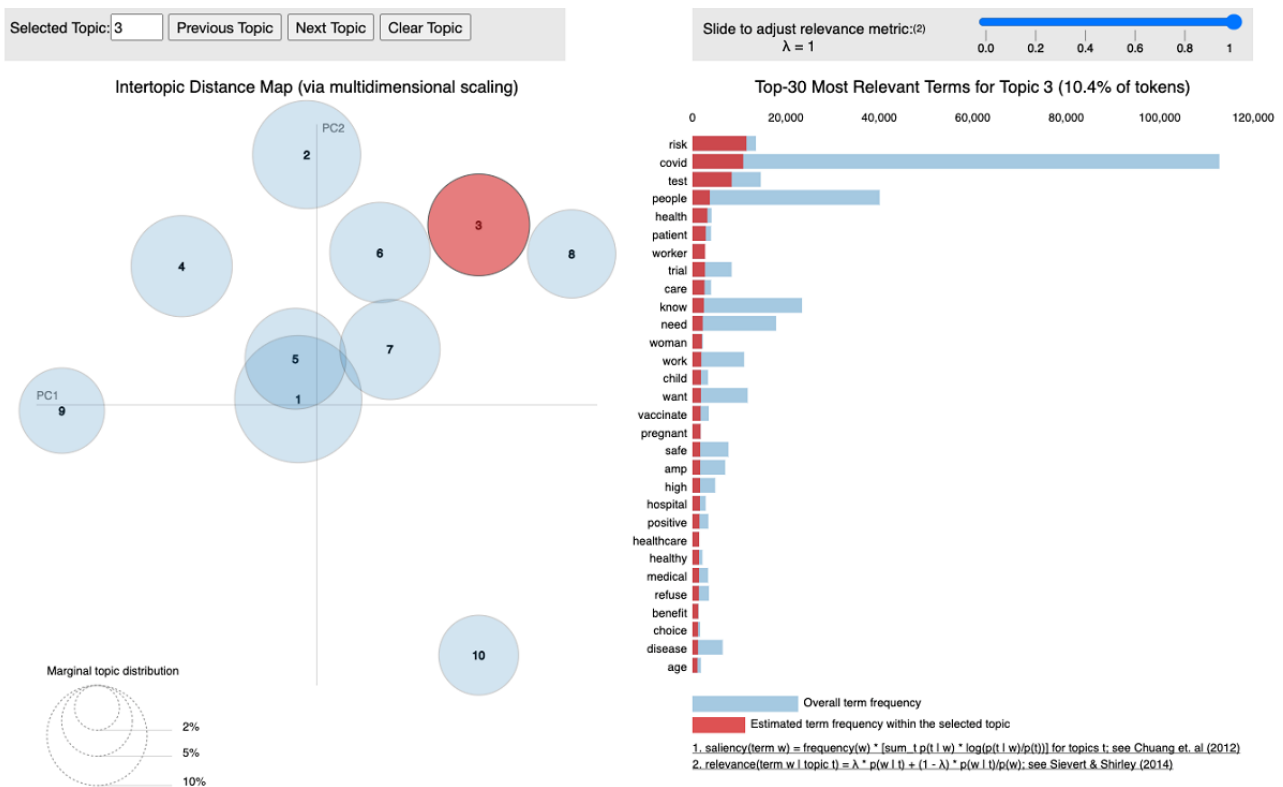
**Attitudes**

Ten topics were extracted among the tweets that contained negative attitudes. The interactive display interface of pyLDavis is shown in Figure 4. The left panel shows the distribution of topics, and we could choose the topic we wanted to analyze by clicking on the bubble (eg, topic 3 highlighted in Figure 4), while the right panel lists the top 30 relevant terms and their weights contributing to the selected topic. Some important keywords contained in topic 3 were “risk,” “test,” “child,” “safe,” “pregnant,” “disease,” and “age.” Topic 3 summarized that users with negative attitudes were concerned about the safety issues of the COVID-19 vaccines, especially about the risks for certain populations such as children, pregnant women, and patients with immune diseases. Other topics reflected concerns about unknown side effects (topic 4) and rushing the

development process (topic 5). Some users even questioned the existence of COVID-19 or COVID-19 vaccines and indicated a lack of trust in the government or scientists (topic 8). In addition, some users feared that the virus mutation would render the vaccine ineffective (topic 7) and thus had negative attitudes toward vaccines.

For tweets containing positive attitudes, in a dominant topic (topic 3), relevant key terms included “safe,” “stay,” “end,” “pandemic,” “news,” “effective,” “trial,” “continue,” and “hope.” This indicates that some positive attitudes might be derived from news of effective trial results and some users hoped that COVID-19 vaccines could end the pandemic. Relevant terms for topic 4 were “hope,” “normal,” “life,” “return,” “start,” “new,” “world,” and “great.” Tweets in topic 4 showed that some users expressed positive attitudes toward vaccines because of the desire to return to a normal life.

**Figure 4.** PyLDavis visualization highlighting the top 30 relevant keywords for a topic found in the tweets that contained negative attitudes toward COVID-19 vaccines.



**Behavioral Intentions**

For tweets containing negative behavioral intentions, topics 8 and 10 clustered independently; however, other topics showed some degree of mutual inclusiveness, indicating that similarities existed in those topics. Key terms for topic 8 were “effect,” “know,” “long,” “term,” “DNA,” “unknown,” and “rush.” This topic reflected that some users’ negative behavioral intentions came from the concerns of the long-term and unknown side effects of COVID-19 vaccines. As another unique topic, the most relevant terms for topic 10 were “force,” “reaction,” “bad,”

“allergic,” “pay,” “adverse,” and “government.” This analysis highlighted that some users mentioned that they would not take the vaccine if it was forced on them by the government. Others worried about the adverse reactions to the COVID-19 vaccines. Some users compared COVID-19 to influenza and mentioned that because they had not previously been vaccinated against influenza, there was also no need to vaccinate against a disease they mistakenly thought had the same low lethality (topic 2). Other users reported that their immune system could naturally help them fight the virus.

For tweets containing positive behavioral intentions, mutual inclusivity existed among topics 1-4 and between topics 9 and 10. Other topics clustered independently. In topic 8, the keywords were “risk,” “immune,” “healthy,” “antibody,” and “immunity.” In this topic, users would like to become immune to the virus causing COVID-19 and stay healthy by being vaccinated.

## Discussion

### Principal Findings

In this study, we provided an annotated dataset with 5000 COVID-19 vaccine-related tweets with labels supporting three classification tasks (opinions, attitudes, and behavioral intentions). We assessed that transfer learning could be used to analyze COVID-19 vaccine content tweets and proved that they outperformed common machine learning models. We analyzed the temporal trends and topics in the COVID-19 vaccine-related tweets posted over a 3-month period (from November 1, 2020 to January 31, 2021). The prevalence of tweets containing positive behavioral intentions increased over time. The word clouds and the LDA analysis proved to be efficient tools to understand topics for tweets in each category.

Transfer learning is now widely used to analyze social media content. Some researchers have applied transfer learning with datasets of tweets related to COVID-19 [36-38] rather than focusing on tweets related to the vaccines developed for this disease. Researchers have analyzed tweets related to other vaccines such as HPV vaccines [25]. However, few studies have annotated tweets containing content about COVID-19 vaccines or developed models to understand public perceptions on COVID-19 vaccines from social media. For example, Levy et al [36] applied cross-lingual transfer learning to model COVID-19 outbreak patterns in one country, and then utilized the model to predict the spread of the disease in another country with a strong Spearman correlation (0.850). A classification model based on transfer learning developed by Spangher et al [37] was able to categorize policy announcements of COVID-19 using event extraction, with an F1 score of 0.770. To identify informative tweets related to COVID-19, Tasneem et al [38] proposed a unified architecture to combine transfer learning with hand-crafted features, achieving an F1 score of 0.820. Du et al [25] used deep learning models to categorize HPV vaccine-related tweets with constructs in the health belief model and theory of planned behavior models, and obtained F1 scores ranging from 0.681 to 0.942. Our study is the first to apply transfer learning models to analyze the public's attitudes and behavioral intentions toward COVID-19 vaccines. Our model also achieved good performance, with F1 scores ranging from 0.579 to 0.792. In addition, we provided an annotated dataset with 5000 tweets, each labeled according to whether the tweet contained users' opinions, attitudes, or behavioral intentions on COVID-19 vaccines. This dataset can be used for further research on social media content related to the COVID-19 vaccines.

Several researchers have applied the Valence Aware Dictionary and Sentiment Reasoner (VADER) tool [39,40], machine learning [41], and deep learning [42] to perform sentiment

analysis on COVID-19-related tweets. Chandrasekaran et al [39] and Yin et al [40] employed the VADER tool to calculate the polarity of sentiment in COVID-19-related tweets posted in the first half of 2020. Both of these studies reported that the proportion of positive tweets was higher than that of negative tweets in general. However, Chandrasekaran et al [39] determined that negative tweets were dominant in the themes of symptoms and spread in cases. Li et al [42] used deep learning to identify fear and sadness emotions mentioned in COVID-19-related tweets to analyze the public's mental health status, and reported area under the receiver operating characteristic curve values ranging from 0.681 to 0.739. Chakraborty et al [41] developed machine learning models with Gaussian membership function-based fuzzy rules to classify sentiment in COVID-19-related tweets, obtaining accuracy values ranging from 0.526 to 0.814. Although these previous studies have classified sentiment in COVID-19-related tweets, our study differs with respect to the task of classifying attitudes toward COVID-19 vaccines. We not only focused on the sentiment of tweets but also simultaneously examined whether the object of the sentiment was the COVID-19 vaccine. During annotation, we noticed that some tweets contained positive words used to describe what would happen after the vaccine rollout but also stated negative attitudes toward the vaccine itself, such as lack of trust and rushing.

Temporal analysis and topic modeling provide an efficient approach to monitor public perceptions of the COVID-19 vaccines on social media platforms. The following events could explain the significant increase in the prevalence of positive behavioral intentions in mid-December. For example, the FDA issued Pfizer-BioNTech COVID-19 vaccines on December 11, 2020, turning the vaccines from a hypothetical situation into a reality. The United States launched its rollout to high-risk health care facilities on December 14, 2020. A large number of health care workers and influential figures such as Joe Biden received COVID-19 vaccines to increase public confidence. This also suggests that more people might be willing to be vaccinated after successful vaccine development and a large-scale rollout. Indeed, social influence has been shown to positively affect the acceptance rate [43]. At the same time, this increase in positive behavioral intentions could also generate a positive social influence, which could lead to a higher vaccine acceptance rate. Therefore, the low acceptance rate of COVID-19 vaccines reported in the surveys conducted prior to December 2020 might not accurately reflect the current situation. Researchers should consider resurveying the public's intention to receive the vaccination. Key terms identified in topic modeling could provide the needed guidance to design or optimize vaccine promotion interventions (eg, education materials). COVID-19 vaccine promotion strategies need to solve concerns on side effects and long-term safety issues, virus mutation, and the difference between COVID-19 and the flu. Moreover, promotion strategies should highlight the chance to return to normal life and stay healthy after being vaccinated for COVID-19.

### Limitations

This study has several limitations. First, users of the Twitter platform are not representative of the entire public. The Twitter platform is usually considered to gather more antivaccinators

and spread misinformation. This group of users is the main subgroup of the population with sentiments of vaccine hesitancy and should therefore be one of the main targets to receive vaccine education. Compared to other populations, they tend to question vaccines from specific perspectives such as the presence of microchips in vaccines [44] and the use of human embryos in the process of developing vaccines. Understanding their perceptions is a necessary step to tailor vaccine promotion education materials, which would provide a better chance of effectively changing their behavior. Second, some topics extracted from the topic modeling might be difficult to infer accurately using relevant terms. In addition, given the complex situation of behavioral intentions toward COVID-19 vaccines mentioned in the tweets, further qualitative studies (eg, content analysis) combined with theoretical models are needed to understand why some people will not get the vaccines from psychological aspects. Third, we applied the “BERT-base-uncased” as the language model. Recently, researchers have developed a transformer-based model COVID-Twitter-BERT (CT-BERT) model, which was pretrained on COVID-19–related tweets [45], and they also expected to obtain performance gains when applying the CT-BERT model on health care content tweets. The impact of using the CT-BERT model on our classification tasks is unknown. Fourth, the annotated corpus included 5000 tweets. If more annotated data could be collected, the performance of the model might be improved.

For future work, we will perform a theory-based content analysis to gain insight into the reasons that led to the changes in

behavioral intentions we noted in the temporal analysis. Using the transfer learning model in this study, researchers can automatically collect tweets containing COVID-19 vaccine–related behavioral intentions and systematically analyze the data through a theoretical model (eg, Capability, Opportunity, Motivation, Behavior model [12,46]) to promote timely promotion strategies. In addition, researchers can extract individual characteristics from the user profile and perform statistical analysis to determine the relationship between individual characteristics and their behavioral intention toward COVID-19 vaccines.

## Conclusion

In this study, we presented an annotated corpus of 5000 tweets and analyzed the potential to use transfer learning with a pretrained BERT model to automatically identify public opinions, behavioral intentions, and attitudes toward COVID-19 vaccines from social media. We demonstrated that transfer learning models outperformed traditional machine learning models in general. In addition, we explored the temporal trends of the public’s change in attitudes and behavioral intentions on a larger dataset with 2,678,372 tweets from November 1, 2020 to January 31, 2021. We found that the LDA technique is useful to extract topics from identified tweets. Overall, we provided an automatic method to analyze the public’s understanding of COVID-19 vaccines from real-time data, which could be used to tailor education programs and other interventions to promote COVID-19 vaccine acceptance urgently.

## Acknowledgments

This work was supported by Sichuan Science and Technology Program (grant number 2020YFS0162).

## Authors' Contributions

JLiu and SL conceived the study. SL, JLiu, and JLi performed the analysis, interpreted the results, and drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

## Conflicts of Interest

None declared.

## References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: <https://covid19.who.int/> [accessed 2021-01-31]
2. Chou WS, Budenz A. Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. *Health Commun* 2020 Dec;35(14):1718-1722. [doi: [10.1080/10410236.2020.1838096](https://doi.org/10.1080/10410236.2020.1838096)] [Medline: [33124475](https://pubmed.ncbi.nlm.nih.gov/33124475/)]
3. Liu J, Liu S. The management of coronavirus disease 2019 (COVID-19). *J Med Virol* 2020 Sep;92(9):1484-1490 [FREE Full text] [doi: [10.1002/jmv.25965](https://doi.org/10.1002/jmv.25965)] [Medline: [32369222](https://pubmed.ncbi.nlm.nih.gov/32369222/)]
4. Liu J, Liu S, Wei H, Yang X. Epidemiology, clinical characteristics of the first cases of COVID-19. *Eur J Clin Invest* 2020 Oct;50(10):e13364. [doi: [10.1111/eci.13364](https://doi.org/10.1111/eci.13364)] [Medline: [32725884](https://pubmed.ncbi.nlm.nih.gov/32725884/)]
5. Mahase E. Covid-19: What do we know about the late stage vaccine candidates? *BMJ* 2020 Nov 24;371:m4576. [doi: [10.1136/bmj.m4576](https://doi.org/10.1136/bmj.m4576)] [Medline: [33234507](https://pubmed.ncbi.nlm.nih.gov/33234507/)]
6. Kreps S, Prasad S, Brownstein JS, Hswen Y, Garibaldi BT, Zhang B, et al. Factors associated with US adults' likelihood of accepting COVID-19 vaccination. *JAMA Netw Open* 2020 Oct 01;3(10):e2025594 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.25594](https://doi.org/10.1001/jamanetworkopen.2020.25594)] [Medline: [33079199](https://pubmed.ncbi.nlm.nih.gov/33079199/)]
7. Chodosh S. Why only half of Americans say they would get a COVID-19 vaccine. *Popular Science*. 2020 Jun 01. URL: <https://www.popsoci.com/story/health/covid-19-vaccine-poll/> [accessed 2021-01-31]

8. D'souza G, Dowdy D. What is herd immunity and how can we achieve it with COVID-19? Johns Hopkins Bloomberg School of Public Health. 2020. URL: <https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html> [accessed 2021-01-31]
9. Kwok KO, Li K, Wei WI, Tang A, Wong SYS, Lee SS. Editor's Choice: Influenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: A survey. *Int J Nurs Stud* 2021 Feb;114:103854 [FREE Full text] [doi: [10.1016/j.ijnurstu.2020.103854](https://doi.org/10.1016/j.ijnurstu.2020.103854)] [Medline: [33326864](https://pubmed.ncbi.nlm.nih.gov/33326864/)]
10. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. *Vaccine* 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: [10.1016/j.vaccine.2015.04.036](https://doi.org/10.1016/j.vaccine.2015.04.036)] [Medline: [25896383](https://pubmed.ncbi.nlm.nih.gov/25896383/)]
11. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Hum Vaccin Immunother* 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: [10.1080/21645515.2020.1780846](https://doi.org/10.1080/21645515.2020.1780846)] [Medline: [32693678](https://pubmed.ncbi.nlm.nih.gov/32693678/)]
12. Liu S, Liu J. Understanding behavioral intentions toward COVID-19 vaccines: theory-based content analysis of tweets. *J Med Internet Res* 2021 May 12;23(5):e28118 [FREE Full text] [doi: [10.2196/28118](https://doi.org/10.2196/28118)] [Medline: [33939625](https://pubmed.ncbi.nlm.nih.gov/33939625/)]
13. Tyson A, Johnson C, Funk C. U.S. public now divided over whether to get COVID-19 vaccine. Pew Research Center. 2020. URL: <https://www.pewresearch.org/science/> [accessed 2021-01-31]
14. COCONEL Group. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. *Lancet Infect Dis* 2020 Jul;20(7):769-770 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30426-6](https://doi.org/10.1016/S1473-3099(20)30426-6)] [Medline: [32445713](https://pubmed.ncbi.nlm.nih.gov/32445713/)]
15. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. *Cureus* 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: [10.7759/cureus.7255](https://doi.org/10.7759/cureus.7255)] [Medline: [32292669](https://pubmed.ncbi.nlm.nih.gov/32292669/)]
16. Brennen J, Simon F, Howard P, Nielsen R. Types, sources, and claims of COVID-19 misinformation key findings. Reuters Institute, University of Oxford. URL: <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation> [accessed 2021-01-30]
17. Chan TM, Dzara K, Dimeo SP, Bhalerao A, Maggio LA. Social media in knowledge translation and education for physicians and trainees: a scoping review. *Perspect Med Educ* 2020 Feb;9(1):20-30 [FREE Full text] [doi: [10.1007/s40037-019-00542-7](https://doi.org/10.1007/s40037-019-00542-7)] [Medline: [31834598](https://pubmed.ncbi.nlm.nih.gov/31834598/)]
18. Lwin MO, Lu J, Sheldenkar A, Schulz PJ. Strategic uses of Facebook in Zika outbreak communication: implications for the crisis and emergency risk communication model. *Int J Environ Res Public Health* 2018 Sep 10;15(9):1974 [FREE Full text] [doi: [10.3390/ijerph15091974](https://doi.org/10.3390/ijerph15091974)] [Medline: [30201929](https://pubmed.ncbi.nlm.nih.gov/30201929/)]
19. Zhang EX, Yang Y, Di Shang R, Simons JJP, Quek BK, Yin XF, et al. Leveraging social networking sites for disease surveillance and public sensing: the case of the 2013 avian influenza A(H7N9) outbreak in China. *Western Pac Surveill Response J* 2015;6(2):66-72 [FREE Full text] [doi: [10.5365/WPSAR.2015.6.1.013](https://doi.org/10.5365/WPSAR.2015.6.1.013)] [Medline: [26306219](https://pubmed.ncbi.nlm.nih.gov/26306219/)]
20. Mourad A, Srour A, Harmanani H, Jenainati C, Arafeh M. Critical impact of social networks infodemic on defeating coronavirus COVID-19 pandemic: Twitter-based study and research directions. *IEEE Trans Netw Serv Manage* 2020 Dec 01;17(4):2145-2155. [doi: [10.1109/tnsm.2020.3031034](https://doi.org/10.1109/tnsm.2020.3031034)]
21. Mavragani A. Tracking COVID-19 in Europe: infodemiology approach. *JMIR Public Health Surveill* 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: [10.2196/18941](https://doi.org/10.2196/18941)] [Medline: [32250957](https://pubmed.ncbi.nlm.nih.gov/32250957/)]
22. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. *JMIR Public Health Surveill* 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: [10.2196/18700](https://doi.org/10.2196/18700)] [Medline: [32293582](https://pubmed.ncbi.nlm.nih.gov/32293582/)]
23. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. *J Med Internet Res* 2020 May 26;22(5):e18796 [FREE Full text] [doi: [10.2196/18796](https://doi.org/10.2196/18796)] [Medline: [32412414](https://pubmed.ncbi.nlm.nih.gov/32412414/)]
24. Fadda M, Galimberti E, Fiordelli M, Romanò L, Zanetti A, Schulz PJ. Effectiveness of a smartphone app to increase parents' knowledge and empowerment in the MMR vaccination decision: A randomized controlled trial. *Hum Vaccin Immunother* 2017 Nov 02;13(11):2512-2521 [FREE Full text] [doi: [10.1080/21645515.2017.1360456](https://doi.org/10.1080/21645515.2017.1360456)] [Medline: [29125783](https://pubmed.ncbi.nlm.nih.gov/29125783/)]
25. Du J, Luo C, Shegog R, Bian J, Cunningham RM, Boom JA, et al. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine. *JAMA Netw Open* 2020 Nov 02;3(11):e2022025 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.22025](https://doi.org/10.1001/jamanetworkopen.2020.22025)] [Medline: [33185676](https://pubmed.ncbi.nlm.nih.gov/33185676/)]
26. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. *PLoS One* 2021;16(2):e0246306 [FREE Full text] [doi: [10.1371/journal.pone.0246306](https://doi.org/10.1371/journal.pone.0246306)] [Medline: [33539390](https://pubmed.ncbi.nlm.nih.gov/33539390/)]
27. Liu S, Ferraro J, Gundlapalli AV, Chapman W, Bucher B. Detection of healthcare-associated infections using electronic health record data. 2018 Nov Presented at: AMIA 2018 Annual Symposium; November 3-7, 2018; San Francisco, CA.
28. Shi J, Liu S, Pruitt LCC, Luppens CL, Ferraro JP, Gundlapalli AV, et al. Using natural language processing to improve EHR structured data-based surgical site infection surveillance. *AMIA Annu Symp Proc* 2019;2019:794-803 [FREE Full text] [Medline: [32308875](https://pubmed.ncbi.nlm.nih.gov/32308875/)]
29. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. 2019 Jun Presented at: NAACL-HLT 2019; 2019; Minneapolis, MN p. 4171-4186.



30. Demsar J. Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 2006 Dec 01;7(1):1-30 [[FREE Full text](#)]
31. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 1979 Jun;74(366a):427-431. [doi: [10.1080/01621459.1979.10482531](https://doi.org/10.1080/01621459.1979.10482531)]
32. Hidayatullah A, Pembrani E, Kurniawan W, Akbar G, Pranata R. Twitter topic modeling on football news. 2018 Apr Presented at: 2018 3rd International Conference on Computer and Communication Systems; 2018; Nagoya, Japan p. 467-471.
33. Blei DM. Probabilistic topic models. *Communications of the ACM* 2012;55(4):77-84.
34. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Machine Learn Res* 2003 Mar 01;3:993-1022. [doi: [10.3156/jsoft.24.4\\_160\\_1](https://doi.org/10.3156/jsoft.24.4_160_1)]
35. Sievert C, Shirley K. LDAvis: A method for visualizing and interpreting topics. 2014 Jun 27 Presented at: Workshop on Interactive Language Learning, Visualization, and Interfaces; 2014; Baltimore, MD p. 63-70.
36. Levy S, Wang W. Cross-lingual transfer learning for COVID-19 outbreak alignment. arXiv. URL: <http://arxiv.org/abs/2006.03202> [accessed 2021-02-11]
37. Spangher A, Peng N, May J, Ferrara E. Enabling low-resource transfer learning across COVID-19 corpora by combining event-extraction and co-training. : Association for Computational Linguistics; 2020 Presented at: ACL 2020 Work NLP-COVID 2020; July 2020; online URL: <https://aclanthology.org/2020.nlpCOVID19-acl.4>
38. Tasneem F, Naim J, Tasnia R, Hossain T, Chy A. CSECU-DSG at WNUT-2020 Task 2: Exploiting ensemble of transfer learning and hand-crafted features for identification of informative COVID-19 English tweets. 2020 Nov 19 Presented at: 2020 EMNLP Workshop W-NUT: The Sixth Workshop on Noisy User-generated Text; November 19, 2020; Online p. 394-398.
39. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. *J Med Internet Res* 2020 Oct 23;22(10):e22624 [[FREE Full text](#)] [doi: [10.2196/22624](https://doi.org/10.2196/22624)] [Medline: [33006937](https://pubmed.ncbi.nlm.nih.gov/33006937/)]
40. Yin H, Yang S, Li J. Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. arXiv. 2020 Jul 05. URL: <http://arxiv.org/abs/2007.02304> [accessed 2021-02-11]
41. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment analysis of COVID-19 tweets by deep learning classifiers-a study to show how popularity is affecting accuracy in social media. *Appl Soft Comput* 2020 Dec;97:106754 [[FREE Full text](#)] [doi: [10.1016/j.asoc.2020.106754](https://doi.org/10.1016/j.asoc.2020.106754)] [Medline: [33013254](https://pubmed.ncbi.nlm.nih.gov/33013254/)]
42. Li I, Li Y, Li T, Alvarez-Napagao S, Garcia D. What are we depressed about when we talk about COVID19: mental health analysis on tweets using natural language processing. arXiv. 2020 Jun 08. URL: <https://arxiv.org/abs/2004.10899> [accessed 2021-02-11]
43. Liu S, Reese TJ, Kawamoto K, Del Fiore G, Weir C. A systematic review of theoretical constructs in CDS literature. *BMC Med Inform Decis Mak* 2021 Mar 17;21(1):102 [[FREE Full text](#)] [doi: [10.1186/s12911-021-01465-2](https://doi.org/10.1186/s12911-021-01465-2)] [Medline: [33731089](https://pubmed.ncbi.nlm.nih.gov/33731089/)]
44. Evanega S, Lynas M, Adams J, Smolenyak K. Coronavirus misinformation: quantifying sources and themes in the COVID-19 "infodemic". Alliance for Science. URL: [https://allianceforscience.cornell.edu/wp-content/uploads/2020/10/Evanega-et-al-Coronavirus-misinformation-submitted\\_07\\_23\\_20-1.pdf](https://allianceforscience.cornell.edu/wp-content/uploads/2020/10/Evanega-et-al-Coronavirus-misinformation-submitted_07_23_20-1.pdf) [accessed 2021-02-11]
45. Müller M, Salathé M, Kummervold P. COVID-Twitter-BERT: a natural language processing model to analyse COVID-19 content on Twitter. arXiv. 2020 May 15. URL: <https://arxiv.org/abs/2005.07503> [accessed 2021-01-28]
46. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011 Apr 23;6:42 [[FREE Full text](#)] [doi: [10.1186/1748-5908-6-42](https://doi.org/10.1186/1748-5908-6-42)] [Medline: [21513547](https://pubmed.ncbi.nlm.nih.gov/21513547/)]

## Abbreviations

**ADF:** Augmented Dickey-Fuller

**BERT:** bidirectional encoder representations from transformers

**CT-BERT:** COVID-Twitter-bidirectional encoder representations from transformers

**FDA:** Food and Drug Association

**HPV:** human papillomavirus vaccine

**LDA:** latent Dirichlet allocation

**VADER:** Valence Aware Dictionary and Sentiment Reasoner

**WHO:** World Health Organization

*Edited by C Basch; submitted 07.05.21; peer-reviewed by J Lei, C Huang, R Poluru; comments to author 09.06.21; revised version received 28.06.21; accepted 11.07.21; published 10.08.21*

*Please cite as:*

*Liu S, Li J, Liu J*

*Leveraging Transfer Learning to Analyze Opinions, Attitudes, and Behavioral Intentions Toward COVID-19 Vaccines: Social Media Content and Temporal Analysis*

*J Med Internet Res 2021;23(8):e30251*

*URL: <https://www.jmir.org/2021/8/e30251>*

*doi: [10.2196/30251](https://doi.org/10.2196/30251)*

*PMID: [34254942](https://pubmed.ncbi.nlm.nih.gov/34254942/)*

©Siru Liu, Jili Li, Jialin Liu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 10.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.