The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1468-4527.htm

The infinity vaccine war: linguistic regularities and audience engagement of vaccine debate on Twitter

Rachel X. Peng and Ryan Yang Wang
School of Communication, College of Liberal Arts, Rochester Institute of Technology, Rochester, New York, USA

Received 4 June 2022
Revised 23 February 2023
27 March 2023
Accepted 27 March 2023

Abstract

Purpose – As public health professionals strive to promote vaccines for inoculation efforts, fervent anti-vaccination movements are marshaling against it. This study is motived by a need to better understand the online discussion around vaccination. The authors identified the sentiments, emotions and topics of pro- and anti-vaxxers' tweets, investigated their change since the pandemic started and further examined the associations between these content features and audiences' engagement.

Design/methodology/approach – Utilizing a snowball sampling method, data were collected from the Twitter accounts of 100 pro-vaxxers (266,680 tweets) and 100 anti-vaxxers (248,425 tweets). The authors are adopting a zero-shot machine learning algorithm with a pre-trained transformer-based model for sentiment analysis and structural topic modeling to extract the topics. And the authors use the hurdle negative binomial model to test the relationships among sentiment/emotion, topics and engagement.

Findings – In general, pro-vaxxers used more positive tones and more emotions of joy in their tweets, while anti-vaxxers utilized more negative terms. The cues of sadness predominantly encourage retweets across the pro- and anti-vaccine corpus, while tweets amplifying the emotion of surprise are more attention-grabbing and getting more likes. Topic modeling of tweets yields the top 15 topics for pro- and anti-vaxxers separately. Among the pro-vaxxers' tweets, the topics of "Child protection" and "COVID-19 situation" are positively predicting audiences' engagement. For anti-vaxxers, the topics of "Supporting Trump," "Injured children," "COVID-19 situation," "Media propaganda" and "Community building" are more appealing to audiences.

Originality/value – This study utilizes social media data and a state-of-art machine learning algorithm to generate insights into the development of emotionally appealing content and effective vaccine promotion strategies while combating coronavirus disease 2019 and moving toward a global recovery.

Peer review – The peer review history for this article is available at https://publons.com/publon/10.1108/OIR-03-2022-0186

Keywords Anti-vaccination, Sentiment analysis, Social media engagement, Topic modeling, Twitter

Paper type Research paper

Introduction

Social media provide a prodigious opportunity to spread a wide scope of health information. Organized groups have quickly capitalized on multiple social media platforms to communicate their messages and beliefs. As scientists and public health professionals strive to promote vaccines for inoculation effort, fervent anti-vaccination movements are marshaling against it (Guidry *et al.*, 2015).

The authors thank the editor and the reviewers for their thoughtful comments and suggestions, which have greatly contributed to the improvement of the quality of the manuscript.

Funding: There was no funding or incentive for conducting this study.

Data availability statement: The dataset collected and analyzed during the study and the codes are available from the corresponding author upon reasonable request.

Conflicts of interest statement: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Online Information Review

1468-4527
DOI 10.1108/OIR-03-2022-0186

Various factors are at play in the reluctance of vaccination, including the proliferation of anti-vaccination movements (Dubé *et al.*, 2013). Studies have shown that even brief exposure to websites that are critical of vaccination can affect attitudes toward vaccine (Stecula *et al.*, 2020). If vaccine-hesitant individuals consistently see information and throw doubts on the safety or effectiveness of vaccines, they may ignore the advice of physicians and health professionals and refuse the recommended vaccination (Ortiz *et al.*, 2019). Thus, anti-vaccine information on social media may push the vaccine hesitancy to the point of disaster. Previous studies showed evidence that the consequences of anti-vaccine information overload included the public losing confidence in vaccines (Larson *et al.*, 2011), lower vaccine acceptability and low immunization rates (Zimet *et al.*, 2013), and even disease outbreaks (Larson, 2018).

A recent poll on vaccine hesitancy reported an increasing hesitancy among the general population to commit to any new vaccine (Kantar, 2021). The data showed that 3 in 4 British and 2 in 3 Americans were willing to receive coronavirus disease 2019 (COVID-19) vaccine, a quarter of people in the United States, Germany and France were hesitant about or squarely against getting the COVID-19 vaccine (Kantar, 2021). This poll, along with previous studies, demonstrated the extraordinary predicament that the global health-care system is facing, and the knife-edge that public health experts have been walking (Walsh *et al.*, 2022). Vaccine hesitancy is a threat to the global recovery from the historical pandemic and also to the well-being of the entire population. With more new transmissible variants emerging, overcoming vaccine hesitancy is critical to worldwide post-COVID recovery. According to the estimation of Centers for Disease Control and Prevention (CDC), California has one of the lowest rates of vaccine hesitancy (11%), which has effectively transformed California from an epicenter of pandemic in early 2020 to a state with the lowest statewide transmission rate today (CDC, 2021). California's secret weapon in COVID-19 success, which is Californians' embrace of COVID-19 vaccines, underscores how critical vaccine uptake is to drive real progress in the war against the virus. Public health and communication specialists need to be continuously actively involved in any efforts encouraging necessary vaccine uptake, building vaccine confidence and also intervening the counterbalance anti-vaccine voice on social media. This also indicates a need to better understand the discussion around vaccination. By doing so, we can provide insights about how to combat widespread vaccination myths and how to effectively communicate vaccine recommendations, and ultimately fuel the public health campaigns on vaccine promotion.

An objective of the current study is to characterize trends in pro- and anti-vaccination discourse on Twitter, which is commonly used as a source of information during public crises or emergencies, such as the global pandemic (Alhayan *et al.*, 2023). We focus on textual data on Twitter where there exists a large archive of discussions about vaccines. We manually label individuals known to publicly post pro- and anti-vaccination content on Twitter, and systematically examine (1) the emotions, sentiments and topics of the messages they convey; (2) the change of sentiments, emotions and topics since the COVID-19 pandemic started; and (3) the engagement of their tweets. In particular, we use the term *pro-vaxxer* to denote those pro-vaccination activists and use *anti-vaxxer* for people who are skeptical about vaccine. Their accounts are created specifically for vaccine promotion and vaccine critics, which may offer details about rhetorical tactics that pro- and anti-vaccine activists frequently use. We are also interested in examining the relationship among sentiments, emotions topics and audiences' engagement. As such, by uncovering meaningful themes of online discussion about vaccines and unpacking the linkage between emotion words, content themes and engagement rate, we hope to provide public health professors with insights to fine-tune their communication strategies and maximize audience engagement while combating COVID-19 and moving toward a global recovery.

Related work

In this paper, we use the term *anti-vaccine* to denote all discourse that does not promote vaccination, such as vaccine resistance (Yaqub *et al.*, 2014), vaccine selective (Ward *et al.*, 2015) and vaccine hesitancy (Larson *et al.*, 2013), which forefronts an action taken against vaccines. Some scholars argued that anti-vaxxers have long had the ascendancy on social media, which is facilitated by the networked properties and the growth of consumer autonomy in health care (Guidry *et al.*, 2015; Smith and Graham, 2019). A number of studies have paid particular attention to discourse underlying anti-vaccinations. Bricker and Justice (2019) theorized the argumentative patterns of anti-vaccine discourse that made vaccine skepticism appealing to undecided individual. Hoffman *et al.* (2019) coded the messages conveyed by anti-vaxxers and conducted social network analysis to examine the connections between these individuals and anti-vaccination topics. Studies listed above have reported a set of common contents and themes of the anti-vaccine websites, including the issue of confidence (e.g., safety concerns, lack of confidence in vaccines or health providers), civil liberties (e.g. non-medical vaccine exemptions), conspiracy theories, beliefs (e.g. preference for natural medicines) and emotion appeals (e.g. stories from parents who believed their children are vaccine-injured). There is a consensus among social scientists that a purely rational response is not sufficient to address the problem of vaccine skepticism and denialism (Grant *et al.*, 2015).

Public health research also focuses on examining how the pro- and anti-vaccine activities developed in the online space. Previous studies have compared the expression of vaccine-skeptical and vaccine-promoting by examining the posts on social media platforms (Guidry *et al.*, 2015; Yiannakoulias *et al.*, 2019), the persuasive features of vaccine websites (Grant *et al.*, 2015), the headlines of webpages (Xu and Guo, 2018) and the content of online articles (Xu *et al.*, 2019). Together, these studies found that the pro-vaccine discourses focused on the accurate transmission of evidence-based scientific research about vaccines, whereas the anti-vaccine arguments were often grounded in anecdotal evidence and challenge the information presented in scientific literature and government documents. Several studies have also assessed the patterns of the conversations about the vaccine debate. Schmidt *et al.* (2018) confirmed the emergence of two polarized communities and proposed that pro- and anti-vaccination attitudes polarized the users on Facebook. Yuan *et al.* (2019) investigated the communication flow of anti-vaccine and pro-vaccine communities on Twitter. Their results showed that anti-vaccine users tended to tightly cluster in a close community. Johnson *et al.* (2020) mapped out the contention surrounding vaccines emerging among three billion Facebook users globally while providing us with a comprehensive depiction of vaccine-related debate from a systematic level.

Researchers also attempted to examine and compare how users engaged with anti- and pro-vaccine content on the social media platforms by investigating responses (i.e. the number of reposting, sharing, liking and comments) from users. Briones *et al.* (2012) analyzed how Human papillomavirus (HPV) vaccine was portrayed on YouTube and found that videos negative in tone had a higher number of average likes than videos positive in tone on YouTube. A quantitative analysis of Facebook likes and comments by Schmidt *et al.* (2018) found that people in conspiracy-like groups showed higher engagement in the community, and anti-vaccine pages received more comments from users. Conversely, Guidry *et al.* (2015) reported pro-vaccine pins elicited consistently more engagement than anti-vaccine pins on Pinterest. Overall, these studies highlight the need for a more in-depth, empirical exploration of social media engagement that is data-driven. And in the midst of the COVID-19 pandemic, it is worth investing the time in studying social media engagement with online vaccine content.

Research questions

Sentiments and emotions

The attitude, views and feelings extracted from social media data constitute an essential part in extracting relevant information from people's opinion among public, which can be referred as the *sentiments* (Daou, 2021). As the content characteristics of media messages, sentiments and emotions have also been recognized as good predictors of audiences' engagement (Buffard and Papasava, 2020). A well-established notion in the news media is that emotions are important in engaging attention (Weber, 2014). Sensational stories whose subject is grim and menacing are likely to get more attention, as expressed in a decades-old mantra, "if it bleeds, it leads." Previous studies have also illustrated that the presence of negative emotion words would lead to more clicks, more replies and more sharing (Heiss *et al.*, 2019). In a study investigating sentiments in the 30 largest spiking events in Twitter posts, Thelwall *et al.* (2011) found that negative sentiment played a key role in the popular events. By constructing and analyzing semantic networks, Kang *et al.* (2017) examined the vaccine sentiments on social media and found that negative sentiment was mainly embedded in anti-vaccine messages, which framed around institutional distrust and vaccine skepticism.

Topics extracted from vaccine tweets

Each document can be represented by distribution over topics and each topic by distribution over words. The topics discovered by unsupervised machine learning techniques are coherent themes, which convey similar semantic meaning in a corpus of documents. Scholars have applied topic modeling approaches to discover what topical issues are being discussed by the populace on COVID-19 pandemic (Abd-Alrazaq *et al.*, 2020) and vaccine (Dyda *et al.*, 2019). These studies indicated that topic modeling is a useful method for automatically identifying vaccine-related discussion topics. We are interested in examining the core topics promoted by pro- and anti-vaxxers on Twitter, and if particular topic resonates with the population.

Social media engagement

Social media yield high reach and elicit engagement from audiences, which poses huge potential for health media campaigns (Pedersen *et al.*, 2020). The term *social media engagement* refers to the level of user's involvement, participation and interaction, including searching, commenting and sharing/forwarding content online (Hargittai and Hsieh, 2010). User engagement is important to the potential success of campaigns or any other social media activities (Joo *et al.*, 2020). Highly engaging messages are more effective in propagating information to a wider scope of audiences and also in augmenting beneficial changes in individuals' health behavior (Lee and Jang, 2010). Creating engaging messages has become a major focus in health communication research. To inform the direction of future vaccine promotion campaigns, it is crucial to study the message characteristics associated with successful audience engagement.

In light of the studies discussed above, we pose our research questions in the context of pro-vaxxers and anti-vaxxers as follows:

RQ1. Which emotions (a) sentiments, (b) and topics (c) are dominant in tweets of pro- and anti-vaxxers, respectively?

RQ2. How do the emotions (a), sentiments (b) and topics (c) change during COVID-19 pandemic?

RQ3. What are the associations between emotion (a), sentiment (b), topic (c) and the engagement of tweets?

Method

Data collection

Adopting a snowball sampling approach, we started with one of the most active pro- or anti-vaxxers on Twitter. The first pro-vaxxer we selected was Seth Berkley (@GaviSeth), who is the CEO of the pro-vaccine organization "Gavi, The Vaccine Alliance." Larry Cook (@stopvaccinating) was the first anti-vaxxer we targeted on Twitter (whose account is terminated by Twitter now). He, as a leading anti-vaccination activist, was well-known for being the admin of the 195 k-member group "Stop Mandatory Vaccination." Then, we searched for another pro- and anti-vaxxer based on the following list of Seth Berkley and Larry Cook. We continued to look for the third pro- and anti-vaxxer according to the second pro/anti-vaxxer's following list and so on. We labeled these accounts as pro- or anti-vaxxer based on that (1) they have listed clearly positive or negative attitude toward vaccine in their Twitter profile biography and (2) their first page has more than 10 tweets about vaccine. They also have to be consistently active (defined by posting more frequently than 75% of users), post within a week prior to initial data collection and tweet in English. Two experienced researchers started with the whole process separately and independently then cross-validated the lists together. Eventually, we selected 100 Twitter accounts with a clear anti-vaccine stance and 100 Twitter accounts with a clear pro-vaccine stance. Also, in order to deal with the potential imbalance of posting frequency and engagement rate, we have also selected half individual accounts and half organization account for each community. Our corpus aims to be comprehensive in collecting as many as possible vaccine-related discussion of pro- and anti-vaccination. Our approach would benefit from relatively concise identification of pro- and anti-vaccination users, which could help shape the foci of further analysis within the specific scope of our interests.

After finalizing the list of the accounts, we scraped the historical tweet data of these 200 identified Twitter accounts through Twitter application programming interface via *rtweet* package (version 0.7.0) on R 3.6.3. Data collection was finished on October 8, 2020. The collected Twitter dataset contains tweet level data including the text, hashtags, URLs, number of favorites, number of retweets, the date the tweet was published, the number of followers, whether the tweet was an original tweet or a retweet and whether the tweet contained any image. The dataset yielded 266,680 tweets from identified pro-vaxxer Twitter accounts and 248,425 tweets from identified anti-vaxxer Twitter accounts. Our dataset did not contain any duplicate tweets based on the status ID. Specifically, the dataset included 172,038 original tweets and 94,642 retweets from pro-vaxxers; and 158,199 original tweets and 90,226 retweets from anti-vaxxers. Also, among pro-vaxxers' tweets, 178,165 were posted before COVID-19 outbreak (February 2020), and 88,515 were posted during the outbreak. For anti-vaxxers' tweets, 83,931 were posted before the outbreak while 164,494 tweets were posted after February 2020. We removed the URLs, numbers, punctuation, hashtags and stop words in English during the text pre-processing stage of topic modeling [1]. The tweet frequency across time is presented on the Part III in the supplemental material.

Data analysis

Tweet engagement

We operationalized engagement as a click-based behavior, which is manifested through actions such as favoriting and retweeting on Twitter. In this paper, we measured tweet engagement using the retweet and like counts of each tweet. As Hwong *et al.* (2017) argued, like (or favorite) could be treated as a virtual endorsement behavior. Meanwhile, retweeting behavior could be regarded as a higher level of engagement as it diffuses the content while all visitors could see it on the account's timeline.

Sentiment/emotion classification

To classify the sentiments and emotions at the tweet level, we employed a zero-shot machine learning algorithm using a python toolkit *pysentimiento*, which is built on a pre-trained transformer-based model *BERTweet* on dataset such as SemEval-2016 and EmoEvent (Nguyen *et al.*, 2020; Pérez *et al.*, 2021). Each tweet is exclusively classified into one of the three sentiments (neutral, negative or positive); and one of the six emotions (anger, disgust, fear, joy, sadness, surprise or others) [2] based on the highest score (between 0 and 1). Our state-of-the-art approach in sentiment analysis outperforms traditional lexicon- and rule-based approaches, as previous literature has demonstrated (Mathew and Bindu, 2020; van Atteveldt *et al.*, 2021). We validated the performance of the pre-trained model using the F1 evaluation metric, based on a random sample of our dataset (600 tweets from pro-vaxxers and anti-vaxxers) with human annotation by the researchers. The pre-trained model has achieved a similar performance in the original paper, and it also significantly outperforms the approach utilizing the Valence Aware Dictionary and sEntiment Reasoner (VADER) lexicon [3].

Structural topic modeling

Topic modeling refers to statistical methods that analyze the words of texts to uncover the underlying themes, how the themes are connected, and how they change over time (Blei, 2012). We performed structural topic modeling with the *stm* package (version 1.3.6). One of the advantages of structural topic modeling comparing to other approaches (e.g. Latent Dirichlet allocation) is that it could incorporate the meta element into the analysis and performs better dealing with short documents like tweets. The structural variable from the metadata we are interested in is whether the tweet was posted before or during the COVID-19 pandemic (February 2020) so that we could examine if topics portions change before and during the pandemic. In order to determine the optimal number of topics for each corpus (i.e. pro-vaxxers and anti-vaxxers), we started with an exploratory approach, by comparing the diagnostic values (i.e. held-out likelihood, lower bound, residuals, semantic coherence and exclusivity) across different *K*-topic models. We have tested various models and identified the optimal topic for pro-vaxxers' corpus was 40 and the one for anti-vaxxers' corpus was 50 [4]. After determining the best model with the optimal number of topics, we extracted the terms for each topic and performed a qualitative thematic analysis to generate labels for each topic. Lastly, we validated, discussed and modified the labels among the researchers in the team.

Statistical analysis

We are interested in investigating if the sentiments and emotions, and also the topics of the tweets posted by pro-vaxxers and anti-vaxxers would impact the engagement (i.e. retweets and likes) of the tweets. We included further factors (as control variables) representing tweet features such as word counts, the inclusion of URL, image and hashtags. Treating the engagement as the dependent variables, we tested out five different regression models, including ordinary least square, Poisson, quasi-Poisson, negative binominal and hurdle model. We concluded that hurdle model is the best model based on its lower value of Akaike information criterion (AIC) [5]. Hurdle models are a class of models for count data, which are better with handling excess zeros and overdispersion like Twitter's engagement (Zeileis *et al.*, 2008). In our pro-vaxxer dataset, 28.38% of the tweets receive zero retweet and 54.18% receive zero likes. And in our anti-vaxxer dataset, 32.99% of them get zero retweet and 56.64% get zero likes. We are using the *pscl* package to fit our final hurdle models. Specifically, a hurdle model we use consists of two parts: (1) a binary logistic regression for the zero-count process (zero portion) models the odds ratio (OR) of the tweet receiving a retweet/like or not; (2) a negative binomial regression of positive-count process (count portion) examines the incident rate ratio (IRR) of the retweet/like count that a tweet obtained. In the following section, only the results of hurdle models are reported and discussed.

Results

Sentiment and emotion analysis

The descriptive statistics of six emotions for these two communities are presented in Table 1. Among the six emotions, pro-vaxxers' tweets have predominantly been identified as expressing Joy while anti-vaxxers' tweets tend to convey disgust, answering our RQ1(a). To answer our RQ2(a) regarding the emotion changes before and during COVID-19 outbreak, we find that both pro-vaxxers and anti-vaxxers expressed more negative emotions, such as anger, disgust, fear and surprise, during the outbreak, while the emotions of joy and sadness were not as prominent.

Similarly, Table 2 presents the descriptive statistics of two types of sentiment for the tweets of each camp. Both pro-vaxxers and anti-vaxxers express more negative sentiment and less positive sentiment during the pandemic. A series of chi-squared tests revealed that in general pro-vaxxers' tweets are more positive than those of anti-vaxxers, regardless of whether the

Category	Emotion type	Time	Tweet count	Average	χ^2	p-value[a]
Pro-vaxxer	Anger		854	0.0032		
		Before COVID-19	522	0.0029	42.272	<0.001
		During COVID-19	332	0.0038		
	Disgust		9,389	0.0352		
		Before COVID-19	4,921	0.0276	21.856	<0.001
		During COVID-19	4,468	0.0505		
	Fear		22,908	0.0859		
		Before COVID-19	13,232	0.0743	552	<0.001
		During COVID-19	9,676	0.1093		
	Joy		36,584	0.1372		
		Before COVID-19	27,337	0.1534	8945.1	<0.001
		During COVID-19	9,247	0.1045		
	Sadness		4,389	0.0165		
		Before COVID-19	3,160	0.0177	849.57	<0.001
		During COVID-19	1,229	0.0139		
	Surprise		1,033	0.0039		
		Before COVID-19	650	0.0036	69.012	<0.001
		During COVID-19	383	0.0043		
Anti-vaxxer	Anger		11,766	0.0701		
		Before COVID-19	3,311	0.0394	2248.9	<0.001
		During COVID-19	8,455	0.0514		
	Disgust		52,007	0.3098		
		Before COVID-19	13,384	0.1595	12,248	<0.001
		During COVID-19	38,623	0.2348		
	Fear		7,093	0.0423		
		Before COVID-19	1863	0.0222	1598.3	<0.001
		During COVID-19	5,230	0.0318		
	Joy		13,065	0.0778		
		Before COVID-19	5,228	0.0623	521	<0.001
		During COVID-19	7,837	0.0476		
	Sadness		5,052	0.0301		
		Before COVID-19	2093	0.0249	148.45	<0.001
		During COVID-19	2,959	0.0180		
	Surprise		1896	0.0113		
		Before COVID-19	515	0.0061	395.55	<0.001
		During COVID-19	1,381	0.0084		

Note(s): [a]The p-value was generated through chi-squared tests between the groups of before COVID-19 outbreak and during COVID-19 outbreak
Source(s): Table by authors

Table 1. The descriptive statistics of emotions for pro- and anti-vaxxers' tweets

Category	Sentiment	Time	Tweet count	Average	χ^2	p-value[a]
Pro-vaxxer	Negative		45,912	0.1722		
		Before COVID-19	28,834	0.1618	3010.2	<0.001
		During COVID-19	17,078	0.1929		
	Positive		94,736	0.3552		
		Before COVID-19	66,837	0.3751	16,004	<0.001
		During COVID-19	27,899	0.3152		
Anti-vaxxer	Negative		113,501	0.6762		
		Before COVID-19	35,307	0.4207	16,205	<0.001
		During COVID-19	78,194	0.4754		
	Positive		34,286	0.2043		
		Before COVID-19	12,376	0.1475	2651.1	<0.001
		During COVID-19	21,910	0.1332		

Note(s): [a]The p-value was generated through chi-squared tests between the groups of before COVID-19 outbreak and during COVID-19 outbreak
Source(s): Table by authors

Table 2.
The descriptive statistics of sentiment score for pro- and anti-vaxxers' tweets

analysis was conducted over the whole timeline or during two separate time frames (χ^2 for positive sentiment = 103.25, $p < 0.001$; χ^2 for negative sentiment = 358.57, $p < 0.001$).

Structural topic modeling

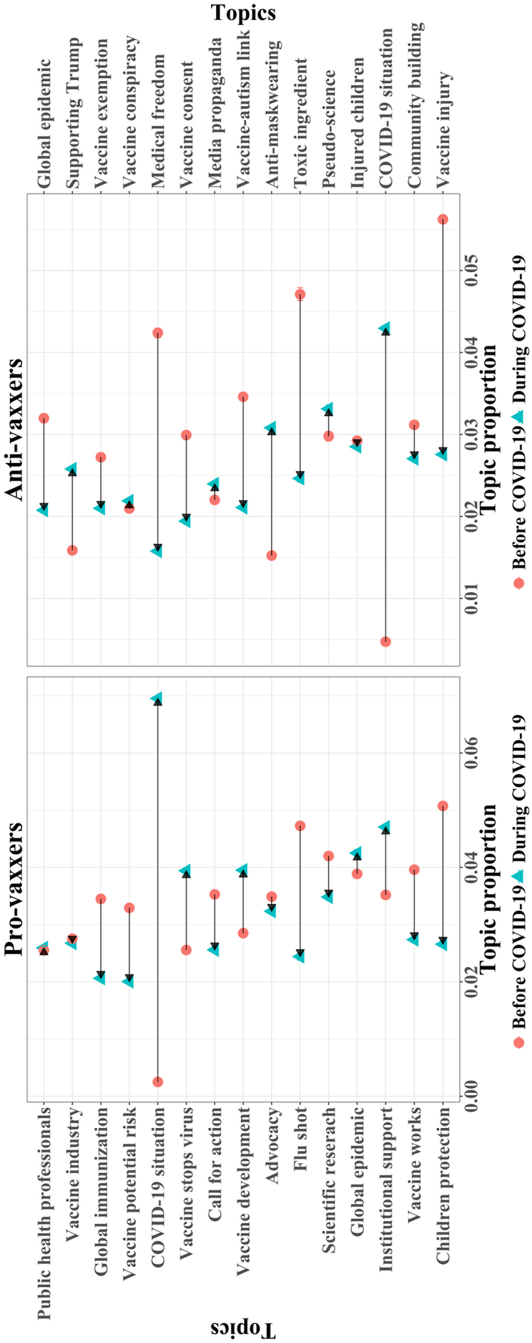
The full list of the 40 topics generated from pro-vaxxers' tweets, including the terms with highest probability and their corresponding aggregated proportion, are presented in supplemental materials. We decided to only examine the top 15 topics among the 40 topics of pro-vaxxers' tweets after excluding some topics that might be less meaningful. We labeled the most popular 15 topics for pro-vaxxers' tweets based on the descending order of topic proportions. Table S1 in the supplemental materials shows the description and the terms of the top 15 topics, answering our RQ1(c). To assess our RQ2 (c), not surprisingly, there were a large volume of pro-vaxxers' tweets talking about "COVID-19 situation" since February 2020, as Figure 1 (left) shows. Meanwhile, they were mentioning more about "Institutional support," "Global epidemic," "Vaccine development," "Vaccine stops virus" and "Public health professionals" during the COVID-19 outbreak.

The full list of the 50 topics generated from anti-vaxxers' tweets, including the terms with highest probability and their corresponding aggregated proportion, are presented in supplemental materials. Similarly, we only investigated the top 15 topics among the 50 topics of anti-vaxxers' tweets after excluding some topics that might be less meaningful. Table S2 in the supplemental materials presents a more detailed description and 20 terms with the highest probabilities of the top 15 topics for anti-vaxxers' tweets.

As Figure 1 (right) shows, anti-vaxxers also talked more about "COVID-19 situation" during the COVID-19 outbreak unsurprisingly [RQ2(c)]. Moreover, during the COVID-19 outbreak, anti-vaxxers discussed more about "Pseudo-science," "Anti-maskwearing," "Media-propaganda," "Supporting Trump" and slightly more about "Vaccine conspiracy" compared to the time period before February 2020.

Engagement

In general, both pro-vaxxers' and anti-vaxxers' tweets were more likely to be retweeted than to be liked. Anti-vaxxers' tweets were getting more retweets since the COVID-19 pandemic. On average, pro-vaxxers received 32.16 likes (median: 0; max: 210,635) and 183.4 retweets (median: 2; max: 2,219,189) out of 266,680 tweets; while anti-vaxxers, with 248,425 tweets,



Source(s): Figure by authors

Figure 1.
Pro-vaxxers (left) and Anti-vaxxers (right) topic change before and during COVID-19 outbreak

attained 19.32 likes (median: 0; max: 40,146) and 685.5 retweets (median: 3; max: 3,088,681) per tweet, which are imbalanced before and during the pandemic [6].

Table 3 shows the results from the hurdle model of pro-vaxxers' tweet as RQ3(a) and (b) asked. For pro-vaxxers' tweets, the results in the zero portion of hurdle model, which tests whether a retweet/like occurs or not, reveal one of the most influential factors is the positive sentiment. If the tweet is a positive, it would increase the odds of getting a retweet by 16.8% (OR = 1.168), and the odds of getting a like by 24.8% (OR = 1.248) compared to a neutral one. For pro-vaxxers' tweets with at least one retweet or like, the count portion of hurdle model, examining the number of retweets and likes, indicates that the emotion of sadness and disgust are two powerful indicators. If the tweet is classified as sadness or disgust, it would retweet count by a factor of 3.155 or 3.128, and increase the number of likes by a factor of 1.210 or 1.570 compared to a neutral tweet, while holding all other variables constant. Across the results from the zero portion and count portion of the hurdle model, we found that the emotion of fear plays a consistent and positive role in predicting getting a retweet as well as the larger number of retweets. The emotion of surprise, disgust and joy are all consistently contributed to the likelihood of getting a like also the increase of like count.

As Table 4 displays, for anti-vaxxer's tweet, the zero portion of the hurdle model indicates negative sentiment is a prominent factor in both predicting getting a retweet (29.7%, OR = 1.297) as well as getting a like (12.5%, OR = 1.125). Also, if a tweet expresses the emotion of fear, it increases the odds of obtaining a retweet by 25.6% (OR = 1.256). Meanwhile, if a tweet expresses other emotions rather than fear, it would increase the probability of getting liked. For the count portion of the hurdle model, surprise is the most

Table 3. The hurdle model results of sentiment and emotions on engagement of pro-vaxxers' tweets

	Retweet[a]				Like[a]			
	Zero portion		Count portion		Zero portion		Count portion	
	Estimate	OR	Estimate	IRR	Estimate	OR	Estimate	IRR
Before/After COVID-19 outbreak	−0.029**	0.971	1.514***	4.545	0.947***	2.577	3.121***	22.663
Wordcount (scale)	0.544***	1.722	−0.276***	0.759	0.160***	1.174	0.612***	1.844
With an URL	−0.536***	0.585	−2.270***	0.103	1.411***	4.101	−0.305***	0.737
With an image	0.088***	1.092	0.003	1.003	1.789***	5.983	0.066***	1.068
With hashtags	0.640***	1.896	−1.343***	0.261	−0.049***	0.952	0.084***	1.087
Sentiment (Baseline: Neutral)								
Negative	−0.036**	0.964	0.174***	1.190	0.182***	1.199	0.398***	1.489
Positive	0.155***	1.168	0.226***	1.254	0.221***	1.248	0.173***	1.188
Emotions (Baseline: Others)								
Anger	−0.531***	0.588	1.684***	5.387	0.247**	1.280	0.061	1.063
Disgust	−0.616***	0.540	1.140***	3.128	0.436***	1.547	0.451***	1.570
Fear	0.190***	1.209	0.046*	1.047	−0.149***	0.862	−0.108***	0.898
Joy	−0.176***	0.838	−0.067**	0.935	0.344***	1.411	0.291***	1.338
Sadness	−0.261***	0.771	1.149***	3.155	0.043	1.043	0.190***	1.210
Surprise	−0.358***	0.699	0.670***	1.954	0.419***	1.521	0.744***	2.104
Constant	0.969***	2.635	−13.518	0.000	−1.622	0.198	−17.046	0.000
N	266,669				266,677			
Log Likelihood	−896,700				−585,700			
AIC	1,793,532				1,171,504			

Note(s): $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$
[a]Eleven outliers (for Retweet) and three outliers (for Like) were removed based on the Cook's distance outlier detection method
Source(s): Table by authors

Table 4. The hurdle model results of sentiment and emotions on engagement of anti-vaxxers' tweets

	Retweet[a]				Like[a]			
	Zero portion		Count portion		Zero portion		Count portion	
	Estimate	OR	Estimate	IRR	Estimate	OR	Estimate	IRR
Before/After COVID-19 outbreak	0.503***	1.653	1.453***	4.276	0.134***	1.144	1.377***	3.962
Wordcount (scale)	0.427***	1.533	−0.371***	0.690	0.000	1.000	0.482***	1.620
With an URL	0.354***	1.425	−2.904***	0.055	1.196***	3.308	0.936***	2.550
With an image	0.655***	1.924	−0.654***	0.520	0.942***	2.565	0.575***	1.776
With hashtags	0.733***	2.081	−1.028***	0.358	0.138***	1.147	0.423***	1.526
Sentiment (Baseline: Neutral)								
Negative	0.260***	1.297	0.057**	1.058	0.117***	1.125	0.031	1.031
Positive	−0.022	0.978	0.358***	1.431	0.267***	1.306	0.059*	1.060
Emotions (Baseline: Others)								
Anger	−0.125***	0.883	0.128***	1.137	0.196***	1.216	−0.061	0.940
Disgust	−0.178***	0.837	0.284***	1.329	0.280***	1.323	−0.039†	0.961
Fear	0.228***	1.256	−0.005	0.995	0.042	1.043	0.193***	1.213
Joy	−0.162***	0.851	0.420***	1.522	0.199***	1.220	0.017	1.017
Sadness	−0.243***	0.784	0.864***	2.372	0.356***	1.428	0.022	1.022
Surprise	0.057	1.059	1.146***	3.145	−0.049	0.952	−0.050	0.952
Constant	−0.021†	0.980	−7.181	0.001	−1.194***	0.303	−16.971	0.000
N	248,422				248,424			
Log Likelihood	−1,020,000				−555,300			
AIC	2,040,061				1,110,702			

Note(s): †$p < 0.1$, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$
[a]Three outliers (for Retweet) and one outlier (for Like) were removed based on the Cook's distance outlier detection method
Source(s): Table by authors

influential emotion as the number of retweets is expected to increase by a factor of 3.145. The results also reveal that fear is the only significant and positive factor associated with like count, as it would increase the number of likes by 1.213. Across two components of hurdle model, only the effect of the negative sentiment is associated with the probability of retweeting as well as the increase in retweet count. The positive sentiment is a positive and consistent factor in predicting the change of getting a like and the like count.

Our RQ3(c) seeks to investigate the relationship between the popular topic proportion and the engagement of the tweets. Among pro-vaxxer's tweets, as Table 5 shows, the result of the zero portion of hurdle model tells us that "COVID-19 situation" is the most influential topic, with 147 times (OR = 147.378) higher odds of obtaining a retweet. "Advocacy" is the most prominent topic with a higher likelihood of getting a like by 592 times (OR = 591.7). According to the count portion of the hurdle model, the topic "COVID-19 situation" is the most influential positive predictor of the retweet count, as one unit of change would result in additional 5,541 retweets. The topic "COVID-19 situation" is the most prominent topic getting like count, as one unit of increase would cause 7,180 more likes. Across both components of hurdle model, we found that two topics, "Child protection" and "COVID-19 situation," stay consistent in predicting the retweet counts of pro-vaxxers' tweets. Two topics, "COVID-19 situation" and "Vaccine industry," consistently predict a higher likelihood of being liked as well as a larger number of likes. To further investigate the potential interaction effect between topic and sentiment, we included the interaction term between topic and both negative and positive sentiment in the hurdle model [7]. When pro-vaxxers used negative sentiment to

Table 5. The hurdle model result of topic proportion on engagement of pro-vaxxers' tweets

	Retweet[a]				Like[a]			
	Zero portion		Count portion		Zero portion		Count portion	
	Estimate	OR	Estimate	IRR	Estimate	OR	Estimate	IRR
Before/During COVID-19 outbreak	−0.195***	0.823	0.424***	1.528	0.568***	1.765	2.413***	11.167
Wordcount	0.442***	1.556	−0.193***	0.824	0.287***	1.332	0.601***	1.824
With an URL	−0.539***	0.583	−2.059***	0.128	1.481***	4.397	−0.204***	0.815
With an image	0.193***	1.213	−0.071***	0.931	1.854***	6.385	0.315***	1.370
With hashtags	0.519***	1.680	−1.125***	0.325	0.006	1.006	−0.164***	0.849
Topics								
Child protection	3.098***	22.154	1.539***	4.660	−1.099***	0.333	0.914***	2.494
Vaccine works	2.337***	10.350	−15.350***	0.000	−1.547***	0.213	−3.120***	0.044
Institutional support	4.663***	105.953	−6.680***	0.001	−1.588***	0.204	−3.432***	0.032
Global epidemic	0.814***	2.257	−1.815***	0.163	−1.446***	0.236	−0.206	0.814
Scientific research	2.394***	10.957	−2.446***	0.087	−0.129	0.879	6.524***	681.298
Flu shot	0.957***	2.604	−4.968***	0.007	−0.775***	0.461	−2.524***	0.080
Advocacy	−0.986***	0.373	−2.653***	0.070	6.383***	591.700	2.495***	12.122
Vaccine development	−1.510***	0.221	−4.244***	0.014	1.771***	5.877	−3.574***	0.028
Call for action	2.244***	9.431	−5.303***	0.005	−1.871***	0.154	−1.869***	0.154
Vaccine stops virus	0.390**	1.477	−3.360***	0.035	1.932***	6.903	−1.187***	0.305
COVID-19 situation	4.993***	147.378	8.620***	5541.386	2.787***	16.232	8.879***	7179.608
Vaccine potential risk	2.285***	9.826	−4.595***	0.010	−1.158***	0.314	−1.856***	0.156
Global immunization	3.272***	26.364	−5.007***	0.007	0.463***	1.589	0.164	1.178
Vaccine industry	0.695***	2.004	−3.770***	0.023	1.472***	4.358	0.667**	1.948
Public health professionals	1.483***	4.406	−7.821***	0.000	−2.407***	0.090	0.090	1.094
Constant	0.160***	1.174	−11.629	0.000	−1.458***	0.233	−16.132	0.000
N	263,586				263,590			
Log Likelihood	−885,168.300				−574,812.700			
AIC	1,770,423				1,149,711			

Note(s): $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$
[a]Five outliers (for Retweet) and One outlier (for Like) were removed
Source(s): Table by authors

discuss the "COVID-19 situation," their tweets received significantly more likes compared to those written in a neutral tone (IRR = 39876.243). Pro-vaxxers' tweets promoting "Scientific research" also received significantly more "likes" when written in a positive sentiment (IRR = 325.784).

For anti-vaxxers, as Table 6 presents, the result of the zero portion of the hurdle model shows "Supporting Trump" is the most influential topic as it would cause a 1,224 times (OR = 1224.148) higher probability of getting a retweet. The topic "Vaccine conspiracy" is the most prominent topic as it leads to a 60 times (OR = 59.979) higher probability of getting a like. Based on the count portion of the hurdle model, "Supporting Trump" is the most influential topic as one unit of change would increase 544,161 more retweets. On the other hand, "Supporting Trump" is the strongest predictor as one unit of change could result in 719 more likes. Across two parts of hurdle model for anti-vaxxers' corpus, we can see that the topics of "COVID-19 situation," "Media propaganda" and "Supporting Trump" are the consistently positive indicators in predicting getting a retweet as well as the number of retweets. Three topics, "Community building," "Injured children" and "Toxic ingredient," consistently and positively predict higher likelihood of getting a like also a larger number of likes. When examining the interaction effect between topics and sentiments, we find that anti-vaxxers' tweets discussing "Vaccine conspiracy" received significantly more retweets when written in a positive tone (IRR = 41753.644). Meanwhile, anti-vaxxers received significantly more likes when discussing topics such as "Vaccine injury," "Injured children," "Pseudo-science," "Vaccine consent," "Medical freedom," "Vaccine exemption," as well as "Supporting Trump" with a positive sentiment (IRR = 20.895, 4982.704, 8.389, 59.317, 14.472, 587.619 and 10870.261, respectively).

Discussion and conclusion

This study sets out to review the discussion around vaccination by identifying the sentiments, emotions and topics of pro- and anti-vaccine tweets using a state-of-art machine learning approach. We further examine the relationships among sentiments, emotions, topics and audience engagement, as well as the potential interaction effects between sentiments and topics, which have seldom been explored in past literature. In sum, our investigation addresses the need and method to explore and manage discussion around vaccine on the social media, also generate fresh insights into the development of more emotionally appealing online content and effective media-based vaccine promotion strategies while combating COVID-19 and moving toward a global recovery.

Sentiments, emotions and engagement

In general, pro-vaxxers used more positive tone and more emotions of joy in their tweets. We found that cues of positive sentiment are linked with higher likelihood of get a retweet and like. People still have much hope in vaccines, leading to the endorsement of tweets with positive sentiment. The emotions of anger, sadness, disgust, surprise and fear would attract more retweets. The cues of surprise, disgust and joy encourage liking as well as more likes.

Anti-vaxxers used more negative tone and utilized more terms containing emotions of anger, disgust, fear, sadness and surprise in their tweets. Their contents with negative sentiment create instant viral traction to resonate with audiences. Their audience are more engaged in sharing the tweets containing emotions of surprise, sadness, joy, disgust and anger. Tweets contain emotion of fear generate more liking from audience.

Sadness is the predominant emotion encouraging retweet across pro- and anti-vaccine corpus, thus tweets with cue of sadness have bigger impact and greater spreading in general. Some scholars proposed that retweets are effectively an endorsement of the originators

Table 6. The hurdle model result of topic proportion on engagement of anti-vaxxers' tweets

	Retweet[a]				Like[a]			
	Zero portion		Count portion		Zero portion		Count portion	
	Estimate	OR	Estimate	IRR	Estimate	OR	Estimate	IRR
Before/During COVID-19 outbreak	0.563***	1.756	0.891***	2.438	0.153***	1.165	1.122***	3.071
Wordcount	0.498***	1.645	−0.180***	0.835	0.045***	1.046	0.556***	1.744
With an URL	0.468***	1.597	−2.355***	0.095	1.316***	3.728	1.141***	3.130
With an image	0.757***	2.132	−0.296***	0.744	0.993***	2.699	0.695***	2.004
With hashtags	0.580***	1.786	−0.734***	0.480	0.136***	1.146	0.457***	1.579
Topics								
Vaccine injury	2.608***	13.572	−6.354***	0.002	−1.144***	0.319	−2.155***	0.116
Community building	−0.897***	0.408	1.035**	2.815	3.186***	24.191	1.002***	2.724
COVID-19 situation	4.502***	90.197	1.880***	6.554	−2.055***	0.128	0.874***	2.396
Injured children	−0.488**	0.614	3.215***	24.903	0.369*	1.446	1.402***	4.063
Pseudo-science	−2.970***	0.051	−4.978***	0.007	−2.220***	0.109	−5.103***	0.006
Toxic ingredient	0.807***	2.241	−3.025***	0.049	0.287***	1.332	0.465**	1.592
Anti-maskwearing	−1.952***	0.142	−0.382	0.682	1.289***	3.629	−0.058	0.944
Vaccine-autism link	1.000***	2.718	−3.011***	0.049	−1.495***	0.224	0.131	1.140
Media propaganda	5.897***	363.944	0.898**	2.455	0.033	1.034	2.826***	16.878
Vaccine consent	2.967***	19.434	−7.823***	0.000	0.702***	2.018	−2.527***	0.080
Medical freedom	3.509***	33.415	−4.719***	0.009	−1.849***	0.157	−3.238***	0.039
Vaccine conspiracy	−3.541***	0.029	−2.300***	0.100	4.094***	59.979	−0.469	0.626
Vaccine exemption	4.877***	131.236	−3.837***	0.022	−1.642***	0.194	2.305***	10.024
Supporting Trump	7.110***	1224.148	13.207***	544160.762	−3.800***	0.022	6.578***	719.100
Global epidemic	0.578***	1.782	−2.544***	0.079	0.384***	1.468	−2.448***	0.086
Constant	−0.525***	0.592	−7.897	0.000	−0.947***	0.388	−10.327	0.000
N		234,251				234,253		
Log Likelihood		−974,530.200				−526,135.700		
AIC		1,949,146				1,052,357		

Note(s): $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

[a]Three outliers (for Retweet) and one outlier (for Like) were removed

Source(s): Table by authors

(Metaxas *et al.*, 2021). Thus, the stronger emotion of sadness surrounding the tweets, the stronger their endorsement. Furthermore, tweets amplifying the emotion of surprise are more attention-grabbing and getting more likes. In controversial and contradictory discussion around vaccine, surprise is a prevalent emotion since scientific skepticism and conspiracy theories are central to anti-vaccine movement. Meanwhile, the role of fear in audience engagement provides us a better understanding on the anti-vaxxers' misconceptions about vaccinations.

Targeting the emotion of fear to promote vaccine safety. As discussed above, the emotion of fear is positively associated with higher retweets of pro-vaccine content and higher likes of anti-vaccine content. The implication here is that on the one hand people were concerned about adverse effects and efficacy, on the other hand, they were motivated to get the vaccine because they wanted to protect others and themselves. This indicated that there is awareness that the vaccine is important despite concerns. This presents an excellent opportunity for public health professionals to design campaigns to promote vaccine acceptance. Campaigns should be developed to normalize health concerns and feelings of uncertainty, and to assist individuals in decision-making. Encouraging individuals to weigh established facts about the vaccine against vague uncertainties that driven by anxiety or intolerance of uncertainty could be an effective strategy for boosting vaccine acceptance. Further investigation is needed to determine ways in which fear may or may not explain vaccine hesitancy across different groups.

Topics and engagement

Among the top 15 topics of pro-vaxxers' tweets, tweets discussing "Child protection" and "COVID-19 situation" are linked with more retweet counts as well as like counts. For anti-vaxxers' corpus, tweets talking about "Supporting Trump," "Injured children," "COVID-19 situation," "Media propaganda" and "Community building" are associated with more retweet counts and like counts.

The most obvious finding to emerge from the analysis is the high engagement on children-related topics among both pro- and anti- vaccine tweets. Parents desire to have more detailed information about the benefits and risks of vaccines so that they can make informed decisions about their child's health care. Anti-vaxxers and their community are also aware of this chance to leverage parents' concerns and hesitations about vaccines. It is found that the narratives of injured children caused by vaccination have become the main anti-vaxxers' emblematic approach to the conversations about vaccine. We postulate that these narratives maintain high traffic and strong appeal because some anti-vaccine audiences seemed to feel a personal responsibility to spread the word about the dangers of vaccination. Anti-vaxxer communities provide a place for this externalization of individualist epistemologies. This also accords with one of our observations, which showed that anti-vaxxers' tweets on "Vaccine injury," "Injured children," "Pseudo-science," "Vaccine consent" and "Medical freedom" received significantly more likes when written in a positive tone. The fear inherent in parenting, as well as the anticipation to protect children, may be one of the important aspects to better understand the refusal of vaccination. Currently, there is less individual storytelling from the pro-vaxxers' side. A possible explanation is that parents who believe vaccines are effective in protecting children may not feel necessary to pay attention to this issue or voice it out online.

Promote partnerships with vaccine-hesitant parents in decision-making. These convincing narratives that highlight the manifold vaccine successes on child protection may be effective to mount a much stronger defense against anti-vaccine messaging. Previous research has established that the parent's motivation to vaccinate their children is influenced by social (Kestenbaum and Feemster, 2015). We should motivate parents who are in favor of vaccines to share their experiences and highlight the benefits of vaccines so that those who are hesitant about vaccines can restore their confidence in vaccination. Prior research has shown that tweets that share personal experiences or stories in health context tend to generate optimum online engagement, compared to factual information disseminated by health experts and journalists (Kothari *et al.*, 2022). For instance, some moms created a Facebook group named *Back to the Vax* to provide support for former antivax and vaccine-hesitant parents looking to start embracing vaccines (Doheny, 2021). We suggest that pediatricians and other health-care providers should acknowledge the parents' concern, prepare themselves to have informative conversations about immunizations and establish better communication to address parents' personal dilemmas and fears with respect and understanding.

We found that anti-vaxxers were putting effort into crafting a sense of community, which have successfully attracted the audiences to get involved. Online communities have been shown to play an increasing role in reinforcing existing beliefs, now are used by anti-vaxxers to cast doubt on vaccine efficacy. Our finding reiterates the importance of communal peer reinforcement as a powerful driver for this individual knowledge building, which could be leveraged by public health practitioners to engage the public in interactive conversations about vaccines. Public health agencies, health-care providers and communities all play a vital role in sharing the benefits and efficacy of vaccination. A recent survey reported that, of the participants who decided to get the vaccine after being initially less than certain, roughly half noted the role of their peers, friends and family members as well as their personal doctors in persuading them to get vaccinated (Kaiser Family Foundation, 2021). Hesitant individuals trust medical practitioners and faith group leaders to provide evidence-based knowledge, but they also rely on their peers/communities to provide emotion-based conviction for decisions.

Building a unified community to boost vaccine confidence. We believe it is important for pro-vaxxer groups to create a more connected and coordinated community around the shared targets toward vaccination and lifesaving. Engaging citizens and local communities will also aid in the development of appropriate vaccination strategies for the local context, overcoming certain logistical challenges and vaccine hesitancy. For example, the United Kingdom's COVID-19 vaccine delivery plan is based on a community-led approach, with collaborations across national government, local authorities, local public health professionals, faith-based leaders and other trusted community voice sectors (Department of Health and Social Care, 2021).

The result from our study shows that anti-vaccination online movement is highly conservative leaning, and the largest influential topic among anti-vaxxers' tweets to boost up the engagement is "Supporting Trump." One possible explanation is that our data collection was conducted close to the 2020 US presidential election. This is consistent with that of Gruzd *et al.* (2022) who reported that the former President Donald Trump plays an important role in the anti-vaccination network. During the COVID-19 outbreak, anti-vaxxers discussed more about "Supporting Trump" and also other topics such as "Pseudo-science," "Anti-maskwearing," "Media-propaganda" and "Vaccine conspiracy" compared to the time period before February 2020. Moreover, anti-vaxxers' tweets on "Supporting Trump" with a positive tone significantly received more likes, while their tweets on "Vaccine Conspiracy" with a positive tone also attracted more endorsement. These results confirmed that the distrust of conspiratorial people may be associated with a generic belief system characterized by negative attitudes toward identified antagonistic groups, such as pharmaceutical corporations, government officials and health-care professionals. Among the crowd, the conspiracy theorists and vaccine skeptics started to mingle with so-called libertarians, while calling out to critic the corporations, elites and experts. Studies across countries have found that vaccine-hesitant people are more inclined to vote for politically extreme parties and to distrust the government (Kennedy, 2019; Qunaibi *et al.*, 2021). These mistrust and conspiracy belief topics have already attracted more audience engagement on tweets, which may potentially amplifying suspicion, mistrust and fear of vaccination, thus could be an obstacle to COVID-19 recovery across the globe.

It is worth noting that COVID-related tweets remained a highly engaged topic since February 2020, as the ongoing global pandemic was new, and many uncertainties remained. People have been actively seeking and engaging with information related to vaccines and other COVID-related topics. Based on our observations, the topic of "Scientific Research" received a large number of likes on Twitter, and the topics of "COVID-19 situation," "Institutional support," "Vaccine development," "Vaccine stops virus" and "Public health professionals" generated tremendous high engagement from audiences among pro-vaxxers' tweets. Together these findings suggested that the popularization of scientific knowledge, such as filling in the data gaps and providing as much timely and accurate information as possible, can be effective in increasing the reach and impact. As we start on the long path to pandemic recovery, more work needs to be done to promote vaccination, to confront science skepticism and conspiracy beliefs, to combat anti-intellectualism, and to restore the trust in medical science and professionals.

Fostering transparency and integrity in vaccine communication. The simplest step public health professionals can take immediately to restore the public trust is making sure information about and results of research into treatments and vaccines are communicated transparently and comprehensively. Previous research has highlighted how a lack of high-quality information and timely data can cause uncertainty in decision-making and foster mistrust in the population during pandemic (Kricorian *et al.*, 2022). We suggest that public health agencies keep paying attention to the pattern, to monitor the online discussions on social media, to reduce the information gap between the needs of the public and the information the agencies have provided and to help people become more fluent about vaccinations. In particular, authorities should ensure that communications regarding potential adverse reactions are handled with care, in order to avoid reinforcing hesitant people's cognitive biases. Overall, the success of pro-vaccine programs depends largely on the public's trust in the effectiveness and safety of the vaccine, the reliability and competence of the institutions responsible for delivering them, and the integrity, openness and fairness in the process.

Limitation and future directions

Our approach has several limitations, and future studies could build and improve on our approach. First, although our snowball sampling approach yields some benefits of providing fine-grained data concentrating on pro- and anti-vaccination discussion, it might lead to the generalization issue. The findings as we addressed above could be amplified by the pro- and anti-vaccination online enclave and echo chamber (Vanderslott, 2019). Second, given the limitation of Twitter's application programming interface, we could not distinguish the retweet or favorite counts of each tweet contributed by either an actual user or by a bot, which could hugely affect the results of our regression modeling. Lastly, we recognize that our emotion classification algorithm achieved a relatively low F1 score (0.647), which can be attributed to the heterogeneity and complexity of emotions across individuals and contexts. In the future, a lexicon-guided hybrid pre-trained model could be employed, taking advantage of their respective strengths. Future research should consider extending the range of selected data and collecting more data to provide a more comprehensive picture of the attitudes and opinions of these two camps. Additionally, studies could also combine the user-generated content from multiple social media platforms to ensure a panoramic coverage of content across the Internet arena instead of only focusing on Twitter. Furthermore, future studies can use human coders to manually examine and compare detailed strategies or narrative used in the popular vaccine topics, as well as investigate if a combination of techniques used in vaccine messages can lead to a higher level of engagement.

Notes

1. To be noted, for the part of sentiment/emotion classification, we did not perform any of the pre-processing, including the removal of emoji.
2. For the description of each type of emotions, please refer to Part I in the supplemental material.
3. For the details of the evaluation metrics, please refer to the Part II in the supplemental material.
4. For the details of selecting and determining the optimal topic model for both corpora, please refer to Part IV in the supplemental materials.
5. For the AIC comparison of each model, please see Part V in the supplemental materials.
6. For the descriptive statistics of the pro-vaxxer's and anti-vaxxer's tweets, please refer to Part II in the supplemental material.
7. The results of the full model are presented in Part VI of Supplemental material.

References

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M. and Shah, Z. (2020), "Top concerns of tweeters during the COVID-19 pandemic: infoveillance study", *Journal of Medical Internet Research*, Vol. 22 No. 4, e19016, doi: 10.2196/19016.

Alhayan, F., Pennington, D. and Ayouni, S. (2023), "Twitter use by the dementia community during COVID-19: a user classification and social network analysis", *Online Information Review*, Vol. 47 No. 1, pp. 41-58, doi: 10.1108/OIR-04-2021-0208.

Blei, D.M. (2012), "Probabilistic topic models", *Communications of the ACM*, Vol. 55 No. 4, pp. 77-84, doi: 10.1145/2133806.2133826.

Bricker, B. and Justice, J. (2019), "The postmodern medical paradigm: a case study of anti-MMR vaccine arguments", *Western Journal of Communication*, Vol. 83 No. 2, pp. 172-189, doi: 10.1080/10570314.2018.1510136.

Briones, R., Nan, X., Madden, K. and Waks, L. (2012), "When vaccines go viral: an analysis of HPV vaccine coverage on YouTube", *Health Communication*, Vol. 27 No. 5, pp. 478-485, doi: 10.1080/10410236.2011.610258.

Buffard, J. and Papasava, A. (2020), "A quantitative study on the impact of emotion on social media engagement and conversion", *Journal of Digital and Social Media Marketing*, Vol. 7 No. 4, pp. 355-375, available at: https://www.jonathanbuffard.com/uploads/1/2/3/9/123956980/socialmediastudy_buffard_and_papasava.pdf (accessed 18 March 2021).

Center for Disease Control and Prevention (2021), "Estimating vaccine hesitancy for COVID-19", available at: https://data.cdc.gov/stories/s/cnd2-a6zw (accessed 18 March 2021).

Daou, H. (2021), "Sentiment of the public: the role of social media in revealing important events", *Online Information Review*, Vol. 45 No. 1, pp. 157-173, doi: 10.1108/OIR-12-2019-0373.

Department of Health and Social Care, United Kingdom (2021), "UK COVID-19 vaccine uptake plan", available at: https://www.gov.uk/government/publications/covid-19-vaccination-uptake-plan/uk-covid-19-vaccine-uptake-plan (accessed 18 March 2021).

Doheny, K. (2021), "Changing minds: what moves the needle for the unvaccinated?", WebMD, 2 September, available at: https://www.webmd.com/special-reports/covid-vaccine-hesitancy/20210902/changing-minds-what-moves-the-needle-for-the-unvaccinated (accessed 18 March 2021).

Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R. and Bettinger, J.A. (2013), "Vaccine hesitancy: an overview", *Human Vaccines and Immunotherapeutics*, Vol. 9 No. 8, pp. 1763-1773, doi: 10.4161/hv.24657.

Dyda, A., Shah, Z., Surian, D., Martin, P., Coiera, E., Dey, A., Leask, J. and Dunn, A.G. (2019), "HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among

Australian Twitter users", *Human Vaccines and Immunotherapeutics*, Vol. 15 Nos 7-8, pp. 1488-1495, doi: 10.1080/21645515.2019.1596712.

Grant, L., Hausman, B.L., Cashion, M., Lucchesi, N., Patel, K. and Roberts, J. (2015), "Vaccination persuasion online: a qualitative study of two provaccine and two vaccine-skeptical websites", *Journal of Medical Internet Research*, Vol. 17 No. 5, p. e133, doi: 10.2196/jmir.4153.

Gruzd, A., Mai, P. and Soares, F.B. (2022), "How coordinated link sharing behavior and partisans' narrative framing fan the spread of COVID-19 misinformation and conspiracy theories", *Social Network Analysis and Mining*, Vol. 12 No. 118, doi: 10.1007/s13278-022-00948-y.

Guidry, J.P., Carlyle, K., Messner, M. and Jin, Y. (2015), "On pins and needles: how vaccines are portrayed on Pinterest", *Vaccine*, Vol. 33 No. 39, pp. 5051-5056, doi: 10.1016/j.vaccine.2015.08.064.

Hargittai, E. and Hsieh, Y.L.P. (2010), "Predictors and consequences of differentiated practices on social network sites", *Information, Communication and Society*, Vol. 13 No. 4, pp. 515-536, doi: 10.1080/13691181003639866.

Heiss, R., Schmuck, D. and Matthes, J. (2019), "What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions", *Information, Communication and Society*, Vol. 22 No. 10, pp. 1497-1513, doi: 10.1080/1369118X.2018.1445273.

Hoffman, B.L., Felter, E.M., Chu, K.H., Shensa, A., Hermann, C., Wolynn, T., Williams, D. and Primack, B.A. (2019), "It's not all about autism: the emerging landscape of anti-vaccination sentiment on Facebook", *Vaccine*, Vol. 37 No. 16, pp. 2216-2223, doi: 10.1016/j.vaccine.2019.03.003.

Hwong, Y.L., Oliver, C., Van Kranendonk, M., Sammut, C. and Seroussi, Y. (2017), "What makes you tick? The psychology of social media engagement in space science communication", *Computers in Human Behavior*, Vol. 68, pp. 480-492, doi: 10.1016/j.chb.2016.11.068.

Johnson, N.F., Velásquez, N., Restrepo, N.J., Leahy, R., Gabriel, N., El Oud, S., Zheng, M., Manrique, P., Wuchty, S. and Lupu, Y. (2020), "The online competition between pro-and anti-vaccination views", *Nature*, Vol. 582, pp. 230-233, doi: 10.1038/s41586-020-2281-1.

Joo, S., Lu, K. and Lee, T. (2020), "Analysis of content topics, user engagement and library factors in public library social media based on text mining", *Online Information Review*, Vol. 44 No. 1, pp. 258-277, doi: 10.1108/OIR-11-2018-0345.

Kaiser Family Foundation (2021), "KFF COVID-19 vaccine monitor: in their own words, six months later", available at: https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-in-their-own-words-six-months-later/ (accessed 18 March 2021).

Kang, G.J., Ewing-Nelson, S.R., Mackey, L., Schlitt, J.T., Marathe, A., Abbas, K.M. and Swarup, S. (2017), "Semantic network analysis of vaccine sentiment in online social media", *Vaccine*, Vol. 35 No. 29, pp. 3621-3638, doi: 10.1016/j.vaccine.2017.05.052.

Kantar (2021), "COVID-19 vaccine faces an increasingly hesitant public", available at: https://www.kantar.com/inspiration/coronavirus/covid-19-vaccine-faces-an-increasingly-hesitant-public (accessed 18 March 2021).

Kennedy, J. (2019), "Populist politics and vaccine hesitancy in Western Europe: an analysis of national-level data", *European Journal of Public Health*, Vol. 29 No. 3, pp. 512-516, doi: 10.1093/eurpub/ckz004.

Kestenbaum, L.A. and Feemster, K.A. (2015), "Identifying and addressing vaccine hesitancy", *Pediatric Annals*, Vol. 44 No. 4, pp. e71-e75, doi: 10.3928/00904481-20150410-07.

Kothari, A., Walker, K. and Burns, K. (2022), "# CoronaVirus and public health: the role of social media in sharing health information", *Online Information Review*, Vol. 46 No. 7, pp. 1293-1312, doi: 10.1108/OIR-03-2021-0143.

Kricorian, K., Civen, R. and Equils, O. (2022), "COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety", *Human Vaccines and Immunotherapeutics*, Vol. 18 No. 1, 1950504, doi: 10.1080/21645515.2021.1950504.

Larson, H.J. (2018), "The biggest pandemic risk? Viral misinformation", *Nature*, Vol. 562, pp. 309-310, doi: 10.1038/d41586-018-07034-4.

Larson, H.J., Cooper, L.Z., Eskola, J., Katz, S.L. and Ratzan, S. (2011), "Addressing the vaccine confidence gap", *The Lancet*, Vol. 378 No. 9790, pp. 526-535, doi: 10.1016/S0140-6736(11)60678-8.

Larson, H.J., Smith, D.M., Paterson, P., Cumming, M., Eckersberger, E., Freifeld, C.C., Ghinai, I., Jarrett, C., Paushter, L., Brownstein, J.S. and Madoff, L.C. (2013), "Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines", *The Lancet Infectious Diseases*, Vol. 13 No. 7, pp. 606-613, doi: 10.1016/S1473-3099(13)70108-7.

Lee, E.J. and Jang, Y.J. (2010), "What do others' reactions to news on internet portal sites tell us? Effects of presentation format and readers' need for cognition on reality perception", *Communication Research*, Vol. 37 No. 6, pp. 825-846, doi: 10.1177/0093650210376189.

Mathew, L. and Bindu, V.R. (2020), "A review of natural language processing techniques for sentiment analysis using pre-trained models", *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 340-345, IEEE, doi: 10.1109/ICCMC48092.2020.ICCMC-00064.

Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M. and Finn, S. (2021), "What do retweets indicate? Results from user survey and meta-review of research", *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 9 No. 1, pp. 658-661, available at: https://ojs.aaai.org/index.php/ICWSM/article/view/14661 (accessed 18 March 2021).

Nguyen, D.Q., Vu, T. and Nguyen, A.T. (2020), "BERTweet: a pre-trained language model for English Tweets", arXiv preprint arXiv:2005.10200, doi: 10.48550/arXiv.2005.10200.

Ortiz, R.R., Smith, A. and Coyne-Beasley, T. (2019), "A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness, knowledge, and attitudes about HPV and HPV vaccination", *Human Vaccines and Immunotherapeutics*, Vol. 15 Nos 7-8, pp. 1465-1475, doi: 10.1080/21645515.2019.1581543.

Pérez, J.M., Giudici, J.C. and Luque, F. (2021), "Pysentimiento: a python toolkit for sentiment analysis and socialnlp tasks", arXiv preprint arXiv:2106.09462, doi: 10.48550/arXiv.2106.09462.

Pedersen, E.A., Loft, L.H., Jacobsen, S.U., Søborg, B. and Bigaard, J. (2020), "Strategic health communication on social media: insights from a Danish social media campaign to address HPV vaccination hesitancy", *Vaccine*, Vol. 38 No. 31, pp. 4909-4915, doi: 10.1016/j.vaccine.2020.05.061.

Qunaibi, E.A., Helmy, M., Basheti, I. and Sultan, I. (2021), "A high rate of COVID-19 vaccine hesitancy in a large-scale survey on Arabs", *elife*, Vol. 10, e68038, doi: 10.7554/eLife.68038.

Schmidt, A.L., Zollo, F., Scala, A., Betsch, C. and Quattrociocchi, W. (2018), "Polarization of the vaccination debate on Facebook", *Vaccine*, Vol. 36 No. 25, pp. 3606-3612, doi: 10.1016/j.vaccine.2018.05.040.

Smith, N. and Graham, T. (2019), "Mapping the anti-vaccination movement on Facebook", *Information, Communication and Society*, Vol. 22 No. 9, pp. 1310-1327, doi: 10.1080/1369118X.2017.1418406.

Stecula, D.A., Kuru, O. and Jamieson, K.H. (2020), "How trust in experts and media use affect acceptance of common anti-vaccination claims", *Harvard Kennedy School Misinformation Review*, Vol. 1 No. 1, pp. 1-11, doi: 10.37016/mr-2020-007.

Thelwall, M., Buckley, K. and Paltoglou, G. (2011), "Sentiment in twitter events", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 2, pp. 406-418, doi: 10.1002/asi.21462.

van Atteveldt, W., Van der Velden, M.A. and Boukes, M. (2021), "The validity of sentiment analysis: comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms", *Communication Methods and Measures*, Vol. 15 No. 2, pp. 121-140, doi: 10.1080/19312458.2020.1869198.

Vanderslott, S. (2019), "Exploring the meaning of pro-vaccine activism across two countries", *Social Science and Medicine*, Vol. 222, pp. 59-66, doi: 10.1016/j.socscimed.2018.12.033.

Walsh, M.J., Baker, S.A. and Wade, M. (2022), "Evaluating the elevation of authoritative health content online during the COVID-19 pandemic", *Online Information Review*, ahead-of-print, doi: 10.1108/OIR-12-2021-0655.

Ward, J.K., Peretti-Watel, P., Larson, H.J., Raude, J. and Verger, P. (2015), "Vaccine-criticism on the internet: new insights based on French-speaking websites", *Vaccine*, Vol. 33 No. 8, pp. 1063-1070, doi: 10.1016/j.vaccine.2014.12.064.

Weber, P. (2014), "Discussions in the comments section: factors influencing participation and interactivity in online newspapers' reader comments", *New Media and Society*, Vol. 16 No. 6, pp. 941-957, doi: 10.1177/1461444813495165.

Xu, Z. and Guo, H. (2018), "Using text mining to compare online pro-and anti-vaccine headlines: word usage, sentiments, and online popularity", *Communication Studies*, Vol. 69 No. 1, pp. 103-122, doi: 10.1080/10510974.2017.1414068.

Xu, Z., Ellis, L. and Umphrey, L.R. (2019), "The easier the better? Comparing the readability and engagement of online pro-and anti-vaccination articles", *Health Education and Behavior*, Vol. 46 No. 5, pp. 790-797, doi: 10.1177/1090198119853614.

Yaqub, O., Castle-Clarke, S., Sevdalis, N. and Chataway, J. (2014), "Attitudes to vaccination: a critical review", *Social Science and Medicine*, Vol. 112, pp. 1-11, doi: 10.1016/j.socscimed.2014.04.018.

Yiannakoulias, N., Slavik, C.E. and Chase, M. (2019), "Expressions of pro-and anti-vaccine sentiment on YouTube", *Vaccine*, Vol. 37 No. 15, pp. 2057-2064, doi: 10.1016/j.vaccine.2019.03.001.

Yuan, X., Schuchard, R.J. and Crooks, A.T. (2019), "Examining emergent communities and social bots within the polarized online vaccination debate in Twitter", *Social Media+ Society*, Vol. 5 No. 3, pp. 1-12, doi: 10.1177/2056305119865465.

Zeileis, A., Kleiber, C. and Jackman, S. (2008), "Regression models for count data in R", *Journal of Statistical Software*, Vol. 27 No. 8, pp. 1-28, doi: 10.18637/jss.v027.i08.

Zimet, G.D., Rosberger, Z., Fisher, W.A., Perez, S. and Stupiansky, N.W. (2013), "Beliefs, behaviors and HPV vaccine: correcting the myths and the misinformation", *Preventive Medicine*, Vol. 57 No. 5, pp. 414-418, doi: 10.1016/j.ypmed.2013.05.013.

Supplemental material

The supplementary material for this article can be found online.

About the authors

Rachel X. Peng's research interests lie at the intersection of media effects, health and strategic communication. Her work to date has focused on using the interdisciplinary application of computer science tools to address health-related questions and problems.

Ryan Yang Wang is a researcher whose work primarily focuses on the intersectionality of political communication, ICT and democracy and computational social science. Ryan Yang Wang is the corresponding author and can be contacted at: ryanyangwang@gmail.com