

RESEARCH ARTICLE

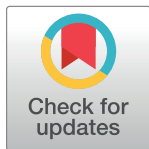
Exploring vaccine hesitancy in digital public discourse: From tribal polarization to socio-economic disparities

Huzeyfe Ayaz¹ , Muhammed Hasan Celik² , Huseyin Zeyd Koytak³ , Ibrahim Emre Yanik⁴ 

1 Department of Informatics, Technical University of Munich, Garching, Munich, Germany, **2** Department of Computer Science Center for Complex Biological Systems, University of California Irvine, Irvine, CA, United States of America, **3** Department of Sociology, Syracuse University, Syracuse, NY, United States of America, **4** Department of Sociology, Ibn Haldun University, Istanbul, Turkey

 These authors contributed equally to this work.

* mcelik@uci.edu (MHC); hkoytak@syr.edu (HZK)



Abstract

This study analyzed online public discourse on Twitter (later rebranded as X) during the COVID-19 pandemic to understand key factors associated with vaccine hesitancy by employing deep-learning techniques. Text classification analysis reveals a significant association between attitudes toward vaccination and the unique socio-economic characteristics of US states, such as education, race, income or voting behavior. However, our results indicate that attributing vaccine hesitancy solely to a single social factor is not appropriate. Furthermore, the topic modeling of online discourse identifies two distinct sets of justifications for vaccine hesitancy. The first set pertains to political concerns, including constitutional rights and conspiracy theories. The second pertains to medical concerns about vaccine safety and efficacy. However, vaccine-hesitant social media users pragmatically use broad categories of justification for their beliefs. This behavior may suggest that vaccine hesitancy is influenced by political beliefs, unconscious emotions, and gut-level instinct. Our findings have further implications for the critical role of trust in public institutions in shaping attitudes toward vaccination and the need for tailored communication strategies to restore faith in marginalized communities.

OPEN ACCESS

Citation: Ayaz H, Celik MH, Koytak HZ, Yanik IE (2024) Exploring vaccine hesitancy in digital public discourse: From tribal polarization to socio-economic disparities. PLoS ONE 19(11): e0308122. <https://doi.org/10.1371/journal.pone.0308122>

Editor: Constantine Dovrolis, Georgia Institute of Technology, UNITED STATES OF AMERICA

Received: October 3, 2023

Accepted: July 14, 2024

Published: November 5, 2024

Copyright: © 2024 Ayaz et al. 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 author and source are credited.

Data Availability Statement: <https://doi.org/10.5281/zenodo.7876072>.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Skepticism about the safety and benefits of vaccines is linked to low trust in institutions [1, 2]. Vaccine hesitancy, along with climate change denial, is part of a larger trend of mistrust in scientific expertise and a decline in trust in public institutions [3–6]. To combat the “infodemic” the World Health Organization (WHO) has worked with major social media platforms to redirect internet users to reliable websites when searching for information related to COVID-19 [7]. This approach is based on the premise that exposing individuals to verified knowledge from authoritative sources can help mitigate vaccine hesitancy. However, social scientists have

long critiqued the deficit perspective—which suggests a direct correlation between scientific literacy and public trust in science—for its implicit assumption of the cultural superiority of scientific knowledge over other forms of knowledge [8]. They argue that this viewpoint neglects the responsibility of scientists to make science more accessible to the public [9]. Furthermore, social scientists emphasize the importance of recognizing that different social groups possess unique cultural perspectives. These perspectives significantly influence how groups interact with social institutions [10].

Restoring confidence in scientific institutions is therefore an essential crucial democratic task for policymakers. This article endeavors to furnish those in positions of responsibility with a comprehensive understanding, while acknowledging the intricate nature of vaccine hesitancy and cautioning against the temptation of oversimplifying the issue by attributing it to a single social parameter, a mistake that many prominent media outlets have made [11–14]. Research on social media content often perceive vaccine attitudes within the digital public sphere as reflecting partisan divides [15, 16]. However, this perspective overlooks the more complex phenomenon of tribal polarization, which encompasses not just political allegiance but a deeper, multifaceted divide. Tribal polarization involves distinct groups that coalesce around shared identities, values, and beliefs, each characterized by different socio-economic variables. This formation leads to significant and expanding rifts in attitudes and behaviors among users within the digital public sphere. These communities, identifiable by their unique socio-economic markers, exemplify how divisions extend beyond mere political affiliations, reflecting deeper, more complex layers of tribal polarization.

Although there is no single universal determinant for vaccine hesitancy, a growing body of literature has identified factors that influence vaccine acceptance [17]. The Strategic Advisory Group of Experts (SAGE) on Immunization has categorized these determinants into three main domains, providing a framework for exploring the complex and multifaceted nature of vaccine hesitancy [18]. Addressing the contextual, individual/social group, and vaccine-specific determinants of vaccine hesitancy can help to overcome mistrust in vaccines and improve population health outcomes. Factors affecting vaccine acceptance are distinguished at different levels in scientific literature [19]. Our research focuses on social determinants of vaccine hesitancy at the population level, rather than on the socio-cultural context that influences individual decision-making.

A substantial body of literature has emerged on vaccine hesitancy to understand the factors that influence the uptake of public health interventions. Researchers have studied vaccine hesitancy to identify the pathways and mediators that contribute to a lack of public consent and resistance to various vaccines [17, 20–23]. A key finding has been that vaccine hesitancy takes various forms and manifests differently [24], and that these vaccine attitudes are unequally distributed among different segments of society [25] and across nations [26], exacerbating existing health disparities [27].

The attitudes toward vaccines are diverse and constantly evolving, making it important to understand their complexity [28, 29]. The views on vaccination have been found to be influenced by various sources of information, including the growing influence of digital public spheres [30–36]. Recent research has highlighted the significance of gathering and analyzing data from social media platforms in order to track the rapidly changing trends [35, 37], polarizations [35, 38], discourses [39–41], and sentiments [42–44] related to vaccine attitudes.

Recent advancements in computational social research have opened up new avenues for social research, particularly in analyzing the competing discourses in digital spheres. Social media data has been used to analyze societal attitudes toward critical social problems, such as immigration, public health, and extremism, due to its cost-effectiveness and ability to eliminate response biases [45]. We focused on analyzing vaccine hesitancy on Twitter (later

rebranded as X), one of the largest social media platforms to discuss public discourse at the time of this study. The mining digital public discourse provides several advantages as a data source in social research compared to conventional surveys. Firstly, it allowed us to analyze attitudes toward vaccines at a relatively low cost. Secondly, computational tools helped us uncover hidden patterns in a large number of online posts messages with minimal human intervention. Unsupervised textual clustering methods provided an insider's perspective on vaccine hesitancy while minimizing potential biases that could arise from the researchers themselves.

Past research has analyzed social media data to identify indicators of vaccine hesitancy within the text of publicly shared posts. These studies then correlated the geographically tagged posts with survey data to assess levels of vaccine hesitancy [46]. Another study leveraging survey data has revealed significant associations between sociodemographic factors and vaccine hesitancy, specifically highlighting the influence of trust in the Trump administration on vaccine hesitancy [47]

Our research aims to unravel the complexity of vaccine-hesitant attitudes and the diversity of vaccine-hesitant populations. Although our study period was marked by COVID-19 related discussions, we intentionally chose keywords that were not confined solely to the COVID-19 vaccine. This approach positions our research to illuminate aspects of vaccine hesitancy that transcend the COVID-19 context, potentially applying to both past and future episodes of vaccine skepticism. We contend that vaccine hesitancy is a complex issue, characterized by the nuanced positions of vaccine-hesitant individuals on social media, who often cite a variety of reasons for their skepticism. Therefore, we advise against simplistic interpretations of media narratives or studies that attribute vaccine hesitancy to a single social factor. Our study aims to contribute fresh perspectives that enhance our understanding of the factors driving tribal polarization in the U.S., re-evaluating the connections between vaccine hesitancy and social determinants to foster better public health strategies.

Materials and methods

All the analyses in the study were conducted in Python (version = 3.8), and PyTorch [48] (version = 1.13.1) deep learning library was used in the text classification and topic modeling. The maps were plotted with GeoPandas [49] and geographic boundaries were obtained from the US Census [50].

Fetching tweets

We utilize the Academic API to access Twitter data. In order to fetch tweets, we utilized searchtweets-v2 [51] and Tweepy [52] python packages. Tweets were fetched using the search endpoint of the Twitter-API (<https://api.twitter.com/2/tweets/search/all>). We fetched all tweets containing the keywords: "Vaccine" or "Vaccination". We limited our analysis to those English tweets posted between 2020 and 2022. We excluded retweets, short, (less than 10 words), and duplicate tweets. Then we used the geo-location identifier of tweets to locate the state where tweets were posted from by using Tweepy library. The use of the data complies with the terms and conditions of the Twitter research grant program.

Dataset and model training for vaccination attitude prediction

To identify individuals who express vaccine hesitancy, tweets were classified into one of three categories of "Positive", "Negative", or "Neutral" by using BERT-based natural language processing (NLP) models. BERT embeddings were extracted using Covid-Twitter-Bert-V2 [53] model from HuggingFace [54] platform. Then, the model was fine-tuned on the dataset by

Pananos et al. (2017) [55] to predict vaccination attitude. The dataset consists of 27,906 tweets which were manually labeled by Amazon Mechanical Turk. The annotators of the crowd-sourced dataset have a consensus for only 16,156 tweets. The dataset is partitioned into the train, validation, and test sets with proportions of 80%, 10%, and 10%, respectively. To create the validation and test sets, we only included tweets with consensus. The remaining tweets, with or without consensus, were used for fine-tuning with SimpleTransformers library [56]. During the fine-tuning, the model was trained for 20 epochs with the early stopping of 5 epochs. To optimize our model's performance, we conducted hyperparameter tuning for several parameters. The key parameters that we hyperparameter-tuned were learning rate, epsilon for Adam optimization, and maximum sequence length. For the learning rate, we experimented with three values: 1e-3, 1e-4, and 1e-5. Similarly, we tried three values for epsilon, 1e-7, 1e-8, and 1e-9. Also, class weights were calculated based on the inverse proportion of class frequencies to ensure balanced training. The weights used were 1.0, 1.43, and 9.2 for Neutral, Positive, and Negative, respectively. The best score was obtained with the parameters of 1e-05 for learning rate, 1e-07 for epsilon, and 128 for maximum sequence length. The performance of our model was evaluated using the f1-macro metric using scikit-learn [57] during the hyperparameter tuning.

Model evaluation of vaccination attitude classification

We evaluated the vaccination attitude classification model using the precision-recall area under the curve (PR-AUC). The area under the precision-recall curve is calculated based on the average precision:

$$auPRC = \sum_n (R_n - R_{n-1})P_n \quad (1)$$

where P_n is precision and R_n is recall at a threshold n .

Prediction of attitudes toward vaccination score

We used the fine-tuned model to predict the label of all fetched tweets and only used high-confidence predictions of the model where the predicted class probability is above 99%. To analyze spatial and temporal dimensions of vaccination hesitancy, we defined the attitudes toward vaccination (ATV) score, which is log odds ratio (OR) positive p to negative n tweets for specific category c (place or date):

$$ATV = \log\left(\frac{p_c/p_n}{p_{other}/n_{other}}\right) \quad (2)$$

To obtain the statistical significance of attitudes toward vaccination scores by the US states, we also employed Fisher's Exact Test [58] with categories of attitudes (positive and negative) and the state tweet was sent from. P-values of Fisher's Exact test were corrected for multiple testing using Bonferroni correction method.

Socio-economic parameters and ATV score

To investigate the relationship between vaccine hesitancy and socio-economic parameters, we used 2020 American Community Survey 1-Year Experimental Estimates data by the U.S. Census Bureau. The parameters included the percentage of Black population (Black percentage), the percentage of people living in poverty in the past 12 months (Poverty), inflation-adjusted median household income in the past 12 months (Median income), employment status for

individuals aged 16 years and over (Unemployment), and citizenship status (Non-native). We also studied the educational attainment of the population aged 25 years and over, specifically the percentage of those with less than a high school degree (Education1) and those with more than a bachelor's degree (Education2). Additionally, we examined the influence of presidential elections (Election) on attitudes toward vaccination using the results of a study [59]. We also included Social Capital (PSU-SC) index [60] and Cat/Dog ownership ratio by the US states [61] as a dichotomous variable ($\log\left(\frac{\text{cat}}{\text{dog}}\right)$).

We calculated Spearman's rank correlation coefficient for each socio-economic parameter. To find the error bars that are corrected for multiple testing, we used the formula in Eq 2:

$$\tanh\left(\operatorname{arctanh}(r) \pm \frac{2.576}{\sqrt{n-3}}\right) \quad (3)$$

where r is the estimate of the correlation and n is the sample size. 2.576 is the adjusted z-score for a 99% confidence interval with the Bonferroni correction.

Partial Least Squares (PLS) Regression

For the multivariate analysis, we used Partial Least Squares (PLS) Regression [62] to avoid multicollinearity between the socio-economic parameters. Bootstrapping was conducted for 10,000 iterations to obtain error bars for regression coefficients with 99% confidence interval. In each bootstrapping iteration, the model was fitted on the random subset of states.

Contextualized Topic Modeling (CTM)

We utilized the Contextualized Topic Modeling (CTM) [63] library, which is a Variational AutoEncoder (VAE) [64] based deep learning model. CTM takes BERT embeddings as an input and predicts a bag of words (BoW) by sampling from the bottleneck layer. The bottleneck size is equal to the number of topics. This approach enabled us to assign a topic to each tweet with an unsupervised method. We conducted topic modeling with only negative tweets. NLTK [65] library is utilized to remove stop words. The number of topics, hidden layer sizes of the encoder and the decoder, and the dropout rate are hyperparameter tuned. We tuned topics between 5 to 10, unit size of the hidden layer of 200, 500, and 700, and dropout rates of 0.2, 0.5, and 0.8. All hyperparameter tuning experiments were done in 2 epochs. The best hyperparameters were chosen based on topic coherence and diversity metrics. We selected the topic number as 8, a hidden layer size as 200, and a dropout as 0.2. Topic coherence is the average normalized pointwise mutual information (NPMI) score between the top keywords in a topic. The formula for topic diversity is defined as follows:

$$TD = \frac{|U|}{k \cdot T} \quad (4)$$

where TD is the topic diversity score, $|U|$ is the number of unique words in the corpus, k is the number of top keywords of each topic, and T is the total number of topics in the model. Top keywords for each topic are generated with 20 steps of post-training sampling from the bottleneck layer. Keywords with high Pointwise Mutual Information (PMI) score (greater than 6) were merged into a phrase. We assigned tweets to topics when the topic probability was marginally greater (at least 12.5%) than other topic probabilities, and then the remaining unassigned tweets were filtered.

Results

The study analyzed tweets from the first two years of the COVID-19 pandemic to gain insights into the online public discourse surrounding vaccination on Twitter. Tweets in the dataset were categorized based on their vaccination attitude, and the findings were presented in three parts: temporal analysis, spatial analysis, and topic modeling. The temporal analysis investigated the stability or variability of attitudes over time. The spatial comparison of geolocated tweets with socio-economic parameters revealed factors associated with vaccine hesitancy. Finally, the study used topic modeling of tweets to explore the most common justifications for vaccine hesitancy.

We utilized the Twitter Search v2 API (Methods) to fetch all English tweets containing the keywords 'vaccine' and 'vaccination' from 2020 to 2022. Our resulting dataset comprises approximately 53 million tweets from around 8.2 million users, with roughly 584 thousand geolocated tweets for spatial analysis Fig 1A. Most of the tweets in our dataset predominantly reflect the Twitter landscape of the United States. During the Eastern Standard Time (EST)

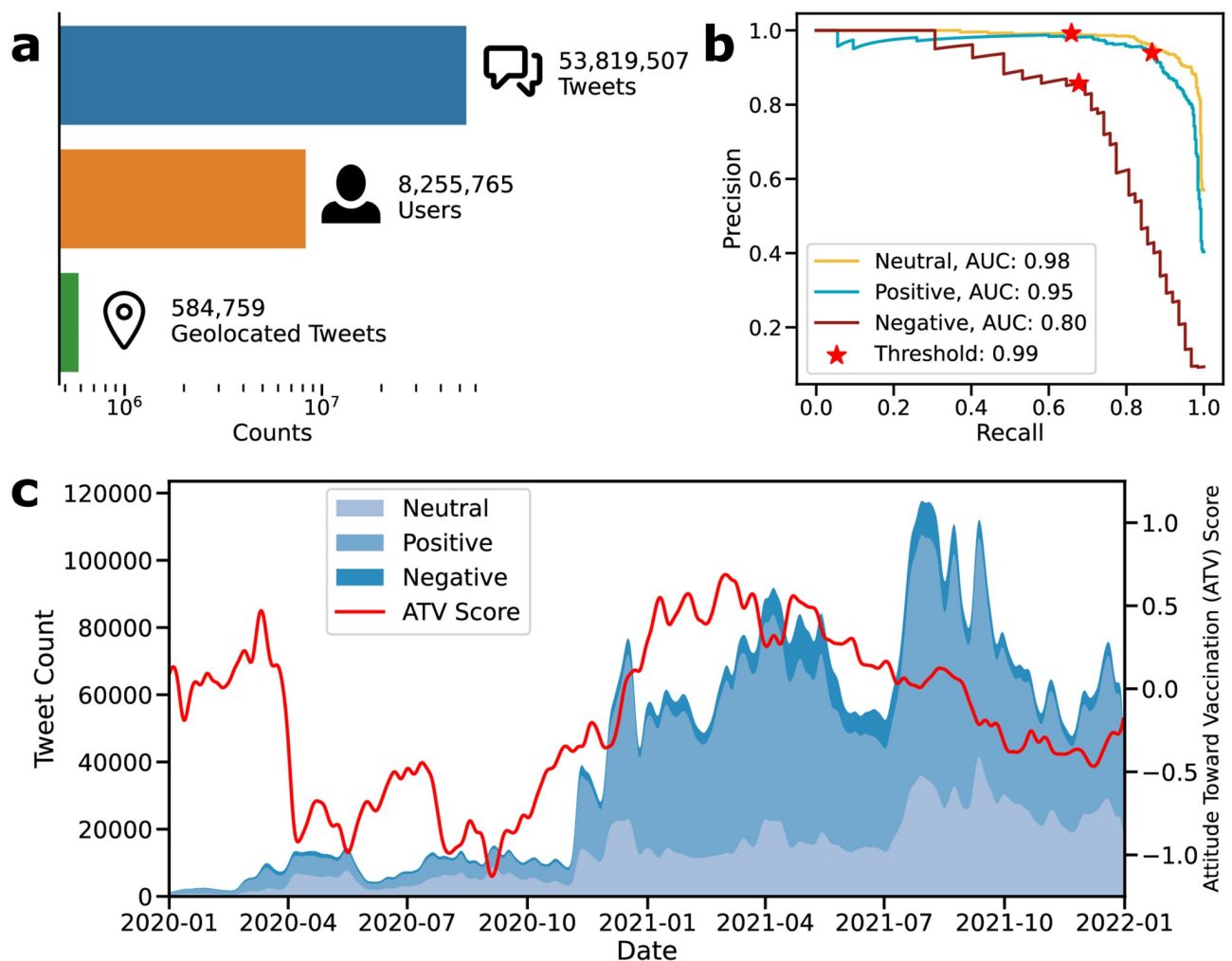


Fig 1. Dataset, attitude prediction. A: Descriptive statistics of the Twitter vaccination dataset, such as the number of tweets, geolocated tweets, and users. B: The precision-recall curve and AUC scores of the model for vaccination attitude classification. The red star indicates the threshold used in this study. C: Temporal distribution of the number of neutral, positive, and negative tweets, and trend of the attitudes toward vaccination (ATV) scores.

<https://doi.org/10.1371/journal.pone.0308122.g001>

daytime hours, we observed a peak in the number of tweets [S1 Fig](#), and a majority of geolocated tweets originated from the United States ([S1 Table in S1 File](#)). To eliminate any possible bias from non-US tweets, we further subset the dataset to solely include geolocated tweets originating from the United States for spatial analysis. We utilized the COVID-Twitter-BERT model [\[53\]](#) to obtain embedding vectors, which are numerical representations of tweets that capture their semantic meaning [\[66\]](#). These embeddings facilitated downstream language tasks, such as classifying vaccination attitudes and conducting topic modeling.

A classification model was trained to predict attitudes toward vaccination by fine-tuning the BERT model on the previously annotated vaccination sentiment dataset [\[55\]](#) ([Methods](#)). The model achieved high performance, with PR-AUC scores of 0.98 for neutral, 0.95 for positive, and 0.8 for negative classes on the test set [Fig 1B](#). Neutral tweets were filtered out, and only positive and negative tweets with a predicted class probability above 99% were used to ensure high precision [S2 Fig](#). At the proposed threshold (indicated with a red star in [Fig 1B](#)), the model achieved a precision of 94% at 87% recall for the positive class and an 86% precision at 68% recall for the negative class ([S2 Table in S1 File](#)). Notably, the classification performance reported in this study outperforms the Fasttext-based approach of Muller et al. (2020) [\[53\]](#), which was reported as 77% for precision and 77% for recall ([S3 Table in S1 File](#)).

The study developed attitudes toward vaccination (ATV) score, which summarizes the proportion of positive and negative tweets in terms of log odds ratio by their location or the specific time period they were posted. The ATV score was calculated per day using all tweets and by the US state using only geolocated tweets. The ATV scores of geolocated tweets are representative of all tweets based on monthly ATV score comparison (Spearman's rank correlation coefficient = 96%, $P = 6 \times 10^{-5}$), as shown in [S3A and S3B Fig](#). The study also investigated the relationship between the vaccination rate of each US state by December 31st, 2021 [\[67\]](#) and state-level ATV scores, finding a moderate correlation (Spearman's rank correlation coefficient = 53%, $P = 6 \times 10^{-5}$) that indicates the proposed ATV score is representative of the vaccination behavior of a broader US population beyond the Twittersphere [S4A and S4B Fig](#).

In order to understand the stability and variability of attitudes toward vaccination, we investigated the timeline of two years [Fig 1C](#). The attitude was defined as stable if the vaccination attitude did not change over time, and as variable if the perspective on vaccination differed over time. While positive tweets about vaccines predominated over negative tweets during the entire timeline, the relative ratio of positive to negative tweets fluctuated. The first quarter of 2020 had a small fraction of all tweets, comprising less than 0.5% of the entire dataset, which could be explained by low COVID-19 death numbers in the USA. However, the ATV score decreased sharply with rapidly increasing COVID-19 death numbers in late March and stayed relatively low until November 2020. This period also had a high unemployment rate, although it peaked in early May at 15.8% [\[68\]](#) [S5 Fig](#). During 2021, vaccination had greater online public attention, and the aggregate attitudes toward vaccination were elevated in any quarter of 2021 compared to 2020 [S6 Fig](#), parallel to the successful vaccine trials [\[69–71\]](#) and the increasing vaccination rate. However, the variability of the aggregate vaccination attitudes may not imply a shift in individuals' views toward vaccination because the composition of users differs over time.

To address this, we conducted a user-level analysis by tracking the vaccination attitudes of users who posted multiple tweets across the studied timeline. We found that 31% of users had more than one tweet in the dataset, and most users were clearly polarized into positive and negative attitude camps by predominantly tweeting with a single attitude ([Fig 2B, S7 Fig](#)). There was no apparent difference between these groups based on the number of tweets they posted ([S6A and S6B Fig](#)). Also, there were only a few outliers with more than thousands of tweets, indicating that few users do not disproportionately influence our analysis ([Fig 2A](#)). The

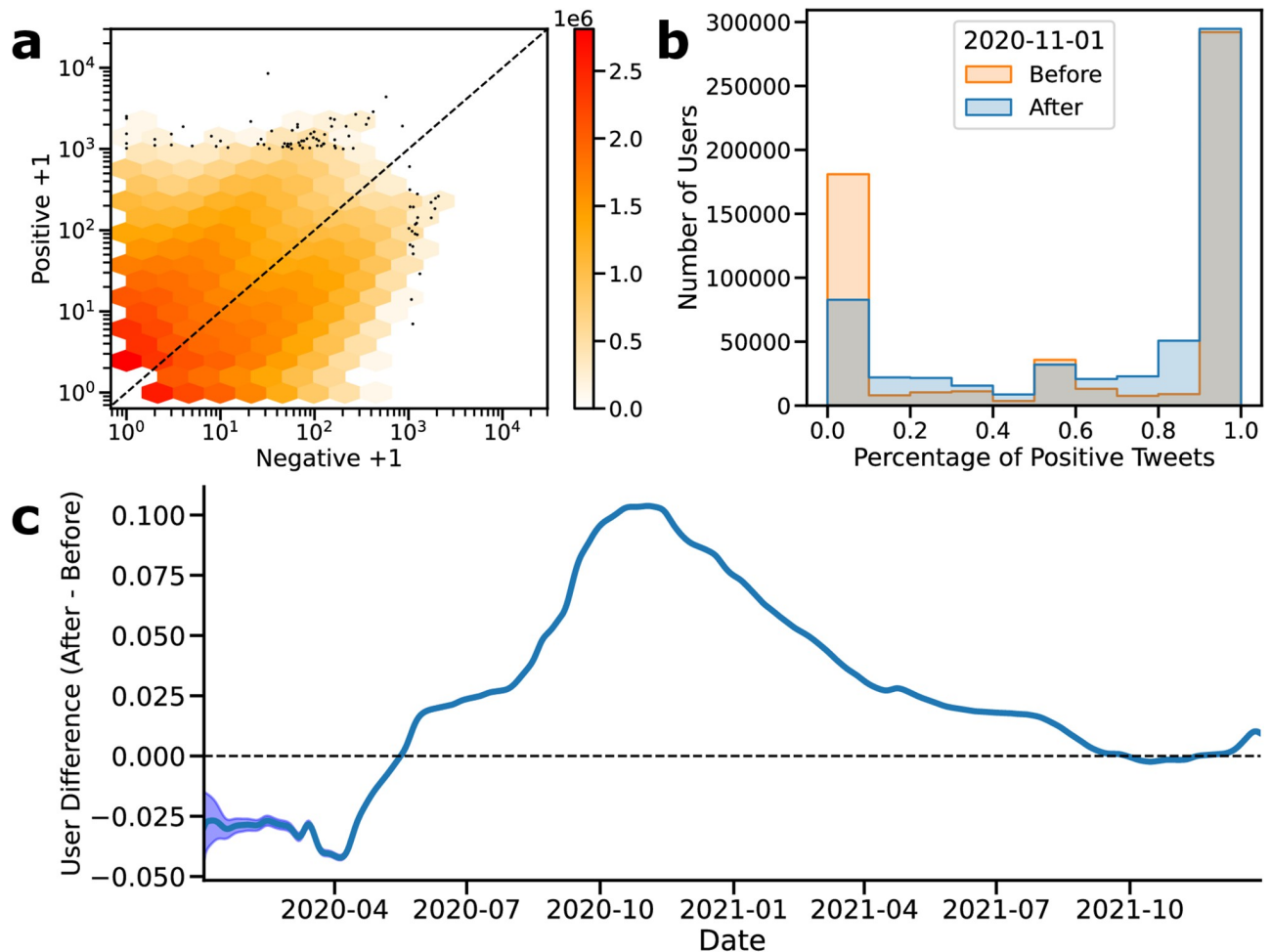


Fig 2. Temporal change of attitudes. A: The number of tweets posted by users B: The percentage of positive tweets of users with at least two tweets before and after November 1st. C: The trend of vaccination attitudes at the user level.

<https://doi.org/10.1371/journal.pone.0308122.g002>

percentage of positive tweet distributions of users before and after November 1st, 2020 is shown in Fig 2B, and there is a statistically significant shift to a more positive opinion. Specifically, users are 11% more likely to post a tweet with a positive attitude after November 1st, 2020, compared to their past selves. We repeated this analysis for all the dates in our dataset and observed that attitudes toward users decreased in the first five months of 2020 and increased afterward, except in the last four months of 2021, when there was no significant change (Fig 2C, S9A and S9B Fig).

We conducted an even simpler analysis with the users who posted two tweets ($n = 906,430$). The comparison of their first tweet against their later tweet shows a similar trend to a more positive attitude, regardless of the date. Only 12% of users with a positive first tweet later posted a negative second tweet, while 40% of users with a negative first tweet posted a positive second tweet (S4 Table in S1 File). Although the majority of users did not change their attitudes toward vaccination, there is a significant minority of negative users who have switched to a more positive attitude. Overall, both aggregate and user-level analysis indicate that a significant minority of users have more positive attitudes; yet the vast majority of users retained that vaccination attitude.

To investigate the socio-economic factors associated with vaccine hesitancy, we calculated the ATV score by the US states based on geolocated tweets and identified distinct regions based on attitude (as shown in Fig 3A). Using Fisher's exact test, we found that 25 out of 50 states have a significant higher vaccine hesitancy score compared to the other 25 states (Fig 3B and S5 Table in S1 File). We collected 10 socio-economic parameters for each state (Methods) and found that 8 out of 10 socio-economic parameters are significantly correlated with vaccination attitude based on univariate correlation analysis (S10A–S10J Fig). For instance, there is a negative correlation of -49% ($p = 3 \times 10^{-4}$) between ATV score and having less than a high school degree (S10B Fig), while there is a positive correlation of 45% ($p = 10^{-3}$) between voting for Trump in 2020 and ATV (S10H Fig). However, we also found that the higher rate of cat ownership compared to the rate of dog ownership has a 46% Spearman's rank correlation coefficient (S10J Fig), indicating the limitation of the univariate analysis, which may not accurately represent the underlying socio-economic factors related to vaccination attitude. It is important to note that socio-economic factors are not easily separable but rather interconnected (as depicted in Fig 3C), and therefore, multicollinearity between socio-economic parameters may lead to spurious correlations.

PCA (principal component analysis) was used to identify orthogonal latent factors from the socio-economic parameters. The first two principal components, which represent the characteristics of each state, were plotted in Fig 3D (S11 and S12 Figs, S6 Table in S1 File). Each state is then colored according to its ATV score. Remarkably, even though the PCA analysis does not utilize the ATV score as a feature, the states are evidently separated based on their attitudes toward vaccination, primarily along the first principal component. This finding strongly suggests that unique state-level characteristics are associated with vaccination attitudes.

We developed a partial-linear (PLS) regression model to predict the ATV score of each state based on the socio-economic parameters (Methods). The PLR regression, similar to PCA, identifies orthogonal latent factors for features and correlates those factors with the ATV score, which overcomes the collinearity that conventional regression models suffer from. The predicted ATV scores based on socio-economic parameters using the PLS model have a Spearman correlation coefficient of 61% ($p = 2.33e - 06$) against ground-truth ATV scores based on the one-leave-out test set (Fig 3E). This result indicates that socio-economic features are predictive of ATV scores. We conducted a bootstrapping analysis to investigate specific socio-economic parameters linked to vaccine hesitancy when other parameters remained constant (Methods). Confidence intervals of coefficients obtained from bootstrapping analysis were adjusted for a false discovery rate of 5% to avoid multiple comparisons (Fig 3F, S13 Fig). The analysis reveals that none of the socio-economic features is significant alone when controlled for other socio-economic parameters.

Overall, these results suggest that states have unique characteristics defined by socio-economic parameters that separate them into two camps and are highly predictive of attitudes toward vaccination. Yet, it is not possible to obtain the definitive link between vaccine attitude and a specific socio-economic parameter through state-level analysis without further research.

The common justifications for vaccine hesitancy were obtained by performing contextual topics modeling analysis (Methods) on the tweets with negative attitudes. As a result, seven main themes for vaccine hesitancy were discovered (Table 1, S14 Fig, S7 Table in S1 File). Similar themes have been reported in the childhood vaccination literature [72]. For instance, Majid and Ahmad (2020) examined 34 qualitative studies to clarify parents' reasons for rejecting or delaying vaccines [73]. Although their examination did not cover the studies on COVID-19 vaccine hesitancy, the fear of side effects and skepticism regarding vaccine effectiveness were prevalent among parents who rejected vaccines, which is consistent with what we found in our topics. Additionally, some of the studies they examined highlighted that

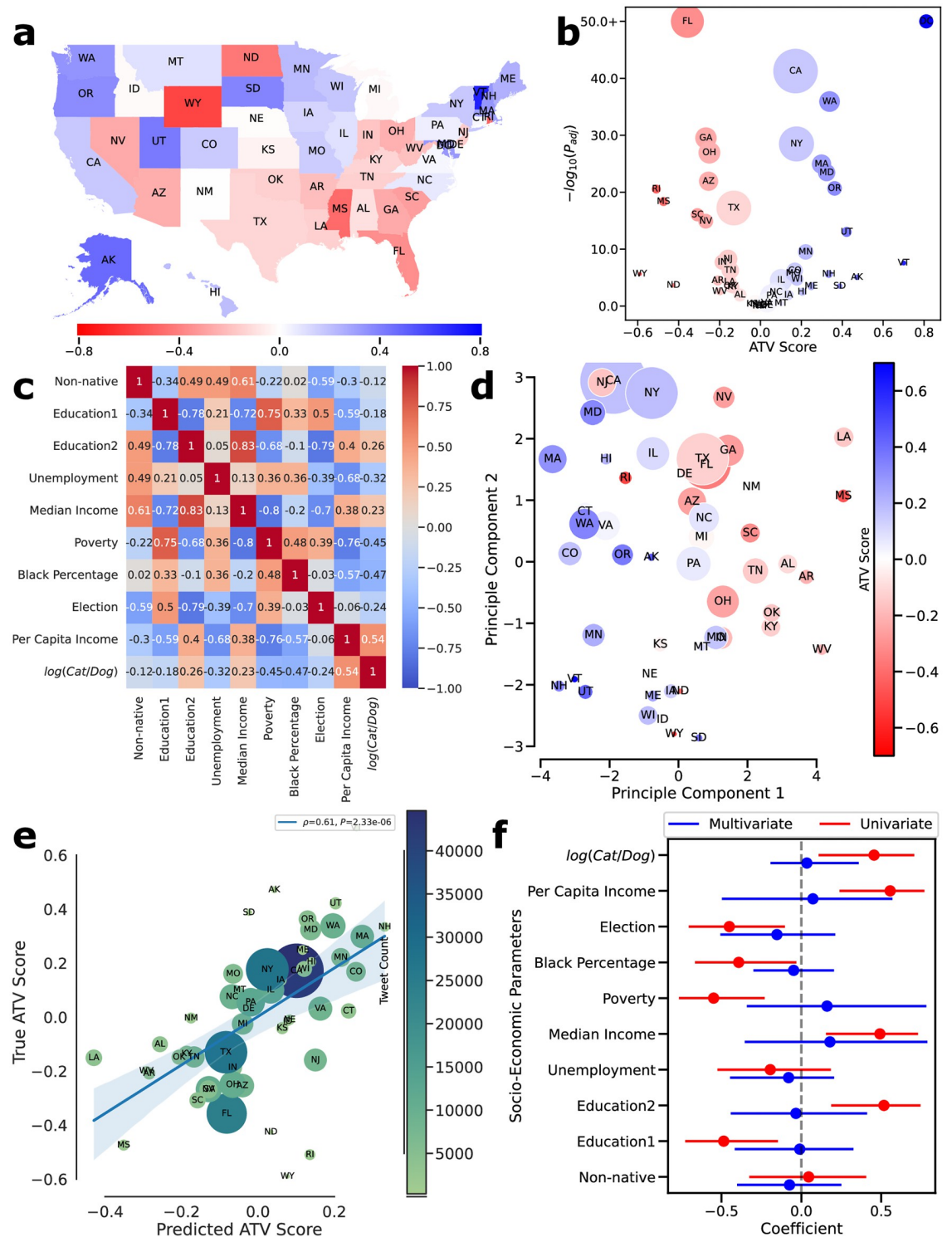


Fig 3. State-level socio-economic parameters and attitudes toward vaccination. A: U.S. states colored by ATV score. B: Volcano plot of ATV score and statistical significance of vaccination attitude of each state. C: Cross-correlation of socio-economic parameters with Spearman’s method. D: The scatterplot of the first two principal components of the US states, which is colored by ATV-score. E: True and predicted ATV scores based on PLS regression. F: Correlation coefficients of socio-economic parameters based on univariate (Spearman’s correlation coefficient) and multivariate (coefficients of PLS regression) analysis with error bars.

<https://doi.org/10.1371/journal.pone.0308122.g003>

Table 1. Themes identified by topic modeling and top keywords describing each theme.

Topic	Keywords
Mandatory Vaccination	choice, want, care, choose, forcing, job, freedom, jobs, force, take, personal, lose, feel, live, someone, right, others, free, anti, else, body, people, choices, forced, decision, life, think, trying, pro, taking, let, away, put, leave, fuck, believe, go, anyone, everyone, chose, rights, dont, tell, say, respect, stupid, something, us, agree, god
Adverse reactions	attacks, heart, cardiac, blood, clots, reported, report, injuries, myocarditis, vaers, reports, related, deaths, inflammation, strokes, reporting, attack, events, caused, injury, following, induced, reactions, increased, adverse, issues, thousands, injured, due, problems, pfizer, stroke, linked, died, number, event, days, pericarditis, within, higher, post, reaction, suffered, dose, death, associated, rare, increase, young, suffering
Constitutionality	mandates, passports, mandate, employees, federal, workers, unconstitutional, fired, mandatory, court, city, illegal, nyc, passport, staff, supreme, requirements, status, state, businesses, requirement, tyranny, policies, id, police, compulsory, proof, digital, law, comply, enforce, tyrannical, york, support, protest, states, policy, requiring, military, florida, courts, oppose, constitution, jobs, enforcing, vote, rights, vaccination, texas, impose
Ineffectiveness	booster, vaxxed, unvaccinated, got, shot, flu, covid, getting, sick, spreading, shots, fully, still, hospital, spread, vaccinated, boosters, get, dying, catch, die, symptoms, jabbed, year, work, protect, jab, people, vaxed, stop, prevent, catching, every, positive, person, know, works, never, cold, keep, vaccine, even, need, last, days, caught, died, double, virus, everyone
Long Term Side Effects	term, long, experimental, therapy, children, fda, kids, gene, effects, safety, side, child, approved, liability, mrna, risks, risk, data, studies, trials, drug, trial, safe, young, emergency, years, zero, use, unknown, definition, eua, healthy, survival, anti, vax, age, benefit, parents, technology, serious, adverse, used, issues, clinical, effect, chance, approval, tested, experiment, known
Politics and Conspiracy Theories	fauci, pharma, big, money, amp, trump, billions, media, profits, pushing, biden, companies, world, boosters, booster, push, gates, said, made, profit, making, us, paid, billion, politicians, trust, dollars, ivermectin, truth, pfizer, real, science, lies, dr, sell, cdc, pandemic, lied, every, told, bill, propaganda, pharmaceutical, fake, control, news, look, china, rich, created
Natural Immunity	transmission, natural, immunity, prevent, infection, variants, variant, unvaccinated, spread, leaky, vaxxed, protection, spreading, antibodies, immune, mutations, mutate, omicron, superior, virus, infected, spike, system, preventing, unvaxxed, rates, prevents, herd, stronger, protect, stopping, mutation, reduce, infections, protein, better, stops, stop, provides, rate, load, viral, effective, catching, contracting, disease, cases, original, reduces, science
Others	assess, certainty, tend, ect, beating, firstly, supplements, prescription, hasnt, ease, noone, additionally, fairly, theyre, readily, vacs, transfer, critically, guarantees, practically, wasnt, historically, complicated, reacts, treatable, versions, incl, bypass, functioning, weather, precautions, covers, depend, foods, stays, knowingly, assure, existed, atleast, advantage, bug, variety, partial, picking, prescribe, importantly, dishonest, imho, shield, smarter

<https://doi.org/10.1371/journal.pone.0308122.t001>

distrust in health system players and mandatory vaccine policies were reasons why parents rejected vaccines for their kids. In our study, we encountered a similar stance on the alleged conflict of interest in the health system and categorized it as politics and conspiracy theories. Similarly, mandatory vaccine policies are a source of frustration for many vaccine-hesitant Twitter users in our research. This demonstrates parallelism between the ways in which people justify their vaccine-hesitant attitudes in the face of the COVID-19 pandemic and previous waves of anti-vaxxer movements.

Using a topic modeling technique, we have identified eight themes related to vaccine hesitancy in our dataset. The first theme covers discussions on mandatory vaccinations and features social media conversations about vaccine administration practices. Many vaccine hesitant Twitter users framed mandatory vaccination as a violation of their individual rights.

Another theme pertains to the constitutionality of vaccine mandates, such as vaccine passports and vaccination requirements for state and federal workers. Our insights from these two themes echo recent research in the history of vaccination, which has documented how activists in US history opposed vaccination on the basis of rights, freedom, and liberty, even when vaccination was not compulsory [74]. This work highlights the ongoing tension between scientific expertise and civic freedom to choose one's own medical practices. In social media, we see similar traces of this tension that led to the *Jacobson v. Massachusetts* case in 1905 [74, 75].

The Twitter discussions focus heavily on the health concerns of users, with several themes emerging. One theme in this category is the alleged adverse reactions of vaccines. Another related theme focuses on the long-term side effects of COVID-19 vaccines. Vaccine hesitant tweets on this topic discuss potential long-term side effects, which may arise from the novel mRNA method or the risks associated with the expedited vaccine approval process. "Ineffectiveness" is another health-related topic that covers skepticism over the efficiency of vaccines or their total denial. A systematic review found that safety and side effects were the two most commonly reported factors influencing COVID-19 vaccine hesitancy in various studies [76]. Another meta-analysis concluded that stronger beliefs in the unsafety of vaccines are among the predictors of vaccine hesitancy globally [77]. The belief that COVID-19 vaccines are unsafe or ineffective is the factor that has been supported by the largest number of studies in high-income countries, while concerns about the rapid development of vaccines follow them on the list [78].

A final set of health-related vaccine hesitancy themes involves social media content that either partially or completely denies the benefits of vaccines. The "natural immunity" theme suggests that natural immunity is superior to vaccines in terms of protection against the virus. Politics and conspiracy theories are closely linked to COVID-19 denialism in health-related tweets. Twitter users discussed the alleged conflict of interest in the production of scientific knowledge and its dissemination by the mass media in line with the political frame of their choice. The relationship between belief in conspiracy theories and vaccine hesitancy has been extensively studied [79, 80]. However, the contribution of conspiracy theories to vaccine hesitancy has been overstated without solid empirical evidence. In fact, an independent research center has found that only a few people were responsible for the vast majority of misleading content about vaccines on social media [81]. In terms of substance, the Big Pharma conspiracy theory predates the COVID-19 pandemic and uses *cui bono* reasoning to identify a small number of elites who benefit from either deliberately manufacturing or not fully curing diseases [82].

The extracted substantial themes, including the seven main themes and the "others" category, reveal the contours of the vaccine hesitant debate on Twitter. Conspiracy theories with political implications represent only one aspect of vaccine hesitant content; there is much more to it. Vaccine administration and health concerns are the two primary branches of vaccine hesitancy in social media discussions. Both of these strands encompass a wide range of themes, from moderate skepticism to extreme denialism. Our model's vaccine hesitant topics are consistent with previous studies on vaccine hesitancy, which have also identified concerns about side effects, lack of trust, and belief in the ineffectiveness of vaccines as subthemes [83].

Earlier in our study, we illustrated that vaccine hesitancy takes on many shades and facets. To gain a better understanding of how vaccine-hesitant individuals rationalize their opposition to COVID-19 vaccines, we explored the extent to which they indiscriminately adopt arguments from other prevalent themes of vaccine hesitancy. Specifically, we examined whether users were solely focused on one particular topic or if their concerns spanned across multiple topics. To achieve this, we categorized users by topics if they had posted at least one tweet on the subject. Then, we measured the topic distribution of their remaining tweets, as illustrated

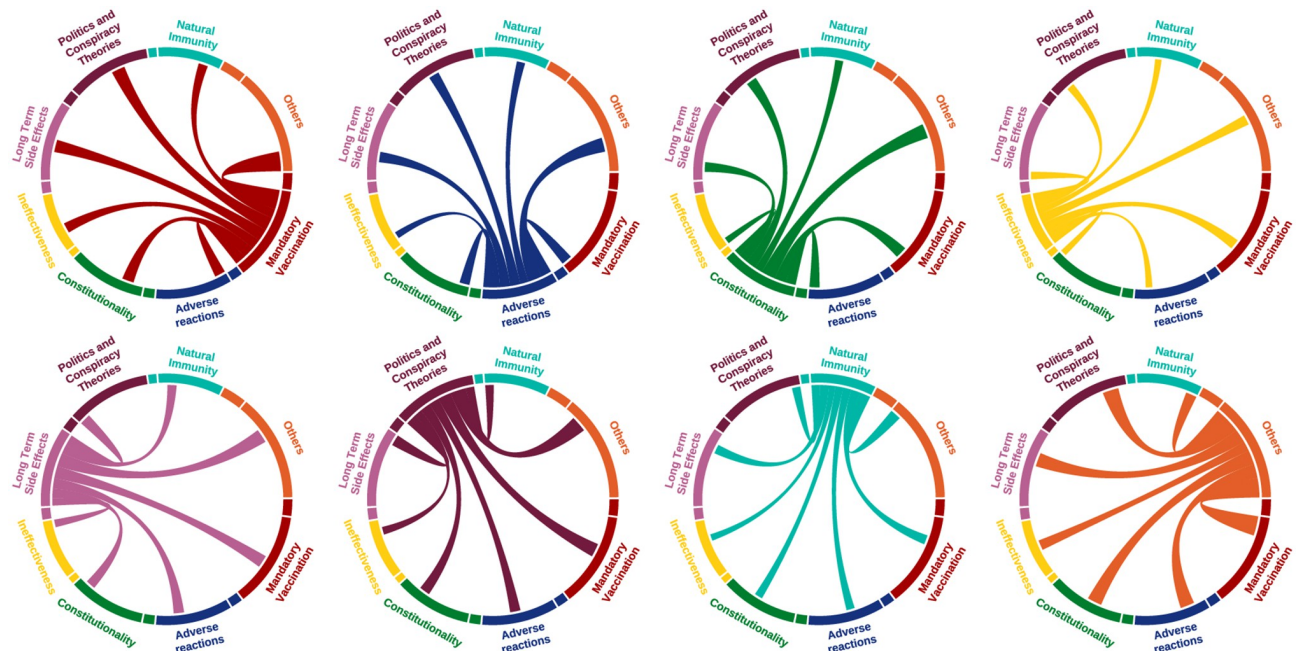


Fig 4. Users' topic distribution. The topic distribution of Twitter posts of users who posted multiple tweets. Each diagram represents a group of users interested in a specific topic and illustrates how their other tweets are distributed.

<https://doi.org/10.1371/journal.pone.0308122.g004>

in Fig 4. We observed that users in our dataset did not exhibit a consistent pattern of opposition based on a single theme within vaccine hesitancy topics. Instead, they demonstrated a high degree of variability across topics, not limiting themselves to a specific theme. These individuals employed a broad range of topics to rationalize their vaccine hesitancy. Fig 4 portrays the co-occurrence of the same users across various vaccine hesitant themes. For instance, a user who posted their first tweet on adverse reactions might subsequently employ an argument from the conspiracy theory theme to bolster their stance. Overall, the desultory use of arguments from a wide range of topics may imply that users are potentially motivated by deeper prejudices, and they utilize any available justification in their Twitter posts.

Discussion

In this study, we investigate skepticism towards vaccination by analyzing public discussions on Twitter using state-of-the-art natural language processing techniques. Our analysis sheds light on the polarization of online communities regarding COVID-19 vaccines.

The temporal aspect of our study design reveals the rigidity of attitudes towards vaccines. Despite the scientific advancements and approval of the first COVID-19 vaccines, only a small minority of Twitter users have shown a positive change in their attitudes towards vaccines. In addition to the composition effect, we also observe a slight optimism regarding the pandemic. However, negative attitudes towards vaccination have mostly persisted within the online community.

To gain a deeper understanding into the attributes of online communities, we used a novel methodology by merging geolocated online content with the conventional socio-economic variables based on the location of Twitter messages. Our association analysis between socio-economic variables and vaccine hesitancy demonstrated that socio-economic parameters can

predict the degree of vaccine hesitancy in US states. The residents of those states suffer from higher rates of unemployment, lower median household income, and poorer educational outcomes. However, the vast disparities in vaccine hesitancy among US states cannot be solely attributed to a single social parameter, such as election results, education, race, or income. Vaccination hesitant online communities have unique characteristics defined by a complex amalgamation of socio-economic parameters.

To gain further insights into discourses of the vaccine-hesitant community, we conducted a textual analysis and identified various themes revolving around two main concerns, political and medical. Political concerns branch out to two different themes: constitutional issues related to vaccine mandates and conspiracy theories. Likewise, medical concerns include two dissimilar themes: medical side effects and the denial of their effectiveness. These concerns are similar to the reasons to refuse vaccines identified by a systematic review of earlier work as (1) medical safety of COVID vaccines, (2) the inefficiency of vaccines, and (3) belief in natural immunity [84]. Yet, Twitter users who posted multiple times are not fixated on a single issue, rather they pragmatically borrowed arguments from a wide range of vaccine hesitant themes and they often justified their stances against vaccines using multiple reasons.

Research has consistently shown that marginalized communities exhibit lower levels of trust in institutions, including vaccines [85, 86]. Several studies have found that vaccination rates are lower among racial and ethnic minorities compared to the general population. [87] Similar socio-economic conditions are shared by some segments of the white working class, aligning them with other disempowered and marginalized communities. [88] This demographic is often portrayed as former President Trump's base [89]. The areas they inhabit are marked by high levels of poverty and elevated rates of premature deaths due to gun violence, suicide, drug overdoses, and alcoholism, compared to the national average [90, 91]. The historical marginalization of communities leads to a higher concern about government intrusion in their personal lives. As a result, those particular segments of society feel fear and anxiety in the face of strict vaccine mandates and protocols.

While our study provides valuable insights into vaccine hesitancy during the COVID period, it comes with certain limitations that warrant consideration. Firstly, the dataset is collected during the COVID pandemic, a time when vaccine attitudes were highly politicized. Secondly, twitter does not provide precise location of the geo-located tweets. Thus geographic resolution of our data does not extend to the county level. Capturing data at the county level would significantly enhance our ability to understand intra-state variability. Lastly, our study does not include individual-level socioeconomic status (SES) data linked to social media attitudes. While linking social media discourse to individual SES on a large scale is impossible due to privacy concerns, such an analysis at the individual level would greatly enhance the precision of our results. The study's focus on aggregate state-level analysis precludes specific conclusions about individual factors influencing vaccine hesitancy.

The moral foundations theory can illuminate our findings regarding the vaccination hesitant community [92]. The theory suggests that people primarily rely on moral intuitions shaped by socio-psychological factors to make political judgments, then they justify their judgments with strategic reasoning [92]. Some Twitter users' attitudes toward vaccines may have been shaped by emotions and gut-level feelings, leading to distrust of institutions that underlies vaccine hesitancy. Research has used moral foundations theory to analyze communication patterns and discursive elements in COVID-related social media content [93, 94]. Our findings regarding the spurious relationship between a solitary socio-economic parameter and vaccine hesitancy as well as the indiscriminate use of different topics may support this idea. This distrust may stem from a heightened moral preference expressed through a series of ad-hoc hypotheses, as classified as diverse vaccine-hesitant themes in our study [95].

To address these concerns, a well-planned public communication strategy is necessary. Understanding the moral foundations of their attitude and establishing empathy should be the initial step. One approach to restore faith and rebuild trust in those communities would be to adopt a persuasive language that emphasizes the common good, with the support of local leaders [96–99]. Without the support of their in-group, public health efforts may be seen as stigmatizing and compulsory measures by marginalized communities [86]. Communication strategies tailored to local contexts may improve to trust in science.

Supporting information

S1 Fig. The hourly tweet count based on Eastern Standard Time (UTC-5). The results show that the number of tweets peaks during the daytime hours of EST.

(TIFF)

S2 Fig. The precision-threshold curve for the proposed attitudes toward vaccination model illustrates the relationship between precision and threshold for classifying tweets as Neutral, Positive, or Negative. Precision is a measure of correctly classified samples out of all samples classified as a specific class, while the threshold is the minimum class probability required for a class assignment. The dashed red line represents a threshold of 0.99.

(TIFF)

S3 Fig. A: This figure displays the daily ATV score for geolocated and non-geolocated tweets between 2020 and 2022. Geolocated and non-geolocated tweets follow the same trend. **B:** Monthly ATV scores correlation between geolocated and non-geolocated tweets. There is a high correlation (R-squared of 87%) between ATV scores of geolocated and non-geolocated tweets.

(TIFF)

S4 Fig. A: The map displays the vaccination percentage of US states as of December 31st, 2021. **B:** This scatter plot presents the correlation between the ATV score of each state and the percentage of eligible population vaccinated in the state. A moderate Spearman's correlation coefficient indicates that attitude on social media is predictive of the actual vaccination rate. The size and color of the points represent the tweet count of each state.

(TIFF)

S5 Fig. This figure displays the trend of the ATV score and user attitude difference, along with other relevant plots such as COVID data (New Cases and vaccination percentage timeline) and socio-economic parameters (Unemployment percentage).

(TIFF)

S6 Fig. The figure depicts how the Attitude Towards Vaccination (ATV) score changes over time at three-month intervals. To accommodate the lack of tweets in 2020, we plotted the ATV scores for that year in a single box plot until 2021. The Mann-Whitney U test was used to calculate the statistical significance of each box plot with Bonferroni correction.

(TIFF)

S7 Fig. The histogram displays the percentage of positive tweets from users who posted multiple tweets, revealing polarization into two stances.

(TIFF)

S8 Fig. A: Number of positive or negative tweets sent by users as a histogram and cumulative numbers. Approximately, half of the users sent only one tweet, while the other half sent multiple tweets. **B:** The percentage of users based on the number of tweets they posted by tweet

attitude. It reveals that users with negative and positive attitudes did not differ in terms of the number of tweets they posted.

(TIFF)

S9 Fig. To examine the monthly difference in user attitudes, we selected users who posted at least one tweet before and after a certain time period and calculated their attitude difference and the standard error of the difference. The figure includes two parts: A: percentage of users' attitude change in each month, with points on the line indicating the change of average attitude towards a more positive (higher values) or negative attitude (lower values) after each month; and B: the average monthly positive ratio of each user, with changes before and after each month presented. In November 2020, there was a roughly 11% shift towards a positive attitude. We calculated the p-values using the Wilcoxon Test and corrected them using the Bonferroni method.

(TIFF)

S10 Fig. Correlation between ATV score and socio-economic parameters.

(TIFF)

S11 Fig. Bi-plot displaying the first two principal components of the socio-economic of the states, where the direction of each socio-economic is indicated with a line.

(TIFF)

S12 Fig. Percentage of variance explained by each principal component displayed as a histogram and cumulative plot. Over 75% of the variance is explained by the top two principal components.

(TIFF)

S13 Fig. The distribution of coefficients for each socio-economic parameter based on bootstrapped analysis of Partial Least Squares Regression (PLSRegression). The figure includes a dashed line indicating the null hypothesis that the coefficient is zero, and red lines representing the confidence intervals for the coefficient. The null hypothesis (zero coefficient) is contained within these intervals.

(TIFF)

S14 Fig. The breakdown of tweets by themes.

(TIFF)

S1 File. S1 Table. Tweet count and ratio of the countries. S2 Table. Precision/Recall table of sentiments. S3 Table. Classification report of sentiments which covers various metrics: Precision, Recall, Accuracy, F1-Micro/Macro etc. S4 Table. Attitude change of users who sent 2 tweets. S5 Table. Tweet count, Attitude Towards Vaccination (ATV) score, and P-values of the US states. S6 Table. First two principle component of socio economic variables. S7 Table. Top 10 tweets of each topic.

(XLSX)

Acknowledgments

Data and code availability

The code used for this project is available on GitHub at https://github.com/twittersphere/twitter_vaccine_hesitancy. The Attitude and Contextualized Topic Modeling predictions generated in this study are available on Zenodo at <https://doi.org/10.5281/zenodo.7876072>.

Author Contributions

Conceptualization: Muhammed Hasan Celik, Huseyin Zeyd Koytak.

Data curation: Huzeyfe Ayaz.

Methodology: Huzeyfe Ayaz, Muhammed Hasan Celik, Huseyin Zeyd Koytak.

Project administration: Muhammed Hasan Celik, Huseyin Zeyd Koytak.

Software: Huzeyfe Ayaz.

Supervision: Muhammed Hasan Celik, Huseyin Zeyd Koytak.

Visualization: Huzeyfe Ayaz.

Writing – original draft: Huzeyfe Ayaz, Muhammed Hasan Celik, Huseyin Zeyd Koytak.

Writing – review & editing: Ibrahim Emre Yanik.

References

1. Goldenberg MJ. Vaccines, values and science. *CMAJ*. 2019; 191(14):E397–E398. <https://doi.org/10.1503/cmaj.181635>
2. Sturgis P, Brunton-Smith I, Jackson J. Trust in science, social consensus and vaccine confidence. *Nature Human Behaviour*. 2021; 5(11):1528–1534. <https://doi.org/10.1038/s41562-021-01115-7> PMID: 34002053
3. Wong-Parodi G, Feygina I. Understanding and countering the motivated roots of climate change denial. *Current Opinion in Environmental Sustainability*. 2020; 42:60–64. <https://doi.org/10.1016/j.cosust.2019.11.008>
4. Gifford R. The dragons of inaction: psychological barriers that limit climate change mitigation and adaptation. *American psychologist*. 2011; 66(4):290. <https://doi.org/10.1037/a0023566> PMID: 21553954
5. Choi Y, Fox AM. Mistrust in public health institutions is a stronger predictor of vaccine hesitancy and uptake than Trust in Trump. *Social Science & Medicine*. 2022; 314:115440. <https://doi.org/10.1016/j.socscimed.2022.115440> PMID: 36332532
6. Jung Y, Lee S. Trump vs. the GOP: Political Determinants of COVID-19 Vaccination. *Political Behavior*. 2023; p. 1–23.
7. Zarocostas J. How to fight an infodemic. *The lancet*. 2020; 395(10225):676. [https://doi.org/10.1016/S0140-6736\(20\)30461-X](https://doi.org/10.1016/S0140-6736(20)30461-X) PMID: 32113495
8. Wynne B, et al. Public understanding of science. *Handbook of science and technology studies*. 1995; 1:361–388. <https://doi.org/10.4135/9781412990127.n17>
9. Hilgartner S. The dominant view of popularization: Conceptual problems, political uses. *Social studies of science*. 1990; 20(3):519–539. <https://doi.org/10.1177/030631290020003006>
10. Gauchat G. Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American sociological review*. 2012; 77(2):167–187. <https://doi.org/10.1177/0003122412438225>
11. Leonhardt D. U.S. Covid Deaths Get Even Redder. *The New York Times*. 2021;.
12. Bump P. Analysis—the ‘what about black people’ defense of Republican vaccine hesitancy; 2021. Available from: <https://www.washingtonpost.com/politics/2021/08/04/what-about-black-people-defense-republican-vaccine-hesitancy/>.
13. Bump P. Analysis—The biggest divide on vaccination isn’t race or income but party—and the divide is growing. *Washington Post*;.
14. Risen J. The Right’s Anti-Vaxxers Are Killing Republicans;. Available from: <https://theintercept.com/2022/10/10/covid-republican-democrat-deaths/>.
15. Ojea Quintana I, Reimann R, Cheong M, Alfano M, Klein C. Polarization and trust in the evolution of vaccine discourse on Twitter during COVID-19. *PLoS One*. 2022; 17(12):e0277292. <https://doi.org/10.1371/journal.pone.0277292> PMID: 36516117
16. Jiang J, Chen E, Yan S, Lerman K, Ferrara E. Political polarization drives online conversations about COVID-19 in the United States. *Human Behavior and Emerging Technologies*. 2020; 2(3):200–211. <https://doi.org/10.1002/hbe2.202> PMID: 32838229

17. Larson HJ, Jarrett C, Eckersberger E, Smith DM, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. *Vaccine*. 2014; 32(19):2150–2159. <https://doi.org/10.1016/j.vaccine.2014.01.081> PMID: [24598724](https://pubmed.ncbi.nlm.nih.gov/24598724/)
18. Jarrett C, Wilson R, O'Leary M, Eckersberger E, Larson HJ, et al. Strategies for addressing vaccine hesitancy—A systematic review. *Vaccine*. 2015; 33(34):4180–4190. <https://doi.org/10.1016/j.vaccine.2015.04.040> PMID: [25896377](https://pubmed.ncbi.nlm.nih.gov/25896377/)
19. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger JA. Vaccine hesitancy: an overview. *Human vaccines & immunotherapeutics*. 2013; 9(8):1763–1773. <https://doi.org/10.4161/hv.24657> PMID: [23584253](https://pubmed.ncbi.nlm.nih.gov/23584253/)
20. Larson HJ, Jarrett C, Schulz WS, Chaudhuri M, Zhou Y, Dube E, et al. Measuring vaccine hesitancy: the development of a survey tool. *Vaccine*. 2015; 33(34):4165–4175. <https://doi.org/10.1016/j.vaccine.2015.04.037> PMID: [25896384](https://pubmed.ncbi.nlm.nih.gov/25896384/)
21. Salmon DA, Dudley MZ, Glanz JM, Omer SB. Vaccine hesitancy: causes, consequences, and a call to action. *Vaccine*. 2015; 33:D66–D71. <https://doi.org/10.1016/j.vaccine.2015.09.035> PMID: [26615171](https://pubmed.ncbi.nlm.nih.gov/26615171/)
22. MacDonald NE, et al. Vaccine hesitancy: Definition, scope and determinants. *Vaccine*. 2015; 33(34):4161–4164. <https://doi.org/10.1016/j.vaccine.2015.04.036> PMID: [25896383](https://pubmed.ncbi.nlm.nih.gov/25896383/)
23. Soares P, Rocha JV, Moniz M, Gama A, Lares PA, Pedro AR, et al. Factors associated with COVID-19 vaccine hesitancy. *Vaccines*. 2021; 9(3):300. <https://doi.org/10.3390/vaccines9030300> PMID: [33810131](https://pubmed.ncbi.nlm.nih.gov/33810131/)
24. Eskola J, Duclos P, Schuster M, MacDonald NE, et al. How to deal with vaccine hesitancy? *Vaccine*. 2015; 33(34):4215–4217. <https://doi.org/10.1016/j.vaccine.2015.04.043> PMID: [25896378](https://pubmed.ncbi.nlm.nih.gov/25896378/)
25. Siddiqui M, Salmon DA, Omer SB. Epidemiology of vaccine hesitancy in the United States. *Human vaccines & immunotherapeutics*. 2013; 9(12):2643–2648. <https://doi.org/10.4161/hv.27243> PMID: [24247148](https://pubmed.ncbi.nlm.nih.gov/24247148/)
26. Solís Arce JS, Warren SS, Meriggi NF, Scacco A, McMurry N, Voors M, et al. COVID-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nature medicine*. 2021; 27(8):1385–1394. <https://doi.org/10.1038/s41591-021-01454-y> PMID: [34272499](https://pubmed.ncbi.nlm.nih.gov/34272499/)
27. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. *EClinicalMedicine*. 2020; 26:100495. <https://doi.org/10.1016/j.eclinm.2020.100495> PMID: [32838242](https://pubmed.ncbi.nlm.nih.gov/32838242/)
28. Dubé E, Gagnon D, Nickels E, Jeram S, Schuster M. Mapping vaccine hesitancy—Country-specific characteristics of a global phenomenon. *Vaccine*. 2014; 32(49):6649–6654. <https://doi.org/10.1016/j.vaccine.2014.09.039> PMID: [25280436](https://pubmed.ncbi.nlm.nih.gov/25280436/)
29. Wiysonge CS, Ndwandwe D, Ryan J, Jaca A, Batoure O, Anya BPM, et al. Vaccine hesitancy in the era of COVID-19: could lessons from the past help in divining the future? *Human vaccines & immunotherapeutics*. 2022; 18(1):1–3. <https://doi.org/10.1080/21645515.2021.1893062> PMID: [33684019](https://pubmed.ncbi.nlm.nih.gov/33684019/)
30. Salathe M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS computational biology*. 2011; 7(10):e1002199. <https://doi.org/10.1371/journal.pcbi.1002199> PMID: [22022249](https://pubmed.ncbi.nlm.nih.gov/22022249/)
31. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm—An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*. 2012; 30(25):3778–3789. <https://doi.org/10.1016/j.vaccine.2011.11.112> PMID: [22172504](https://pubmed.ncbi.nlm.nih.gov/22172504/)
32. Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine*. 2012; 30(25):3727–3733. <https://doi.org/10.1016/j.vaccine.2012.02.025> PMID: [22365840](https://pubmed.ncbi.nlm.nih.gov/22365840/)
33. Witteman HO, Zikmund-Fisher BJ. The defining characteristics of Web 2.0 and their potential influence in the online vaccination debate. *Vaccine*. 2012; 30(25):3734–3740. <https://doi.org/10.1016/j.vaccine.2011.12.039> PMID: [22178516](https://pubmed.ncbi.nlm.nih.gov/22178516/)
34. 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. *Human vaccines & immunotherapeutics*. 2020; 16(11):2586–2593. <https://doi.org/10.1080/21645515.2020.1780846> PMID: [32693678](https://pubmed.ncbi.nlm.nih.gov/32693678/)
35. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. *BMJ global health*. 2020; 5(10):e004206. <https://doi.org/10.1136/bmjgh-2020-004206> PMID: [33097547](https://pubmed.ncbi.nlm.nih.gov/33097547/)
36. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine*. 2010; 28(7):1709–1716. <https://doi.org/10.1016/j.vaccine.2009.12.022> PMID: [20045099](https://pubmed.ncbi.nlm.nih.gov/20045099/)
37. Zhang W, Wang R, Liu H. Moral expressions, sources, and frames: Examining COVID-19 vaccination posts by facebook public pages. *Computers in Human Behavior*. 2023; 138:107479. <https://doi.org/10.1016/j.chb.2022.107479> PMID: [36091923](https://pubmed.ncbi.nlm.nih.gov/36091923/)

38. Fridman A, Gershon R, Gneezy A. COVID-19 and vaccine hesitancy: A longitudinal study. *PloS one*. 2021; 16(4):e0250123. <https://doi.org/10.1371/journal.pone.0250123> PMID: 33861765
39. Johnson NF, Velasquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro-and anti-vaccination views. *Nature*. 2020; 582(7811):230–233. <https://doi.org/10.1038/s41586-020-2281-1> PMID: 32499650
40. Kornides ML, Badlis S, Head KJ, Putt M, Cappella J, Gonzalez-Hernandez G. Exploring content of misinformation about HPV vaccine on twitter. *Journal of Behavioral Medicine*. 2022; p. 1–14. <https://doi.org/10.1007/s10865-022-00342-1> PMID: 35896853
41. Wawrzuta D, Jaworski M, Gotlib J, Panczyk M. What arguments against COVID-19 vaccines run on Facebook in Poland: content analysis of comments. *Vaccines*. 2021; 9(5):481. <https://doi.org/10.3390/vaccines9050481> PMID: 34068500
42. Zhang W, Mukerjee S, Qin H. Topics and sentiments influence likes: A study of Facebook public pages' posts about COVID-19 vaccination. *Cyberpsychology, Behavior, and Social Networking*. 2022; 25(9):552–560. <https://doi.org/10.1089/cyber.2022.0063> PMID: 35969378
43. Monselise M, Chang CH, Ferreira G, Yang R, Yang CC. Topics and sentiments of public concerns regarding COVID-19 vaccines: Social media trend analysis. *Journal of medical Internet research*. 2021; 23(10):e30765. <https://doi.org/10.2196/30765> PMID: 34581682
44. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. *Journal of medical Internet research*. 2021; 23(6):e24435. <https://doi.org/10.2196/24435> PMID: 34115608
45. Koytak HZ, Celik MH. A Text Mining Approach to Determinants of Attitude Towards Syrian Immigration in the Turkish Twittersphere. *Social Science Computer Review*. 2022; p. 08944393221117460.
46. Melotte S, Kejriwal M. Predicting zip code-level vaccine hesitancy in US Metropolitan Areas using machine learning models on public tweets. *PLOS Digital Health*. 2022; 1(4):e0000021. <https://doi.org/10.1371/journal.pdig.0000021> PMID: 36812517
47. Shen K, Kejriwal M. Using conditional inference to quantify interaction effects of socio-demographic covariates of US COVID-19 vaccine hesitancy. *PLOS Global Public Health*. 2023; 3(5):e0001151. <https://doi.org/10.1371/journal.pgph.0001151> PMID: 37172006
48. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*. 2019; 32.
49. Jordahl K, den Bossche JV, Fleischmann M, Wasserman J, McBride J, Gerard J, et al. geopandas/geopandas: v0.8.1; 2020. Available from: <https://doi.org/10.5281/zenodo.3946761>.
50. Bureau USC. Cartographic boundary files—shapefile; 2021. Available from: <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>.
51. SearchTweets contributors. searchtweets-v2 1.1.1; 2023. Available from: <https://pypi.org/project/searchtweets-v2/>.
52. Roesslein J. tweepy Documentation. Online] <http://tweepy.readthedocs.io/en/v3>. 2009;5:724.
53. Müller M, Salathé M, Kummervold PE. COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter. *arXiv preprint arXiv:200507503*. 2020;.
54. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:191003771*. 2019;.
55. Pananos AD, Bury TM, Wang C, Schonfeld J, Mohanty SP, Nyhan B, et al. Critical dynamics in population vaccinating behavior. *Proceedings of the National Academy of Sciences*. 2017; 114(52):13762–13767. <https://doi.org/10.1073/pnas.1704093114> PMID: 29229821
56. Rajapakse TC. Simple Transformers; 2019. <https://github.com/ThilinaRajapakse/simpletransformers>.
57. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011; 12:2825–2830.
58. Upton GJ. Fisher's exact test. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 1992; 155(3):395–402. <https://doi.org/10.2307/2982890>
59. of Technology (MIT) MI, Data E, Lab S. 'US President 1976-2016. Data and Science MIT Election Lab 2017. 2019;.
60. Rupasingha A, Goetz SJ, Freshwater D. The production of social capital in US counties (2006 with updates). *The journal of socio-economics*. 2006; 35(1):83–101. <https://doi.org/10.1016/j.socce.2005.11.001>
61. King E, Mueller MK, Dowling-Guyer S, McCobb E. Financial fragility and demographic factors predict pet owners' perceptions of access to veterinary care in the United States. *Journal of the American Veterinary Medical Association*. 2022; 1(aop): 1–8. <https://doi.org/10.2460/javma.21.11.0486> PMID: 35482560

62. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Analytica chimica acta*. 1986; 185:1–17. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
63. Bianchi F, Terragni S, Hovy D, Nozza D, Fersini E. Cross-lingual contextualized topic models with zero-shot learning. arXiv preprint arXiv:200407737. 2020;.
64. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:13126114. 2013;.
65. Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the natural language toolkit. "O'Reilly Media, Inc."; 2009.
66. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018;.
67. Mathieu E, Ritchie H, Rodes-Guirao L, Appel C, Giattino C, Hasell J, et al. Coronavirus Pandemic (COVID-19). Our World in Data. 2020;.
68. Economic Research Service, U S Department of Agriculture. Unemployment and median household income for the U.S., States, and counties, 2000–2021; 2021. <https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/>.
69. Thomas SJ, Moreira ED Jr, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine through 6 months. *New England Journal of Medicine*. 2021; 385(19):1761–1773. <https://doi.org/10.1056/NEJMoa2110345> PMID: 34525277
70. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England journal of medicine*. 2020; 383(27):2603–2615. <https://doi.org/10.1056/NEJMoa2034577> PMID: 33301246
71. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England journal of medicine*. 2021; 384(5):403–416. <https://doi.org/10.1056/NEJMoa2035389> PMID: 33378609
72. Jordan Z, Rowland E. Parental perceptions of chickenpox and the varicella vaccine: A qualitative systematic review. *Vaccine*. 2023; . PMID: 38129287
73. Majid U, Ahmad M. The factors that promote vaccine hesitancy, rejection, or delay in parents. *Qualitative Health Research*. 2020; 30(11):1762–1776. <https://doi.org/10.1177/1049732320933863> PMID: 32597313
74. Colgrove J, Samuel SJ. Freedom, rights, and vaccine refusal: The history of an idea. *American Journal of Public Health*. 2022; 112(2):234–241. <https://doi.org/10.2105/AJPH.2021.306504> PMID: 35080944
75. Mariner WK, Annas GJ, Glantz LH. Jacobson v Massachusetts: it's not your great-great-grandfather's public health law. *American Journal of Public Health*. 2005; 95(4):581–590. <https://doi.org/10.2105/AJPH.2004.055160> PMID: 15798113
76. Roy DN, Biswas M, Islam E, Azam MS. Potential factors influencing COVID-19 vaccine acceptance and hesitancy: A systematic review. *PloS one*. 2022; 17(3):e0265496. <https://doi.org/10.1371/journal.pone.0265496> PMID: 35320309
77. Pires C. Global predictors of COVID-19 vaccine hesitancy: A systematic review. *Vaccines*. 2022; 10(8):1349. <https://doi.org/10.3390/vaccines10081349> PMID: 36016237
78. Aw J, Seng JJB, Seah SSY, Low LL. COVID-19 vaccine hesitancy—A scoping review of literature in high-income countries. *Vaccines*. 2021; 9(8):900. <https://doi.org/10.3390/vaccines9080900> PMID: 34452026
79. Farhart CE, Douglas-Durham E, Trujillo KL, Vitriol JA. Vax attacks: How conspiracy theory belief undermines vaccine support. *Progress in Molecular Biology and Translational Science*. 2022; 188(1):135–169. <https://doi.org/10.1016/bs.pmbts.2021.11.001> PMID: 35168741
80. Hornsey MJ, Harris EA, Fielding KS. The psychological roots of anti-vaccination attitudes: A 24-nation investigation. *Health psychology*. 2018; 37(4):307. <https://doi.org/10.1037/hea0000586> PMID: 29389158
81. Bond S. Just 12 people are behind most vaccine hoaxes on social media, research shows. NPR News. 2021;.
82. Blaskiewicz R. The big Pharma conspiracy theory. *Medical Writing*. 2013; 22(4):259–261. <https://doi.org/10.1179/2047480613Z.000000000142>
83. Fard TM, Shokri M, Mirfazeli FS, Mohammadsadeghi H, Shafiei N, Salehian R, et al. COVID-19 vaccination hesitancy is not all a conspiracy theory: A qualitative study from Iran. *Acta Psychologica*. 2023; p. 103839. <https://doi.org/10.1016/j.actpsy.2023.103839>
84. Troiano G, Nardi A. Vaccine hesitancy in the era of COVID-19. *Public health*. 2021; 194:245–251. <https://doi.org/10.1016/j.puhe.2021.02.025> PMID: 33965796
85. De Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling

- study. *The Lancet*. 2020; 396(10255):898–908. [https://doi.org/10.1016/S0140-6736\(20\)31558-0](https://doi.org/10.1016/S0140-6736(20)31558-0) PMID: 32919524
86. Pertwee E, Simas C, Larson HJ. An epidemic of uncertainty: rumors, conspiracy theories and vaccine hesitancy. *Nature medicine*. 2022; 28(3):456–459. <https://doi.org/10.1038/s41591-022-01728-z> PMID: 35273403
 87. Granade CJ, Lindley MC, Jatlaoui T, Asif AF, Jones-Jack N. Racial and ethnic disparities in adult vaccination: a review of the state of evidence. *Health Equity*. 2022; 6(1):206–223. <https://doi.org/10.1089/heq.2021.0177> PMID: 35402775
 88. Gest J. The white working-class minority: a counter-narrative. *Politics, Groups, and Identities*. 2016; 4(1):126–143. <https://doi.org/10.1080/21565503.2015.1053950>
 89. Carnes N, Lupu N. The white working class and the 2016 election. *Perspectives on Politics*. 2021; 19(1):55–72. <https://doi.org/10.1017/S1537592720001267>
 90. Case A, Deaton A. Deaths of Despair and the Future of Capitalism. In: *Deaths of Despair and the Future of Capitalism*. Princeton University Press; 2020.
 91. Metzl JM. *Dying of whiteness: How the politics of racial resentment is killing America's heartland*. Hachette UK; 2019.
 92. Haidt J. *The righteous mind: Why good people are divided by politics and religion*. Vintage; 2012.
 93. Borghouts J, Huang Y, Gibbs S, Hopfer S, Li C, Mark G. Understanding underlying moral values and language use of COVID-19 vaccine attitudes on twitter. *PNAS nexus*. 2023; 2(3):pgad013. <https://doi.org/10.1093/pnasnexus/pgad013> PMID: 36896130
 94. Jiang J, Luceri L, Ferrara E. Moral Values Underpinning COVID-19 Online Communication Patterns. *arXiv preprint arXiv:240108789*. 2024;.
 95. Rossen I, Hurlstone MJ, Dunlop PD, Lawrence C. Accepters, fence sitters, or rejecters: Moral profiles of vaccination attitudes. *Social Science & Medicine*. 2019; 224:23–27. <https://doi.org/10.1016/j.socscimed.2019.01.038> PMID: 30735925
 96. Padmawati RS, Heywood A, Sitaresmi MN, Atthobari J, MacIntyre CR, Soenarto Y, et al. Religious and community leaders' acceptance of rotavirus vaccine introduction in Yogyakarta, Indonesia: a qualitative study. *BMC Public Health*. 2019; 19:1–6. <https://doi.org/10.1186/s12889-019-6706-4> PMID: 30943929
 97. Dhaliwal BK, Seth R, Thankachen B, Qaiyum Y, Closser S, Best T, et al. Leading from the frontlines: community-oriented approaches for strengthening vaccine delivery and acceptance. In: *BMC proceedings*. vol. 17. Springer; 2023. p. 5.
 98. Kaufman J, Overmars I, Leask J, Seale H, Chisholm M, Hart J, et al. Vaccine champions training program: empowering community leaders to advocate for COVID-19 vaccines. *Vaccines*. 2022; 10(11):1893. <https://doi.org/10.3390/vaccines10111893> PMID: 36366401
 99. Airoidi EM, Christakis NA. Induction of social contagion for diverse outcomes in structured experiments in isolated villages. *Science*. 2024; 384(6695):eadi5147. <https://doi.org/10.1126/science.adi5147> PMID: 38696582