
FROM TRIBAL POLARIZATION TO SOCIO-ECONOMIC DISPARITIES: EXPLORING THE LANDSCAPE OF VACCINE HESITANCY ON TWITTER

 **Huzeyfe Ayaz***

Department of Informatics
Technical University of Munich
Garching, Munich, Germany

 **Muhammed Hasan Celik* †**

Department of Computer Science
Center for Complex Biological Systems
University of California Irvine
Irvine, CA, USA
mcelik@uci.edu

 **Huseyin Zeyd Koytak* †**

Department of Sociology
Syracuse University
Syracuse, NY, USA
hkoytak@syr.edu

Ibrahim Emre Yanik

Department of Sociology
Syracuse University
Syracuse, NY, USA

ABSTRACT

1 This study analyzed Twitter posts related to vaccine hesitancy and its association with socio-economic
2 variables in the US at the state level. The unique socio-economic characteristics of US states, such
3 as education, race, or income, are significantly associated with attitudes toward vaccination. Our
4 results indicate that vaccine hesitancy is a multifaceted phenomenon shaped by a complex interplay
5 of factors. Furthermore, the research identifies two distinct sets of justifications for vaccine hesitancy.
6 The first set pertains to political concerns, including constitutional rights and conspiracy theories.
7 The second pertains to medical concerns about vaccine safety and efficacy. However, vaccine-hesitant
8 Twitter users pragmatically use broad categories of justification for their beliefs. This behavior may
9 suggest that vaccine hesitancy is influenced by political beliefs, unconscious emotions, and gut-level
10 instinct. Our findings have further implications for the critical role of trust in shaping attitudes
11 toward vaccination and the need for tailored communication strategies to restore faith in marginalized
12 communities.

13 **Keywords** Vaccine hesitancy · Trust in institutions · COVID-19 · political beliefs · Twitter

14 1 Introduction

15 Skepticism about the safety and benefits of vaccines is linked to low trust in institutions [1, 2]. Vaccine hesitancy, along
16 with climate change denial, is part of a larger trend of mistrust in scientific expertise and a decline in trust in public

*All authors contributed equally

† Author to whom correspondence should be addressed

Exploring the Landscape of Vaccine Hesitancy on Twitter

17 institutions [3, 4, 5]. To combat the "infodemic," the World Health Organization (WHO) has worked with major social
18 media platforms to redirect internet users to reliable websites when searching for information related to COVID-19
19 [6]. Restoring confidence in scientific institutions is therefore an essential crucial democratic task for policymakers.
20 This article endeavors to furnish those in positions of responsibility with a comprehensive understanding, while
21 acknowledging the intricate nature of vaccine hesitancy and cautioning against the temptation of oversimplifying the
22 issue by attributing it to a single social parameter, a mistake that many prominent media outlets have made. [7, 8, 9, 10].

23 Although there is no single universal determinant for vaccine hesitancy, a growing body of literature has identified
24 factors that influence vaccine acceptance [11]. The Strategic Advisory Group of Experts (SAGE) on Immunization
25 has categorized these determinants into three main domains, providing a framework for exploring the complex and
26 multifaceted nature of vaccine hesitancy [12]. Addressing the contextual, individual/social group, and vaccine-specific
27 determinants of vaccine hesitancy can help to overcome mistrust in vaccines and improve population health outcomes.
28 Factors affecting vaccine acceptance are distinguished at different levels in scientific literature [13]. Our research
29 focuses on social determinants of vaccine hesitancy at the population level, rather than on the socio-cultural context that
30 influences individual decision-making.

31 A substantial body of literature has emerged on vaccine hesitancy to understand the factors that influence the uptake of
32 public health interventions. Researchers have studied vaccine hesitancy to identify the pathways and mediators that
33 contribute to a lack of public consent and resistance to various vaccines [11, 14, 15, 16, 17]. A key finding has been
34 that vaccine hesitancy takes various forms and manifests differently [18], and that these self-destructive attitudes are
35 unequally distributed among different segments of society [19] and across nations [20], exacerbating existing health
36 disparities [21].

37 The attitudes toward vaccines are diverse and constantly evolving, making it important to understand their complexity
38 [22, 23]. The views on vaccination have been found to be influenced by various sources of information, including the
39 growing influence of digital public spheres [24, 25, 26, 27, 28, 29, 30]. Recent research has highlighted the significance
40 of gathering and analyzing data from social media platforms in order to track the rapidly changing trends [29, 31],
41 polarizations [32, 29], discourses [33, 34, 35], and sentiments [36, 37, 38] related to vaccine attitudes.

42 Recent advancements in computational social research have opened up new avenues for social research, particularly in
43 analyzing the competing discourses in digital spheres. Social media data has been used to analyze societal attitudes
44 toward critical social problems, such as immigration, public health, and extremism, due to its cost-effectiveness and
45 ability to eliminate response biases [39]. In this study, we focused on analyzing vaccine hesitancy on Twitter, which
46 offers easy access to large amounts of relevant content from millions of users through text mining techniques. Twitter
47 provides several advantages as a data source in social research compared to conventional surveys. Firstly, it allowed us
48 to analyze attitudes toward vaccines at a relatively low cost. Secondly, computational tools helped us uncover hidden
49 patterns in a large number of Twitter messages with minimal human intervention. Unsupervised textual clustering
50 methods provided an insider's perspective on vaccine hesitancy while minimizing potential biases that could arise from
51 the researchers themselves.

52 Our research demonstrates the complexity of vaccine-hesitant attitudes and the diversity of vaccine-hesitant populations.
53 Our findings indicate that vaccine hesitancy is a multifaceted phenomenon, and the vast majority of vaccine-hesitant
54 Twitter users pragmatically use different grounds to justify their stance. Additionally, we found that vaccine-hesitant
55 content on Twitter comes from diverse geographical locations. While socio-economic factors do play a role in vaccine
56 hesitancy at the state level in the U.S., our study suggests that attributing vaccine hesitancy solely to a single social
57 factor, such as education, income, race, or voting behavior, is not appropriate. It is crucial to exercise caution when
58 interpreting media reports and correlational studies that emphasize a single social determinant driving vaccine hesitancy.
59 Our study provides new insights that can help to deepen our understanding of the underlying factors contributing to
60 tribal polarization in the U.S. and revisit the relationships between vaccine hesitancy and social factors to improve
61 public health outcomes.

62 2 Results

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

69 We utilized the Twitter Search v2 API (Methods) to fetch all English tweets containing the keywords 'vaccine' and
70 'vaccination' from 2020 to 2022. Our resulting dataset comprises approximately 53 million tweets from around 8.2
71 million users, with roughly 584 thousand geolocated tweets for spatial analysis (Fig. 1.a). The vast majority of tweets
72 within our dataset are representative of the United States. During the Eastern Standard Time (EST) daytime hours, we
73 observed a peak in the number of tweets (Fig. S1), and a majority of geolocated tweets originated from the United
74 States (Table S1). To eliminate any possible bias from non-US tweets, we further subset the dataset to solely include
75 geolocated tweets originating from the United States for spatial analysis. We utilized the COVID-Twitter-BERT model
76 [40] to obtain embedding vectors, which are numerical representations of tweets that capture their semantic meaning
77 [41]. These embeddings facilitated downstream language tasks, such as classifying vaccination attitudes and conducting
78 topic modeling.

79 A classification model was trained to predict attitudes toward vaccination by fine-tuning the BERT model on the
80 previously annotated vaccination sentiment dataset [42] (Methods). The model achieved high performance, with
81 PR-AUC scores of 0.98 for neutral, 0.95 for positive, and 0.8 for negative classes on the test set (Fig. 1.b). Neutral
82 tweets were filtered out, and only positive and negative tweets with a predicted class probability above 99% were used
83 to ensure high precision (Fig. S2). At the proposed threshold (indicated with a red star in Fig. 1.b), the model achieved
84 a precision of 94% at 87% recall for the positive class and an 86% precision at 68% recall for the negative class
85 (Table S2). Notably, the classification performance reported in this study outperforms the Fasttext-based approach of
86 Muller et al. (2019), which was reported as 77% for precision and 77% for recall (Table S3).

87 The study developed attitudes toward vaccination (ATV) score, which summarizes the proportion of positive and
88 negative tweets in terms of log odds ratio by their location or the specific time period they were posted. The ATV
89 score was calculated per day using all tweets and by the US state using only geolocated tweets. The ATV scores of
90 geolocated tweets are representative of all tweets based on monthly ATV score comparison (Spearman's rank correlation
91 coefficient = 96%, $P = 6 * 10^{-5}$), as shown in Fig. S3.a,b. The study also investigated the relationship between the
92 vaccination rate of each US state by December 31st, 2021 [43] and state-level ATV scores, finding a high correlation
93 (Spearman's rank correlation coefficient = 53%, $P = 6 * 10^{-5}$) that indicates the proposed ATV score is representative
94 of the vaccination behavior of a broader US population beyond the Twittersphere (Fig. S4.a,b).

95 In order to understand the stability and variability of attitudes toward vaccination, we investigated the timeline of two
96 years (Fig. 1.c). The attitude was defined as stable if the vaccination attitude did not change over time, and as variable if
97 the perspective on vaccination differed over time. While positive tweets about vaccines predominated over negative
98 tweets during the entire timeline, the relative ratio of positive to negative tweets fluctuated. The first quarter of 2020
99 had a small fraction of all tweets, comprising less than 0.5% of the entire dataset, which could be explained by low
100 COVID-19 death numbers in the USA. However, the ATV score decreased sharply with rapidly increasing COVID-19
101 death numbers in late March and stayed relatively low until November 2020. This period also had a high unemployment
102 rate, although it peaked in early May at 15.8% [44] (Fig. S5). During 2021, vaccination had greater online public
103 attention, and the aggregate attitudes toward vaccination were elevated in any quarter of 2021 compared to 2020 (Fig.
104 S6), parallel to the successful vaccine trials and the increasing vaccination rate. However, the variability of the aggregate
105 vaccination attitudes may not imply a shift in individuals' views toward vaccination because the composition of users
106 differs over time.

Exploring the Landscape of Vaccine Hesitancy on Twitter

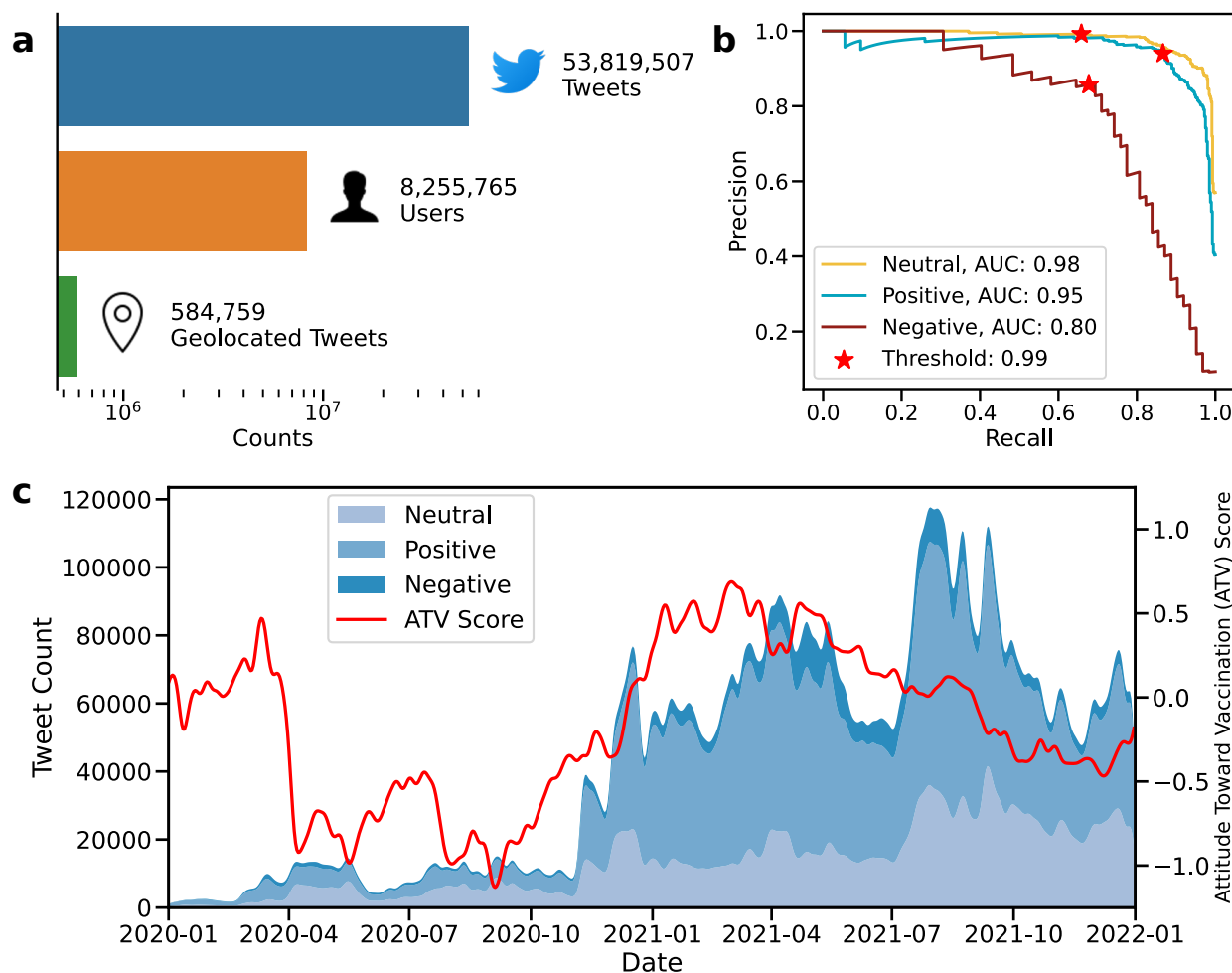


Figure 1: **Dataset, attitude prediction** (a) Descriptive statistics of the Twitter vaccination dataset, such as the number of tweets, geolocated tweets, and users. (b) The precision-recall curve and AUC scores of the model for vaccination attitude classification. The red star indicates the threshold 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.

107 To address this, we conducted a user-level analysis by tracking the vaccination attitudes of users who posted multiple
108 tweets across the studied timeline. We found that 31% of users had more than one tweet in the dataset, and most users
109 were clearly polarized into positive and negative attitude camps by predominantly tweeting with a single attitude (Fig.
110 2.b, Fig. S7). There was no apparent difference between these groups based on the number of tweets they posted (Fig.
111 S8.a,b). Also, there were only a few outliers with more than thousands of tweets, indicating that few users do not
112 disproportionately influence our analysis (Fig. 2.a). The percentage of positive tweet distributions of users before and
113 after November 1st, 2020 is shown in Fig. 2.b, and there is a statistically significant shift to a more positive opinion.
114 Specifically, users are 11% more likely to post a tweet with a positive attitude after November 1st, 2020, compared
115 to their past selves. We repeated this analysis for all the dates in our dataset and observed that attitudes toward users
116 decreased in the first five months of 2020 and increased afterward, except in the last four months of 2021, when there
117 was no significant change (Fig. 2.c, Fig. S9.a,b).

118 We conducted an even simpler analysis with the users who posted two tweets ($n=906,430$). The comparison of their first
119 tweet against their later tweet shows a similar trend to a more positive attitude, regardless of the date. Only 12% of users
120 with a positive first tweet later posted a negative second tweet, while 40% of users with a negative first tweet posted a
121 positive second tweet (Table S4). Although the majority of users did not change their attitudes toward vaccination, there

Exploring the Landscape of Vaccine Hesitancy on Twitter

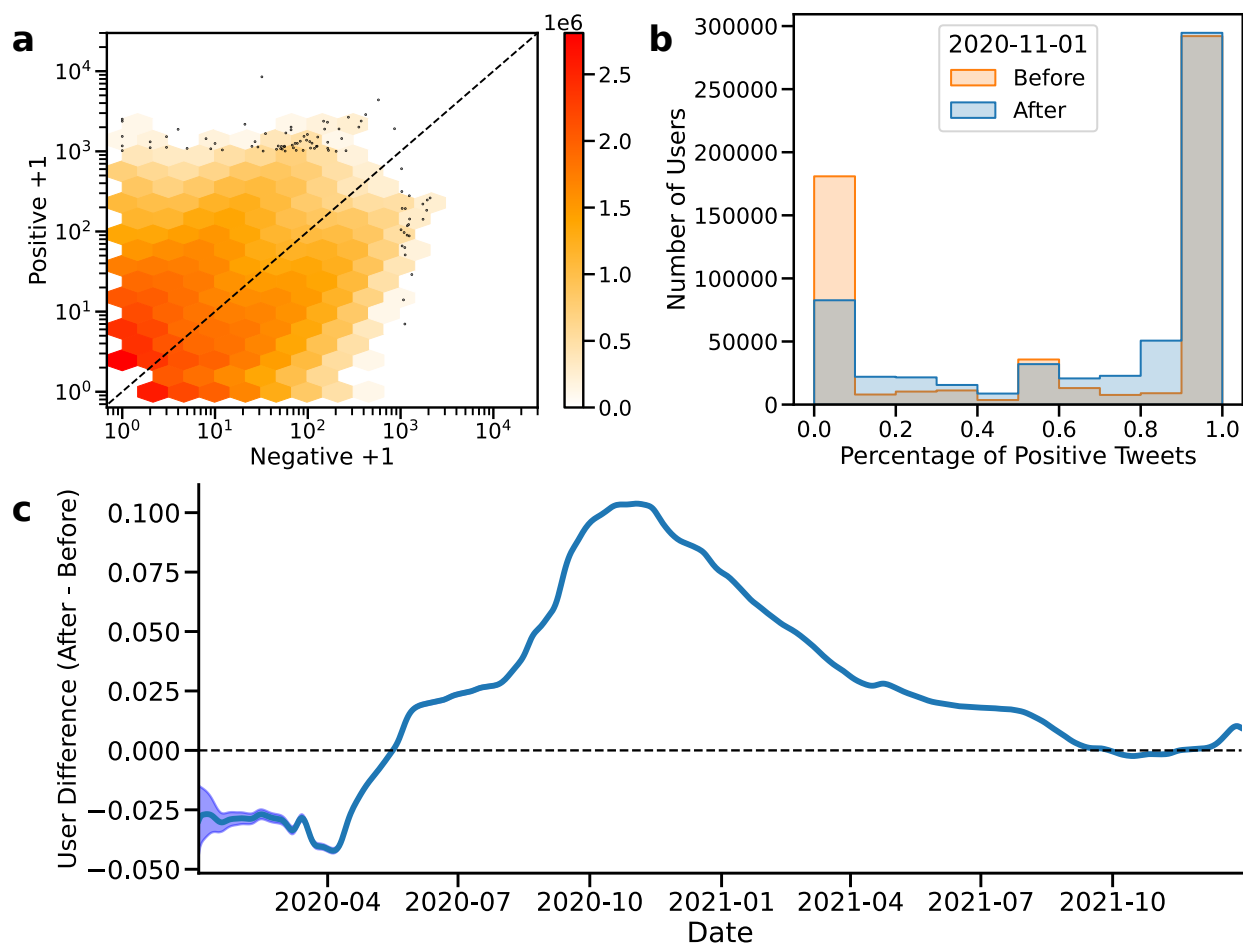


Figure 2: **Temporal change of attitudes** (a) The number of tweets posted by users (b) The percentage of positive tweets of users with at least two tweets before and after November 1st. (c) The trend of vaccination attitudes at the user level.

122 is a significant minority of negative users who have switched to a more positive attitude. Overall, both aggregate and
 123 user-level analysis indicate that a significant minority of users have more positive attitudes; yet the vast majority of
 124 users retained that vaccination attitude.

125 To investigate the socio-economic factors associated with vaccine hesitancy, we calculated the ATV score by the
 126 US states based on geolocated tweets and identified distinct regions based on attitude (as shown in Fig. 3.a). Using
 127 Fisher's exact test, we found that 25 out of 50 states have a significant pro-vaccine stance, while the other 25 have
 128 an anti-vaccine stance (Fig. 3.b and Table S5). We collected 10 socio-economic parameters for each state (Methods)
 129 and found that 8 out of 10 socio-economic parameters are significantly correlated with vaccination attitude based on
 130 univariate correlation analysis (Fig. S10.a-j). For instance, there is a negative correlation of -49% ($p = 3 * 10^{-4}$)
 131 between ATV score and having less than a high school degree (Fig. S10.b), while there is a positive correlation of 45%
 132 ($p = 10^{-3}$) between voting for Trump in 2020 and ATV (Fig. S10.h). However, we also found that the higher rate of
 133 cat ownership compared to the rate of dog ownership has a 46% Spearman's rank correlation coefficient (Fig. S10.j),
 134 indicating the limitation of the univariate analysis, which may not accurately represent the underlying socio-economic
 135 factors related to vaccination attitude. It is important to note that socio-economic factors are not easily separable but
 136 rather interconnected (as depicted in Fig. 3.c), and therefore, multicollinearity between socio-economic parameters may
 137 lead to spurious correlations.

Exploring the Landscape of Vaccine Hesitancy on Twitter

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

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

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

158 The common justifications for vaccine hesitancy were obtained by performing contextual topics modeling analysis
159 (Methods) on the tweets with negative attitudes. As a result, seven main themes for vaccine hesitancy were discovered
160 (Table 1, Fig. S14, Table S7). Similar themes have been reported in the childhood vaccination literature. For instance,
161 Majid and Ahmad (2020) examined 34 qualitative studies to clarify parents' reasons for rejecting or delaying vaccines
162 [45]. Although their examination did not cover the studies on COVID-19 vaccine hesitancy, the fear of side effects and
163 skepticism regarding vaccine effectiveness were prevalent among parents who rejected vaccines, which is consistent
164 with what we found in our topics. Additionally, some of the studies they examined highlighted that distrust in health
165 system players and mandatory vaccine policies were reasons why parents rejected vaccines for their kids. In our study,
166 we encountered a similar stance on the alleged conflict of interest in the health system and categorized it as politics and
167 conspiracy theories. Similarly, mandatory vaccine policies are a source of frustration for many vaccine-hesitant Twitter
168 users in our research. This demonstrates parallelism between the ways in which people justify their vaccine-hesitant
169 attitudes in the face of the COVID-19 pandemic and previous waves of anti-vaxxer movements.

170 Using a topic modeling technique, we have identified eight themes related to vaccine hesitancy in our dataset. The
171 first theme covers discussions on mandatory vaccinations and features social media conversations about vaccine
172 administration practices. Many vaccine hesitant Twitter users framed mandatory vaccination as a violation of their
173 individual rights. Another theme pertains to the constitutionality of vaccine mandates, such as vaccine passports and
174 vaccination requirements for state and federal workers. Our insights from these two themes echo recent research in
175 the history of vaccination, which has documented how activists in US history opposed vaccination on the basis of
176 rights, freedom, and liberty, even when vaccination was not compulsory. [46] This work highlights the ongoing tension
177 between scientific expertise and civic freedom to choose one's own medical practices. In social media, we see similar
178 traces of this tension that led to the Jacobson v. Massachusetts case in 1905. [46, 47]

179 The Twitter discussions focus heavily on the health concerns of users, with several themes emerging. One theme in this
180 category is the alleged adverse reactions of vaccines. Another related theme focuses on the long-term side effects of
181 COVID-19 vaccines. Vaccine hesitant tweets on this topic discuss potential long-term side effects, which may arise
182 from the novel mRNA method or the risks associated with the expedited vaccine approval process. "Ineffectiveness" is

Exploring the Landscape of Vaccine Hesitancy on Twitter

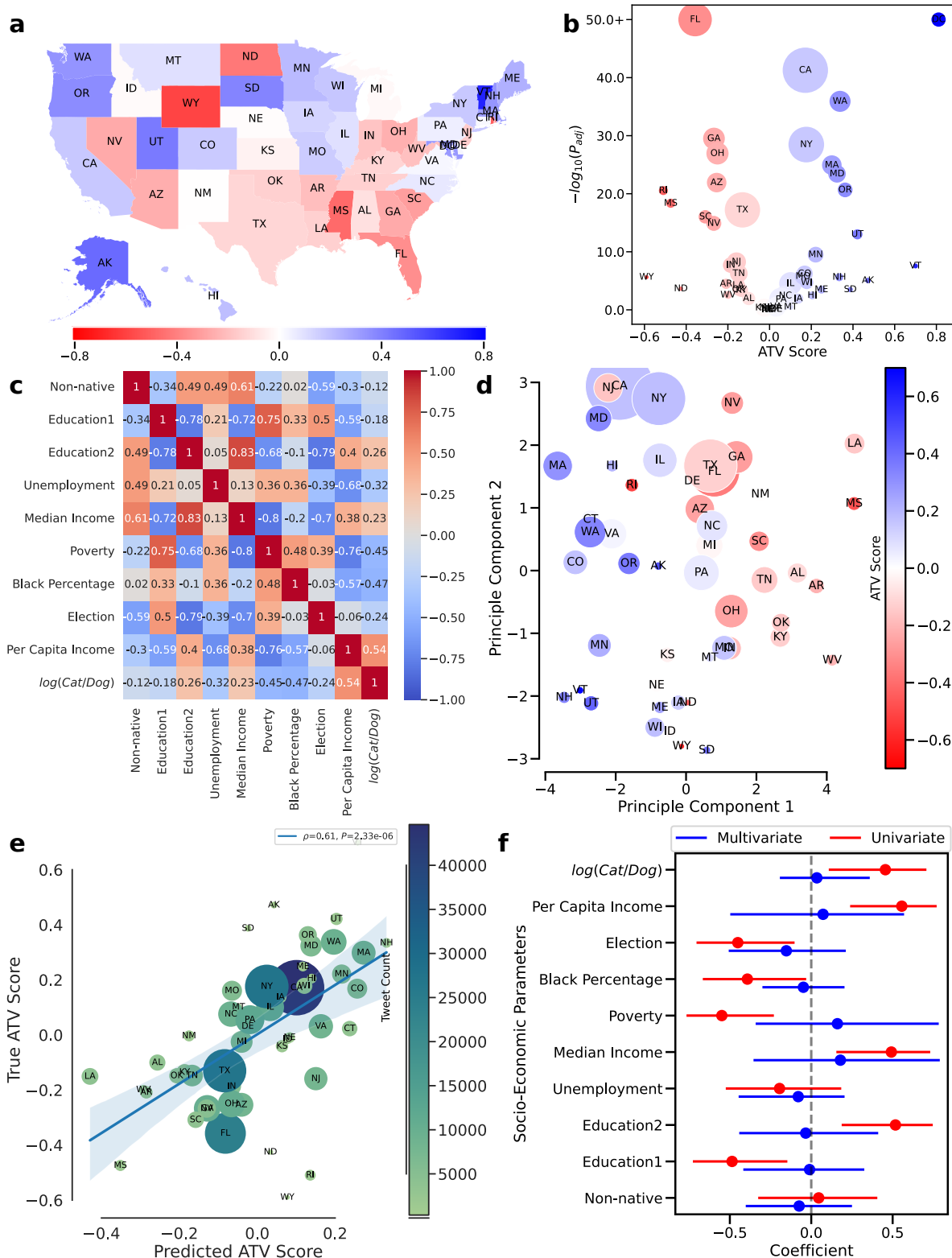


Figure 3: **State-level socio-economic parameters and attitudes toward vaccination** (a) U.S. states colored by ATV score. (b) Volcano plot of ATV score and statistical significance of vaccination attitude of each state (c) Cross-correlation of socio-economic parameters with Spearman's method (d) The 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.

Exploring the Landscape of Vaccine Hesitancy on Twitter

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

183 another health-related topic that covers skepticism over the efficiency of vaccines or their total denial. A systematic
 184 review found that safety and side effects were the two most commonly reported factors influencing COVID-19 vaccine
 185 hesitancy in various studies [48]. Another meta-analysis concluded that stronger beliefs in the unsafety of vaccines are
 186 among the predictors of vaccine hesitancy globally [49]. The belief that COVID-19 vaccines are unsafe or ineffective is
 187 the factor that has been supported by the largest number of studies in high-income countries, while concerns about the
 188 rapid development of vaccines follow them on the list [50].

189 A final set of health-related vaccine hesitancy themes involves social media content that either partially or completely
 190 denies the benefits of vaccines. The "natural immunity" theme suggests that natural immunity is superior to vaccines in
 191 terms of protection against the virus. Politics and conspiracy theories are closely linked to COVID-19 denialism in
 192 health-related tweets. Twitter users discussed the alleged conflict of interest in the production of scientific knowledge
 193 and its dissemination by the mass media in line with the political frame of their choice. The relationship between
 194 belief in conspiracy theories and vaccine hesitancy has been extensively studied [51, 52]. However, the contribution of

Exploring the Landscape of Vaccine Hesitancy on Twitter

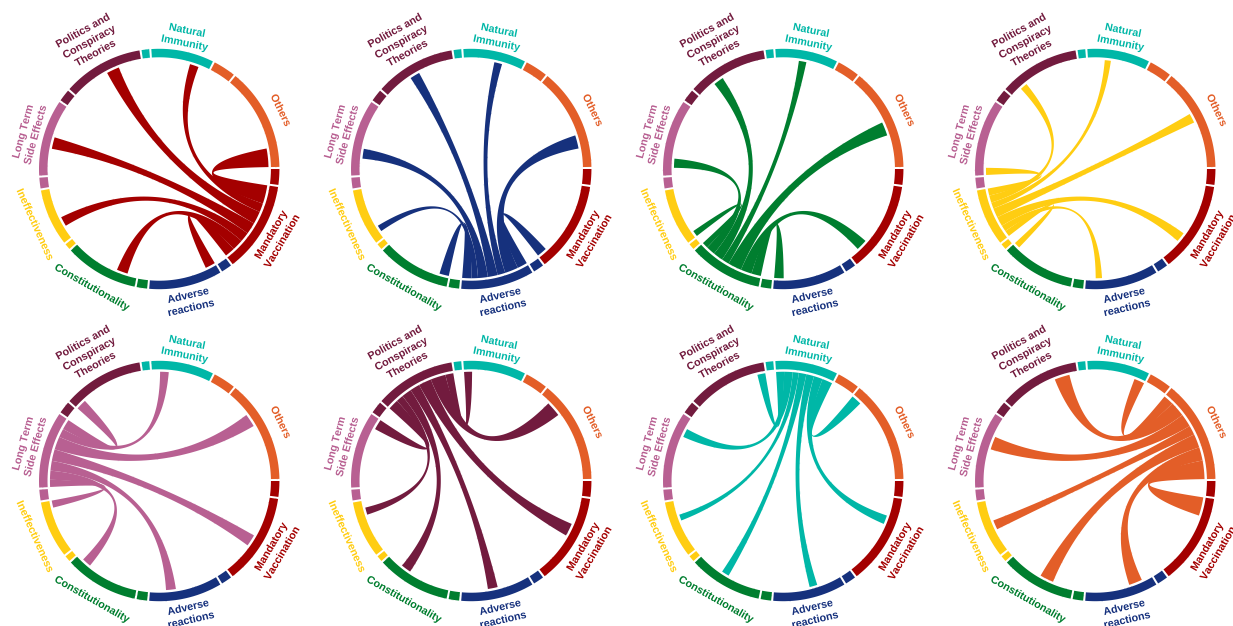


Figure 4: The topic distribution of Twitter posts of users who posted multiple tweets. Each diagram represents a group of users interested in a specific topic and illustrates how their other tweets are distributed.

195 conspiracy theories to vaccine hesitancy has been overstated without solid empirical evidence. In fact, an independent
196 research center has found that only a few people were responsible for the vast majority of misleading content about
197 vaccines on social media [53]. In terms of substance, the Big Pharma conspiracy theory predates the COVID-19
198 pandemic and uses *cui bono* reasoning to identify a small number of elites who benefit from either deliberately
199 manufacturing or not fully curing diseases [54].

200 The extracted substantial themes, including the seven main themes and the "others" category, reveal the contours of the
201 vaccine hesitant debate on Twitter. Conspiracy theories with political implications represent only one aspect of vaccine
202 hesitant content; there is much more to it. Vaccine administration and health concerns are the two primary branches of
203 vaccine hesitancy in social media discussions. Both of these strands encompass a wide range of themes, from moderate
204 skepticism to extreme denialism. Our model's vaccine hesitant topics are consistent with previous studies on vaccine
205 hesitancy, which have also identified concerns about side effects, lack of trust, and belief in the ineffectiveness of
206 vaccines as subthemes [55].

207 Earlier in our study, we illustrated that vaccine hesitancy takes on many shades and facets. To gain a better understanding
208 of how vaccine-hesitant individuals rationalize their opposition to COVID-19 vaccines, we explored the extent to which
209 they indiscriminately adopt arguments from other prevalent themes of vaccine hesitancy. Specifically, we examined
210 whether users were solely focused on one particular topic or if their concerns spanned across multiple topics. To achieve
211 this, we categorized users by topics if they had posted at least one tweet on the subject. Then, we measured the topic
212 distribution of their remaining tweets, as illustrated in Fig. 4. We observed that users in our dataset did not exhibit a
213 consistent pattern of opposition based on a single theme within vaccine hesitancy topics. Instead, they demonstrated a
214 high degree of variability across topics, not limiting themselves to a specific theme. These individuals employed a broad
215 range of topics to rationalize their vaccine hesitancy. Fig. 4 portrays the co-occurrence of the same users across various
216 vaccine hesitant themes. For instance, a user who posted their first tweet on adverse reactions might subsequently
217 employ an argument from the conspiracy theory theme to bolster their stance. Overall, the desultory use of arguments
218 from a wide range of topics may imply that users are potentially motivated by deeper prejudices, and they utilize any
219 available justification in their Twitter posts.

220 **3 Discussion**

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

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

229 To delve deeper into the attributes of the polarized online communities, we used a novel methodology by merging
230 geolocated online content with the conventional socio-economic variables based on the location of Twitter messages.
231 Our association analysis between socio-economic variables and vaccine hesitancy demonstrated that socio-economic
232 parameters can predict the degree of vaccine hesitancy in US states. The residents of those states suffer from higher rates
233 of unemployment, lower median household income, and poorer educational outcomes. However, the vast disparities in
234 vaccine hesitancy among US states cannot be solely attributed to a single social parameter, such as election results,
235 education, race, or income. Vaccination hesitant online communities have unique characteristics defined by a complex
236 amalgamation of socio-economic parameters.

237 To gain further insights into discourses of the vaccine-hesitant community, we conducted a textual analysis and identified
238 various themes revolving around two main concerns, political and medical. Political concerns branch out to two different
239 themes: constitutional issues related to vaccine mandates and conspiracy theories. Likewise, medical concerns include
240 two dissimilar themes: medical side effects and the denial of their effectiveness. These concerns are similar to the
241 reasons to refuse vaccines identified by a systematic review of earlier work as (1) medical safety of COVID vaccines,
242 (2) the inefficiency of vaccines, and (3) belief in natural immunity [56]. Yet, Twitter users who posted multiple times
243 are not fixated on a single issue, rather they pragmatically borrowed arguments from a wide range of vaccine hesitant
244 themes and they often justified their stances against vaccines using multiple reasons.

245 Several studies have indicated that marginalized communities are less likely to trust institutions, including vaccines
246 [57, 58]. For example, working-class Whites are one such group, often portrayed as former President Trump's base [59].
247 They inhabit areas characterized by high levels of poverty and elevated rates of premature deaths due to gun violence,
248 suicide, drug overdoses, and alcoholism, compared to the national average [60, 61]. The historical marginalization
249 of those communities leads to a higher concern about government intrusion in their personal lives. As a result, those
250 particular segments of society feel fear and anxiety in the face of strict vaccine mandates and protocols.

251 The moral foundations theory can illuminate our findings regarding the vaccination hesitant community [62]. The
252 theory suggests that people primarily rely on moral intuitions shaped by socio-psychological factors to make political
253 judgments, then they justify their judgments with strategic reasoning [62]. Some Twitter users' attitudes toward vaccines
254 may have been shaped by emotions and gut-level feelings, leading to distrust of institutions that underlies vaccine
255 hesitancy. Our findings regarding the spurious relationship between a solitary socio-economic parameter and vaccine
256 hesitancy as well as the indiscriminate use of different topics may support this idea. This distrust may stem from a
257 heightened moral preference expressed through a series of ad-hoc hypotheses, as classified as diverse vaccine-hesitant
258 themes in our study [63].

259 To address these concerns, a well-planned public communication strategy is necessary. Understanding the moral
260 foundations of their attitude and establishing empathy should be the initial step. One approach to restore faith and
261 rebuild trust in those communities would be to adopt a persuasive language that emphasizes the common good, with
262 the support of local leaders. Without the support of their in-group, public health efforts may be seen as stigmatizing

263 and compulsory measures by marginalized communities [58]. Communication strategies tailored to local contexts may
264 improve to trust in science.

265 **4 Methods**

266 **4.1 Fetching Tweets**

267 In order to fetch tweets, we utilized searchtweets-v2 [64] and Tweepy [65] python packages. Tweets were fetched using
268 the search endpoint of the Twitter-API (<https://api.twitter.com/2/tweets/search/all>). We fetched all tweets containing the
269 keywords: “Vaccine” or “Vaccination”. We limited our analysis with those English tweets posted between 2020 and
270 2022. We excluded retweets, short, (less than 10 words) and duplicate tweets. Then we used the geo-location identifier
271 of tweets to locate the state where tweets were posted from by using Tweepy library.

272 **4.2 Dataset and Model Training For Vaccination Attitude Prediction**

273 To identify individuals who express vaccine hesitancy, tweets were classified into one of three categories of "Positive",
274 "Negative", or "Neutral" by using BERT-based natural language processing (NLP) models. BERT embeddings were
275 extracted using Covid-Twitter-Bert-V2 [66] model from HuggingFace [67] platform. Then, the model was fine-tuned
276 on the dataset by Pananos et. al. [42] to predict vaccination attitude. The dataset is consists of 27,906 tweets which
277 were manually labeled by Amazon Mechanical Turk. The annotators of the crowdsourced dataset have a consensus for
278 only 16,156 tweets. The dataset is partitioned into the train, validation, and test sets with proportions of 80%, 10%, and
279 10%, respectively. To create the validation and test sets, we only included tweets with consensus. The remaining tweets,
280 with or without consensus, were used for fine-tuning with SimpleTransformers library [68]. During the fine-tuning,
281 the model was trained for 20 epochs with the early stopping of 5 epochs. To optimize our model’s performance, we
282 conducted hyperparameter tuning for several parameters. The key parameters that we hyperparameter-tuned were
283 learning rate, epsilon for Adam optimization, and maximum sequence length. For the learning rate, we experimented
284 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
285 weights were calculated based on the inverse proportion of class frequencies to ensure balanced training. The weights
286 used were 1.0, 1.43, and 9.2 for Neutral, Positive, and Negative, respectively. The best score was obtained with the
287 parameters of 1e-05 for learning rate, 1e-07 for epsilon, and 128 for maximum sequence length. The performance of
288 our model was evaluated using the f1-macro metric using scikit-learn [69] during the hyperparameter tuning.

289 **4.3 Prediction of Attitudes Toward Vaccination Score**

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

$$ATV = \log\left(\frac{p_c/p_n}{p_{other}/n_{other}}\right) \quad (1)$$

294 To obtain the statistical significance of attitudes toward vaccination scores by the US states, we also employed Fisher’s
295 Exact Test [70] with categories of attitudes (positive and negative) and the state tweet was sent from. P-values of
296 Fisher’s Exact test were corrected for multiple testing using Bonferroni correction method.

297 4.4 Socio-Economic Parameters and ATV Score

298 To investigate the relationship between vaccine hesitancy and socio-economic parameters, we used 2020 American
299 Community Survey 1-Year Experimental Estimates data by the U.S. Census Bureau. The parameters included the
300 percentage of Black population (Black percentage), the percentage of people living in poverty in the past 12 months
301 (Poverty), inflation-adjusted median household income in the past 12 months (Median income), employment status
302 for individuals aged 16 years and over (Unemployment), and citizenship status (Non-native). We also studied the
303 educational attainment of the population aged 25 years and over, specifically the percentage of those with less than a
304 high school degree (Education1) and those with more than a bachelor's degree (Education2). Additionally, we examined
305 the influence of presidential elections (Election) on attitudes toward vaccination using the results of a study [71]. We
306 also included Social Capital (PSU-SC) index (Rupasingha et al. (2006 with updates) [72]) and Cat/Dog ownership ratio
307 by the US states [73] as a dichotomous variable ($\log(\frac{cat}{dog})$).

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

$$\tanh\left(\operatorname{arctanh}(r) \pm \frac{2.576}{\sqrt{n-3}}\right) \quad (2)$$

310 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
311 interval with the Bonferroni correction.

312 4.5 Partial Least Squares (PLS) Regression

313 For the multivariate analysis, we used Partial Least Squares (PLS) Regression [74] to avoid multicollinearity between
314 the socio-economic parameters. Bootstrapping was conducted for 10,000 iterations to obtain error bars for regression
315 coefficients with 99% confidence interval. In each bootstrapping iteration, the model was fitted on the random subset of
316 states.

317 4.6 Contextualized Topic Modeling (CTM)

318 We utilized the Contextualized Topic Modeling (CTM) [75] library, which is a Variational AutoEncoder (VAE) [76]
319 based deep learning model. CTM takes BERT embeddings as an input and predicts a bag of words (BoW) by sampling
320 from the bottleneck layer. The bottleneck size is equal to the number of topics. This approach enabled us to assign a
321 topic to each tweet with an unsupervised method. We conducted topic modeling with only negative tweets. NLTK [77]
322 library is utilized to remove stop words. The number of topics, hidden layer sizes of the encoder and the decoder, and
323 the dropout rate are hyperparameter tuned. We tuned topics between 5 to 10, unit size of the hidden layer of 200, 500,
324 and 700, and dropout rates of 0.2, 0.5, and 0.8. All hyperparameter tuning experiments were done in 2 epochs. The best
325 hyperparameters were chosen based on topic coherence and diversity metrics. We selected the topic number as 8, a
326 hidden layer size as 200, and a dropout as 0.2. Topic coherence is the average normalized pointwise mutual information
327 (NPMI) score between the top keywords in a topic. The formula for topic diversity is defined as follows:

$$TD = \frac{|U|}{k \cdot T} \quad (3)$$

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

333 5 Data and Code Availability

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

337 References

- 338 [1] Maya J Goldenberg. Vaccines, values and science. *CMAJ*, 191(14):E397–E398, 2019.
- 339 [2] Patrick Sturgis, Ian Brunton-Smith, and Jonathan Jackson. Trust in science, social consensus and vaccine
340 confidence. *Nature Human Behaviour*, 5(11):1528–1534, 2021.
- 341 [3] Gabrielle Wong-Parodi and Irina Feygina. Understanding and countering the motivated roots of climate change
342 denial. *Current Opinion in Environmental Sustainability*, 42:60–64, 2020.
- 343 [4] Robert Gifford. The dragons of inaction: psychological barriers that limit climate change mitigation and adaptation.
344 *American psychologist*, 66(4):290, 2011.
- 345 [5] Yongjin Choi and Ashley M Fox. Mistrust in public health institutions is a stronger predictor of vaccine hesitancy
346 and uptake than trust in trump. *Social Science & Medicine*, 314:115440, 2022.
- 347 [6] John Zarocostas. How to fight an infodemic. *The lancet*, 395(10225):676, 2020.
- 348 [7] David Leonhardt. U.s. covid deaths get even redder. *The New York Times*, Nov 2021.
- 349 [8] Philip Bump. Analysis | the 'what about black people' defense of republican vaccine hesitancy, Aug 2021.
- 350 [9] Philip Bump. Analysis | the biggest divide on vaccination isn't race or income but party — and the divide is
351 growing. *Washington Post*.
- 352 [10] James Risen. The right's anti-vaxxers are killing republicans.
- 353 [11] Heidi J Larson, Caitlin Jarrett, Elisabeth Eckersberger, David MD Smith, and Pauline Paterson. Understanding
354 vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published
355 literature, 2007–2012. *Vaccine*, 32(19):2150–2159, 2014.
- 356 [12] Caitlin Jarrett, Rose Wilson, Maureen O'Leary, Elisabeth Eckersberger, Heidi J Larson, et al. Strategies for
357 addressing vaccine hesitancy—a systematic review. *Vaccine*, 33(34):4180–4190, 2015.
- 358 [13] Eve Dubé, Caroline Laberge, Maryse Guay, Paul Bramadat, Réal Roy, and Julie A Bettinger. Vaccine hesitancy:
359 an overview. *Human vaccines & immunotherapeutics*, 9(8):1763–1773, 2013.
- 360 [14] Heidi J Larson, Caitlin Jarrett, William S Schulz, Mohuya Chaudhuri, Yuqing Zhou, Eve Dube, Melanie Schuster,
361 Noni E MacDonald, Rose Wilson, et al. Measuring vaccine hesitancy: the development of a survey tool. *Vaccine*,
362 33(34):4165–4175, 2015.
- 363 [15] Daniel A Salmon, Matthew Z Dudley, Jason M Glanz, and Saad B Omer. Vaccine hesitancy: causes, consequences,
364 and a call to action. *Vaccine*, 33:D66–D71, 2015.
- 365 [16] Noni E MacDonald et al. Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34):4161–4164,
366 2015.
- 367 [17] Patricia Soares, João Victor Rocha, Marta Moniz, Ana Gama, Pedro Almeida Laires, Ana Rita Pedro, Sónia Dias,
368 Andreia Leite, and Carla Nunes. Factors associated with covid-19 vaccine hesitancy. *Vaccines*, 9(3):300, 2021.
- 369 [18] Juhani Eskola, Philippe Duclos, Melanie Schuster, Noni E MacDonald, et al. How to deal with vaccine hesitancy?
370 *Vaccine*, 33(34):4215–4217, 2015.
- 371 [19] Mariam Siddiqui, Daniel A Salmon, and Saad B Omer. Epidemiology of vaccine hesitancy in the united states.
372 *Human vaccines & immunotherapeutics*, 9(12):2643–2648, 2013.

Exploring the Landscape of Vaccine Hesitancy on Twitter

- 373 [20] Julio S Solís Arce, Shana S Warren, Niccolò F Meriggi, Alexandra Scacco, Nina McMurry, Maarten Voors,
374 Georgiy Syunyaev, Aryn Abdul Malik, Samya Aboutajdine, Opeyemi Adejo, et al. Covid-19 vaccine acceptance
375 and hesitancy in low-and middle-income countries. *Nature medicine*, 27(8):1385–1394, 2021.
- 376 [21] Aryn A Malik, SarahAnn M McFadden, Jad Elharake, and Saad B Omer. Determinants of covid-19 vaccine
377 acceptance in the us. *EClinicalMedicine*, 26:100495, 2020.
- 378 [22] Eve Dubé, Dominique Gagnon, Emily Nickels, Stanley Jeram, and Melanie Schuster. Mapping vaccine
379 hesitancy—country-specific characteristics of a global phenomenon. *Vaccine*, 32(49):6649–6654, 2014.
- 380 [23] Charles Shey Wiysonge, Duduzile Ndwandwe, Jill Ryan, Anelisa Jaca, Oumarou Batouré, Blanche-
381 Philomene Melanga Anya, and Sara Cooper. Vaccine hesitancy in the era of covid-19: could lessons from
382 the past help in divining the future? *Human vaccines & immunotherapeutics*, 18(1):1–3, 2022.
- 383 [24] Marcel Salathé and Shashank Khandelwal. Assessing vaccination sentiments with online social media: implications
384 for infectious disease dynamics and control. *PLoS computational biology*, 7(10):e1002199, 2011.
- 385 [25] Anna Kata. Anti-vaccine activists, web 2.0, and the postmodern paradigm—an overview of tactics and tropes used
386 online by the anti-vaccination movement. *Vaccine*, 30(25):3778–3789, 2012.
- 387 [26] Cornelia Betsch, Noel T Brewer, Pauline Brocard, Patrick Davies, Wolfgang Gaissmaier, Niels Haase, Julie Leask,
388 Frank Renkewitz, Britta Renner, Valerie F Reyna, et al. Opportunities and challenges of web 2.0 for vaccination
389 decisions. *Vaccine*, 30(25):3727–3733, 2012.
- 390 [27] Holly O Witteman and Brian J Zikmund-Fisher. The defining characteristics of web 2.0 and their potential
391 influence in the online vaccination debate. *Vaccine*, 30(25):3734–3740, 2012.
- 392 [28] Neha Puri, Eric A Coomes, Hourmazd Haghbayan, and Keith Gunaratne. Social media and vaccine hesitancy:
393 new updates for the era of covid-19 and globalized infectious diseases. *Human vaccines & immunotherapeutics*,
394 16(11):2586–2593, 2020.
- 395 [29] Steven Lloyd Wilson and Charles Wiysonge. Social media and vaccine hesitancy. *BMJ global health*,
396 5(10):e004206, 2020.
- 397 [30] Anna Kata. A postmodern pandora’s box: anti-vaccination misinformation on the internet. *Vaccine*, 28(7):1709–
398 1716, 2010.
- 399 [31] Weiyu Zhang, Rong Wang, and Haodong Liu. Moral expressions, sources, and frames: Examining covid-19
400 vaccination posts by facebook public pages. *Computers in Human Behavior*, 138:107479, 2023.
- 401 [32] Ariel Fridman, Rachel Gershon, and Ayelet Gneezy. Covid-19 and vaccine hesitancy: A longitudinal study. *PloS*
402 *one*, 16(4):e0250123, 2021.
- 403 [33] Neil F Johnson, Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud,
404 Minzhang Zheng, Pedro Manrique, Stefan Wuchty, and Yonatan Lupu. The online competition between pro-and
405 anti-vaccination views. *Nature*, 582(7811):230–233, 2020.
- 406 [34] Melanie L Kornides, Sarah Badlis, Katharine J Head, Mary Putt, Joseph Cappella, and Graciela Gonzalez-
407 Hernandez. Exploring content of misinformation about hpv vaccine on twitter. *Journal of Behavioral Medicine*,
408 pages 1–14, 2022.
- 409 [35] Dominik Wawrzuta, Mariusz Jaworski, Joanna Gotlib, and Mariusz Panczyk. What arguments against covid-19
410 vaccines run on facebook in poland: content analysis of comments. *Vaccines*, 9(5):481, 2021.
- 411 [36] Weiyu Zhang, Subhayan Mukerjee, and Huazhi Qin. Topics and sentiments influence likes: A study of facebook
412 public pages’ posts about covid-19 vaccination. *Cyberpsychology, Behavior, and Social Networking*, 25(9):552–
413 560, 2022.

Exploring the Landscape of Vaccine Hesitancy on Twitter

- 414 [37] Michal Monselise, Chia-Hsuan Chang, Gustavo Ferreira, Rita Yang, and Christopher C Yang. Topics and
415 sentiments of public concerns regarding covid-19 vaccines: Social media trend analysis. *Journal of medical*
416 *Internet research*, 23(10):e30765, 2021.
- 417 [38] Joanne Chen Lyu, Eileen Le Han, and Garving K Luli. Covid-19 vaccine-related discussion on twitter: topic
418 modeling and sentiment analysis. *Journal of medical Internet research*, 23(6):e24435, 2021.
- 419 [39] Huseyin Zeyd Koytak and Muhammed Hasan Celik. A text mining approach to determinants of attitude towards
420 syrian immigration in the turkish twittersphere. *Social Science Computer Review*, page 08944393221117460,
421 2022.
- 422 [40] Martin M Müller and Marcel Salathé. Crowdbreaks: tracking health trends using public social media data and
423 crowdsourcing. *Frontiers in public health*, 7:81, 2019.
- 424 [41] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional
425 transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- 426 [42] A Demetri Pananos, Thomas M Bury, Clara Wang, Justin Schonfeld, Sharada P Mohanty, Brendan Nyhan, Marcel
427 Salathé, and Chris T Bauch. Critical dynamics in population vaccinating behavior. *Proceedings of the National*
428 *Academy of Sciences*, 114(52):13762–13767, 2017.
- 429 [43] Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie
430 Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina, and Max Roser. Coronavirus pandemic
431 (covid-19). *Our World in Data*, 2020. <https://ourworldindata.org/coronavirus>.
- 432 [44] Economic Research Service, U.S. Department of Agriculture. Unemployment and median household in-
433 come for the U.S., States, and counties, 2000-2021. [https://www.ers.usda.gov/data-products/](https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/)
434 [county-level-data-sets/county-level-data-sets-download-data/](https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/), 2021. Data as of June 2, 2021.
435 Contact: Kathleen Kassel, kathleen.kassel@usda.gov.
- 436 [45] Umair Majid and Mobeen Ahmad. The factors that promote vaccine hesitancy, rejection, or delay in parents.
437 *Qualitative Health Research*, 30(11):1762–1776, 2020.
- 438 [46] James Colgrove and Sara J Samuel. Freedom, rights, and vaccine refusal: The history of an idea. *American*
439 *Journal of Public Health*, 112(2):234–241, 2022.
- 440 [47] Wendy K Mariner, George J Annas, and Leonard H Glantz. Jacobson v massachusetts: it’s not your great-great-
441 grandfather’s public health law. *American Journal of Public Health*, 95(4):581–590, 2005.
- 442 [48] Debendra Nath Roy, Mohitosh Biswas, Ekramul Islam, and Md Shah Azam. Potential factors influencing covid-19
443 vaccine acceptance and hesitancy: A systematic review. *PloS one*, 17(3):e0265496, 2022.
- 444 [49] Carla Pires. Global predictors of covid-19 vaccine hesitancy: A systematic review. *Vaccines*, 10(8):1349, 2022.
- 445 [50] Junjie Aw, Jun Jie Benjamin Seng, Sharna Si Ying Seah, and Lian Leng Low. Covid-19 vaccine hesitancy—a
446 scoping review of literature in high-income countries. *Vaccines*, 9(8):900, 2021.
- 447 [51] Christina E Farhart, Ella Douglas-Durham, Krissy Lunz Trujillo, and Joseph A Vitriol. Vax attacks: How
448 conspiracy theory belief undermines vaccine support. *Progress in Molecular Biology and Translational Science*,
449 188(1):135–169, 2022.
- 450 [52] Matthew J Hornsey, Emily A Harris, and Kelly S Fielding. The psychological roots of anti-vaccination attitudes:
451 A 24-nation investigation. *Health psychology*, 37(4):307, 2018.
- 452 [53] Shannon Bond. Just 12 people are behind most vaccine hoaxes on social media, research shows. *NPR News*, 2021.
- 453 [54] Robert Blaskiewicz. The big pharma conspiracy theory. *Medical Writing*, 22(4):259–261, 2013.
- 454 [55] Tina Moghadam Fard, MohammadAli Shokri, Fatemeh Sadat Mirfazeli, Homa Mohammadsadeghi, Neda Shafiei,
455 Razieh Salehian, and Taghi Riahi. Covid-19 vaccination hesitancy is not all a conspiracy theory: A qualitative
456 study from iran. *Acta Psychologica*, page 103839, 2023.

Exploring the Landscape of Vaccine Hesitancy on Twitter

- 457 [56] Gianmarco Troiano and Alessandra Nardi. Vaccine hesitancy in the era of covid-19. *Public health*, 194:245–251,
458 2021.
- 459 [57] Alexandre De Figueiredo, Clarissa Simas, Emilie Karafillakis, Pauline Paterson, and Heidi J Larson. Mapping
460 global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective
461 temporal modelling study. *The Lancet*, 396(10255):898–908, 2020.
- 462 [58] Ed Pertwee, Clarissa Simas, and Heidi J Larson. An epidemic of uncertainty: rumors, conspiracy theories and
463 vaccine hesitancy. *Nature medicine*, 28(3):456–459, 2022.
- 464 [59] Nicholas Carnes and Noam Lupu. The white working class and the 2016 election. *Perspectives on Politics*,
465 19(1):55–72, 2021.
- 466 [60] Anne Case and Angus Deaton. Deaths of despair and the future of capitalism. In *Deaths of Despair and the
467 Future of Capitalism*. Princeton University Press, 2020.
- 468 [61] Jonathan M Metzl. *Dying of whiteness: How the politics of racial resentment is killing America’s heartland*.
469 Hachette UK, 2019.
- 470 [62] Jonathan Haidt. *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.
- 471 [63] Isabel Rossen, Mark J Hurlstone, Patrick D Dunlop, and Carmen Lawrence. Accepters, fence sitters, or rejecters:
472 Moral profiles of vaccination attitudes. *Social Science & Medicine*, 224:23–27, 2019.
- 473 [64] SearchTweets contributors. searchtweets-v2 1.1.1, 2023. [Online; accessed 10-April-2023].
- 474 [65] Joshua Roesslein. tweepy documentation. *Online*] <http://tweepy.readthedocs.io/en/v3>, 5:724, 2009.
- 475 [66] Martin Müller, Marcel Salathé, and Per E Kummervold. Covid-twitter-bert: A natural language processing model
476 to analyse covid-19 content on twitter. *arXiv preprint arXiv:2005.07503*, 2020.
- 477 [67] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac,
478 Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface’s transformers: State-of-the-art natural language
479 processing. *arXiv preprint arXiv:1910.03771*, 2019.
- 480 [68] T. C. Rajapakse. Simple transformers. <https://github.com/ThilinaRajapakse/simpletransformers>,
481 2019.
- 482 [69] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel,
483 Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python.
484 *the Journal of machine Learning research*, 12:2825–2830, 2011.
- 485 [70] Graham JG Upton. Fisher’s exact test. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*,
486 155(3):395–402, 1992.
- 487 [71] Massachusetts Institute of Technology (MIT), Election Data, and Science Lab. “us president 1976-2016. *Data
488 and Science MIT Election Lab 2017*, 2019.
- 489 [72] Anil Rupasingha, Stephan J Goetz, and David Freshwater. The production of social capital in us counties. *The
490 journal of socio-economics*, 35(1):83–101, 2006.
- 491 [73] Erin King, Megan K Mueller, Seana Dowling-Guyer, and Emily McCobb. Financial fragility and demographic
492 factors predict pet owners’ perceptions of access to veterinary care in the united states. *Journal of the American
493 Veterinary Medical Association*, 1(aop):1–8, 2022.
- 494 [74] Paul Geladi and Bruce R Kowalski. Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185:1–17,
495 1986.
- 496 [75] Federico Bianchi, Silvia Terragni, Dirk Hovy, Debora Nozza, and Elisabetta Fersini. Cross-lingual contextualized
497 topic models with zero-shot learning. *arXiv preprint arXiv:2004.07737*, 2020.
- 498 [76] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Exploring the Landscape of Vaccine Hesitancy on Twitter

- 499 [77] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the*
500 *natural language toolkit*. " O'Reilly Media, Inc.", 2009.

