

Original Paper

Cross-Cultural Differences in Public Discourse on COVID-19 Vaccination in the United States and South Korea: Cross-Sectional Analysis Using Natural Language Processing

Sangpil Youm¹, MS; Sou Hyun Jang², PhD; Haewoon Kwak³, PhD; Jaeyoung Choi⁴, BS; Yong Jeong Yi⁵, PhD

¹Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, United States

²Department of Sociology, Korea University, Seoul, Republic of Korea

³Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, United States

⁴Aml&Regtech Team, Naver Financial, Seongnam, Republic of Korea

⁵Department of Applied Artificial Intelligence, College of Computing & Informatics, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:

Yong Jeong Yi, PhD

Department of Applied Artificial Intelligence

College of Computing & Informatics

Sungkyunkwan University

25-2 Seonggyunwan-ro, Jongno District

Seoul, 03063

Republic of Korea

Phone: 82 2 760 0259

Email: yjyi@g.skku.edu

Abstract

Background: The COVID-19 vaccine was introduced as a crucial tool to combat the pandemic. However, concerns about its effectiveness, side effects, and misinformation spread remain. Prior research largely relied on survey-based approaches with limited populations. To address these limitations, social media offers a broader, more naturalistic lens into public discourse on COVID-19 vaccination. Accordingly, our study leverages social media data to identify factors shaping vaccine-related information needs, perceptions, and communication dynamics.

Objective: This study investigated public discourse about COVID-19 vaccines on community-driven question-and-answer sites in the United States (Quora; Quora, Inc) and South Korea (Naver Knowledge-iN; Naver Corp) to identify cross-national similarities and differences in vaccine-related information needs, sentiment patterns, and public perceptions over time.

Methods: We analyzed publicly available COVID-19 vaccine-related questions and answers posted between June 27, 2020, and June 27, 2021, on 2 community-driven question-and-answer platforms: Quora (United States) and Naver Knowledge-iN (South Korea). After preprocessing and sample-size matching, the dataset included 3952 question-answer pairs per platform, with one community-selected (most upvoted) answer analyzed per question. Natural language processing (NLP) techniques were applied for topic classification and sentiment analysis. Questions were categorized using a hybrid topic modeling approach combining Latent Dirichlet Allocation (LDA) and Top2Vec, identifying 5 topics on Quora and 7 topics on Naver Knowledge-iN. Answer sentiments were classified using an ensemble of Bidirectional Encoder Representations from Transformers (BERT; Google LLC)– and Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA; Google LLC)–based transformer models, and temporal sentiment trends were examined using monthly aggregation.

Results: Five shared information needs emerged, including effects of vaccines, variants, government policy, visiting overseas, and different vaccines, while South Korea uniquely exhibited vaccination appointments (711/3952, 18%) and school and education (513/3592, 13%). Negative sentiment predominated in US (Quora) answers across 4 of 5 topics, whereas positive sentiment exceeded 50% (498/790, 337/474, 367/592, 218/316, 348/553, 562/711, and 364/513) across all 7 topics on Naver Knowledge-iN. Temporally, US sentiment exhibited multiple positive-negative crossovers, whereas Korean sentiment stabilized toward positivity after February 2021, coinciding with the national vaccine rollout. Question-answer sentiment pairs showed contrasting interaction patterns, including negative-negative pairs dominated in the United States (eg, 504/978, 51.5% for different vaccines), while in South Korea, positive-positive and negative-positive pairs accounted for more than 63% (498/790, 337/474, 367/592, 218/316,

348/553, 562/711, and 364/513) of interactions in 7 topics, with positive-positive pairs most prevalent in 6 of 7 topics, except for variants.

Conclusions: Public perceptions of COVID-19 vaccines and related information needs differ between the 2 countries, shaped by cultural context, trust in government, and information-seeking environments. Analysis of social question and answer data from the 2 countries reveals shared information needs but divergent sentiment patterns. These findings highlight the value of social media data for public health research and the need for culturally and platform-specific communication strategies.

(*JMIR Form Res* 2026;10:e84791) doi: [10.2196/84791](https://doi.org/10.2196/84791)

KEYWORDS

COVID-19 vaccine; natural language processing; topic detection; sentiment analysis; temporal dynamics; health communication

Introduction

Since the recognition of COVID-19 in December 2019, many efforts have been made to control it. Before the development and distribution of the COVID-19 vaccine, the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) recommended passive preventive measures like mask-wearing, hand hygiene, and social distancing [1,2]. The COVID-19 vaccine, released in 2020 [3], was anticipated to be a pivotal tool in ending the pandemic. However, the scientific evidence supporting its effectiveness remains limited, and concerns about potential side effects persist [4,5]. Furthermore, misinformation about the vaccine has fueled public uncertainty and apprehension [6-8], contributing to vaccine hesitancy [4,9-11].

These concerns were reflected in real-world vaccination uptake, which varied across populations and national contexts. Despite the availability of COVID-19 vaccines in both the United States and Korea, vaccination rates stalled at some point, with 68.7% fully vaccinated in the United States [12] and 85.6% in Korea [13] as of October 2022. Public perceptions and intentions toward vaccination varied significantly in both countries [14-17].

In the United States, numerous studies have focused on vaccine hesitancy, analyzing factors like sociodemographic characteristics [5,14,15,18,19]. Vaccination hesitancy was notably higher among those experiencing social injustices, including race and religious discrimination, with hesitancy rates at 68.9% for non-Hispanic Whites, 74% for African Americans, and 59.5% for Hispanics or Latinos [15]. Lower-income Americans showed less fear of COVID-19 and less trust in vaccines, while higher-income Whites were more willing to get vaccinated [18,20]. Political affiliations further highlighted disparities, with Republicans exhibiting the highest vaccine hesitancy [21,22]. Racial injustice exhibited higher vaccine hesitancy due to limited access to information and greater mistrust in the government [19,23].

Taken together, existing studies have primarily relied on self-reported online surveys to characterize vaccine hesitancy, emphasizing the roles of personal attributes, demographic characteristics, and political orientations, while offering limited insight into how concerns and information needs are expressed in natural settings [24,25]. However, many studies are limited by focusing on specific populations rather than the general public [26]. Survey data limitations can be mitigated by analyzing public opinions expressed on social media [27]. Social

media is a natural setting for understanding public perceptions and information needs regarding the COVID-19 vaccine, potentially uncovering new factors influencing vaccine uptake [28].

To contextualize these vaccine-related discussions, social media serves as a central channel for exchanging health information, enabling users to share advice, knowledge, and experiences [27,29] while also playing a critical role in public health communication during health crises such as Hemagglutinin Type 1 and Neuraminidase Type 1 (H1N1), Middle East Respiratory Syndrome (MERS), and the COVID-19 pandemic [30,31].

During the COVID-19 pandemic, social media significantly impacted information sharing and vaccine hesitancy [32-35]. These studies underscore social media's role in shaping public perceptions and providing social and informational support, fostering a sense of connectedness during lockdowns [36,37].

Although social media offers many benefits, previous studies have suggested its potential risks associated with the spread of misinformation, which may cause severe consequences [33,38]. In the case of vaccine information, negative content has been widely disseminated through social media channels, contributing to vaccine hesitancy [39-41].

On Parler (Parler LLC), conservative-leaning hashtags like #NoCovidVaccine have supported negative vaccine perceptions [42]. YouTube's (Google LLC) user-generated content often includes misinformation due to limited regulations, and its recommendation systems promote conspiracy theories [43].

On Twitter (X Corp), sentiment toward vaccines varied by country; Switzerland had more positive tweets (11,621/16,268, 71.43%) than negative tweets, but the United States had only 10.48% positive tweets [44]. Only a small portion (< 10%, 9/107) of the most influential profiles engaging in the COVID-19 vaccine conversations originated from the medical professionals, indicating the need for public health stakeholders to address misinformation on social media [42,45].

Among these platforms, community-driven question and answer (Q&A) sites stand out as valuable online resources. Q&A sites like Quora (Quora, Inc) in the United States and Naver Knowledge-iN (Naver KiN; Naver Corporation) in South Korea (hereafter Korea) allow users to express their health information needs in natural language while fostering a sense of community [46]. During the COVID-19 pandemic, Q&A sites served as forums for discussing vaccines, offering valuable insights into

public information needs. Understanding these needs and vaccine perceptions is crucial for creating effective vaccination campaigns [9,47]. To achieve this understanding, a strategic approach and the use of information and communication technology are essential [9].

Though not fully representative of Western and Eastern countries, the United States and Korea are frequently compared in health communication studies due to their distinct cultural traits, highlighting regional differences [48]. Throughout the COVID-19 pandemic, these countries exhibited disparities in vaccine-related aspects such as case proportions, quarantine regulations, and vaccine availability [49,50]. Nevertheless, there are few studies examining vaccine information needs and perceptions across multiple countries [51,52].

Consequently, we investigated the information needs, perceptions, and attitudes toward the COVID-19 vaccine expressed on social media by people in the United States and Korea. Our findings emphasize the complexity of public perceptions of vaccines and the critical need for targeted public health communication strategies. The unique information needs and varying sentiments across topics and Q&A pairs over time in both the United States and Korea illustrate the nuanced challenges faced by public health initiatives. Addressing those challenges through targeted, adaptive, and culturally aware communication efforts is essential for overcoming vaccine hesitancy, enhancing public trust, and ultimately achieving herd immunity against COVID-19.

Methods

Overview

This study conducted a cross-sectional analysis of vaccine-related questions and associated information needs on Quora and Naver KiN between June 2020 and June 2021. We followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology; [Multimedia Appendix 1](#)) reporting guidelines. This study design is aligned with recent cross-sectional analyses of social media use for health-related issues [53].

To examine the perceptions and information needs of Q&A site users regarding the COVID-19 vaccine, we collected data from Q&A sites with representative US and Korean user bases. For this purpose, we chose Quora, a US platform hosting millions of questions and answers [54], and Naver KiN, the foremost social Q&A site in Korea [55]. Data were collected using Selenium (Software Freedom Conservancy, Inc), an open-source framework that allows automatic interaction with web browsers using Python 3.8.5 (Python Software Foundation) [56]. We set intervals during data collection to avoid overusing the service. For each question, we collected the URL, upload date, and answers. The data collection ran from June 27, 2020, to June 27, 2021, based on the upload dates for both websites, and the inclusion criteria were questions and answers pertinent to COVID-19 and any morphological form of the keyword “vaccine” (in Korean, 백신), such as “vaccines” and words containing “vaccine.” By applying those criteria and excluding unanswered questions, we identified and subsequently removed questions falling outside our specified criteria.

As a result, after sampling the data, we collected 3952 sets of questions and answers from Quora and Naver KiN, as depicted in [Figure 1](#). Rather than analyzing all the answers to the questions, we opted to examine only the most upvoted answers so that we could analyze the characteristics of the answers that users deemed most favorable [48]. Accordingly, only one answer for each question was used from an average of 27.4 and 2.7 answers detected on Quora and Naver KiN, respectively. All data and source code files can be found in the repository.

In this study, we used 2 major approaches to analyze the questions and answers ([Figure 2](#)), including topic categorization for questions based on Latent Dirichlet Allocation (LDA) [57] and Top2Vec [58], and sentiment analysis for answers using ensemble-bagging with Bidirectional Encoder Representations from Transformers (BERT) [59] and Efficiently Learning an Encoder that Classifies Token Replacement Accurately (ELECTRA) [60]. For answers, this study used the most upvoted answers to focus on the characteristics of answers that users deemed most favorable. These natural language processing (NLP) techniques have been shown to be effective for analyzing large-scale data [61].

Figure 1. Data preprocessing. Data from Quora and Naver Knowledge-iN (Naver KiN) were collected between June 27, 2020, and June 27, 2021, using the keyword “Vaccine” (in Korean, 백신). After preprocessing, the Naver KiN data were randomly sampled to match the Quora sample size per platform (N= 3952).

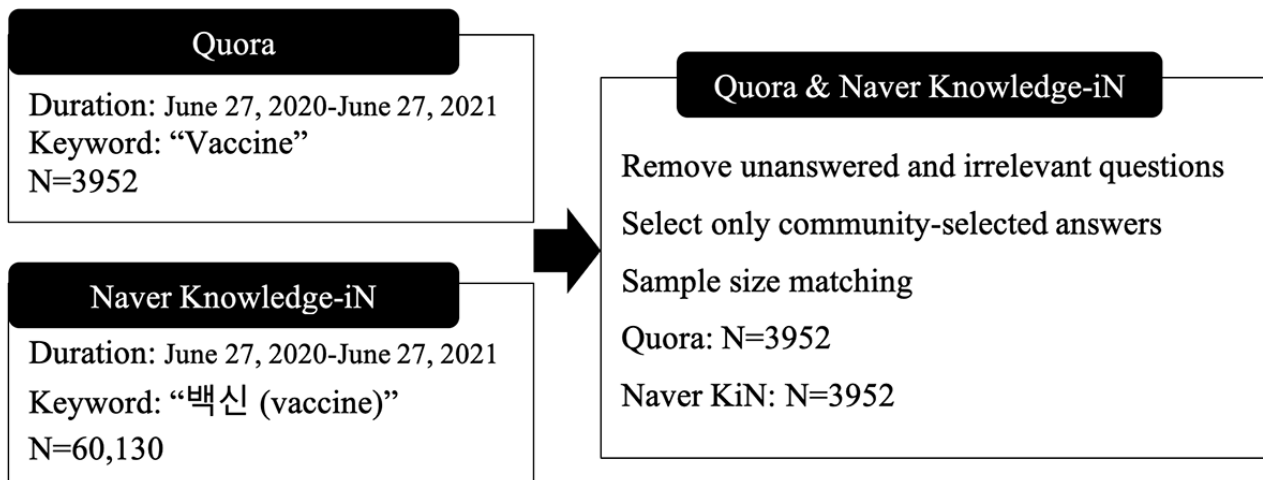
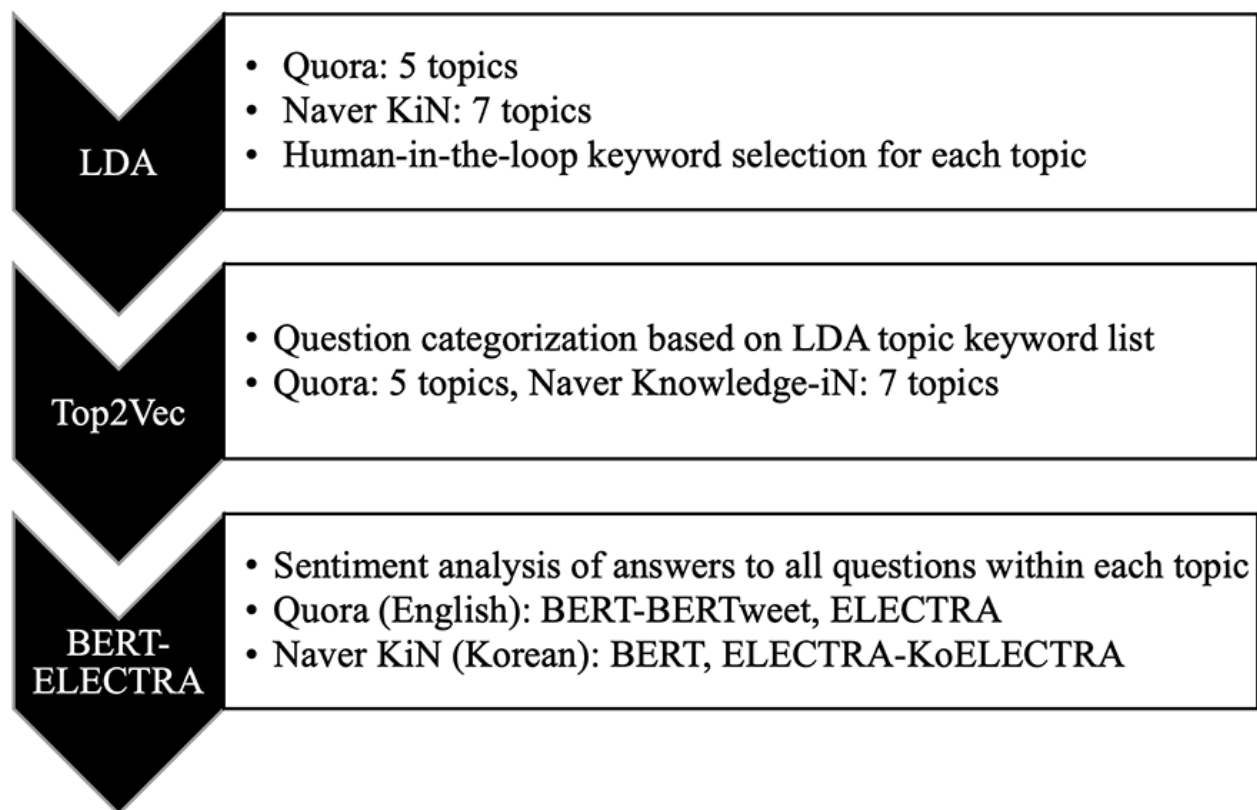


Figure 2. Overview of data analysis framework. Questions from Quora and Naver Knowledge-iN (Naver KiN) are grouped into topics using LDA and Top2Vec, with human-in-the-loop keyword selection. Sentiment analysis of answers within each topic is then performed using BERT- and ELECTRA-based models. BERT: Bidirectional Encoder Representations from Transformers; ELECTRA: Efficiently Learning an Encoder that Classifies Token Replacements Accurately; LDA: Latent Dirichlet Allocation.



Categorization of Questions: LDA and Top2Vec

In this study, we combined machine learning predictions with human-in-the-loop validation to categorize questions from Quora and Naver KiN. To this, we applied topic modeling to identify key topics across the documents, offering guidance for strengthening public health communication strategies [62]. Specifically, LDA was used to categorize the questions on Quora into 5 topics and those on Naver KiN into 7 topics. We initially

extracted 50 topics, which is the optimal number of topics determined by the coherence score, and manually merged them into 5 topics for Quora and 7 topics for Naver KiN. Researchers (SY, SHJ, JC, and YJY) discussed and decided on each topic together. While merging topics, we discovered that Naver-KiN covered more topics than Quora.

LDA uses documents, words, and latent variables that represent the latent topics associated with word sequences. Although LDA

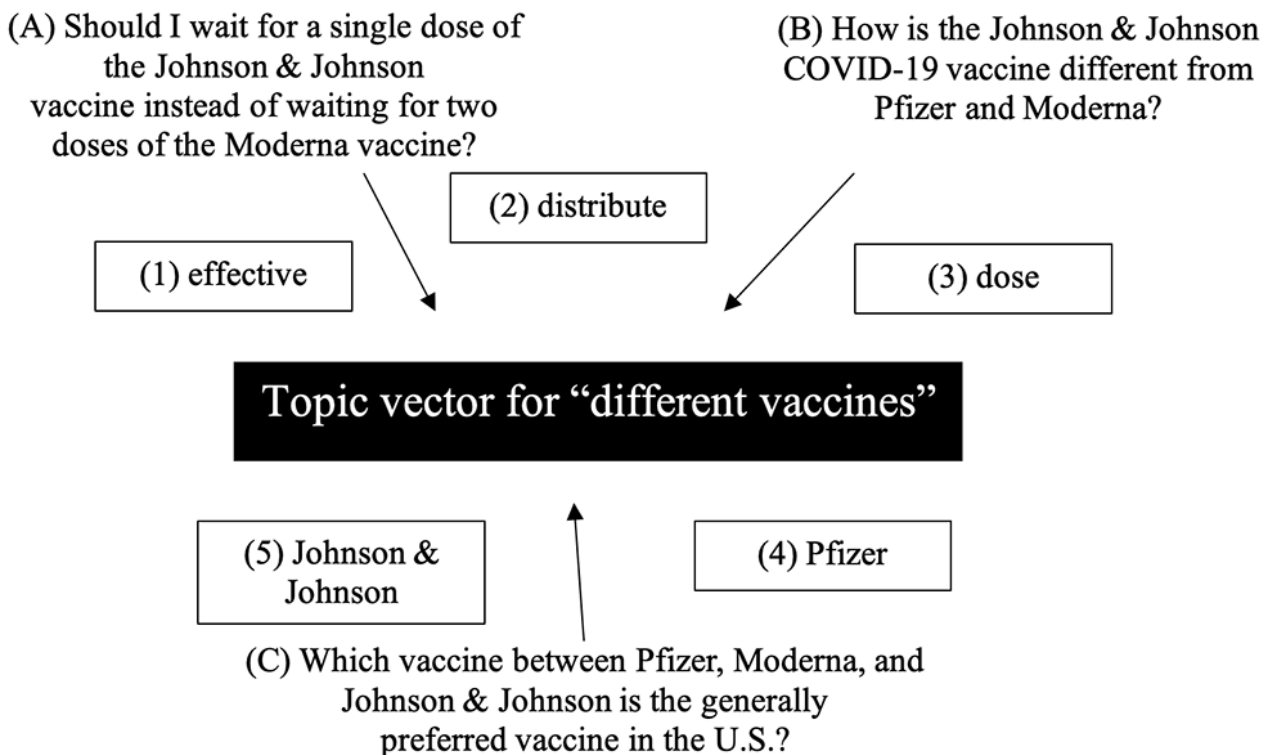
is effective in restricting the number of topics and offering keywords associated with each topic, it does not quantify the similarity between each keyword and its topic (it mainly uses document-word frequency when categorizing topics). That makes it challenging to determine which words are most relevant to the topic when relying solely on human decisions for keyword selection. To address this limitation, we incorporate Top2Vec, which preserves word semantics by representing words and documents in a shared embedding space and derives topic representations directly from these semantic structures.

Building on this motivation, we adopt a hybrid LDA-Top2Vec pipeline to combine probabilistic topic structure with semantic embedding-based topic modeling. While Top2Vec is often applied directly to raw documents, we instead use LDA to first identify topic-relevant keywords, which serve as a coarse topical filter. This step constrains the semantic space Top2Vec explores, enabling it to focus on semantically meaningful regions associated with each topic. As a result, this hybrid approach yields more interpretable and stable topic representations than

applying Top2Vec alone to the full document corpus. While formal coherence scores were not used as a selection criterion, this hybrid approach was chosen to prioritize interpretability and topic stability.

Operationally, the LDA-extracted keywords were represented as word vectors, and the questions were represented as document vectors in the vector space. Figure 2 illustrates the process of question analysis using the mixed LDA and Top2Vec method. Figure 3 further visualizes a semantic space that consists of word vectors (enclosed numbers 1, 2, 3, 4, and 5) and document vectors (labeled A, B, and C). The centroid of the document vectors is a topic vector (black central square labeled topic vector for different vaccines). The word vectors in Figure 3 are the words with the closest semantic relationship to the topic vector, whereas the LDA keywords indicate the word distribution per topic solely by frequency. In other words, LDA delineates the statistical relationship among topics, but it fails to capture semantic information about those topics [63].

Figure 3. Illustration of topic vector construction for “different vaccines.” The topic vector is defined as the centroid of the document (question) vectors (A, B, and C) and the associated keywords (1, 2, 3, 4, and 5). Keywords that are semantically closest to this centroid are used to characterize the topic.



Sentiment Analysis of the Answers: Transformer-Based Models

Because COVID-19 greatly affected the sentiments expressed on social media [64], we used a document-level sentiment analysis to extract and classify perceptions and attitudes toward the topics discussed in the documents [32,65]. This analysis served as an indicator of public perception regarding COVID-19 concerns [66]. We used 2 transformer-based models, BERT and ELECTRA, that are capable of fine-tuning with pretrained models. The pretraining and fine-tuning mechanism enhances

each model’s ability to effectively grasp the intricacies of downstream NLP tasks, such as sentiment analysis.

BERT is a widely used language model that uses pretrained data and a self-attention mechanism to generate contextualized vector outputs. It is applied to a range of downstream tasks, including sentiment analyses. BERT uses the masked language model; some tokens in the sentence are replaced with (MASK), and then the model is trained to regenerate the original tokens [59]. On the other hand, ELECTRA, which is another pretrained transformer model, uses a discriminator to predict the replaced tokens by generating token output instead of relying on a masked language model [60]. Both models have been widely used for

various NLP downstream tasks [67], and we used models fine-tuned for an English sentiment analysis and a Korean sentiment analysis.

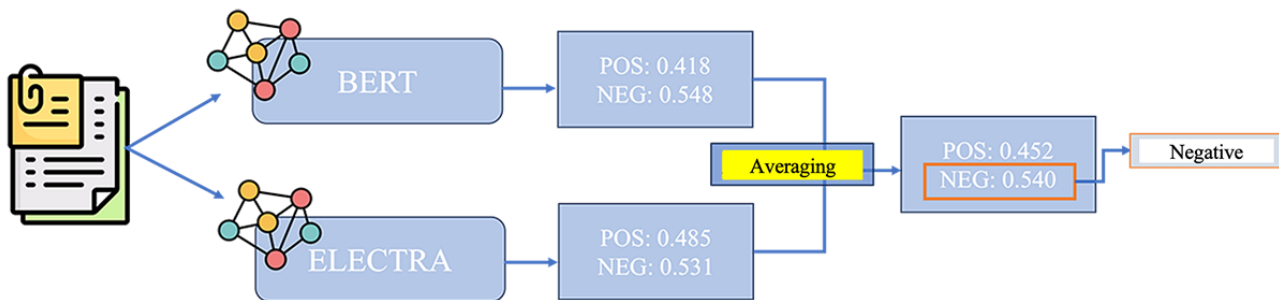
For the English language model, both BERT-based [68] and ELECTRA-based [60] were specifically fine-tuned for a sentiment analysis using Twitter datasets. While Twitter and Quora have posts of different lengths (a Twitter post consists of 8 words on average and a Quora post consists of 40 words on average) [69], both Twitter and Quora texts are social media data originating from public interactions on a social media platform. They share similar traits, and the model fine-tuned on Twitter was effectively applied to the Quora dataset [70], and we conducted a sentiment analysis of Quora data with 66% (33/50) accuracy. For the Korean language model, both BERT-based [71] and ELECTRA-based [72] were fine-tuned using Naver KiN shopping reviews, Steam game reviews, and movie reviews. We evaluated the accuracy of these models for Naver KiN data and achieved 72% (36/50) accuracy in the sentiment analysis. This choice thus mirrored the strategy we used for the English model.

One limitation of using transformer models is the constraint of the input tensor size, which is restricted to a length of 128

tokens. This limitation can sometimes introduce biases because only the first 128 tokens are considered, potentially leading to inaccurate sentiment analysis results. For instance, the beginning of a post might convey a positive sentiment by starting with a greeting, for example, that is not representative of the main content of the post. To address that issue, we divided each post into chunks of 128 tokens, measured the sentiment of each chunk, and then averaged the sentiment values of all the chunks to determine the overall sentiment result.

To further enhance the validity of the sentiment analysis, we used an ensemble method called the bagging or bootstrap aggregating technique, which is known for its effectiveness in classification tasks. This method generates an ensemble classification output that outperforms the accuracy of a single prediction value [73,74]. We applied it to predict improved sentiment values for each answer using both BERT and ELECTRA. An overview of the process is provided in Figure 4. The outcomes from both models were the probabilities of positive and negative sentiments. Subsequently, we computed the average probabilities of positive and negative sentiments. Once we obtained the averaged positive and negative probabilities, we compared them and used the higher one.

Figure 4. Ensemble averaging for sentiment analysis. Positive and negative sentiment scores are independently predicted using BERT and ELECTRA. The corresponding scores are then averaged across models, and the sentiment label with the higher averaged score is selected as the final prediction. BERT: Bidirectional Encoder Representations from Transformers; ELECTRA: Efficiently Learning an Encoder that Classifies Token Replacements Accurately; Neg: negative; POS: positive.



Ethical Considerations

Ethical review was not required for this study because it exclusively used publicly available social media content shared in accordance with platform policies. As the data were publicly accessible, Institutional Review Board approval and a waiver of informed consent were not required [75,76]. This study did not use or infer individuals' identities but instead examined vaccine-related opinions at an aggregate level to provide an overall picture of public perceptions. To protect user privacy, all identifiable information was removed during preprocessing, and all questions and comments presented in this manuscript were paraphrased. The final aggregated dataset contains no personally identifiable information and cannot be traced back to individual users.

Results

Topic Analysis of Questions

Table 1 presents the topic modeling results for Quora, where 5 primary topics were identified. The most frequently discussed topic was different vaccines (978/3952, 25%), with users asking

about characteristics and comparisons of various vaccine brands such as AstraZeneca (AstraZeneca plc) and Moderna (Moderna, Inc). This was followed by government policy (821/3952, 21%), where questions centered on administrative decisions and vaccine distribution, with keywords including "require," "Trump," and "supply." Variants (782/3952, 20%) included discussions focused on the severity of new COVID-19 strains and vaccine efficacy. The topic of visiting overseas (743/3952, 19%) comprised questions about travel eligibility based on vaccination status. The least frequent topic, effects of vaccine (626/3952, 16%), involved concerns about immunity, side effects, and infection risks.

Table 2 shows the topic modeling outcomes for Naver KiN, which yielded 7 topics. The most common was effects of vaccine (790/3952, 20%), marked by terms like "side effects" and "uptake," where users expressed hopes for effective vaccination and sought information on its impact. Vaccination appointments followed (711/3952, 18%), with users asking about eligibility and scheduling logistics. The variants topic (592/3952, 15%) reflected interest in vaccine efficacy against emerging strains. Government policy (553/3952, 14%) included questions about

national vaccination strategies. School and education (513/3952, 13%) highlighted questions about children's school attendance in relation to family vaccination status. Visiting overseas (474/3952, 12%) addressed international travel feasibility amid

varying vaccination policies. The least common topic was different vaccines (316/3952, 8%), where users asked about vaccine suitability for individual health conditions.

Table 1. Example question for each topic in Quora. For each topic identified in the Quora dataset, the table reports the proportion of questions assigned to the topic and provides an example question illustrating the topic content (N=3952).

Label	Topic proportion, n (%)	Example question
Different vaccines	978 (25)	"How come the other vaccines, including the Pfizer and Moderna vaccines, require 2 doses; however, the Johnson & Johnson COVID-19 vaccine requires only one dose?"
Government policy	821 (21)	"On his tonight show, Sean Hannity credits Donald Trump with developing the COVID-19 vaccine rather than President Biden. Who in your opinion deserves the credit?"
Variant	782 (20)	"Does the delta variant of COVID-19 pose the highest threat to individuals who have not yet received vaccinations because it is more contagious, causes a more serious illness, and is not affected by the current vaccines?"
Visiting overseas	743 (19)	"Would you get vaccinated only to travel if many of the largest airlines in the world require verification of vaccination status in your passport and prevent unvaccinated people from traveling?"
Effects of vaccine	626 (16)	"Is it acceptable to state that you do not want the vaccine because everyone you know who has received the COVID-19 vaccine has become ill shortly after vaccination?"

Table 2. Example questions for each topic in Naver Knowledge-iN (Naver KiN). For each topic identified in the Naver KiN dataset, the table reports the proportion of questions assigned to the topic and provides an example question illustrating the topic content (N=3952).

Label	Topic proportion, n (%)	Example question
Effects of vaccine	790 (20)	"I heard that individuals under 18 years of age cannot be vaccinated but why? I also want to get vaccinated. COVID-19 will come to an end if everyone gets vaccinated quickly and becomes immune to the virus. I want to take off my mask and go somewhere."
Vaccine appointment	711 (18)	"Can civil defense corps members also get vaccinated? When are they eligible to get vaccinated? Tell me the vaccination schedule for both the 1st and 2nd dose."
Variant	592 (15)	"They say that COVID-19 vaccines are effective but is it too early to come to conclusions? For vaccines to be effective, antibodies should last for a long period. However, for COVID-19 vaccines, antibodies do not last even 1 year, indicating they are effective only for a short period of time. (...) Thus, is it too early to determine the effectiveness of the vaccines?"
Government policy	553 (14)	"I have recently taken interest in our government. I believe our government has responded well to the COVID-19 outbreak compared with other countries, and the delay in vaccine supply appears reasonable because vaccines are prioritized to countries that have not managed the virus well. (...) However, why are there so many negative comments regarding the government on Naver KiN? I am just wondering. Maybe I am wrong but I am leaving a comment because I want to know why people are so angry."
School and education	513 (13)	"My parents were vaccinated. Can I go to school?"
Visiting overseas	474 (12)	"How serious is the COVID-19 situation in Japan? Has Japan started vaccinations? Please tell me specifically about this situation. I am planning on traveling to Japan and was wondering whether it would be possible to go to Japan after COVID-19 comes to an end. I am planning my trip for next year. Will COVID-19 end in both countries by then?"
Different vaccines	316 (8)	"I am in my 20s and I have hypothyroidism. Are people with hypothyroidism considered chronically ill patients? When can people with chronic diseases get vaccinated? Which vaccine will I receive?"

Sentiment Analysis of Answers

A sentiment analysis was used to examine the answers given on both Quora and Naver KiN. Figure 5 illustrates the proportions of sentiments in Quora answers to questions from each topic. Across 4 of the topics—effects of vaccine, visiting overseas, variants, and different vaccines—the number of answers with negative sentiment exceeded the number of positive ones. However, within the topic of government policy,

answers with positive sentiment were more prevalent than negative ones.

In Figure 5, positive versus negative sentiments were observed as follows: effects of vaccines (306/626, 48.79% vs 320/626, 51.12%), visiting overseas (335/743, 45.05% vs 408/743, 54.95%), variants (362/782, 46.29% vs 420/782, 53.71%), different vaccines (386/978, 39.47% vs 592/978, 60.53%), and government policy (437/821, 53.23% vs 384/821, 46.77%),

where the first value represents positive sentiment and the second represents negative sentiment.

We analyzed sentiment outcomes on a monthly basis from June 2020 to June 2021, focusing on topic popularity over time (see Figure 6). While finer-grained temporal analyses are possible, we adopted monthly aggregation to reduce sparsity and short-term volatility and to emphasize broader discourse trends

rather than reactions to specific events. Across all topics except different vaccines, positive-negative sentiment crossovers occurred at least 3 times. After May 2021, positive sentiment exceeded 50%, indicating its dominance across topics. Notably, only government policy consistently showed a surplus of positive sentiment over negative for 4 consecutive months after February 2021.

Figure 5. Sentiment proportions per topic in Quora.

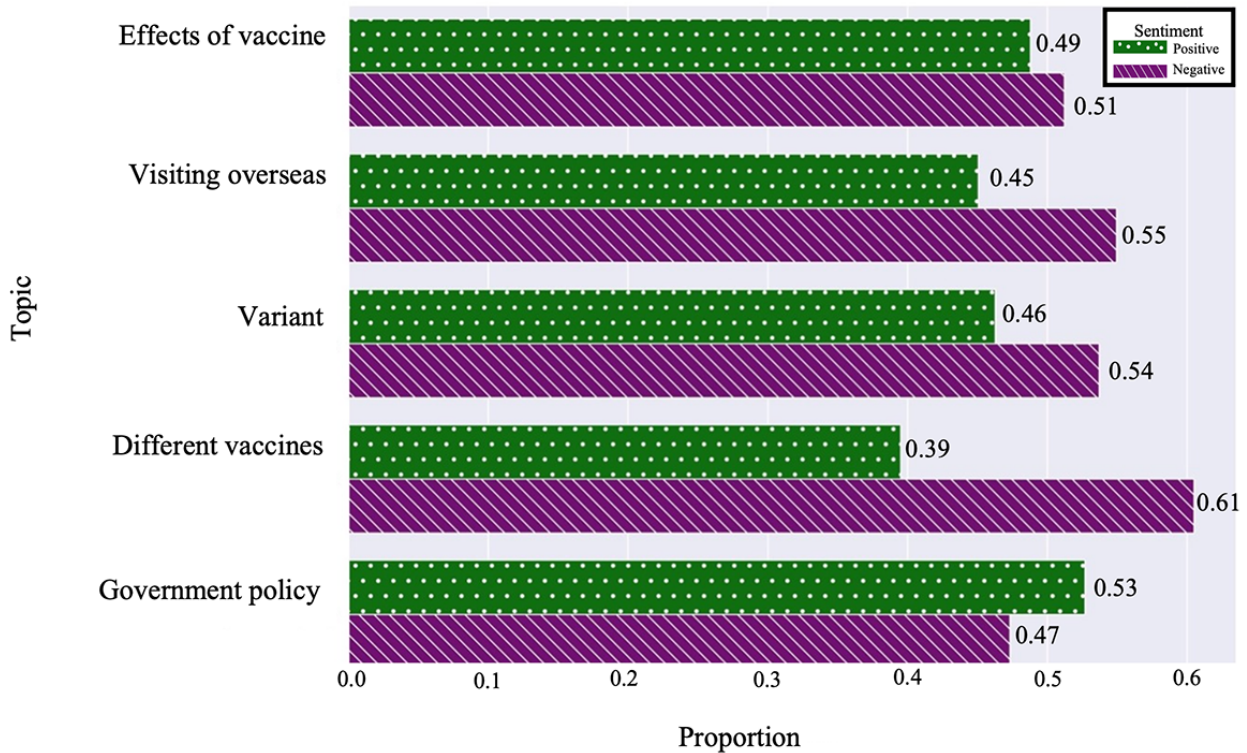


Figure 6. Temporal analysis of sentiments per topic in Quora. Due to limited data availability, 4 days at the end of June 2020 were merged into July 2020 for the analysis.



During this period, the vaccination rate increased by 15.3%, from 17.45% to 32.78% [52], coinciding with heightened public attention to vaccine rollout and access. Positive sentiment toward government policy appears to reflect evaluations of these concrete policy outcomes rather than a broader shift in

institutional trust. In particular, the temporary increase in positive sentiment after February 2021 aligns with major administrative transitions and policy announcements, including changes in federal coordination, alongside continued public

recognition of vaccine development efforts initiated under Trump, as reflected in the example in Table 1.

Prior research similarly suggests that during public health crises, shifts in public trust toward government are closely tied to public perception of policy responses and outcomes. Notably, policy-specific achievements (eg, visible progress in crisis management) are often associated with temporary increases in positive evaluations even amid persistent skepticism toward government institutions [77]. Thus, the observed temporal positivity in “government policy” sentiment likely captures situational approval of pandemic response measures rather than a reversal of widespread distrust in US governance.

The results of the sentiment analysis of answers to questions on Naver KiN are depicted in Figure 7. Across all 7 topics—different vaccines, government policy, schools and education, vaccine appointments, variants, effects of vaccines, and visiting overseas—positive sentiments were predominant over negative sentiments. This overarching trend is also highlighted in Figure 8, which shows that positive sentiment consistently exceeded 50% throughout the year except for the topic of different vaccines.

In Figure 7, the proportions of positive versus negative sentiment were as follows: effects of vaccines (498/790, 63.04% vs 292/790, 36.96%), visiting overseas (337/474, 71.10% vs 137/474, 28.90%), variants (367/592, 61.99% vs 225/592, 38.01%), different vaccines (218/316, 68.99% vs 98/316, 31.01%), government policy (348/553, 62.93% vs 205/553, 37.07%), vaccine appointment (562/711, 79.04% vs 149/711, 20.96%), and school and education (364/513, 70.96% vs 149/513, 29.04%).

For the topic of different vaccines, 4 positive-negative sentiment crossovers occurred between June 2020 and January 2021, often driven by concerns about side effects and individual health conditions. During this period, confirmed COVID-19 deaths per million rose sharply by 313% [52]. As shown in Table 2, a question regarding vaccine eligibility for individuals with chronic illnesses received a relevant response. After February 2021, coinciding with the launch of South Korea’s vaccination program on February 26, positive sentiment prevailed, with no further crossovers observed [52].

Figure 7. Sentiment proportions per topic in Naver Knowledge-iN (Naver KiN).

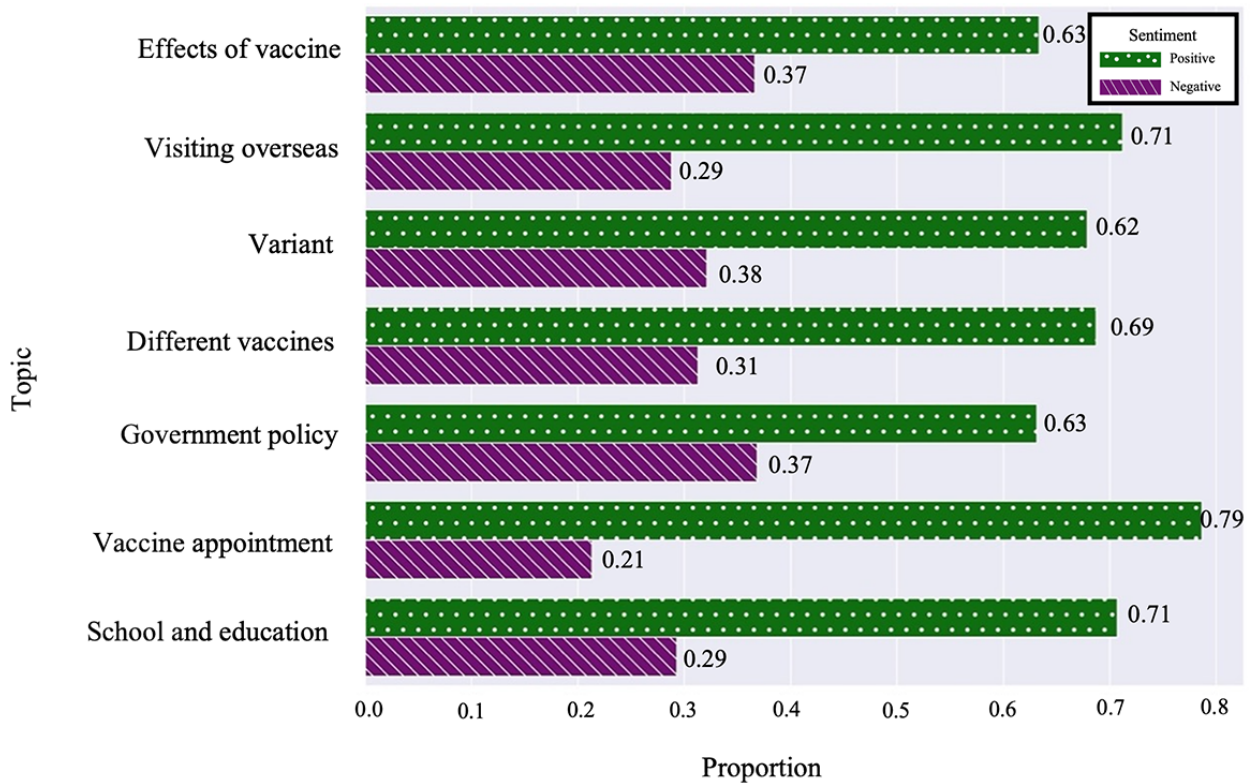


Figure 8. Temporal analysis of sentiment per topic in Naver KiN. Due to limited data availability, 4 days at the end of June 2020 were merged into July 2020 for the analysis.



Sentiment Analysis of Q&A Pairs

After analyzing the sentiments of the answers, we examined question–answer sentiment pairs within each topic to see if the question’s sentiment might affect the answer. Table 3 shows the sentiments of question-answer pairs on Quora. In 4 topics—effects of vaccine, visiting overseas, variants, and different vaccines—the predominant sentiment pairs are negative-negative, followed by negative-positive, positive-positive, and positive-negative. For the topic of government policy, negative-positive sentiment pairs are as

prevalent as negative-negative, with positive-positive and positive-negative following.

Table 4 illustrates the sentiment for question-answer pairs on Naver KiN. Across all topics, positive-positive and negative-positive pairs constituted more than 63% of the sentiment. For 6 topics—effects of vaccines, visiting overseas, different vaccines, government policy, vaccine appointments, and school and education—positive-positive pairs were the most prevalent, followed by negative-positive pairs. In contrast, for the topic of variants, negative-positive pairs were the most prevalent, followed by positive-positive pairs.

Table 3. Question-answer sentiment pairs for each topic on Quora.

Topic	Positive-positive, n (%)	Positive-negative, n (%)	Negative-positive, n (%)	Negative-negative, n (%)
Effects of vaccines (N=626)	101 (16.07)	49 (7.85)	205 (32.71)	271 (43.36)
Visiting overseas (N=743)	84 (11.3)	68 (9.11)	251 (33.75)	340 (45.84)
Variants (N=782)	123 (15.78)	91 (11.61)	239 (30.51)	329 (42.11)
Different vaccines (N=978)	114 (11.66)	88 (8.97)	272 (27.86)	504 (51.52)
Government policy (N=821)	123 (14.87)	73 (8.78)	314 (37.82)	311 (37.53)

Table 4. Question-answer sentiment pairs in each topic on Naver Knowledge-iN (Naver KiN).

Topic	Positive-positive, n (%)	Positive-negative, n (%)	Negative-positive, n (%)	Negative-negative, n (%)
Effects of vaccines (N=790)	279 (35.32)	187 (23.67)	219 (27.72)	105 (13.29)
Visiting overseas (N=474)	239 (50.42)	101 (21.31)	98 (20.68)	36 (7.59)
Variants (N=592)	178 (30.07)	102 (17.23)	189 (31.92)	123 (20.78)
Different vaccines (N=316)	125 (39.56)	49 (15.51)	93 (29.43)	49 (15.5)
Government policy (N=553)	177 (32.01)	90 (16.27)	171 (30.92)	115 (20.8)
Vaccine appointment (N=711)	374 (52.6)	110 (15.47)	188 (26.44)	39 (5.49)
School and education (N=513)	207 (40.35)	85 (16.57)	157 (30.61)	64 (12.47)

Discussion

Principal Findings

Social media provides valuable insights for policymakers developing health policies [27]. Our study pioneers exploring public discourse on vaccines in 2 countries using advanced NLP

techniques, surpassing traditional survey-based methods. This approach reveals hidden aspects of vaccine perceptions beyond conventional demographics. We analyzed overall perceptions (via topic analysis) and communication within each topic (via sentiment analysis). In this discussion, we consider platform characteristics and acknowledge the study’s limitations.

Comparative Discourse on the COVID-19 Vaccine

The United States and Korea both have concerns about 5 topics, including effects of the vaccine, visiting overseas, variants, different vaccines, and government policy; however, Korea uniquely highlights vaccine appointments and school and education. Vaccine appointments became a significant issue in Korea due to delays in vaccine distribution [49], with the government implementing an age-based distribution policy to respect the cultural emphasis on older adults and their high risk of serious disease. Although prioritizing the older adults might have effectively reduced COVID-19-related mortality and morbidity within this age group, other vulnerable populations, such as prisoners and the homeless, experienced relative disadvantages. Therefore, the equitable distribution of vaccinations is a topic requiring open discussion.

Naver KiN also revealed a unique topic about school and education that did not appear on Quora. This topic highlights how deeply education is valued in Korean culture, which made vaccinating school-aged children and their parents crucial [78]. This contrasts with the United States, where, despite the CDC's recommendations [79], vaccine hesitancy remained a notable barrier among many parents [80]. Thus, the Naver KiN results underscore that cultural values and societal priorities, such as Korea's strong commitment to educational success, played a significant role in shaping public attitudes toward COVID-19 vaccination.

Based on those topics, we investigated the sentiments expressed in question-answer pairs and gleaned insights into the public perceptions of COVID-19 vaccines expressed on social Q&A sites in the United States and Korea. Our findings highlight distinct sentiment patterns reflective of the cultural and informational contexts of these countries. Quora and Naver KiN both have a similar tendency of consistency between a question's sentiment and the answers it receives (eg, positive [Q]-positive [A] or negative [Q]-negative [A]). However, users of the 2 platforms expressed different sentiments toward the COVID-19 vaccine.

The sentiment analysis of Quora data reveals that in the United States, negative sentiment was predominant in discussions about the effects of the vaccine, visiting overseas, variants, and different vaccines. In those topics, negative-negative (question-answer) pairs are the most common, indicating widespread skepticism and concern among users [15,25]. Unlike Quora, the majority of answers were positive in the Naver KiN data from Korea, with positive-positive (question-answer) and negative-positive pairs constituting more than 63% of the total. This difference may be attributed to varying levels of trust in the government between the 2 countries.

At the time we collected data from Naver and Quora, South Korea was effectively managing the pandemic, and public trust in the government was high, consistent with previous research showing high compliance with COVID-19 government policies [81]. As a result, both questions and answers were generally positive. The only topic that saw a higher occurrence of negative questions paired with answers with positive sentiment was related to variants, likely because the Korean government could not control the emergence of variants, leading to more negative

sentiment. Even in discussions about variants, where there was more negative sentiment toward the question, the sentiment toward the answers remained largely positive, indicating resilience and hope [82,83]. In contrast, in the United States, the situation was markedly different. Public trust in the US government was lower, and adherence to COVID-19 policies was less consistent [81]. The skepticism and concerns expressed in the US discussions may have been driven by widespread dissatisfaction with how the pandemic was being handled, leading to a pervasive negative sentiment across various topics [81-83].

These contrasting attitudes suggest that the differences observed in the public discourse on the COVID-19 vaccine in the United States and South Korea may stem from cultural and situational factors rather than true variations in public perceptions. Given that the unique social context of each country shapes public discourse and sentiment, it can be inferred that communication strategies targeted to cultural contexts are required to address vaccine hesitancy. In the United States, public health campaigns should counter misinformation [32] and address specific safety concerns, engaging trusted community leaders and positive narratives about vaccination benefits. In Korea, on the other hand, maintaining positive public discourse through transparency and timely updates is crucial. The unique focus on vaccine appointments and education in Korean discussions highlights specific information needs that should be addressed.

Effects of the 2 Platforms on Sentiment

Platform-specific user dynamics also play a role in shaping public discourse [84,85], affecting the sentiments of each platform's answers. The presence of health care professionals on Naver KiN, who provide expert opinions, likely contributes to the positive sentiment and higher likelihood of their answers being perceived as trustworthy, in contrast to the diverse user base on Quora. Overall, understanding and leveraging these distinct sentiment patterns can help public health authorities craft more effective communication strategies to promote a positive and informed discourse about COVID-19 vaccination.

To select the responses for our analysis, we used a methodology intended to mirror the typical user experience on these platforms. Responses were chosen based on specific criteria designed to reflect a balance between popularity and relevance and, above all, to reflect which answers the users deemed most effective. This approach aligns with methodologies used in previous user-generated content analyses on platforms such as Quora and Naver KiN [48].

Limitations of This Study

This study has several limitations. First, there was a notable variance in the average length of answers between Quora and Naver KiN. This discrepancy might have introduced potential bias when comparing the answers. Future studies could mitigate this by standardizing the length of answers or selecting questions with a similar number of words for a more meaningful comparison. In addition, platform-specific user demographics may partially confound cross-cultural comparisons. Naver KiN serves as a dominant information-seeking platform in South Korea, characterized by large-scale daily participation and broad

distribution of user expertise [48,86]. In contrast, Quora users in the United States tend to be younger and more educated [87]. These structural differences may shape how questions are shaped, how answers are evaluated, and how sentiment is expressed. As a result, the observed sentiment asymmetry between the 2 platforms may reflect platform participation norms and demographic composition, as well as national-level attitudes.

Second, although we used models already trained on our target social media outlets, Quora and Naver KiN, we applied both topic modeling and sentiment analysis models without fine-tuning them with our data. This may result in less optimal and lower-performing outputs from the experiments. To mitigate this limitation, fine-tuning each model should be performed to yield more reliable results in future studies.

Third, our analysis focuses on the most favorable answers associated with each question. This design choice may limit the generalizability of findings regarding public perception. By prioritizing community-validated information, often in the form of expert-driven responses, this analysis may also introduce a positive bias toward answers. Thus, the observed association between question sentiment and answer sentiment should be interpreted as reflecting how question framing relates to the community-validated information, rather than the full range of public perception, which may include greater uncertainty, disagreement, or negative affect.

Finally, we acknowledge as a limitation that no explicit bot-detection or automated content-authenticity filtering algorithm was applied. Given the highly politicized and time-sensitive nature of vaccine-related discourse during 2020-2021, bot-generated content may have contributed to the observed sentiment patterns. Although we partially mitigated this risk by excluding unanswered questions and constraining questions that include the keyword “vaccine” (in Korean, 백신), these measures did not guarantee the removal of automated or coordinated accounts. As such, some sentiment estimates may reflect amplified or artificial signals rather than purely user opinions.

Conclusions

This study has provided critical insights into the dynamics of vaccine discourse, underscoring the need for context-specific strategies for promoting vaccination. By melding state-of-the-art

NLP technologies with a cross-cultural analysis, we have forged a path toward more informed, efficacious, and targeted efforts in public health communication, both in the ongoing battle against the COVID-19 pandemic and beyond. Above all, this study underscores the paramount importance of harnessing social media as a prolific source of public sentiment and information needs, particularly within the sphere of vaccine hesitancy. Analyzing public discourse on platforms such as Quora and Naver KiN provides researchers with an unfiltered view of individuals' concerns and perceptions, offering insights invaluable for crafting responsive public health strategies.

Our findings reaffirm the profound effects of cultural and societal factors on public attitudes toward vaccines. The stark contrast observed between the United States and Korea in terms of vaccine perceptions underscores the need for culturally sensitive approaches in health communication and vaccination campaigns. This study advances the academic landscape by demonstrating the efficacy of NLP technologies for comprehending public perceptions and information needs related to vaccines. NLP methods transcend traditional survey methodologies, providing a more naturalistic perspective rooted in real-world discussions on community-driven social media platforms. Our innovative amalgamation of LDA, Top2Vec, and BERT- and ELECTRA-based sentiment analyses sets a pioneering benchmark for future research in this domain. This multifaceted approach augments our comprehension of the intricacies encircling vaccine discourse and can be adapted to explore analogous public health issues across diverse countries and cultures.

As for practical insights, our findings have implications for policymakers and health communication practitioners in both countries. The discernment of disparities in vaccine-related information needs and perceptions between the United States and Korea underscores the necessity of tailoring public health campaigns to the specific nuances of each nation's context. Effectively addressing vaccine hesitancy—a complex interplay of safety concerns, access barriers, and misinformation—demands a nuanced, multifaceted approach. Our study underscores the need for policy interventions that take these underlying factors into account, while also recognizing the profound influence of cultural and societal values on vaccination decisions.

Acknowledgments

During manuscript preparation, we used the generative artificial intelligence tool ChatGPT [88] exclusively to improve language clarity and readability. All content was reviewed and edited by the authors. The authors take full responsibility for the final manuscript.

Funding

This research was supported by the “Regional Innovation System & Education (RISE)” through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government (2025-RISE-01-018-05).

Data Availability

The data that support the findings of this study are openly available in GitHub repository.

Authors' Contributions

SY, SHJ, JC, and YJY contributed to the conceptualization of the study. SY and JC contributed to data curation, methodology, and visualization. SY, SHJ, HK, JC, and YJY contributed to formal analysis and manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

STROBE checklist.

[\[PDF File \(Adobe PDF File\), 116 KB-Multimedia Appendix 1\]](#)

References

1. COVID-19 how to protect yourself and others. Centers for Disease Control and Prevention (CDC). 2023. URL: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html> [accessed 2023-04-02]
2. Advice for the public: coronavirus disease (COVID-19). World Health Organization. 2023. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public> [accessed 2023-04-02]
3. FDA. FDA approves first COVID-19 vaccine. U.S. Food & Drug. 2023. URL: <https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine> [accessed 2023-04-02]
4. Troiano G, Nardi A. Vaccine hesitancy in the era of COVID-19. *Public Health*. 2021;194:245-251. [FREE Full text] [doi: [10.1016/j.puhe.2021.02.025](https://doi.org/10.1016/j.puhe.2021.02.025)] [Medline: [33965796](https://pubmed.ncbi.nlm.nih.gov/33965796/)]
5. Menni C, May A, Polidori L, Louca P, Wolf J, Capdevila J, et al. COVID-19 vaccine waning and effectiveness and side-effects of boosters: a prospective community study from the ZOE COVID Study. *Lancet Infect Dis*. 2022;22(7):1002-1010. [FREE Full text] [doi: [10.1016/S1473-3099\(22\)00146-3](https://doi.org/10.1016/S1473-3099(22)00146-3)] [Medline: [35405090](https://pubmed.ncbi.nlm.nih.gov/35405090/)]
6. Hayawi K, Shahriar S, Serhani MA, Taleb I, Mathew SS. ANTi-Vax: a novel twitter dataset for COVID-19 vaccine misinformation detection. *Public Health*. 2022;203:23-30. [FREE Full text] [doi: [10.1016/j.puhe.2021.11.022](https://doi.org/10.1016/j.puhe.2021.11.022)] [Medline: [35016072](https://pubmed.ncbi.nlm.nih.gov/35016072/)]
7. Kricorian K, Civen R, Equils O. COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety. *Hum Vaccin Immunother*. 2022;18(1):1950504. [FREE Full text] [doi: [10.1080/21645515.2021.1950504](https://doi.org/10.1080/21645515.2021.1950504)] [Medline: [34325612](https://pubmed.ncbi.nlm.nih.gov/34325612/)]
8. Apuke OD, Omar B, Tunca EA, Gever CV. Information overload and misinformation sharing behaviour of social media users: testing the moderating role of cognitive ability. *J Inf Sci*. 2022;50(6):1371-1381. [doi: [10.1177/01655515221121942](https://doi.org/10.1177/01655515221121942)]
9. Cardenas N. 'Europe and United States vaccine hesitancy': leveraging strategic policy for 'Infodemic' on COVID-19 vaccines. *J Public Health (Oxf)*. 2022;44(2):e315-e316. [FREE Full text] [doi: [10.1093/pubmed/fdab228](https://doi.org/10.1093/pubmed/fdab228)] [Medline: [34137434](https://pubmed.ncbi.nlm.nih.gov/34137434/)]
10. van der Linden S, Dixon G, Clarke C, Cook J. Inoculating against COVID-19 vaccine misinformation. *EClinicalMedicine*. 2021;33:100772. [FREE Full text] [doi: [10.1016/j.eclinm.2021.100772](https://doi.org/10.1016/j.eclinm.2021.100772)] [Medline: [33655205](https://pubmed.ncbi.nlm.nih.gov/33655205/)]
11. Grover H, Nour R, Zary N, Powell L. Online interventions addressing health misinformation: scoping review. *J Med Internet Res*. 2025;27:e69618. [FREE Full text] [doi: [10.2196/69618](https://doi.org/10.2196/69618)] [Medline: [40906516](https://pubmed.ncbi.nlm.nih.gov/40906516/)]
12. See how vaccinations are going in your county and state. *The New York Times*. URL: <https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html> [accessed 2024-01-18]
13. Korea Disease Control and Prevention Agency. 2022. URL: <https://www.kdca.go.kr/eng/index.do> [accessed 2026-02-07]
14. Liu R, Li GM. Hesitancy in the time of coronavirus: temporal, spatial, and sociodemographic variations in COVID-19 vaccine hesitancy. *SSM Popul Health*. 2021;15:100896. [FREE Full text] [doi: [10.1016/j.ssmph.2021.100896](https://doi.org/10.1016/j.ssmph.2021.100896)] [Medline: [34414255](https://pubmed.ncbi.nlm.nih.gov/34414255/)]
15. Hildreth JEK, Alcendor DJ. Targeting COVID-19 vaccine hesitancy in minority populations in the US: implications for herd immunity. *Vaccines (Basel)*. 2021;9(5):489. [FREE Full text] [doi: [10.3390/vaccines9050489](https://doi.org/10.3390/vaccines9050489)] [Medline: [34064726](https://pubmed.ncbi.nlm.nih.gov/34064726/)]
16. Hwang SE, Kim WH, Heo J. Socio-demographic, psychological, and experiential predictors of COVID-19 vaccine hesitancy in South Korea, October-December 2020. *Hum Vaccin Immunother*. 2022;18(1):1-8. [FREE Full text] [doi: [10.1080/21645515.2021.1983389](https://doi.org/10.1080/21645515.2021.1983389)] [Medline: [34614382](https://pubmed.ncbi.nlm.nih.gov/34614382/)]
17. Kweon SS, Yun I, Choi C, Ryu SY, Cho JH, Shin MH. Factors associated with COVID-19 vaccine hesitancy in Korea. *Chonnam Med J*. 2022;58(1):43-47. [FREE Full text] [doi: [10.4068/cmj.2022.58.1.43](https://doi.org/10.4068/cmj.2022.58.1.43)] [Medline: [35169559](https://pubmed.ncbi.nlm.nih.gov/35169559/)]
18. Willis DE, Andersen JA, Bryant-Moore K, Selig JP, Long CR, Felix HC, et al. COVID-19 vaccine hesitancy: race/ethnicity, trust, and fear. *Clin Transl Sci*. 2021;14(6):2200-2207. [FREE Full text] [doi: [10.1111/cts.13077](https://doi.org/10.1111/cts.13077)] [Medline: [34213073](https://pubmed.ncbi.nlm.nih.gov/34213073/)]
19. Savoia E, Piltch-Loeb R, Goldberg B, Miller-Idriss C, Hughes B, Montrond A, et al. Predictors of COVID-19 vaccine hesitancy: socio-demographics, co-morbidity, and past experience of racial discrimination. *Vaccines (Basel)*. 2021;9(7):767. [FREE Full text] [doi: [10.3390/vaccines9070767](https://doi.org/10.3390/vaccines9070767)] [Medline: [34358184](https://pubmed.ncbi.nlm.nih.gov/34358184/)]
20. Green AL, Stewart AL, Nápoles AM, Strassle PD. COVID-19 vaccination willingness and uptake among low-income Black/African American, Latino, and White adults living in the U.S. *Prev Med Rep*. 2023;35:102367. [FREE Full text] [doi: [10.1016/j.pmedr.2023.102367](https://doi.org/10.1016/j.pmedr.2023.102367)] [Medline: [37638353](https://pubmed.ncbi.nlm.nih.gov/37638353/)]

21. Piltch-Loeb R, Silver DR, Kim Y, Norris H, McNeill E, Abramson DM. Determinants of the COVID-19 vaccine hesitancy spectrum. *PLoS One*. 2022;17(6):e0267734. [FREE Full text] [doi: [10.1371/journal.pone.0267734](https://doi.org/10.1371/journal.pone.0267734)] [Medline: [35648748](https://pubmed.ncbi.nlm.nih.gov/35648748/)]
22. Fridman A, Gershon R, Gneezy A. COVID-19 and vaccine hesitancy: a longitudinal study. *PLoS One*. 2021;16(4):e0250123. [FREE Full text] [doi: [10.1371/journal.pone.0250123](https://doi.org/10.1371/journal.pone.0250123)] [Medline: [33861765](https://pubmed.ncbi.nlm.nih.gov/33861765/)]
23. Taylor-Clark KA, Viswanath K, Blendon RJ. Communication inequalities during public health disasters: Katrina's wake. *Health Commun*. 2010;25(3):221-229. [doi: [10.1080/10410231003698895](https://doi.org/10.1080/10410231003698895)] [Medline: [20461607](https://pubmed.ncbi.nlm.nih.gov/20461607/)]
24. Cordina M, Lauri MA, Lauri J. Attitudes towards COVID-19 vaccination, vaccine hesitancy and intention to take the vaccine. *Pharm Pract (Granada)*. 2021;19(1):2317. [FREE Full text] [doi: [10.18549/PharmPract.2021.1.2317](https://doi.org/10.18549/PharmPract.2021.1.2317)] [Medline: [33828623](https://pubmed.ncbi.nlm.nih.gov/33828623/)]
25. Kreps S, Prasad S, Brownstein JS, Hswen Y, Garibaldi BT, Zhang B, et al. Factors associated with us adults' likelihood of accepting COVID-19 vaccination. *JAMA Netw Open*. 2020;3(10):e2025594. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.25594](https://doi.org/10.1001/jamanetworkopen.2020.25594)] [Medline: [33079199](https://pubmed.ncbi.nlm.nih.gov/33079199/)]
26. Ceron A, Curini L, Iacus SM, Porro G. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media Soc*. 2013;16(2):340-358. [doi: [10.1177/1461444813480466](https://doi.org/10.1177/1461444813480466)]
27. Jafar Z, Quick JD, Larson HJ, Venegas-Vera V, Napoli P, Musuka G, et al. Social media for public health: reaping the benefits, mitigating the harms. *Health Promot Perspect*. 2023;13(2):105-112. [FREE Full text] [doi: [10.34172/hpp.2023.13](https://doi.org/10.34172/hpp.2023.13)] [Medline: [37600540](https://pubmed.ncbi.nlm.nih.gov/37600540/)]
28. Savolainen R. Assessing the credibility of COVID-19 vaccine mis/disinformation in online discussion. *J Inf Sci*. 2023;49(4):1096-1110. [FREE Full text] [doi: [10.1177/01655515211040653](https://doi.org/10.1177/01655515211040653)] [Medline: [37461399](https://pubmed.ncbi.nlm.nih.gov/37461399/)]
29. Choi DH. The multifaceted impact of social media on risk, behavior, and negative emotions during the COVID-19 outbreak in South Korea. *Asian J Commun*. 2021;31(5):337-354. [doi: [10.1080/01292986.2021.1968447](https://doi.org/10.1080/01292986.2021.1968447)]
30. Westerman D, Spence PR, Van Der Heide B. Social media as information source: recency of updates and credibility of information. *J Comput Mediat Commun*. 2013;19(2):171-183. [doi: [10.1111/jcc4.12041](https://doi.org/10.1111/jcc4.12041)]
31. Yazdavar AH, Mahdavinjad MS, Bajaj G, Romine W, Sheth A, Monadjemi AH, et al. Multimodal mental health analysis in social media. *PLoS One*. 2020;15(4):e0226248. [FREE Full text] [doi: [10.1371/journal.pone.0226248](https://doi.org/10.1371/journal.pone.0226248)] [Medline: [32275658](https://pubmed.ncbi.nlm.nih.gov/32275658/)]
32. Jang SH, Gerend MA, Youm S, Yi YJ. Understanding coronavirus disease 2019 (COVID-19) vaccine hesitancy: evidence from the community-driven knowledge site Quora. *Digit Health*. 2022;8:20552076221145426. [FREE Full text] [doi: [10.1177/20552076221145426](https://doi.org/10.1177/20552076221145426)] [Medline: [36544537](https://pubmed.ncbi.nlm.nih.gov/36544537/)]
33. Jang SH, Jung EK, Yi Y. The power of fake news: big data analysis of discourse about COVID-19-related fake news in South Korea. *Int J Commun*. 2023;17:27. [FREE Full text]
34. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. *Lancet Digit Health*. 2021;3(3):e175-e194. [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30315-0](https://doi.org/10.1016/S2589-7500(20)30315-0)] [Medline: [33518503](https://pubmed.ncbi.nlm.nih.gov/33518503/)]
35. Sasse K, Mahabir R, Gkountouna O, Crooks A, Croitoru A. Understanding the determinants of vaccine hesitancy in the United States: a comparison of social surveys and social media. *PLoS One*. 2024;19(6):e0301488. [FREE Full text] [doi: [10.1371/journal.pone.0301488](https://doi.org/10.1371/journal.pone.0301488)] [Medline: [38843170](https://pubmed.ncbi.nlm.nih.gov/38843170/)]
36. Hamilton JL, Nesi J, Choukas-Bradley S. Teens and social media during the COVID-19 pandemic: staying socially connected while physically distant. *PsyArXiv*. Preprint posted online April 29, 2020. [doi: [10.31234/osf.io/5stx4](https://doi.org/10.31234/osf.io/5stx4)]
37. Zhang YB, Han M, Terigele T, Jewell A, Liu N, Byrd G. Social media intergroup expression and White Americans' attitudes toward ethnic Chinese during COVID-19. *Asian J Commun*. 2024;34(4):479-498. [doi: [10.1080/01292986.2024.2337190](https://doi.org/10.1080/01292986.2024.2337190)]
38. Elishar V, Ariel Y, Weimann Saks D. Psychological dynamics and media influence on rumor propagation during the 2023-2024 Israel-Hamas war: a comparative analysis of early- and late-stage conflict. *Asian J Commun*. 2025:1-21. [doi: [10.1080/01292986.2025.2531498](https://doi.org/10.1080/01292986.2025.2531498)]
39. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Hum Vaccin Immunother*. 2020;16(11):2586-2593. [FREE Full text] [doi: [10.1080/21645515.2020.1780846](https://doi.org/10.1080/21645515.2020.1780846)] [Medline: [32693678](https://pubmed.ncbi.nlm.nih.gov/32693678/)]
40. Muric G, Wu Y, Ferrara E. COVID-19 vaccine hesitancy on social media: building a public twitter data set of antivaccine content, vaccine misinformation, and conspiracies. *JMIR Public Health Surveill*. 2021;7(11):e30642. [FREE Full text] [doi: [10.2196/30642](https://doi.org/10.2196/30642)] [Medline: [34653016](https://pubmed.ncbi.nlm.nih.gov/34653016/)]
41. Lachyan A, Kumar V, Ingale S, Gade K. Analyzing the impact of COVID-19 vaccine infodemic on social media: a study of human behaviour in state of Uttar Pradesh. *International Journal of Research and Analytical Reviews (IJRAR)*. 2021. URL: <https://papers.ssrn.com/abstract=3967172> [accessed 2023-01-29]
42. Baines A, Ittefaq M, Abwao M. #Scamdemic, #Plandemic, or #Scaredemic: what parler social media platform tells us about COVID-19 vaccine. *Vaccines (Basel)*. 2021;9(5):421. [FREE Full text] [doi: [10.3390/vaccines9050421](https://doi.org/10.3390/vaccines9050421)] [Medline: [33922343](https://pubmed.ncbi.nlm.nih.gov/33922343/)]
43. Jennings W, Stoker G, Bunting H, Valgarðsson VO, Gaskell J, Devine D, et al. Lack of trust, conspiracy beliefs, and social media use predict COVID-19 vaccine hesitancy. *Vaccines (Basel)*. 2021;9(6):593. [FREE Full text] [doi: [10.3390/vaccines9060593](https://doi.org/10.3390/vaccines9060593)] [Medline: [34204971](https://pubmed.ncbi.nlm.nih.gov/34204971/)]

44. Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera JR, Guillem-Saiz P, Bermejo-Higuera JR, Guillem-Saiz J, et al. Vaccine hesitancy on social media: sentiment analysis from June 2011 to April 2019. *Vaccines (Basel)*. 2021;9(1):28. [FREE Full text] [doi: [10.3390/vaccines9010028](https://doi.org/10.3390/vaccines9010028)] [Medline: [33430428](https://pubmed.ncbi.nlm.nih.gov/33430428/)]
45. Hernandez RG, Hagen L, Walker K, O'Leary H, Lengacher C. The COVID-19 vaccine social media : healthcare providers' missed dose in addressing misinformation and vaccine hesitancy. *Hum Vaccin Immunother*. 2021;17(9):2962-2964. [FREE Full text] [doi: [10.1080/21645515.2021.1912551](https://doi.org/10.1080/21645515.2021.1912551)] [Medline: [33890838](https://pubmed.ncbi.nlm.nih.gov/33890838/)]
46. Lee H, Hyun M, Baek J. Development of consumer health information seeking model through an analysis of social Q&A service . *International Information Institute (Tokyo)*. 2015;18(8):3361-3374. [FREE Full text]
47. Yin F, Ji M, Yang Z, Wu Z, Xia X, Xing T, et al. Exploring the determinants of global vaccination campaigns to combat COVID-19. *Humanit Soc Sci Commun*. 2022;9(1):1-13. [doi: [10.1057/s41599-022-01106-7](https://doi.org/10.1057/s41599-022-01106-7)]
48. Bae BJ, Yi YJ. Identification and Comparison of the Persuasive Elements Present in “Best Answers” to STD-Related Questions on Social Q&A Sites: Yahoo! Answers (United States) Versus Knowledge-iN (South Korea). *International Journal of Communication*. 2019;13:19. [FREE Full text]
49. Choe SH. Want a vaccine reservation in South Korea? Try waiting 111 hours. *New York Times*. 2021. URL: <https://www.nytimes.com/2021/07/28/world/asia/south-korea-vaccine-covid.html> [accessed 2023-04-02]
50. Mellish TI, Luzmore NJ, Shahbaz AA. Why were the UK and USA unprepared for the COVID-19 pandemic? The systemic weaknesses of neoliberalism: A comparison between the UK, USA, Germany, and South Korea. *J Glob Faultlines*. 2020;7(1):9-45. [doi: [10.13169/jglobfaul.7.1.0009](https://doi.org/10.13169/jglobfaul.7.1.0009)]
51. Kerr JR, Schneider CR, Recchia G, Dryhurst S, Sahlin U, Dufouil C, et al. Correlates of intended COVID-19 vaccine acceptance across time and countries: results from a series of cross-sectional surveys. *BMJ Open*. 2021;11(8):e048025. [FREE Full text] [doi: [10.1136/bmjopen-2020-048025](https://doi.org/10.1136/bmjopen-2020-048025)] [Medline: [34341047](https://pubmed.ncbi.nlm.nih.gov/34341047/)]
52. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. *Nat Hum Behav*. May 10, 2021;5(7):947-953. [doi: [10.1038/s41562-021-01122-8](https://doi.org/10.1038/s41562-021-01122-8)]
53. Alasfour M, Alqahtani R, Amri M, Alsultan S, Hobani S, Almufaireej K, et al. Use patterns and challenges of the social media platform x among physiotherapists in Saudi Arabia: cross-sectional study. *JMIR Form Res*. 2026;10:e84471. [FREE Full text] [doi: [10.2196/84471](https://doi.org/10.2196/84471)] [Medline: [41576292](https://pubmed.ncbi.nlm.nih.gov/41576292/)]
54. Chandra M, Rodrigues A, George J. An enhanced deep learning model for duplicate question detection on quora question pairs using siamese LSTM. *IEEE*; 2022. Presented at: IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE); 2022, April 23-24; Ballari, India. [doi: [10.1109/icdcece53908.2022.9792906](https://doi.org/10.1109/icdcece53908.2022.9792906)]
55. Gazan R. *J Am Soc Inf Sci*. 2011;62(12):2301-2312. [doi: [10.1002/asi.21562](https://doi.org/10.1002/asi.21562)]
56. García B, Gallego M, Gortázar F, Muñoz-Organero M. A survey of the selenium ecosystem. *Electronics*. 2020;9(7):1067. [doi: [10.3390/electronics9071067](https://doi.org/10.3390/electronics9071067)]
57. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimed Tools Appl*. 2018;78(11):15169-15211. [doi: [10.1007/s11042-018-6894-4](https://doi.org/10.1007/s11042-018-6894-4)]
58. Angelov D. Top2Vec: distributed representations of topics. *arXiv:2008.09470*. 2020. [doi: [10.5860/choice.189890](https://doi.org/10.5860/choice.189890)]
59. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2019. Presented at: Proceedings of NAACL-HLT 2019; 2019 June 2 -7:4171-4186; Minneapolis, Minnesota. URL: <https://arxiv.org/abs/1810.04805v2> [doi: [10.18653/v1/n19-1423](https://doi.org/10.18653/v1/n19-1423)]
60. Clark K, Luong MT, Le QV, Manning CD. ELECTRA: pre-training text encoders as discriminators rather than generators. *ArXiv*. Preprint posted online March 23, 2020. [FREE Full text]
61. Pampapura Madali N, Alsaïd M, Hawamdeh S. The impact of social noise on social media and the original intended message: BLM as a case study. *J Inf Sci*. 2022;50(1):89-103. [doi: [10.1177/01655515221077347](https://doi.org/10.1177/01655515221077347)]
62. Zhan X, Song M, Shrader CH, Forbes CE, Algarin AB. Decoding HIV discourse on social media: large-scale analysis of 191,972 tweets using machine learning, topic modeling, and temporal analysis. *J Med Internet Res*. 2025;27:e76745. [FREE Full text] [doi: [10.2196/76745](https://doi.org/10.2196/76745)] [Medline: [40882217](https://pubmed.ncbi.nlm.nih.gov/40882217/)]
63. Niu L, Dai X, Zhang J, Chen J. Topic2Vec: learning distributed representations of topics. *IEEE*; 2015. Presented at: International Conference on Asian Language Processing (IALP); 2015 October 24-25:193-196; Suzhou. [doi: [10.1109/ialp.2015.7451564](https://doi.org/10.1109/ialp.2015.7451564)]
64. Luu TJP, Follmann R. The relationship between sentiment score and COVID-19 cases in the United States. *J Inf Sci*. 2022;49(6):1615-1630. [doi: [10.1177/01655515211068167](https://doi.org/10.1177/01655515211068167)]
65. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J*. 2014;5(4):1093-1113. [doi: [10.1016/j.asej.2014.04.011](https://doi.org/10.1016/j.asej.2014.04.011)]
66. Zhang K, Guo Z, Ai Y, Li AR, Li A, Liu Z, et al. Surveillance of twitter data on COVID-19 symptoms during the omicron variant period: a sentiment analysis. *JMIR Form Res*. 2025;9:e66237. [FREE Full text] [doi: [10.2196/66237](https://doi.org/10.2196/66237)] [Medline: [40986849](https://pubmed.ncbi.nlm.nih.gov/40986849/)]
67. Cortiz D. Exploring transformers models for emotion recognition: a comparison of BERT, DistilBERT, RoBERTa, XLNET and ELECTRA. *Association for Computing Machinery*; 2022. Presented at: CCRIS '22: Proceedings of the 2022 3rd International Conference on Control, Robotics and Intelligent System; 2022, August 26-28:230-234; Virtual event China. [doi: [10.1145/3562007.3562051](https://doi.org/10.1145/3562007.3562051)]

68. Camacho-Collados J, Rezaee K, Riahi T, Ushio A, Loureiro D, Antypas D, et al. TweetNLP: cutting-edge natural language processing for social media. 2022. Presented at: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; 2026, February 06:38-49; Abu Dhabi, UAE. [doi: [10.18653/v1/2022.emnlp-demos.5](https://doi.org/10.18653/v1/2022.emnlp-demos.5)]
69. Viswanathan S, Damodaran N, Simon A, George A, Anand KM, Soman KP. Detection of duplicates in quora and twitter corpus. In: Advances in Intelligent Systems and Computing. Singapore. Springer; 2019:519-528.
70. Md M, Qudar A, Mago V. TweetBERT: A Pretrained Language Representation Model for Twitter Text Analysis. Kaggle. URL: <https://www.kaggle.com/parulpandey/100-mostfollowed-twitter-accounts-as-of-dec2019> [accessed 2024-04-01]
71. SKTBrain/KoBERT. GitHub. 2021. URL: <https://github.com/SKTBrain/KoBERT> [accessed 2024-07-02]
72. Park J. monologg/KoELECTRA. GitHub. 2020. URL: <https://github.com/monologg/KoELECTRA> [accessed 2024-07-02]
73. Ganaie M, Hu M, Malik A, Tanveer M, Suganthan P. Ensemble deep learning: a review. Eng Appl Artif Intell. Oct 2022;115:105151. [doi: [10.1016/j.engappai.2022.105151](https://doi.org/10.1016/j.engappai.2022.105151)]
74. Yang Y, Lv H, Chen N. A survey on ensemble learning under the era of deep learning. Artif Intell Rev. 2022;56(6):5545-5589. [doi: [10.1007/s10462-022-10283-5](https://doi.org/10.1007/s10462-022-10283-5)]
75. Blake L, Tucker P, Vanderloo LM. Mothers' perspectives of the barriers and facilitators to reducing young children's screen time during COVID-19: A reddit content analysis. PLoS One. 2024;19(3):e0301089. [FREE Full text] [doi: [10.1371/journal.pone.0301089](https://doi.org/10.1371/journal.pone.0301089)] [Medline: [38536885](https://pubmed.ncbi.nlm.nih.gov/38536885/)]
76. Shah HA, Househ M. Understanding loneliness through analysis of twitter and reddit data: comparative study. Interact J Med Res. 2025;14:e49464. [FREE Full text] [doi: [10.2196/49464](https://doi.org/10.2196/49464)] [Medline: [40085832](https://pubmed.ncbi.nlm.nih.gov/40085832/)]
77. You Y, Ma D, Chen C. Public trust during a public health crisis: evaluating the immediate effects of the pandemic on institutional trust. J Chin Polit Sci. 2023;29(1):1-29. [doi: [10.1007/s11366-023-09874-y](https://doi.org/10.1007/s11366-023-09874-y)]
78. Shim E. Optimal allocation of the limited COVID-19 vaccine supply in South Korea. J Clin Med. 2021;10(4):591. [FREE Full text] [doi: [10.3390/jcm10040591](https://doi.org/10.3390/jcm10040591)] [Medline: [33557344](https://pubmed.ncbi.nlm.nih.gov/33557344/)]
79. COVID-19: stay up to date with COVID-19 vaccines including boosters. Centers for Disease Control and Prevention. 2023. URL: <https://stacks.cdc.gov/view/cdc/126909> [accessed 2026-02-07]
80. Alfieri NL, Kusma JD, Heard-Garris N, Davis MM, Golbeck E, Barrera L, et al. Parental COVID-19 vaccine hesitancy for children: vulnerability in an urban hotspot. BMC Public Health. 2021;21(1):1662. [FREE Full text] [doi: [10.1186/s12889-021-11725-5](https://doi.org/10.1186/s12889-021-11725-5)] [Medline: [34517848](https://pubmed.ncbi.nlm.nih.gov/34517848/)]
81. Rieger MO, Wang M. Trust in government actions during the COVID-19 crisis. Soc Indic Res. 2022;159(3):967-989. [FREE Full text] [doi: [10.1007/s11205-021-02772-x](https://doi.org/10.1007/s11205-021-02772-x)] [Medline: [34456450](https://pubmed.ncbi.nlm.nih.gov/34456450/)]
82. Yi YJ. Sexual health information-seeking behavior on a social media site: predictors of best answer selection. Online Inf Rev. 2018;42(6):880-897. [doi: [10.1108/oir-06-2017-0204](https://doi.org/10.1108/oir-06-2017-0204)]
83. Bae BJ, Yi YJ. Internet Res. 2017;27(5):1104-1121. [doi: [10.1108/intr-08-2016-0245](https://doi.org/10.1108/intr-08-2016-0245)]
84. Forrest E, Duff AS. Ethical ePunditry? The role of expertise in online opinion-making. J Inf Ethics. 2017;26(1):66-80. [FREE Full text]
85. Kong D, Chen A, Zhang J, Xiang X, Lou WQV, Kwok T, et al. Public discourse and sentiment toward dementia on chinese social media: machine learning analysis of weibo posts. J Med Internet Res. 2022;24(9):e39805. [FREE Full text] [doi: [10.2196/39805](https://doi.org/10.2196/39805)] [Medline: [36053565](https://pubmed.ncbi.nlm.nih.gov/36053565/)]
86. Jang M, Kim S. Humanit Soc Sci Commun. 2024;11(1):744. [doi: [10.1057/s41599-024-03239-3](https://doi.org/10.1057/s41599-024-03239-3)]
87. Jiang H, Qiang M, Zhang D, Wen Q, Xia B, An N. Sustainability. May 10, 2018;10(5):1509. [FREE Full text] [doi: [10.3390/su10051509](https://doi.org/10.3390/su10051509)]
88. OpenAI. ChatGPT. OpenAI. URL: <https://openai.com/index/introducing-gpt-5/> [accessed 2026-02-07]

Abbreviations

- BERT:** Bidirectional Encoder Representations from Transformers
- CDC:** Centers for Disease Control and Prevention
- ELECTRA:** Efficiently Learning an Encoder that Classifies Token Replacement Accurately
- H1N1:** Hemagglutinin Type 1 and Neuraminidase Type 1
- LDA:** Latent Dirichlet Allocation
- MERS:** Middle East Respiratory Syndrome
- Naver KiN:** Naver Knowledge-iN
- NLP:** natural language processing
- Q&A:** question and answer
- STROBE:** Strengthening the Reporting of Observational Studies in Epidemiology
- WHO:** World Health Organization

Edited by A Mavragani; submitted 25.Sep.2025; peer-reviewed by A Afzalian, Z Liu; comments to author 06.Jan.2026; revised version received 30.Jan.2026; accepted 02.Feb.2026; published 05.Mar.2026

Please cite as:

Youm S, Jang SH, Kwak H, Choi J, Yi YJ

Cross-Cultural Differences in Public Discourse on COVID-19 Vaccination in the United States and South Korea: Cross-Sectional Analysis Using Natural Language Processing

JMIR Form Res 2026;10:e84791

URL: <https://formative.jmir.org/2026/1/e84791>

doi: [10.2196/84791](https://doi.org/10.2196/84791)

PMID:

©Sangpil Youm, Sou Hyun Jang, Haewoon Kwak, Jaeyoung Choi, Yong Jeong Yi. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 05.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.