Social Network Analysis and Mining (2025) 15:104
https://doi.org/10.1007/s13278-025-01532-w

ORIGINAL ARTICLE

Cross-platform spatiotemporal sentiment trends analysis of COVID-19 vaccine discourse

Abdulrahman Alharbi[1,2] · Rafaa Aljurbua[1,3] · Shelly Gupta[1] · Hussain Otudi[1,2] · Jovan Andjelkovic[1] · Zoran Obradovic[1]

Received: 30 May 2025 / Revised: 11 September 2025 / Accepted: 12 September 2025

Abstract
The COVID-19 pandemic has sparked intense global discussions about vaccine safety, efficacy, and distribution on social media. It underscored the need to analyze how vaccine-related sentiments propagate across social media and interact with news media articles. Despite extensive research on COVID-19 vaccines, most existing studies examine the sentiment of the COVID-19 vaccine by focusing on social media or news articles in isolation. This study bridges the gap by exploring correlations between these sources through a hierarchical spatiotemporal sentiment analysis framework that integrates social media discussions and mainstream news across global, national (US), and regional (Pennsylvania and Philadelphia) scales. Leveraging over 7 million English tweets and 6,500 news articles alongside physical events, official government records, and demographic data collected between January 2020 and June 2022, we introduce a user location inference method to approximate geographic context. Our approach leverages TriLex, a multi-lexicon sentiment method, and BERTopic to extract nuanced topics, further refined by ChatGPT for enhanced interpretability. The study period was divided into six key intervals, ranging from the beginning of the COVID-19 pandemic to the emergence of the Delta and Omicron variants. The results indicate distinct sentiment patterns in different regions and periods, partially aligning with the NYT's vaccine-related articles. Although no causal link has been established, our findings highlight the value of correlating multi-scale social media analysis with news articles to address vaccine hesitancy, refine public health messaging, and guide future research on information diffusion in global crises.

Keywords Sentiment analysis · Spatiotemporal analysis · Topic modeling · ChatGPT interpretation · COVID-19 vaccine

✉ Abdulrahman Alharbi
adalharbi@temple.edu

✉ Zoran Obradovic
zoran.obradovic@temple.edu

Rafaa Aljurbua
rafaa.aljurbua@temple.edu

Shelly Gupta
tuk40762@temple.edu

Hussain Otudi
tuk81560@temple.edu

Jovan Andjelkovic
tul67382@temple.edu

1 Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, USA

2 Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia

3 Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia

1 Introduction

The COVID-19 pandemic has significantly influenced public discourse on social media, particularly regarding vaccine development and distribution. Research indicates that vaccine acceptance varied considerably across regions and countries, with users in the World Health Organization's South-East Asia, Eastern Mediterranean, and Western Pacific Regions expressing higher acceptance than other areas (Zhou et al. 2023, 2022). For instance, India, Indonesia, and Pakistan demonstrated elevated acceptance rates, whereas high-income nations such as South Korea, Japan, and the Netherlands exhibited lower acceptance (Zhou et al. 2022). Cultural beliefs and socioeconomic factors also shaped vaccine sentiments: in Indonesia, cultural myths and religious stances led to heightened hesitancy and resistance (Ida et al. 2024), while socioeconomically disadvantaged groups in the United States were likelier to hold polarized

Published online: 06 November 2025

opinions, being either strongly pro-vaccine or anti-vaccine (Lyu et al. 2022). Furthermore, an analysis of tweets in multiple languages underscored diverse concerns; one study of Japanese tweets highlighted evolving topics of fear, anxiety, and personal experiences (Takedomi et al. 2024), whereas English-language tweets often revolved around side effects, safety, and efficacy (Takedomi et al. 2024; Ahmad et al. 2024).

Social media platforms, such as Twitter, provide a rich source of data for studying public opinion and sentiment on various topics, including vaccines (Baj-Rogowska 2021; Du et al. 2017). Social media has significantly shaped and reflected public opinion toward COVID-19 vaccines over time due to regional and cultural differences. Early optimism about vaccine development led to skepticism linked to possible side effects and disparate distribution across low and high-income countries (Zhou et al. 2023, 2022; Verma et al. 2023). Concerns about government transparency, fairness of vaccine allocation, and inconsistent public health guidelines contributed to increased distrust (Slavik et al. 2023; Nogara et al. 2024). Such skepticism and frustration were further magnified by online conspiracy theories and misinformation, often fueled by influential accounts and politically biased media (Lenti et al. 2023; Jiang et al. 2024). In contrast, efforts to moderate the vaccine-related content had mixed effects. Although platforms like Twitter successfully reduced misinformation by suspending certain accounts, the spread of low credibility information persisted through alternative platforms such as Telegram and YouTube (Lenti et al. 2023; Nogara et al. 2024). These complex and evolving dynamics underscore the need for systematic, multilevel analyses, especially those capturing spatiotemporal variations, to guide targeted interventions and inform public health strategies. By examining large-scale social media data in conjunction with news media coverage, researchers and policymakers can gain deeper insight into how vaccine sentiments emerge, evolve, and might be effectively addressed.

Sentiment analysis and topic modeling, which are referred to as opinion mining, involve the automatic identification of sentiments expressed in text to study the emotional tone of text, and thematic discussions have emerged as a powerful tool for estimating public sentiment on social media. It categorizes text into positive, negative, or neutral sentiments, providing valuable insights into public opinions. The rise of social media platforms like Twitter has made sentiment analysis a critical tool for understanding real-time public sentiment on various topics, events, or brands (Challapalli 2024; Singh et al. 2023). On the other hand, news articles provide structured and formal text, which can be analyzed to gauge public sentiment on broader topics or events. The combination of sentiment analysis on tweets and news articles offers a comprehensive view of public opinion, enabling businesses and policymakers to make informed decisions (Manisha and Acharya 2023; Prasad et al. 2023). Recent studies have increasingly underscored the value of sentiment analysis of cross-platforms in tracing COVID-19 vaccine discussions and monitoring public reactions. The authors of Alipour et al. (2024) examine the social dynamics of public opinion concerning the COVID-19 vaccines and the release of ChatGPT across multiple online platforms, including Twitter, Facebook, Instagram, Reddit, YouTube, and GDELT. Their findings reveal that both platform design and user interests play significant roles in shaping the content and pace of conversations related to vaccines. Another study highlights the impact of politically biased news on vaccine sentiments within social media, demonstrating that users with moderate or uncertain stances are especially susceptible to politicized content that news feeds can magnify hesitancy or polarized attitudes toward vaccination (Jiang et al. 2024). Other studies spotlight the link between social media news and real-world behavioral outcomes. For example, a study (Zhang et al. 2023) argues that exposure to vaccine risk news can increase individuals' anxieties and fuel hesitancy, whereas frequent contact with safety news bolsters confidence in the COVID-19 vaccine. This interplay between informational cues and public behavior is further emphasized by Bai and Lee (2024), which examines the coverage of news media alongside the sentiments on social media regarding the stock prices of major manufacturers of the COVID-19 vaccine. These studies illustrate that analyzing vaccine discourse across various media channels, including news articles and social media platforms, can offer valuable insights into the evolving patterns of public sentiment, news dynamics, and policy challenges.

In our previous study (Alharbi et al. 2024), we utilized the VADER lexicon for sentiment classification and ChatGPT with BERTopic to interpret the results of topic modeling and sentiment analysis, thereby enhancing the contextual understanding of public discourse on social media. However, the sentiment classification relies on a single lexicon. Based on this, our subsequent study introduced TriLex (Alharbi et al. 2025), a novel unsupervised sentiment analysis approach that combines the strengths of TextBlob, VADER, and AFINN with BERT embeddings and an LSTM layer to deliver more robust polarity estimates for short and noisy text. However, neither of these studies examined how sentiment evolves over time and space. Building on our previous work and the observations from our literature review, this study introduces a multiscale spatiotemporal analysis framework that captures both global trends and local nuances in public sentiment regarding the COVID-19 vaccines across two platforms: Twitter and news articles. Although prior work has explored vaccine acceptance from various angles, few efforts have integrated a detailed spatiotemporal approach that contextualizes sentiments alongside pivotal real-world events, demographic trends, and news outlets' reporting.

We hypothesize that it is challenging to estimate people's reactions to COVID-19 vaccines based on social media activities, as reactions and emotions may change in response to recent events and COVID-19 variants. Moreover, there is a measurable correlation between the sentiment expressed in news articles and the sentiment on social media. We aspire to gain deeper insights into the underlying dynamics that shape societal reactions and responses by demonstrating the interplay between public sentiment and news articles. The following research questions guide our research:

- **RQ1:** Could news media shape public discourse about COVID-19 vaccines?
- **RQ2:** Do people in different states of the USA react similarly or differently to various COVID-19 vaccines?
- **RQ3:** Why do people refuse or accept COVID-19 vaccines in Pennsylvania and Philadelphia? This question is studied by addressing the following aspects:
  a) What factors affect the acceptance or rejection of COVID-19 vaccines in Philadelphia?
  b) Does social level or race affect the attitude toward the COVID-19 vaccine in Philadelphia?

This study examines public reactions to COVID-19 vaccines from a spatiotemporal perspective. The main contributions of this paper are summarized as follows:

- Introduce a multiscale spatiotemporal analysis approach to demonstrate how global trends differ from local nuances. This in-depth examination of hierarchical spatiotemporal analysis offers comprehensive insights into public opinion across space and time.
- Collect over 7 million English tweets and 6, 581 news titles from the NYT related to COVID-19 vaccines published between January 2020 and June 2022. For Philadelphia, we took a step further by collecting and labeling tweets from all zip codes within the city (West, East, South, and North). This allowed us to gain a comprehensive understanding of sentiments in different areas.
- Utilize TriLex and BERTopic clusters on top of ChatGPT to address interpretability challenges in large-scale textual analysis. We feed ChatGPT with the keywords of the topic extracted by BERTopic and the sentiment category achieved by TriLex, then ask ChatGPT to assign a topic title, keywords, description, and sentiment category over time and space.
- Integrate multiple physical events with SDI demographic information to explore how these factors influence sentiment and vaccination rate in Philadelphia. We collect physical events, such as death rates, case numbers, and vaccination rates, from Google Analytics, Bloomberg, and government websites for Pennsylvania and Philadelphia. Additionally, we obtained vaccination and demography data from each zip code in Philadelphia.

The remainder of this paper is organized as follows. Section 2 discusses related works and previous studies. Section 3 describes the methodology, including data collection, preprocessing, geographical location extractions, and sentiment analysis approaches. Section 4 shows the results, findings, and discussions of the COVID-19 vaccination tweets and news articles. Finally, Sect. 6 presents the conclusion and future works.

2 Related works

Social media has become integral to professional and societal communication channels (Wong et al. 2021; Aljurbua et al. 2023, 2024, 2025). Such platforms were critical in spreading health information and misleading information regarding COVID-19 (Tsao et al. 2021). Previous studies investigated how social platforms can enhance public trust in the COVID-19 vaccine (Puri et al. 2020).

2.1 Sentiment analysis

Sentiment analysis often leverages machine learning, deep learning, and lexicon-based approaches, supported by NLP techniques for data preprocessing and feature extraction. Supervised machine learning approaches (e.g., SVM, Logistic Regression, Naive Bayes) are popular due to their good performance with labeled datasets, achieving up to 78% testing accuracy in some tweet analysis tasks (Vladic et al. 2024). However, deep learning models (LSTM, BiLSTM, BERT) have proven especially powerful for capturing complex language patterns, with testing accuracies reaching as high as 90% and F1 scores of 98% (Challapalli 2024; Kumar et al. 2023). Meanwhile, the lexicon-based approach provides an effective solution and reduces the sources and time complexity for large and unsupervised datasets. The unsupervised TriLex approach, combined with BERT embedding and LSTM, yielded the best performance, achieving the highest accuracy and F1 score across different domains. Specifically, it achieved an accuracy of 98.78% and an F1 score of 98.83% for the COVID-19 tweets dataset (Alharbi et al. 2025). Additionally, some machine learning-based work focused on adapting the Latent Dirichlet Allocation (LDA) for topic extraction and VADER for sentiment analysis (Abadah et al. 2023). Social media platforms like

Twitter are known to host a mix of human and automated (bot) accounts, which may influence the structure and sentiment of public discourse. For instance, Ng and Carley (2025) provided a global comparison of bot and human behaviors, showing that global event discussion in social media comes from 20% bots, which may amplify specific narratives and distort sentiment trends across regions. While our study focuses on sentiment and topic modeling from user-generated content, it is essential to acknowledge that some content may originate from automated sources, which could potentially introduce bias into sentiment distributions. Addressing this challenge remains an important direction for future work.

2.2 Sentiment analysis using COVID-19 vaccines tweets

Sentiment analysis on tweets related to COVID-19 vaccines has been the focus of several recent studies. In 2021, Mehedi Shamrat et al. (2021) used natural language processing and a supervised KNN classification algorithm to analyze 30, 000 tweets from three COVID-19 vaccine hashtags #Pfizer, #Moderna, and #AstraZeneca consisting of 10, 000 tweets associated with each hashtag. The study results indicated that 43.53% of the tweets were neutral, 33.04% were positive, and 23.42% were negative. In a larger study, Liu and Liu (2021) collected 2, 678, 372 tweets and used a VADER sentiment analysis tool to categorize the sentiment of tweets associated with COVID-19 vaccines. The study's results indicated that people expressed positive sentiments when discussing the efficacy of vaccines, the benefits of vaccination, and the importance of immunization. However, people expressed negative sentiments when talking about vaccine hesitancy, the safety of vaccines, and the politics surrounding vaccines.

Recent studies have emphasized a strong connection between the spread of COVID-19 vaccine misinformation on social media and growing public vaccine hesitancy, raising concerns about its impact on vaccination efforts (Skafle et al. 2022). Another study (Blane et al. 2023) has also shown that both pro- and anti-vaccine communities engage in similar approaches. However, while anti-vaccine tweets often rely on explanations via anecdotes, pro-vaccine tweets emphasize facts and statistical reasoning to support vaccine efficacy and safety. Building on this, researchers found that anti-vaccine tweets often use moral and emotional language, and that certain emotions, such as anger, are linked to consistent stances, highlighting how moral principles can be beneficial to counter misinformation (Phillips et al. 2024). Other studies have examined how public sentiment evolves across social media and sociopolitical contexts. For example, Hwang et al. (2022) employed Structural Topic Modeling (STM) to uncover key topics discussed within positive and negative sentiments on Twitter conversations during the early phase of the COVID-19 pandemic. Their findings underscore the importance of monitoring evolving public discourses to inform interventions aimed at combating vaccine hesitancy. Ebeling et al. (2022) investigated how political polarization influenced vaccination stances in Brazil, revealing that vaccine sentiment often aligned with political ideologies, thereby shaping public behavior and discourse. These studies complement our work by illustrating the broader socio-political dynamics embedded in vaccine-related sentiment and highlight the need for scalable, automated sentiment frameworks to uncover spatiotemporal and thematic trends across large datasets.

2.3 Spatiotemporal sentiment analysis

Several studies have applied spatiotemporal sentiment analysis in social media across various topics and domains, such as politics, tourism, and health. Ali, GG Md Ali et al. (2021) performed sentiment analysis on vaccine tweets to analyze public reaction in the USA. They conducted a spatiotemporal analysis at the state level on 14056 geo-tagged posts from early February to late March 2021. The finding indicated that positive sentiments were higher than negative sentiments in both periods, with a notable decline in positive sentiment in late March 2021 and an increase in negative and neutral sentiments. Moreover, Li et al. (2022) proposed a spatiotemporal framework for analyzing over 1 million sentiments towards COVID-19 vaccination posted on Chinese social media. That research investigated the potential influence of public opinion by unveiling the spatiotemporal changes in public discourse volume, emotional expressions, and topics in three stages. Results indicated less discussion in the early pandemic stage.

In contrast, there were more discussions with positive sentiments in the late stage, which were influenced by confirmed cases, local epidemic outbreaks, and news dissemination. Moreover, spatial statistics and hotspot analysis were successfully applied to COVID-19 cases to identify the spatio-temporal patterns and trends (Neşe and Bakir 2022). The results revealed significant variations in COVID-19 incidence rates across regions and over time, highlighting the potential of spatio-temporal analysis for understanding and controlling disease outbreaks. Further, public sentiment of users' tweets in the United States was estimated by a spatio-temporal aspect integrating geographic and temporal information (Hu et al. 2021).

2.4 Cross-platform topic and sentiment modeling

Recent research has explored the dynamic interplay between formal news media and social media discourse, particularly through cross-platform sentiment and topic modeling. These studies highlight how content dissemination, user behavior, and credibility assessments differ between platforms and shape public discourse. Nanayakkara et al. (2024) conducted a large-scale topic modeling study using BERTopic and RoBERTa embeddings across Twitter and online news media related to the Black Lives Matter movement. Their cross-platform framework revealed platform-specific discourse, showing how Twitter captured emotionally charged narratives while news media conveyed more formal, structured themes. The study demonstrated the importance of aligning computational models with platform-specific communication styles. Another study by Hua et al. (2016) introduced the News and Twitter Interaction Topic (NTIT) model, a hierarchical Bayesian approach that jointly models and aligns topics across Twitter and news articles, capturing platform-specific topics, temporal dynamics, and directional topic influences, thus revealing how events emerge and propagate differently on each platform. This study establishes a methodological precedent for our cross-platform alignment. Kalogeropoulos et al. (2017) provide comparative insights into the behavioral divergence in news engagement across platforms, finding that social media users exhibit higher degrees of commenting and sharing driven by demographic factors and participatory cultures distinct from those in traditional news environments. Lepird et al. (2024) further introduce a network-based non-credibility scoring (NCS) model to detect misleading or agenda-driven news sites on Facebook, which parallels our use of sentiment and topic coherence metrics to assess the quality of discourse in vaccine narratives. These works provide the theoretical and methodological foundation for cross-platform analyses, supporting our study, which examines the interaction between Twitter (X) and The New York Times (NYT) articles within high-stakes contexts, such as COVID-19 vaccine discourse.

Our study distinguishes itself from previous research by adding a new hierarchical spatiotemporal perspective to the conversation around COVID-19 vaccine discourse by systematically linking multilevel social media sentiment (global, national, and local scales) with news articles coverage. Whereas prior studies often concentrate on either cross-platform comparison, the impacts of politically biased news, localized behavioral responses to social media cues, or correlations between coverage and market outcomes, our work synthesizes these dimensions under a single framework. Specifically, we combine large-scale Twitter data with NYT articles, apply TriLex for sentiment classification, and BERTopic on top of ChatGPT-assisted topic modeling to capture the temporal, regional, and contextual factors shaping vaccine sentiments. This multilevel, cross-platform approach gives us a nuanced view of how news events and real-world policies reverberate through social media dialogue.

3 Materials and methods

The proposed methodology shown in Fig. 1 presents a multiscale spatiotemporal analysis framework for analyzing sentiment and topic trends in public discourse related to COVID-19 vaccines. The workflow begins by leveraging two datasets: (i) over 7 million English tweets that reference COVID-19 vaccines and (ii) 6,581 vaccine-related articles

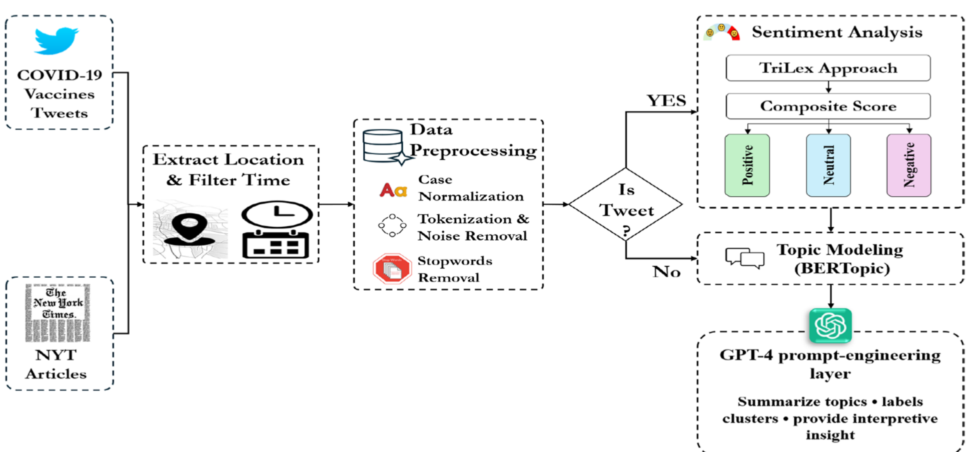

Fig. 1 Workflow of the proposed hierarchical spatiotemporal framework for analyzing sentiment and topics related to COVID-19 vaccines. The process integrates data from Twitter and The New York Times, extracts temporal and geographic metadata, and applies standardized preprocessing. Tweets are analyzed using the TriLex sentiment classification model to generate composite sentiment scores. Both platforms are subjected to topic modeling via BERTopic and GPT-4 to summarize and enhance interpretability across multiple spatiotemporal scales

Table 1 Summary of the collected datasets across Twitter (X) and news articles

Data Source	Volume	Time Frame
Twitter (X) English	7, 217, 663 Posts	Jan 2020 - Jun 2022
Filtered Tweets Mentioning Four Vaccine (Pfizer, Moderna, AstraZeneca, Johnson & Johnson)	1, 037, 650 Posts	Jan 2020 - Jun 2022
NYT Vaccine Articles	6, 851 Articles	Jan 2020 - Jun 2022

Table 2 Summary of the number of tweets corresponding to 6, 851 news articles used in each spatiotemporal scale related to COVID-19 vaccines between Jan 2020 and Jun 2022

Scale	Data Source	# Tweets	Key Characteristics
Global (206 countries)	Twitter (English only)	810,097 (location-tagged)	Reflects a broad overview of sentiment with location inference
US States (50 states)	Twitter (subset of above)	356,443	Focus on US states, merges official data from CDC & local
Pennsylvania	Twitter (Pennsylvania-located)	17,558	State-level detail, correlated with PA Dept. of Health records
Philadelphia	Twitter (Philadelphia-located)	3,347	Zip-code granularity, matched with demographic data

from The New York Times (NYT). We extract geographical location from the user's profile metadata for Twitter and filter timestamps according to six pandemic milestones defined in Sect. 3.6. NYT articles are aligned to the same six temporal windows, but no geographic filtering is applied. Both datasets were then preprocessed considering case normalization, tokenization, noise reduction, and the removal of stopwords to ensure textual consistency and analytical quality. Then, we build the TriLex sentiment estimation approach and the BERTopic modeling approach on top of ChatGPT to generate thematic trends titles and interpretive summaries for each interval across different locations. In the Twitter branch, we prompt ChatGPT to utilize TriLex to assign sentiment polarity scores as positive, neutral, or negative, and limit BERTopic to $\leq$ 20 themes per interval and location. We then feed each cluster's keywords and TriLex polarity into ChatGPT to provide human-like insights. In the NYT branch, we skip TriLex-focusing instead on topic extraction with BERTopic, followed by the ChatGPT summarization. This approach yields sentiment-rich, temporally aligned topic narratives for social media and news articles, providing a rigorous foundation for systematic comparison between user-generated discourse and mainstream news coverage.

3.1 Datasets

The data collection involved scraping data from social media posts and news articles about COVID-19 vaccines. We utilized Twitter as a popular social media platform providing massive public opinions since people tend to use it to express and share their thoughts and feelings. We utilized the snscrape scraper to collect a sizable corpus of English-language tweets related to COVID-19 vaccines based on specific hashtags, including (#Pfizer, #Moderna, #AstraZeneca, and #johnsonandjohnson). Moreover, we collected tweets from other relevant hashtags (#COVID, #COVID-19, #COVIDVaccine) using case-insensitive keywords such as Pfizer, BioNTech, Pfizer-BioNTech, AstraZeneca, Oxford, AstraZeneca/Oxford, Moderna, mRNA, Spikevax, Johnson, Janssen, J&J, and J&J/Janssen. By employing these hashtags and keywords, we obtained 7, 217, 663 English tweets related to COVID-19 vaccinations worldwide between January 1st, 2020, and June 1st, 2022, after removing all duplicate tweets. News articles were retrieved from The New York Times outlet utilizing the developer API. We scraped 6, 581 news articles related to COVID-19 vaccines in the same period corresponding to the tweet time using similar keywords to hashtags. This dual-source approach comprehensively represents informal public opinions and formal media narratives, providing a rich sentiment and topic analysis dataset. Table 1 summarizes the retrieved data related to the COVID-19 vaccines. Additionally, Table 2 summarizes the datasets used in each spatiotemporal scale.

3.2 Data preprocessing

Data preprocessing is essential in transforming raw, unstructured text into a clean and structured format suitable for analysis. This process ensures consistency and enhances the quality of insights derived from the raw data. Leveraging a text mining approach is widely recognized for its effectiveness in understanding user sentiments on social media. The process includes the following steps:

- Case Normalization: All text was converted to lowercase to eliminate discrepancies caused by case sensitivity. For instance, ill and iLl appear to have the same meaning. Therefore, we changed the format to lowercase to avoid confusion.

- Stopword and Noise Removal: Stopwords were removed by implementing the method from the Python NLTK package. Moreover, we remove noise from tweets, including mentions, hashtags, punctuations, square brackets, numbers, URLs, new lines, and retweets.
- Stemming: The stemming technique was applied to reduce words to their root forms. For example, "vaccinations" was stemmed to "vaccine," allowing for more consistent and efficient analysis of related terms.
- Sentiment Categorization: Composite sentiment polarity scores were computed using the TriLex approach, which integrates the strength of three lexicons to calculate composite sentiment scores. This step enables the categorization of text into positive, neutral, or negative sentiments, forming the foundation for subsequent analysis.

Applying these preprocessing steps transformed the dataset into a clean and structured format, facilitating robust sentiment and topic analysis. This structured data is essential for deriving accurate insights and supporting data-driven decision-making.

3.3 Geographical location extraction

We employed geographical location extraction to investigate the spatial distribution of user sentiments. However, extracting accurate location information from user profiles is a challenging task due to the incompleteness and inaccuracy of the location attribute. An alternative approach is to use the tweet's geotagging feature, which provides precise latitude and longitude coordinates. Nevertheless, this feature is often disabled by Twitter users for privacy reasons. To address this challenge, we utilized the GeonamesCache Python package, which provides functions to retrieve names and ISO and FIPS codes of continents, countries, and US states and counties. Additionally, it offers dictionaries for 252 countries, all US states, 3234 US counties, and 25286 cities to map tweet locations accurately.

The mapping process involved two key processes: preparing geographic locations Fig 2 and extracting geographical locations for sentiment Fig 3. Figure 2 illustrates the initialization and organization of geographic location data into structured dictionaries for efficient access and processing. It begins by initializing a GeonamesCache object, which serves as the source for geographic information. Then, we retrieve four geographic data dictionaries for countries, US states, cities, and US counties. This step ensures the data availability for subsequent analyses or operations.

The second process involves extracting location information for each tweet and matching it against the prepared geographic dictionaries. Figure 3 demonstrates steps for extracting spatial information from the tweet. We identify and assign corresponding geographic attributes to the dataset. The goal is to populate new columns in the DataFrame, including the country's name and code, the US state's name and code, the US county, and the city's name.

We observed that the order and case sensitivity of the dictionaries used in the mapping process affect the results. For instance, mapping the location “Bristol, UK” using the order of cities, US counties, US states, and countries resulted in an incorrect mapping of Bristol City in Pennsylvania. In contrast, using the order of Countries, US States, US Counties, and Cities incorrectly mapped to the United Kingdom. Therefore, the procedure begins by iterating over each location entry and performs multiple nested loops to match the location with relevant geographic entities from the provided dictionaries. First, country data is matched, and the country's name and code columns are updated accordingly. Subsequently, US state information is processed to update the state's name and code columns and ensure the country's name and code are marked as "United States" and "US," respectively. Then, the US counties' data is checked, and the US county column is updated while maintaining

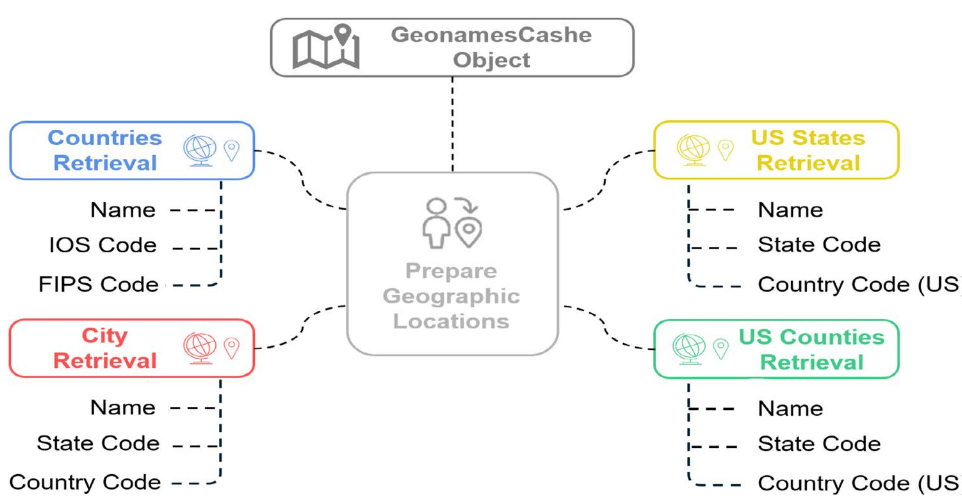

Fig. 2 Overview of the geographic location preparation pipeline using the GeoNamesCache object. The process retrieves location metadata for multiple spatial dictionaries, including countries, U.S. states, counties, and cities, along with relevant geographic codes (e.g., FIPS, ISO, state, and country codes). This hierarchical metadata prepares accurate geotagging dictionaries for the location extraction mapping process

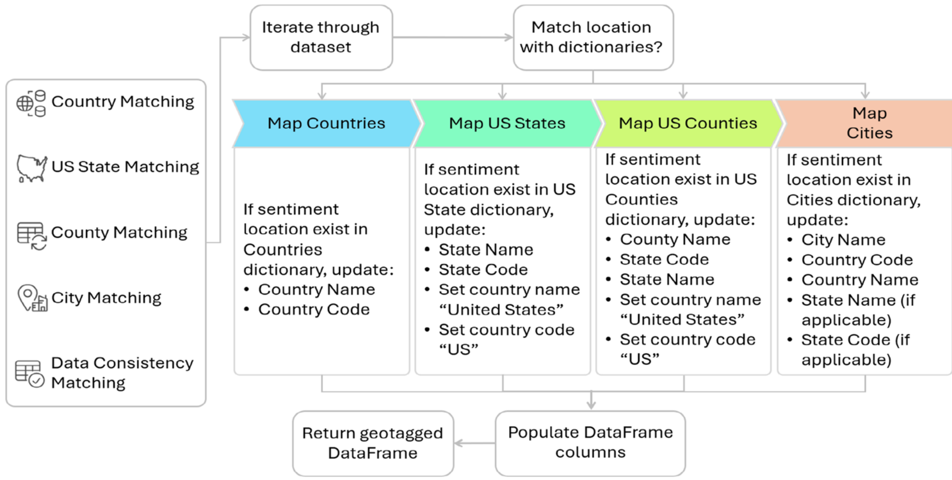

Fig. 3 Workflow for extracting and standardizing geographic locations from X data. The system iteratively matches location strings with dictionaries at four levels (country, U.S. state, county, and city) and updates location fields in the X dataset with consistent names and codes. This process ensures accurate mapping and enables spatio-temporal sentiment analysis across global, national, and regional levels

consistency in state and country-level attributes. Next, the city names assigned to the city column and corresponding state and country codes are updated for consistency. The process ensures that overlapping matches are correctly resolved and the dataset is systematically populated with complete geographic metadata. This process provides detailed location attributes that facilitate the spatiotemporal aspect of sentiment analyses. These processes are beneficial in sentiment analysis workflows, where geographical context is critical for understanding regional variations in sentiment trends.

3.4 Sentiment analysis

This study utilized the Alharbi et al. (2025) unsupervised sentiment analysis approach for short text. This approach helps automate clustering opinions, emotions, and attitudes from short texts such as tweets. The process involves clustering opinions expressed in the text as positive, negative, and neutral.

TriLex leveraged the strength of the three commonly used lexicons, including TextBlob, VADER, and AFINN. It applies the majority vote mechanism among these lexicons to enhance the accuracy of the sentiment labeling. The method identifies strong labels where the tools agree and treats disagreements as weak labels. Moreover, it normalizes sentiment scores across all lexicons and computes a composite sentiment score using a dynamic threshold. The following Eqs (1) and (2) summarize the process of calculating the composite sentiment score:

$$NSS_{t_n^{L_m}} = \frac{SS_n + SS_{\text{Max}}}{SS_{\text{Max}} - SS_{\text{Min}}} \tag{1}$$

$$CSS_i^k = \sum_{m=1}^{3} NSS_{t_n^{L_m}} \times W_{L_m} \tag{2}$$

Table 3 Threshold values for the TriLex sentiment analysis approach

	Sentiment Category		
Lexicon Approach	Positive	Neutral	Negative
TriLex	Score >0.5	0.4 ≤ Score ≤ 0.5	Score < 0.4

To generate a sentiment label, we utilize the standard threshold of the TriLex as shown in Table 3.

3.5 Topic modeling

Integrating BERTopic and ChatGPT proved to play a significant role in enhancing the topic modeling and interpretation without relying on domain experts (Alharbi et al. 2024). Therefore, we employed TriLex and BERTopic on top of ChatGPT to uncover the thematic structures within the discourse on COVID-19 vaccines across tweets and news articles. TriLex categorizes the tone of the sentiment as positive, neutral, or negative. BERTopic leverages transformer-based embeddings denoted as $E = \{e_1, e_2, \ldots, e_n\}$, where e_i represents the embedding of the $i-th$ document, and clustering techniques to generate coherent topics. Given a set of embeddings E, BERTopic applies a clustering algorithm, such as Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), to group embeddings into clusters $C = \{c_1, c_2, \ldots, c_n\}$, where c_j denotes the $j-th$ topic cluster. Each cluster c_j is defined as a subset of embeddings $c_j \subseteq E$, and the centroid μ_j of each cluster is used to represent the semantic center of the topic.

ChatGPT was utilized using the API for GPT-4 for automated summarization and contextualization to enhance the interpretability and contextual understanding of the generated topics. We feed ChatGPT with tweets, sentiment categories based on TriLex, and BERTopic keywords. We ask ChatGPT to generate no more than 20 topics for Twitter discussion and news articles at each interval. The intervals are illustrated in Sect. 3.6 below. For each topic c_j, ChatGPT provided a summary S_j derived from the set of documents D_j associated with c_j. The summary can be expressed as $S_j = \mathrm{Summarize}_{\mathrm{ChatGPT}}(D_j)$, where $\mathrm{Summarize}_{\mathrm{ChatGPT}}$ is the summarization function applied to the documents within a topic cluster.

To enhance interpretability and thematic clarity, we utilized ChatGPT to generate concise and representative topic labels for each group of documents. Specifically, we directed ChatGPT to assign a topic title for each tweet cluster or news article cluster based on the content and the top keywords extracted using BERTopic in each time interval. The prompt was structured as follows:

You are provided with a list of keywords [top_n_keywords] and a set of representative texts [content] representing a topic cluster from a group of documents identified by BERTopic at [Interval] of the COVID-19 pandemic. Please generate a clear and informative title that best summarizes the main theme. Be careful, BERTopic may extract thousands of preliminary topics. Therefore, please generate no more than 20 concise and informative topic titles with interpretation for this [Interval], prioritizing semantic coherence and representation quality.

This prompt format was designed to provide the model with sufficient context for accurate summarization while maintaining a focused and interpretable output. We manually reviewed and validated the generated titles to ensure alignment with the underlying content in Sect. 3.7.

This dual approach facilitated the analysis of spatiotemporal variations in vaccine-related sentiment and public discourse. Integrating TriLex, BERTopic, and ChatGPT derived detailed topic representations that enhanced our understanding of the narratives that shape public opinion across social media and news platforms.

3.6 Analysis intervals

We segmented the study period into six intervals defined by critical events in the vaccine timeline to capture sentiment trends corresponding to key developments. We begin with the early outbreak phase of the pandemic (ending in March 10th, 2020), preceding widespread interventions and marking the initial global response. This is followed by the vaccine development and clinical trials interval (March 11th–November 8th, 2020), capturing research progress, preliminary authorizations, and heightened public anticipation. Next, the vaccine announcements window (November 9th–December 13th, 2020) covers high-profile efficacy disclosures from major manufacturers. Overlapping with portions of the announcements period is the US presidential election phase (October 1st, 2020–January 20th, 2021), which includes intensified political debates around vaccination. Subsequently, a vaccine distribution and access phase (December 14th, 2020 - August 31st, 2021), focusing on logistical and policy challenges as vaccination campaigns ramped up. Finally, the Delta and Omicron variants' emergence (September 1st, 2021–May 31st, 2022) encompasses renewed public discourse regarding booster doses, vaccine efficacy, and evolving public attitudes. By organizing data into these six intervals, we can perform a spatiotemporal sentiment analysis that aligns with significant milestones and contextual shifts in COVID-19 vaccine discussions over time. Table 4 provides an overview of the number of tweets and news articles, and the discussed topics for each interval.

Table 4 Overview of the number of tweets, news articles, and topics corresponding to the six intervals from each platform

	Twitter (X)		News Articles (NYT)	
Interval	# Tweets	# Topics	# Articles	# Topics
Beginning of the pandemic	1, 267	20	13	13
Vaccine Development and Clinical Trials	44, 058	20	822	20
Vaccine Announcements	74, 664	20	280	10
US Presidential Election	144, 734	20	799	20
Vaccine Distributions and Access	532, 919	20	4012	20
Delta and Omicron Variants Emergence	216, 481	20	1454	20

The number of topics generated per interval in Table 4 was not fixed but determined adaptively by both the semantic coherence and density of the data during each interval. We instructed ChatGPT to refine and consolidate the raw outputs from BERTopic, which may initially produce thousands of topic representations into no more than 20 semantically distinct and interpretable topics per interval. This dynamic generation ensures that the extracted topics meaningfully reflect the variations in discourse volume and complexity over time. Notably, there were only 13 articles in the NYT articles dataset at the beginning of the pandemic, which naturally constrained the number of coherent topics.

3.7 Experimental evaluation

To assess the reliability and validity of the proposed approach, we conducted a manual evaluation with two independent researchers serving as human annotators. A random sample of 120 tweets from the six defined periods was selected. Each annotator independently assigned sentiment labels ("Positive", "Neutral", or "Negative") to each tweet.

For the topic modeling evaluation task, the same annotators reviewed 120 randomly selected topic outputs from both datasets (Twitter and NYT articles). Each sample included the original text, extracted keywords, and the generated topic title generated by our approach. Annotators were asked to indicate whether they agreed or disagreed with the semantic coherence and relevance of the assigned topic.

The evaluation results indicate strong agreement between the annotators and the proposed method. Specifically, sentiment labels assigned by annotators matched TriLex outputs with an overall agreement rate of 89%, computed as the proportion of instances (107 out of 120 tweets) where both annotators agreed with the TriLex label or where at least one annotator matched the TriLex label and the other did not contradict it. This finding represents the direct match rate between TriLex predictions and the majority or individual human judgment, not an average of annotator agreement scores.

Annotator 1 agreed with the TriLex model on 81% of the samples, while Annotator 2 showed 76% agreement. For topic modeling, Annotator 1 agreed with the assigned topics for 96% of the samples, and Annotator 2 agreed with 94%. Inter-annotator agreement was 77% for sentiment classification and 90% for topic interpretation.

Table 5 summarizes these evaluation outcomes, highlighting the consistency between human judgments and automated predictions. These findings support the robustness of our approach for both sentiment classification and topic assignment across platforms and data sources. This validation supports the methodological choices described in Sect. 3, demonstrating that our automated techniques align well with human judgment and can be reliably used for large-scale spatiotemporal analysis.

4 Results and discussion

This section presents the results of our hierarchical spatiotemporal analysis of findings derived from Twitter (X) data and the New York Times (NYT) articles. Specifically, sentiment trends are primarily extracted from X data due to the high volume and user-level granularity of the content. Topic modeling is conducted separately on both datasets using

Table 5 Annotator agreement summary

	Twitter (X)		News Articles (NYT)	
	Agree	Disagree	Agree	Disagree
Annotators.vs Proposed approach	107 89%	13 11%	–	–
Annotator 1.vs Proposed approach	97 81%	23 19%	115 96%	5 4%
Annotator 2.vs Proposed approach	91 76%	29 24%	113 94%	7 6%
Annotator 1.vs Annotator 2	92 77%	28 23%	108 90%	12 10%

the integration of BERTopic and the ChatGPT framework, allowing for examination of the influence of cross-platform in shaping public reaction, and direct comparison of discourse patterns across platforms. In cross-platform analyses, such as shared topic detection, temporal alignment, and sentiment correlations, both datasets are integrated. For each subsection, we explicitly specify the data source to ensure clarity and highlight the unique insights contributed by social media and news media content.

4.1 Cross-platform temporal relationship analysis

The rapid spread of information is driven by endogenous factors, where users share content within their social network, and exogenous factors, where information originates from external sources like news media and is introduced into the network (Hua et al. 2016). To investigate the temporal dynamics between formal media and public discourse, we conducted a cross-correlation analysis between sentiment trends derived from the NYT and X platforms.

Our findings reveal two primary patterns:

1) **Sharing Some Common Topics:** In each interval, both datasets exhibit shared themes (e.g., COVID-19 vaccine information, vaccine approval, and registration and availability). Particularly, sentiment expressed in NYT articles often precedes shifts in X sentiment by approximately 1–2 weeks and vice versa, implying a potential directional influence of formal media on public discourse. For instance, the NYT published some articles in mid-October 2020 regarding the temporary suspension of certain vaccines due to blood clot concerns. This topic subsequently gained traction on X approximately a week later, accompanied by a sentiment decline from positive to negative between October 25th and December 10th, 2020. In contrast, we also observed instances where public discourse appeared to influence media coverage. Notably, X users began sharing their personal vaccination experiences in late January 2021, which corresponded with a surge of related NYT articles in February 2021.
2) **Platform-Specific Topics:** Despite areas of sharing, each platform also demonstrated distinct thematic focuses. For example, vaccine lottery and mandatory vaccination were prominent only in X discussions. In contrast, the NYT featured unique topics such as addressing discrimination in vaccine distribution and the impact of vaccination efforts on pharmaceutical stock performance. These differences highlight the complementary roles that social media and traditional news play in shaping and reflecting public discourse.

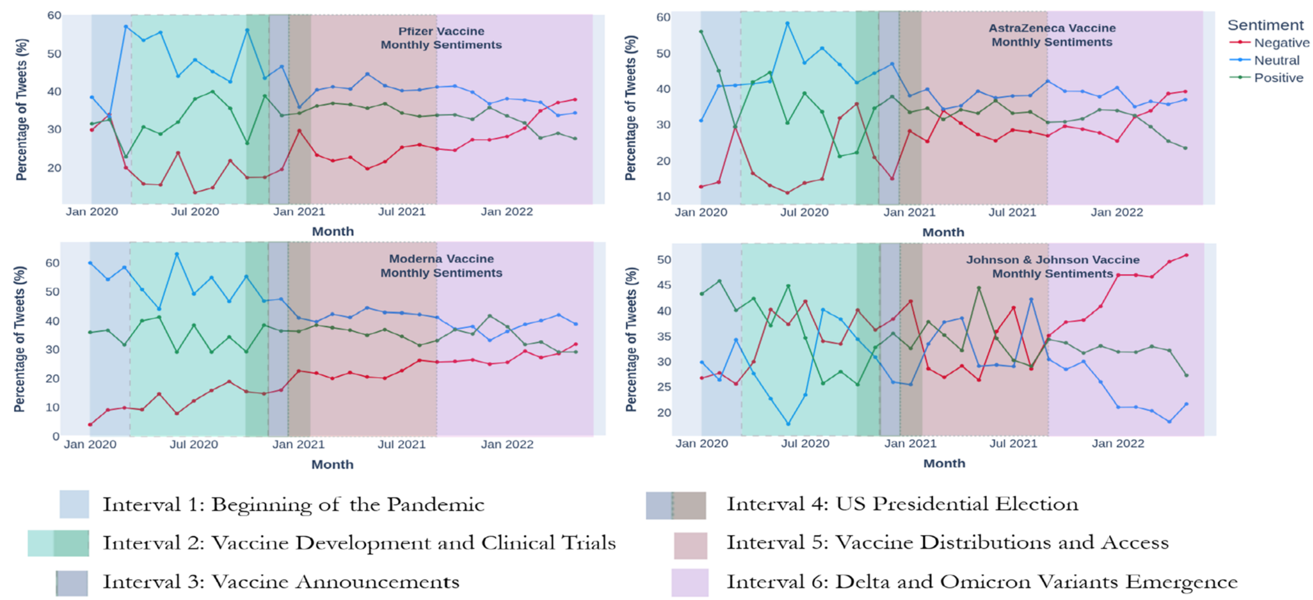

Fig. 4 Temporal sentiment distribution of global Twitter discussions for four major COVID-19 vaccines (Pfizer, AstraZeneca, Moderna, Johnson & Johnson) from January 2020 to May 2022. Sentiment is categorized using the TriLex model as positive (green), neutral (blue), or negative (red). Each subplot corresponds to a vaccine, showing changes in public sentiment across six pandemic intervals, marked by shaded regions. Notable shifts in sentiment align with key events such as vaccine approvals, pauses due to side effects, and the emergence of Delta and Omicron variants

Overall, the temporal and thematic analysis highlights a bidirectional relationship between social media and news media. In other words, both media contribute to framing and influence user-driven discourse.

4.2 Spatiotemporal analysis

We conducted a spatiotemporal analysis using the TriLex and BERTopic with ChatGPT to estimate and compare the global and local public sentiment toward four COVID-19 vaccines. The study was performed on Twitter discourse and news articles from the New York Times between January 1st, 2020, and June 1st, 2022. Current research has primarily focused on analyzing social media platforms or news articles separately, but there is a need for more comprehensive studies that examine the interplay between these sources.

4.2.1 Spatiotemporal sentiment analysis for X data

We consider a large corpus of tweets written in English related to the COVID-19 vaccines to scale down from global to local. Specifically, we focus on hierarchical spatiotemporal analysis encompassing the national (United States) and regional (Pennsylvania and Philadelphia).

We successfully extracted location information for 810, 097 tweets mentioning the four vaccines: Pfizer, Moderna, AstraZeneca, and Johnson & Johnson. First, we investigate the public attitude toward four COVID-19 vaccines over the six intervals as shown in Fig 4.

Table 6 Sentiment shift detection for X sentiment at global level for each of the time periods

Periods	*chi–Square*	p-value	Significant ($p < 0.05$)
Interval 1 vs Interval 2	39.10	0.00	✓
Interval 2 vs Interval 3	380.02	0.00	✓
Interval 3 vs Interval 4	157.74	0.00	✓
Interval 4 vs Interval 5	569.84	0.00	✓
Interval 5 vs Interval 6	462.10	0.00	✓

The analysis of the sentiments expressed in collected tweets revealed that public opinion regarding vaccines changes over time and is influenced by various events, including changes in case and death rates, vaccination announcements and availability, the emergence of COVID-19 variants, and news media.

To substantiate the observed temporal variations in public sentiment, we conducted a series of chi-square tests of independence to examine whether sentiment distributions (positive, neutral, negative) differed significantly across adjacent time intervals, as shown in Table 6. At the global level, all five pairwise comparisons between intervals yielded highly significant results ($p < 0.001$). For example, the sentiment distribution between Interval 1 and Interval 2 resulted in a chi-square value of 39.10 ($p = 0.000$), while a much larger shift was detected between Interval 4 and Interval 5, with a chi-square value of 569.84 ($p = 0.000$), indicating a strong shift in global sentiment during that phase.

To assess the consistency, we repeated this analysis for the top five countries with the highest volume of geolocated

Table 7 Sentiment shift detection for X sentiment at global level, including the top 5 countries: United States (US), United Kingdom (UK), Canada (CA), India (IN), and Australia (AU) for each of the time periods

	Interval 1 vs 2		Interval 2 vs 3		Interval 3 vs 4		Interval 4 vs 5		Interval 5 vs 6	
	chi^2	p-value	chi^2	p-value	chi^2	p-value	chi^2	p-value	chi^2	p-value
US	43.47	0.00	242.38	0.00	41.97	0.00	91.90	0.00	491.01	0.00
UK	7.73	0.02	154.91	0.00	46.85	0.00	31.55	0.00	12.58	0.00
CA	6.53	0.04	72.02	0.00	24.58	0.00	47.93	0.00	19.89	0.00
IN	21.34	0.00	73.63	0.00	17.41	0.00	192.80	0.00	50.12	0.00
AU	3.72	0.16	3.18	0.20	5.88	0.05	2.25	0.32	36.56	0.00

Table 8 Sentiment shift detection for X sentiment at national level (United States), including the top 10 states: California, New York, Texas, Florida, District of Columbia (D.C), Indiana, Massachusetts, Illinois, Washington, Pennsylvania for each of the time periods

US State	Interval 1 vs 2	Interval 2 vs 3	Interval 3 vs 4	Interval 4 vs 5	Interval 5 vs 6
California	✔	✔	✔	✔	✔
New York	✔	✔	✔	✔	✔
Texas	✔	✖	✔	✔	✔
Florida	✔	✔	✔	✔	✔
D.C	✔	✔	✔	✔	✔
Indiana	✔	✔	✔	✔	✔
Massachusetts	✔	✖	✖	✔	✔
Illinois	✔	✖	✔	✔	✔
Washington	✔	✔	✔	✔	✔
Pennsylvania	✔	✔	✔	✔	✔

tweets: the United States (US), the United Kingdom (UK), Canada (CA), India (IN), and Australia (AU). Table 7 demonstrates that sentiment shifts were statistically significant ($p < 0.05$) in most comparisons. Notably, the United States exhibited highly substantial changes across all intervals, including Interval 2 vs. 3 with a chi-square value of 242.38 ($p < 0.001$) and Interval 5 vs. Interval 6 with a chi-square value of 491.01 ($p < 0.001$). In contrast, Australia displayed weaker and less consistent shifts, with non-significant results in Interval 1 vs. Interval 2 with a chi-square of 3.72 ($p = 0.16$) and Interval 2 vs. Interval 3 chi-square of 3.18 ($p = 0.20$), reflecting more stable sentiment patterns.

At the national level, we analyzed sentiment changes across ten U.S. states with the highest tweet volumes: California, New York, Texas, Florida, the District of Columbia (D.C.), Indiana, Massachusetts, Illinois, Washington, and Pennsylvania. As shown in Table 8, the majority of these states demonstrated consistent and statistically significant sentiment shifts across all five transitions. For example, California showed significant changes in all intervals ($p < 0.05$), underscoring the dynamic nature of public discourse in response to evolving pandemic conditions. Meanwhile, states such as Massachusetts and Texas exhibited non-significant shifts in specific intervals, suggesting spatial variability in public reaction.

These statistical results provide robust support for our temporal analysis and confirm that the observed sentiment trends are not simply descriptive fluctuations but rather reflect significant changes in public opinion during critical stages of the COVID-19 vaccine rollout.

4.2.2 Temporal interplay between news media and Twitter discourse

Since news media is one of the crucial sources of information, we collected 6, 581 news articles from the NYT outlet to explore the following research question:

- **RQ1:** Could news media shape public discourse about COVID-19 vaccines?

News articles often serve as primary sources of information that are then disseminated and discussed on social media platforms (Malik et al. 2024; Skafle et al. 2022). In this study, we consider the temporal aspect to examine the time between the publication of NYT articles and the corresponding discussions on X. We also compare the thematic coverage of both sources using topic modeling enhanced by BERTopic and ChatGPT to investigate the influence of news articles on social media discussions. Figure 5 presents Venn diagrams that compare topics discussed in NYT articles and vaccine-related tweets across six pandemic intervals, as defined in Sect. 3.6. The diagrams reveal both shared themes, such as vaccine efficacy, side effects, and approval, and platform-specific focuses. For instance, "vaccine lottery" appeared only in tweets, while "addressing discrimination in vaccine distribution" was uniquely emphasized in NYT coverage. This comparison highlights similarities and differences in thematic focus between public social media discourse and formal news media.

The interplay between social media and news articles demonstrates that Twitter discourse was dominated by neutral sentiment, reflecting widespread uncertainty and limited information about vaccines at the beginning of the pandemic. Positive and negative feelings were minimal, as discussions primarily centered on the outbreak's initial impacts and lockdown. In contrast, NYT articles addressed broader societal impacts like economic consequences and

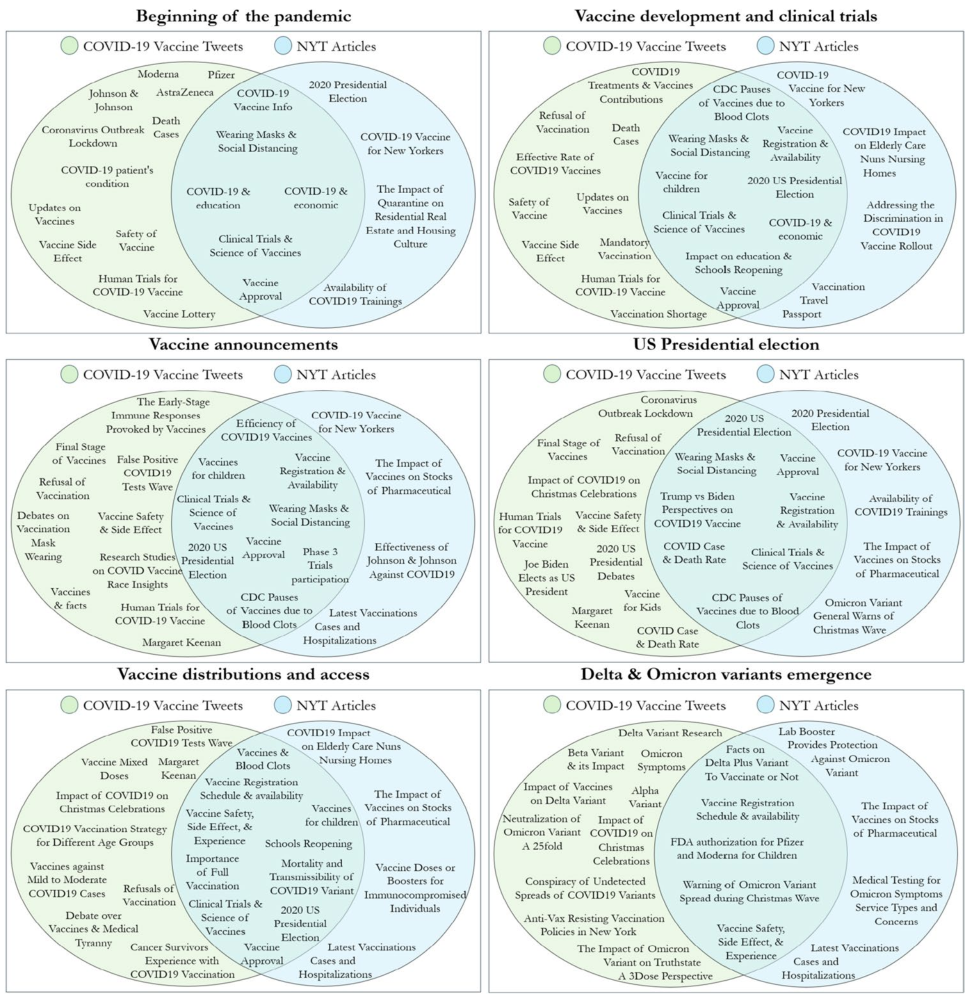

Fig. 5 Topic modeling comparison between COVID-19 vaccine-related tweets and NYT articles across six pandemic intervals. Each Venn diagram presents shared and platform-specific topics extracted using BERTopic and ChatGPT, revealing both shared topics and distinct topics

vaccine science, with shared themes such as death rates and education.

As vaccine development progressed, Twitter sentiment shifted toward optimism due to encouraging news about vaccine progress and registration by the NYT. However, negative sentiment also increased, likely driven by concerns over vaccine safety, efficacy, and the ethical implications of accelerated timelines in Twitter. Despite these changes, a significant proportion of neutral sentiment persisted, indicating cautious public optimism.

In the vaccine announcements phase, NYT articles often contextualized the vaccine rollout policy frameworks, while Twitter reactions focused more on personal experiences and trust issues, with slightly higher positive reactions. However, the NYT coverage of vaccine pauses due to blood clot concerns, followed by tweets one to two weeks later, expressed increased skepticism, with sentiment turning more negative.

The US presidential election interval was marked by heightened polarization in sentiment. Positive sentiment declined slightly, while negative sentiment grew. This finding reflects politically charged news around vaccine distribution plans and the politicization of the pandemic through Twitter discussions and NYT posts. This period underscores the influence of political dynamics on public trust in vaccines.

Finally, positive sentiment sharply declined in the Delta and Omicron variants' emergence phase on Twitter, while negative sentiment surged, particularly for Johnson & Johnson. This shift reflects growing concerns about vaccine efficacy against variants, breakthrough infections, and evolving public frustration on both platforms.

Overall, this cross-platform analysis supports the hypothesis posed in RQ1. The temporal and thematic alignment between NYT coverage and Twitter discussions suggests that news media can shape and amplify public discourse on social media. While Twitter captures real-time and emotional reactions, NYT articles offer a more structured and analytical narrative. The alignment and divergence between the two sources provide valuable insights into how public sentiment, media framing, and real-world events interact to shape the public reaction over time.

Table 9 Overall COVID-19 vaccine sentiments distribution in the United States for four vaccines

COVID-19	Sentiment Category			
Vaccine	Total	Positive	Neutral	Negative
Pfizer-BioNTech	153, 271	37%	50%	13%
Moderna	81, 982	36%	50%	14%
AstraZeneca	78, 417	30%	43%	27%
Johnson & Johnson	42, 773	29%	37%	34%

4.2.3 United States

The United States was the leading contributor of tweets in our dataset, accounting for approximately 44% of the total COVID-19 vaccine-related discussion. Moreover, it is the most affected country by the COVID-19 pandemic. Therefore, we conducted a more detailed analysis of public opinion within the United States compared to the global context. In addition, we explored the similarities and differences in reactions towards COVID-19 vaccines across states in the United States. To accomplish this, we applied a geolocation filter to retrieve tweets originating from the United States, resulting in a dataset of 356, 443 tweets mentioning the four vaccines. The distribution of these tweets across the United States is presented in Table 9, and the associated sentiment distributions are visualized in Fig 6. By examining the sentiment of tweets within each state, we can effectively address our third research question:

- **RQ2:** Do people in different states of the USA react similarly or differently to various COVID-19 vaccines?

The US Food and Drug Administration (FDA) has authorized Pfizer-BioNTech, Moderna, and Johnson & Johnson vaccines for emergency use against COVID-19. As a result, they have been extensively distributed across the United States. However, the availability of specific vaccines varied by state, which should be considered when analyzing state-level sentiment. The variety of distribution was affected by many factors, such as local vaccination policies and vaccine supplies. Consequently, our analysis specifically focused on tweets referencing these three vaccines. Our findings reveal a convergence between public opinions in the United States and global sentiments. Nonetheless, we found that reactions towards Pfizer-BioNTech, Moderna, and Johnson & Johnson vaccines differed from one state to another. The geographical representation of the United States, as depicted in Fig 6, illustrates varying levels of engagement in discussions about the three vaccines.

We found that users from California produced most discussions on Twitter about Pfizer-BioNTech with a total of 12, 486 posts. Most California posts were neutral, 38%, followed by positive, 37%, and 25% were negative. Users from Wyoming, Colorado, Connecticut, Kentucky, New Jersey, Pennsylvania, Maryland, Ohio, Alabama, West Virginia, New York, and South Carolina had more positive attitudes toward Pfizer-BioNTech, with more than 40% positive. In contrast, the highest negative sentiments regarding this vaccine were retrieved from Mississippi, Rhode Island, and Montana, with higher than 30% negatives. The feelings of posts related to the Pfizer-BioNTech vaccine from Wyoming

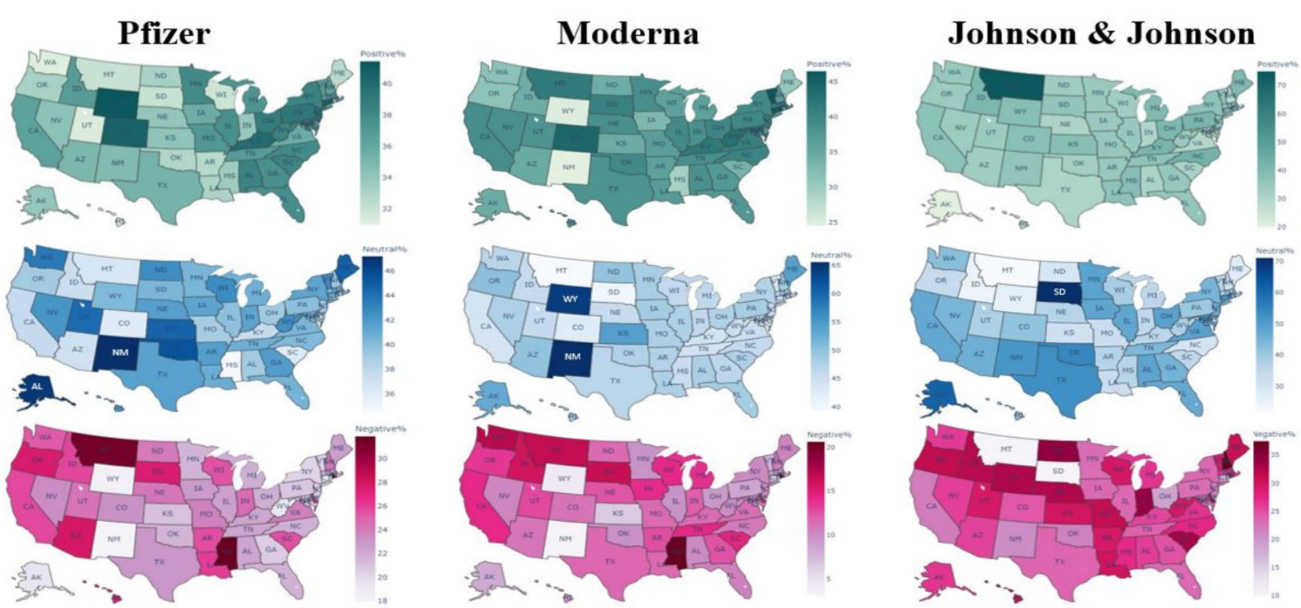

Fig. 6 Sentiment distribution across the United States for three major COVID-19 vaccines: Pfizer, Moderna, and Johnson & Johnson on the X data. Each row represents a sentiment category: positive (green), neutral (blue), and negative (red). Choropleth maps illustrate the proportion of sentiment-labeled tweets per state, revealing distinct geographic patterns in public perception

were the most optimistic (42%), but a very small number of tweets were georeferenced to that state (only 82).

California and New York contributed the most to Moderna-related discussions, with 7, 933 and 7, 034 tweets respectively. In California, the sentiment was 39% positive, 37% neutral, and 24% negative, while New York's sentiment was 40% positive, 42% neutral, and 18% negative. Across 18 states, including Pennsylvania, Illinois, and Virginia, Moderna received over 40% positive sentiment, which is higher than both Pfizer and Johnson & Johnson in those regions.

For the Johnson & Johnson vaccine, New York and California users produced the most discussion (2, 571 and 2, 234 tweets, respectively). We found that 38% of posts in both states were positive, while neutral posts were 37% and 39%, and negative posts were 25% and 23%, respectively. Only posts from Montana were highly satisfied with this vaccine, with 75% positives, which is the highest satisfaction among all states and vaccines. Posts toward the Johnson & Johnson vaccine from other states except Montana and South Dakota were predominantly negative and dissatisfied. Negative sentiment was specifically present in New Hampshire, North Dakota, Indiana, Delaware, and South Carolina posts.

In summary, Twitter-based sentiment across the United States suggests that users from the East Coast of the United States are predominantly optimistic and have more positive discussions about Pfizer and Moderna. In contrast, the West Coast users have more negative discussions. On the other hand, Johnson & Johnson received more negative sentiment in most US states except Montana and South Dakota. However, the state-level analysis reveals more localized patterns of sentiment driven by political, demographic, and logistical factors unique to each state, underscoring the importance of tailoring public health strategies to regional contexts.

4.2.4 Pennsylvania

To enhance the depth of our analysis and assess the utility of localized sentiment trends, we selected Pennsylvania and its largest city, Philadelphia, as a case study for localized spatiotemporal sentiment analysis using X data. This zoomed-in approach enables us to investigate whether more granular geographic resolution yields deeper insights that may be ambiguous at broader scales. Pennsylvania was chosen due to its strategic significance as a politically diverse swing state, its elevated COVID-19 case and vaccination rates, and its representativeness of broader national dynamics. Within Pennsylvania, Philadelphia offers a uniquely valuable case study given its demographic heterogeneity, encompassing a wide range of racial, ethnic, social, and political backgrounds. Furthermore, Philadelphia is one of five U.S. cities that received vaccines directly from the federal government, providing a distinctive policy context for examining public response. The availability of demographic and ZIP-code-level data in Philadelphia supports robust spatial analysis, allowing us to capture nuanced variations in public sentiment across sub-regional populations.

Pennsylvania is ranked among the top 10 U.S. states severely affected by the virus. Consequently, the Pennsylvania government has administered over 24.5 million doses of COVID-19 vaccines from Pfizer-BioNTech, Moderna, and Johnson & Johnson. However, only approximately 69% of the vaccine supply has been utilized. Therefore, we analyzed the sentiments of the tweets from Pennsylvania and linked them to the physical events to understand the reasons behind individuals' acceptance or refusal of COVID-19 vaccinations by exploring the following question:

- **RQ3:** Why do people refuse or accept COVID-19 vaccines in Pennsylvania and Philadelphia?

Information on the physical events related to COVID-19 vaccination in Pennsylvania and Philadelphia was collected from multiple sources, including Google Analytics, Bloomberg ,[1] and the Commonwealth of Pennsylvania Dataset from the Pennsylvania Department of Health .[2] We believe that integrating physical events with sentiment analysis may provide valuable insights in capturing the reasons behind the acceptance or rejection of COVID-19 vaccines.

Since the Pennsylvania government administered the three vaccines mentioned in our Twitter dataset, we analyzed posts from Pennsylvania that mentioned Pfizer-BioNTech, Moderna, and Johnson & Johnson. Pennsylvania users generated a total of 17, 558 tweets related to the administered vaccines. Among these tweets, Pfizer-BioNTech accounted for the majority at 53%, Moderna at 33%, and Johnson & Johnson at 14%. Fig 7 illustrates the monthly sentiment trends for Pfizer, Moderna, and Johnson & Johnson vaccines in Pennsylvania, segmented into the same six intervals defined earlier. Across all vaccines, the pandemic's beginning is dominated by neutral sentiment due to limited information. Positive sentiment peaks during the vaccine announcements interval for Pfizer and Moderna, reflecting strong public optimism. However, Johnson & Johnson struggles to achieve similar levels of positive sentiment. Sentiment polarization is observed during the US presidential election interval, highlighting the role of political discourse in shaping public opinion. The Delta and Omicron variants' emergence interval shows the most significant

[1] https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/

[2] https://data.pa.gov/Covid-19/COVID-19-Vaccinations-by-Day-by-County-of-Residenc/bicw-3gwi

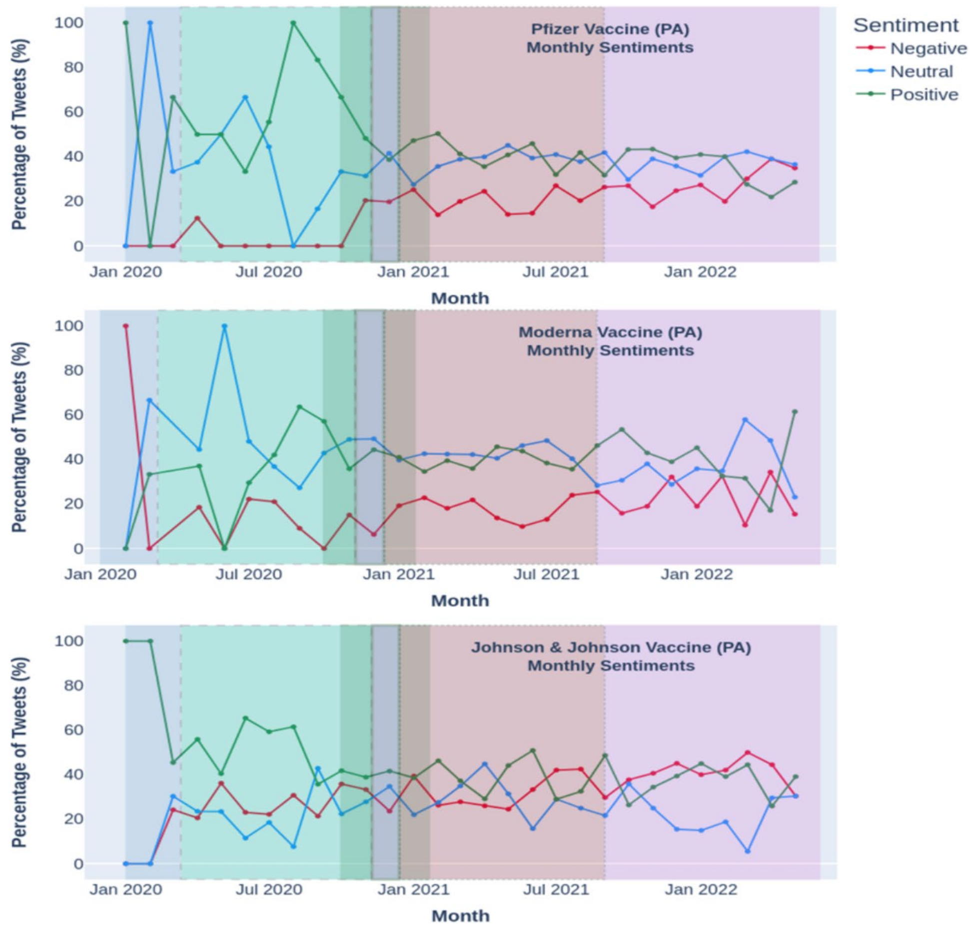

Fig. 7 Temporal sentiment trends toward three major COVID-19 vaccines: Pfizer, Moderna, and Johnson & Johnson based on X discussions geolocated in Pennsylvania. Sentiment categories (positive in green, neutral in blue, and negative in red) are shown monthly from January 2020 to May 2022. The timeline is divided into six pandemic intervals, reflecting critical milestones such as clinical trials, vaccine approvals, and variant surges

negative sentiment for all vaccines, underscoring public concerns about vaccine efficacy against variants.

Findings on Pennsylvania align closely with the sentiment trends and thematic narratives observed globally and nationally. Moreover, it adds depth to the previous results and shows that Pennsylvania follows similar global and national sentiment trends. Local dynamics, such as vaccine distribution challenges and region-specific political discourse, may expand or moderate these trends. Therefore, we analyzed real-time Pennsylvania's vaccination rates, COVID-19 cases, and death rates from Google Analytics. By correlating this information with the sentiment analysis of tweets shown in Fig. 7, we observed that sentiments are affected by multiple factors, including approval, availability, booster shots of vaccines, the appearance of COVID-19 variants, and cases and death rates.

To study our question comprehensively, we narrowed our analysis to Philadelphia, the largest city in Pennsylvania. The subsection below describes data from Philadelphia.

4.2.5 Philadelphia

Philadelphia is the most populous city in Pennsylvania and was severely affected by the COVID-19 pandemic. The city's vaccination effort differed from the state of Pennsylvania's vaccine program since Philadelphia is one of the five US cities getting vaccines directly from the federal government. According to the City of Philadelphia Government vaccine dataset, a total of 3, 331, 184 doses of COVID-19 vaccines were distributed, with more than 63% Pfizer-BioNTech, 35% Moderna, and 2% Johnson & Johnson.

To obtain a more comprehensive understanding of the sentiments posted by Philadelphia's residents on Twitter regarding the vaccines, we collected and labeled tweets from all zip codes within Philadelphia that mentioned any of the vaccines. We divided the zip codes within Philadelphia into four regions: East, West, South, and North. Due to the lack of tweets associated with specific zip codes, we converted zip codes into latitude and longitude coordinates utilizing the FreeMapTools website.[3] The new dataset was collected based on each zip code's latitude and longitude coordinates.

[3] https://www.freemaptools.com/convert-us-zip-code-to-lat-lng.htm

Consequently, we obtained 3, 347 tweets distributed as follows: 1, 675 related to Pfizer-BioNTech, 1, 141 mentioned Moderna, and 531 about Johnson & Johnson.

Our study revealed that users from Philadelphia produced 41% neutral sentiments, 33% positive, and 26% negative sentiments about the three vaccines. Sentiment from Philadelphia aligned with the findings of this study in the national (United States) and regional (Pennsylvania) sentiments, but with more positive reactions toward all vaccines. Fig 8 presents the monthly sentiment trends and provides a localized view of public opinion dynamics in Philadelphia. The Twitter sentiments produced by Philadelphia users illustrate a more neutral sentiment at the pandemic's beginning, reflecting limited public information. Positive sentiment peaks during the vaccine announcement interval for Pfizer and Moderna, indicating strong public confidence in their efficacy announcements, while Johnson & Johnson shows only a modest rise. The US presidential election phase introduces polarization, with increased negative sentiment for all vaccines due to politically charged discussions. In the vaccine distribution and access phase, negative sentiment rises steadily, while positive sentiment declines. The sentiments shift is driven by distribution inequities, logistical issues, and reports of side effects published by the NYT. Finally, the Delta and Omicron variants' emergence interval shows a significant surge in negative sentiment for all vaccines. This finding highlights widespread concerns about vaccine efficacy against emerging variants. Moreover, individuals from East and West Philadelphia produced higher positive sentiments. In contrast, people from North Philadelphia have more negative sentiments than people from other regions, as shown in Fig 9. These findings emphasize the significant influence of local dynamics and external events, such as demographic composition and urban-specific challenges, in shaping vaccine sentiment in Philadelphia.

These observations on Twitter's sentiments have prompted the formulation of the following research questions:

a) What factors affect the acceptance or rejection of COVID-19 vaccines in Philadelphia?
b) Does social level or race affect the attitude toward the COVID-19 vaccine in Philadelphia?

A comprehensive analysis utilized data from multiple sources to address the above research questions. We obtained vaccination and demographic information from each zip code within Philadelphia by referring to authoritative sources

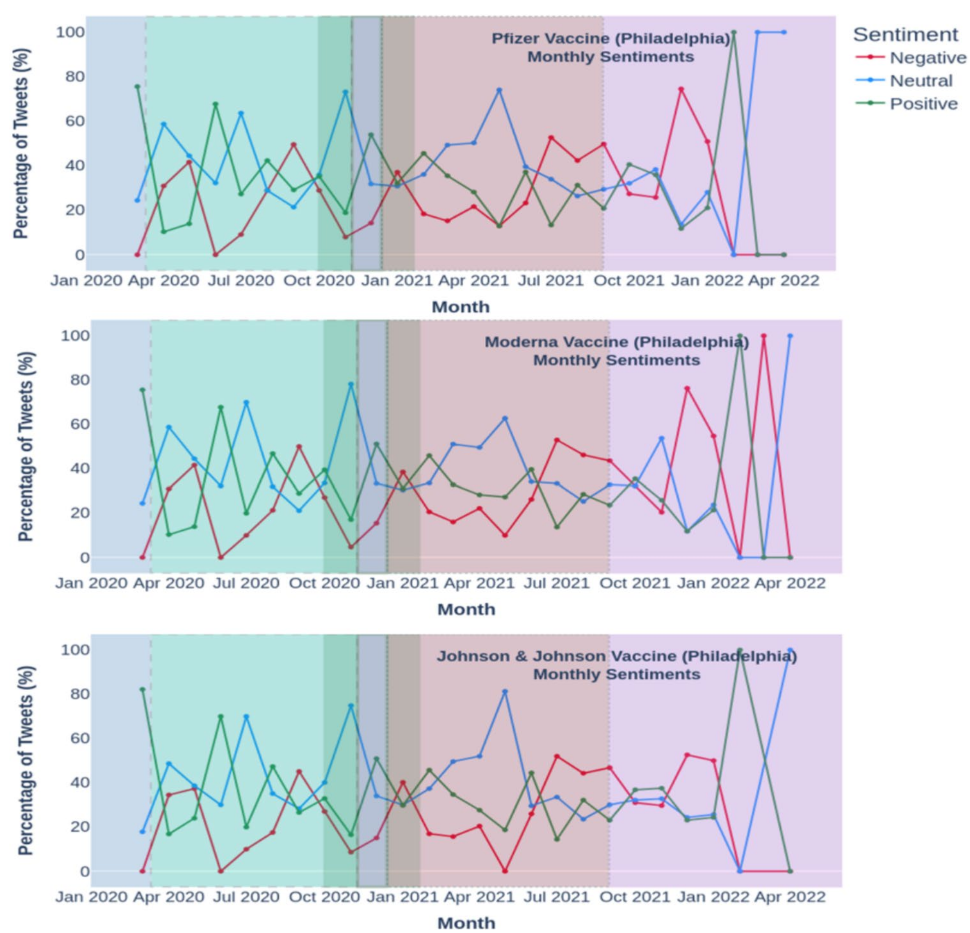

Fig. 8 Temporal sentiment dynamics toward Pfizer, Moderna, and Johnson & Johnson COVID-19 vaccines in Philadelphia based on geolocated X data. Monthly sentiment percentages positive (green), neutral (blue), and negative (red) are tracked across six key pandemic intervals

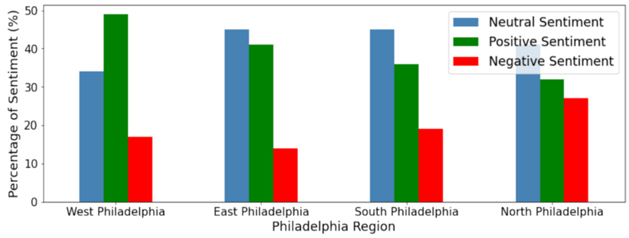

Fig. 9 Regional sentiment distribution toward COVID-19 vaccines across Philadelphia's four major areas: West, East, South, and North. The bar chart illustrates the relative proportions of positive (green), neutral (blue), and negative (red) sentiments expressed on X. Notably, North Philadelphia exhibits the highest proportion of negative sentiment, while West Philadelphia demonstrates a stronger leaning toward positive sentiment

Table 10 Philadelphia Vaccination Status and Demography

	% of Fully Vaccinated	Average Income	% of White American	% of African American
East Philadelphia	86%	69, 698	74%	9%
South Philadelphia	80%	44, 410	70%	30%
North Philadelphia	63%	33, 737	30%	55%
West Philadelphia	61%	28, 626	16%	76%

such as the City of Philadelphia Government and the United States Zip Codes Organization. This information was then appended to the respective regions of Philadelphia: East, North, South, and West, as outlined in Table 10. We gained valuable insights by correlating the patterns derived from these data sources with our sentiment analysis results.

Our analysis further correlated these sentiment patterns with data from Google Analytics and the Philadelphia government. The findings reveal a corresponding increase in the death rate between December 2020 and January 2021. Therefore, the public sentiment in Philadelphia reflects optimism during the vaccine announcement phase. Moreover, the interval between March and April 2021 witnessed an increase in the number of individuals who had been fully vaccinated. Consequently, this period in Philadelphia saw a rise in neutral and negative sentiments alongside a decline in positive sentiment. It reflects the transition from early optimism to a more polarized and cautious public response as vaccines were rolled out. Despite the shift in the sentiments, the data analysis from Pennsylvania and Philadelphia revealed a noteworthy trend: When the number of COVID-19 cases and the death rate increased, individuals tended to seek vaccination, including second and booster doses.

Additionally, our findings indicate that more residents in the East and South regions of Philadelphia are fully vaccinated compared to the other areas of Philadelphia, with percentages of 86% vs 80%, respectively. Furthermore, the East and South Philadelphia displayed a lower prevalence of negative sentiment towards COVID-19 vaccines vs the rest of the city. According to the demographic information, these two regions have the highest average annual income and a larger proportion of white people. Consequently, it appears that individuals belonging to the white demographic group and those with higher incomes are more likely to receive complete vaccination and take booster doses compared to individuals from other racial backgrounds and lower incomes.

5 Limitations and future work

Despite the valuable insights gained through our comprehensive framework for spatiotemporal sentiment and topic analysis across social and news media platforms, several limitations should be acknowledged, which also suggest promising directions for future research.

- This study relies exclusively on articles from The New York Times (NYT) as the representative source for news media. Although NYT provides structured, high-quality content, it also reflects a specific editorial and political stance, which may introduce media bias. Future research could address this limitation by incorporating news outlets across the political spectrum (left, right, and center) to explore the influence of media bias on public perception. This would provide a more balanced, comprehensive, and representative view of the evolution of public discourse and sentiment.
- Our sentiment analysis primarily leverages geographic metadata extracted from user profiles on X, without incorporating additional demographic attributes such as age, gender, or occupation. Including these demographic attributes could enrich the analysis with demographic dimensions and offer deeper insight into sentiment variations among different population segments, as demonstrated by Priyam et al. (2024). However, inference of demographic information from social media presents challenges in data availability, accuracy, and privacy, and thus warrants cautious and ethical implementation in future extensions.
- The current work focuses solely on English tweets as the social media source. While Twitter offers rich, time-sensitive, and location-tagged data, it represents only one mode of online discourse. Other platforms, such as Reddit, Facebook, and YouTube, differ in terms of user demographics, content formats, and engagement mechanisms. Cross-platform analyses (Ruan et al. 2022; Ruan and Lv 2023) have shown the potential for uncovering complementary or diverse public perceptions across media ecosystems. Additionally, the focus on English tweets may not represent the views of the entire population. Future research could build on our

framework by comparing sentiment and topic dynamics across multiple languages and platforms to capture a broader social view.
- In this study, sentiment analysis and topic modeling were performed as separate stages to maintain clarity and interpretability. However, integrating sentiment and topics by assigning sentiment scores with specific themes or using joint models could provide valuable insights into public attitudes across different themes. Ruan et al. (Ruan and Lv 2022) demonstrated how such integrated models can uncover evolving sentiment trends tied to individual topics, offering a more fine-grained understanding of public discourse. Future work could adopt joint spatiotemporal sentiment-topic modeling approaches to capture the co-evolution of sentiment and topics over space and time.
- While we uncover consistent correlations between news outlets and social media sentiment, we acknowledge that the question of causality remains open for future research.

6 Conclusion

This study provides a comprehensive hierarchical spatiotemporal analysis of public sentiment and discourse on COVID-19 vaccines across global, national, and regional levels. By leveraging over 7 million tweets, 6500 news articles, and demographic data, our findings reveal how news articles shaped and reflected public opinion on platforms like Twitter. Furthermore, the study highlights the shift in sentiment driven by key events such as vaccine announcements, political developments, the emergence of variants, and demographic composition. Moreover, our findings revealed that Twitter users have quite positive attitudes toward Pfizer-BioNTech and Moderna vaccines, contrasted with a predominantly negative view of AstraZeneca and Johnson & Johnson. Notably, positive attitudes were higher among East Coast users in the United States, while adverse reactions were more pronounced on the West Coast. Additionally, positive sentiment and vaccine uptake rate correlated with increases in COVID-19 cases and deaths, especially within communities with higher incomes and a larger white population. This research underlines the efficiency of spatiotemporal sentiment analysis as a cost-effective approach and provides a starting point for multi-level cross-platform sentiment and topic analysis to understand public opinion and provides valuable insights.

Acknowledgements The research was sponsored by the DEVCOM Analysis Center and was accomplished under Cooperative Agreement Number W911NF-22-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the DEVCOM Analysis Center or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Author contributions Author contributions Conceptualization, A.A., R.A., S.G., H.O., J.A., and Z.O.; Methodology, A.A., R.A., S.G., and Z.O.; Formal Analysis, A.A., R.A., S.G., and Z.O.; Investigation, A.A. and R.A.; Data Curation, A.A.; Writing – Original Draft Preparation, A.A. and R.A.; Writing – Review & Editing, S.G., H.O., J.A., and Z.O.; Supervision, Z.O. All authors have read and agreed to the published version of the manuscript.

Data availability The data used in this study were collected from Twitter (X) in accordance with the platform's terms of service and developer policies. Due to Twitter's (X) data sharing restrictions, we are unable to share the raw tweet content publicly. However, we can make available the corresponding tweet IDs and relevant metadata (e.g., timestamps, user location, sentiment labels, and topic annotations) upon request.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

Abadah M, Keikhosrokiani P, Zhao X (2023) Analytics of public reactions to the COVID-19 vaccine on Twitter using sentiment analysis and topic modelling. Handbook of research on applied artificial intelligence and robotics for government processes. pp 156-188

Ahmad H, Azhar A, Algosaibi A, Hussain J (2024) Understanding fluctuations in public opinion toward COVID-19 vaccines: insights from social media analysis. SAGE Open 14:21582440241296964

Alharbi A, Hai A, Aljurbua R, Obradovic Z (2024) AI-driven sentiment trend analysis: enhancing topic modeling interpretation with ChatGPT. IFIP international conference on artificial intelligence applications and innovations. pp 3-17

Alharbi A, Aljurbua R, Gupta S, Obradovic Z (2025) Trilex: a fusion approach for unsupervised sentiment analysis of short texts. PLoS ONE 20:e0317100

Ali G, Rahman M, Hossain M, Rahman M, Paul K, Thill J, Samuel J (2021) Public perceptions of COVID-19 vaccines: policy implications from us spatiotemporal sentiment analytics. Healthcare 9:1110

Alipour S, Galeazzi A, Sangiorgio E, Avalle M, Bojic L, Cinelli M, Quattrociocchi W (2024) Cross-platform social dynamics: an analysis of chatgpt and COVID-19 vaccine conversations. Sci Rep 14:2789

Aljurbua R, Alshehri J, Alharbi A, Power W, Obradovic Z (2023) Social media sensors for weather-caused outage prediction based on spatio-temporal multiplex network representation. IEEE Access 11:125883–125896

Aljurbua R, Alshehri J, Gupta S, Alharbi A, Obradovic Z (2024) Early prediction of power outage duration through hierarchical spatio-temporal multiplex networks. International conference on complex networks and their applications. pp 320-334

Aljurbua R, Alshehri J, Gupta S, Alharbi A, Obradovic Z (2025) Leveraging multi-modal data for early prediction of severity in forced transmission outages with hierarchical spatiotemporal multiplex networks. PLoS ONE 20:e0326752

Bai S, Lee E (2024) Examining media's coverage of covid-19 vaccines and social media sentiments on vaccine manufacturers' stock prices. Front Public Health 12:1411345

Baj-Rogowska A (2021) Mapping of the COVID-19 vaccine uptake determinants from mining twitter data. IEEE Access 9:134929–134944

Blane J, Ng L, Carley K (2023) Analyzing social-cyber maneuvers for spreading COVID-19 pro-and anti-vaccine information. Vaccine communication online: counteracting misinformation, rumors and lies. pp 57-80

Challapalli S (2024) Sentiment analysis of the twitter dataset for the prediction of sentiments. J Sens IoT Health Sci (JSIHS ISSN 2584-2560) 2:1–15

Du J, Xu J, Song H et al (2017) Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. J Biomed Semant 8:9

Ebeling R, Sáenz C, Nobre J, Becker K (2022) Analysis of the influence of political polarization in the vaccination stance: the Brazilian COVID-19 scenario. Proc Int AAAI Conf Web Soc Med 16:159–170

Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, Liu R, Ly K, Kacker V, She B et al (2021) Revealing public opinion towards COVID-19 vaccines with twitter data in the United States: spatio-temporal perspective. J Med Internet Res 23:e30854

Hua T, Ning Y, Chen F, Lu C, Ramakrishnan N (2016) Topical analysis of interactions between news and social media. *Proceedings of the AAAI conference on artificial intelligence*. **30**

Hwang J, Su M, Jiang X, Lian R, Tveleneva A, Shah D (2022) Vaccine discourse during the onset of the COVID-19 pandemic: topical structure and source patterns informing efforts to combat vaccine hesitancy. PLoS ONE 17:e0271394

Ida R, Kinasih S, Febriyanti S, Puspa R, Saud M, Bakar M (2024) Covid-19 vaccine rumour and resistance in Indonesia: analysing vaccine hesitancy and cultural myths circulated on online platforms. Vacunas (English Edition) 25:313–322

Jiang B, Cheng L, Tan Z, Guo R, Liu H (2024) Media bias matters: Understanding the impact of politically biased news on vaccine attitudes in social media. 2024 IEEE 11th international conference on data science and advanced analytics (DSAA). pp 1-10

Kalogeropoulos A, Negredo S, Picone I, Nielsen R (2017) Who shares and comments on news?: A cross-national comparative analysis of online and social media participation. Soc Med Soc 3:2056305117735754

Kumar I, Chauhan R, Negi A, Singh T, Mohd N (2023) Tweets sentiment analysis using machine learning and modified BERT methods. 2023 International conference on sustainable emerging innovations in engineering and technology (ICSEIET). pp 395-400

Lenti J, Mejova Y, Kalimeri K, Panisson A, Paolotti D, Tizzani M, Starnini M (2023) Global misinformation spillovers in the vaccination debate before and during the COVID-19 pandemic: multilingual twitter study. JMIR Infodemiol 3:e44714

Lepird C, Ng L, Carley K (2024) Non-credibility scores: relative ranking of news sites shared on social media to identify new pink slime sites. *First Monday*

Li T, Zeng Z, Sun J, Sun S (2022) Using data mining technology to analyse the spatiotemporal public opinion of covid-19 vaccine on social media. Electron Libr 40:435–452

Liu S, Liu J (2021) Public attitudes toward COVID-19 vaccines on English-language twitter: a sentiment analysis. Vaccine 39(39):5499–5505

Lyu H, Wang J, Wu W, Duong V, Zhang X, Dye T, Luo J (2022) Social media study of public opinions on potential covid-19 vaccines: informing dissent, disparities, and dissemination. Intell Med 2:1–12

Malik M, Bauer-Maison N, Guarna G, D'Souza R (2024) Social media misinformation about pregnancy and COVID-19 vaccines: a systematic review. Med Princ Pract 33:232–241

Manisha R, Acharya K (2023) The impact of artificial intelligence on news curation and distribution: a review literature. J Commun Manag 2:23–26

Mehedi Shamrat FM, Chakraborty S, Imran MM, Muna JN, Billah MM, Das P, Rahman MO (2021) Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. Indones J Electri Eng Comput Sci 23(1):463

Nanayakkara A, Kumara B, Rathnayaka R (2024) Cross-platform topic analysis: identifying trends and divergences in social media discourse. *2024 8th SLAAI international conference on artificial intelligence (SLAAI-ICAI)*. pp 1-6

Neşe A, Bakir H (2022) Spatio-temporal analysis of COVID-19 in turkey. Sustain Cities Soc 76:103421

Ng L, Carley K (2025) A global comparison of social media bot and human characteristics. Sci Rep 15:10973

Nogara G, Pierri F, Cresci S, Luceri L, Giordano S (2024) Misinformation and Polarization around COVID-19 vaccines in France, Germany, and Italy. Proceedings of the 16th ACM web science conference. pp 119-128

Phillips S, Ng L, Zhou W, Carley K (2024) Moral and emotional influences on attitude stability towards COVID-19 vaccines on social media. International conference on social computing behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. pp 216–225

Prasad O, Nandi S, Dogra V, Diwakar D (2023) A systematic review of NLP methods for Sentiment classification of Online News Articles. 2023 14th international conference on computing communication and networking technologies (ICCCNT). pp 1-9

Priyam T, Ruan T, Lv Q (2024) Demographic-based public perception analysis of electric vehicles on online social networks. Sustainability 16:305

Puri N, Coomes E, Haghbayan H, Gunaratne K (2020) Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccines Immunother 16:2586–2593

Ruan T, Lv Q (2022) Public perception of electric vehicles on reddit over the past decade. Commun Transp Res 2:100070

Ruan T, Lv Q (2023) Public perception of electric vehicles on reddit and twitter: a cross-platform analysis. Transp Res Interdiscip Perspect 21:100872

Ruan T, Kong Q, McBride S, Sethjiwala A, Lv Q (2022) Cross-platform analysis of public responses to the 2019 ridgecrest earthquake sequence on twitter and reddit. Sci Rep 12:1634

Singh K, Kumar A, Kumar G, Choudhury T, Singh T, Kotecha K (2023) Sentiment analysis in social media marketing: leveraging natural language processing for customer insights. International conference on information and communication technology for competitive strategies. pp 457-467

I, Nordahl-Hansen A, Quintana D, Wynn R, Gabarron E (2022) Misinformation about COVID-19 vaccines on social media: rapid review. J Med Internet Res 24:e37367

Slavik C, Yiannakoulias N, Buttle C, Darlington J (2023) vaccinfluencers: a study of influential voices criticizing COVID-19 vaccination efforts and negative vaccine information discourse on twitter. Commun Rev 26:300–322

Takedomi Y, Nakayama Y, Suda T, Uno T, Hashimoto T, Toyoda M, Yoshinaga N, Kitsuregawa M, Kobayashi R (2024) Evolution of COVID-19 vaccination-related topics on twitter. Transac Jpn Soc Artif Intell 39:C-N93

Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt Z (2021) What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 3:e175–e194

Verma M, Moudgil N, Goel G, Pardeshi P, Joseph J, Kumar N, Singh K, Singh H, Kodali P (2023) People's perceptions on COVID-19 vaccination: an analysis of twitter discourse from four countries. Sci Rep 13:14281

Vladic E, Novalic M, Mehanovic B, Mehanovic D, Kečo D (2024) Sentiment classification of Tweets using machine learning and NLP techniques. *Available At SSRN 5001926*

Wong A, Ho S, Olusanya O, Antonini M, Lyness D (2021) The use of social media and online communications in times of pandemic COVID-19. J Intensiv Care Soc 22:255–260

Zhou X, Zhang X, Larson H, Figueiredo A, Jit M, Fodeh S, Vermund S, Zang S, Lin L, Hou Z (2023) Spatiotemporal trends in covid-19 vaccine sentiments on a social media platform and correlations with reported vaccine coverage. Bull World Health Organ 102:32

Zhang Q, Zhang R, Wu W, Liu Y, Zhou Y (2023) Impact of social media news on COVID-19 vaccine hesitancy and vaccination behavior. Telemat Inform 80:101983

Zhou X, Zhang X, Larson H, Figueiredo A, Jit M, Fodeh S, Vermund S, Zang S, Lin L, Hou Z (2022) Global spatiotemporal trends and determinants of COVID-19 vaccine acceptance on Twitter: a multilingual deep learning study in 135 countries and territories. MedRxiv. pp 2022-11