# Analysis of Social Media Post Sentiment Shifts Before and After the COVID-19 Pandemic



## Executive Summary

The COVID-19 pandemic precipitated profound global changes, significantly impacting public sentiment as reflected across social media platforms. A consistent trend observed in academic research is an initial, statistically significant increase in negative sentiment during the pandemic's onset and peak, followed by varying degrees of partial emotional recovery or shifts towards more neutral expressions. These shifts were not uniform, exhibiting nuanced variations depending on the specific topics discussed, such as public health measures, online learning, or even coping mechanisms like alcohol consumption.

The studies employed a diverse array of sophisticated methodologies, ranging from traditional Natural Language Processing (NLP) tools to advanced deep learning models like BERT and Graph Attention Networks (GATs), and more recently, Large Language Models (LLMs). These computational approaches allowed researchers to process vast datasets from platforms such as Reddit, Twitter, and Instagram, providing granular insights into collective emotional states.

The findings carry significant implications for public and mental health. Social media sentiment analysis emerged as a powerful, near real-time indicator of societal psychological distress, correlating with documented increases in anxiety, depression, and stress. It also highlighted the critical role of social media in disseminating information, both accurate and misleading, which directly influenced public adherence to health measures and overall community resilience. However, the field faces inherent challenges, including the complexity of human emotion, the noise and biases within social media data, and the proliferation of misinformation, all of which necessitate careful interpretation and continuous methodological refinement. These studies collectively underscore social media's role as a dynamic barometer of public psychological states and a vital channel for crisis communication and intervention.

## 1. Introduction

The COVID-19 pandemic represented an unprecedented global health crisis, profoundly impacting public health, economic stability, and individual well-being worldwide.[1] The rapid onset and evolving nature of the pandemic introduced widespread uncertainty, fear, and disruption to daily life, compelling fundamental shifts in human activities, including the transition from in-person interactions to online modalities for learning and communication.[3]

During this period of immense upheaval, social media platforms emerged as crucial conduits for real-time communication, information dissemination, and the expression of public sentiment.[3] These digital spaces served not only as primary sources of news and updates but also as vital outlets for individuals to express emotional responses, share coping mechanisms, and organize community support efforts.[5] The sheer volume of user-generated content across platforms like Twitter, Reddit, and Instagram presented an invaluable, albeit complex, dataset for researchers seeking to understand collective emotional and psychological states during an unfolding crisis.[2]

The reliance on social media data for understanding public sentiment during a crisis such as COVID-19 marks a notable evolution in public health surveillance. Traditionally, public health monitoring has relied on methods such as surveys, clinical data, and reported case numbers. However, the rapid, global spread of the pandemic and the immediate need for dynamic public insight often outpaced the responsiveness of these conventional methods. Social media, with its real-time nature and pervasive reach, naturally filled this void. This shift indicates that social media platforms are no longer merely communication tools but have, by necessity, become a form of public health observatory, albeit one that generates complex and sometimes noisy data. This development has significant implications for how public health agencies gather intelligence and formulate responses to future crises, encouraging a move towards more agile, data-driven approaches.

This report aims to synthesize and critically analyze academic research studies that specifically compared social media post sentiment before and after the onset of the COVID-19 pandemic. By examining the observed sentiment shifts, the diverse methodological approaches employed, their implications for public and mental health, and the inherent challenges in this nascent field, this report seeks to provide a comprehensive understanding of the pandemic's digital emotional footprint.

## 2. Observed Sentiment Shifts Across Social Media Platforms

This section details the specific changes in social media sentiment, distinguishing between general trends and thematic variations, and providing concrete data points from the research.

### 2.1. General Sentiment Trends and Recovery Patterns

A consistent finding across multiple studies is a significant increase in negative sentiment during the initial phases of the pandemic when compared to pre-pandemic levels. For instance, an analysis of Reddit data collected from 128 U.S. university communities revealed that the odds of negative sentiment in 2020, representing the peak period of the pandemic, were 25% higher than in 2019, which served as the pre-pandemic baseline.[2] This increase was statistically significant, with a p-value of less than 0.001.[2] In absolute terms, the percentage of negative sentiment on Reddit rose from 35.93% in 2019 to 41.75% in 2020, marking a 16.20% relative increase.[2] Similarly, a study tracking Instagram discourse from January 2020 to September 2024 observed a notable shift: positive sentiment on the platform decreased from 38.35% to 28.69%, while neutral sentiment simultaneously rose from 44.19% to 58.34%.[9] This suggests a move away from overtly positive expressions towards a more cautious, uncertain, or less emotionally charged neutral stance as the pandemic progressed.

While the initial surge in negativity was pronounced, some research indicates a partial recovery in sentiment composition over time. The Reddit study, for example, showed that although negative sentiment remained elevated in 2021 (7.3% higher odds) and 2022 (6.3% higher odds) compared to 2019, these increases were substantially lower than the 25% observed during the peak in 2020.[8] This pattern suggests a gradual, albeit incomplete, return towards pre-pandemic sentiment levels, indicating a "partial recovery in the sentiment composition".[8] This observed partial recovery in sentiment suggests that societal emotional responses to prolonged crises are not linear but rather adaptive. They typically involve phases of acute distress followed by a gradual normalization, even if a full return to baseline emotional states is not fully achieved. The initial spike in negative sentiment, characterized by a 25% higher likelihood in 2020, directly reflects the acute shock and widespread disruption brought about by the pandemic. The subsequent reduction in the odds of negative sentiment in 2021 and 2022 implies a process of societal adjustment. This adjustment does not necessarily equate to a return to pre-pandemic levels of happiness or optimism, but rather a process of normalizing the crisis, where the most extreme emotional reactions subside as individuals and communities adapt to the evolving "new normal." This adaptive phase is crucial for understanding the long-term psychological impacts and the resilience of society, and it carries implications for policy responses, which must evolve from immediate crisis management to supporting sustained psychological well-being.

However, not all sentiment shifts were uniformly negative, with some studies revealing contradictory findings in specific contexts. For instance, a study examining Twitter communication related to alcohol during the onset of the COVID-19 pandemic (comparing 40 days before and 40 days after the World Health Organization's declaration on March 11, 2020) found that tweets in the post-declaration period exhibited a more positive emotional tone.[10] These tweets also contained more authentic and affiliation-oriented language.[10] This counter-intuitive finding suggests that specific topics or the use of certain coping mechanisms might lead to different, even contrasting, sentiment trajectories compared to overall public discourse.

### 2.2. Thematic Sentiment Variations

Sentiment shifts on social media were highly thematic, reflecting diverse public reactions to specific aspects of the pandemic. Analysis of the Belgian Reddit community from January 2020 to June 2022 revealed that the volume of posts concerning COVID-19 mitigation measures, such as lockdowns, mask mandates, and vaccinations, correlated strongly with external events.[11] Crucially, the sentiment expressed on these topics was influenced by previous posts, leading to phenomena like homophily—the tendency for users to engage with content sharing similar sentiment—and polarization within discussions.[11] Furthermore, news coverage related to pandemic situations, anti-pandemic measures, and supportive actions was found to reduce the public's negative emotions, while comments mentioning central government entities (e.g., the Government of Hong Kong) tended to reveal more negativity.[1] This highlights the critical role of governmental communication and the perceived effectiveness of policy measures in shaping public emotional responses.

The abrupt shift to online learning during the pandemic also significantly impacted the sentiment of students and teachers. A study analyzing Twitter data on "online learning" between April 2020 and December 2021 found a prevalent negativity, with 37.19% of posts expressing negative sentiment, contrasting with 29.37% positive and 33.43% neutral sentiments.[4] This negativity was primarily attributed to technological challenges and the strain placed on online learning platforms.[4] This finding stands in stark contrast to the more positive sentiment observed in alcohol-related tweets, indicating that the specific context and associated practical challenges heavily influence the emotional tone of online discourse.

Regarding alcohol-related discourse, the Twitter study noted an increase in authentic and positive emotional language in tweets after the pandemic's declaration.[10] Language related to personal concerns was consistently higher in alcohol-related tweets, suggesting that alcohol was perceived or used as a coping mechanism for insecurities and anxieties during the crisis.[10] A comparative analysis of Twitter and Reddit concerning loneliness found similar thematic patterns across both platforms. However, Reddit discussions predominantly centered on personal-emotional themes, while Twitter exhibited a broader range of categories.[7] Both platforms aligned with psychological linguistic features associated with the self-expression of mental health issues, underscoring social media's role in individuals articulating such concerns.[7]

A particularly significant finding within the pandemic period (2020) from the Reddit study was that the odds of negative sentiments associated with *in-person learning* were 48.3% higher than with remote learning ($p = 0.029$).[2] This suggests that the perceived risks, restrictions, or general disruptions inherent in maintaining in-person activities during the pandemic significantly contributed to heightened negativity among the affected population.

The differential emotional responses to specific topics, such as the more positive sentiment observed for alcohol-related tweets, the prevalent negativity surrounding online learning, and the polarization concerning mitigation measures, indicates that the pandemic's emotional impact was highly granular and context-dependent, rather than a monolithic wave of negativity. If the pandemic had simply caused universal negativity, all topics would likely reflect a similar downward trend. However, the data reveals nuanced reactions: the increased positive sentiment in alcohol-related tweets might point to shared coping mechanisms or a sense of collective experience, while the predominantly negative sentiment in online learning discussions is clearly linked to practical challenges and frustrations. The observed polarization around mitigation measures further reflects underlying sociopolitical divisions and differing perceptions of governmental responses. This granularity underscores that effective interventions or public communications require a deep, topic-specific understanding of public sentiment, rather than a generalized approach. It also suggests that individual and collective coping strategies manifested differently across various aspects of daily life.

**Table 1: Comparative Overview of Key Sentiment Shifts by Platform and Timeframe**

| Study ID | Social Media Platform(s) | Timeframes Compared | Specific Topics (if applicable) | Key Sentiment Shift | Statistical Significance (p-value) |
|---|---|---|---|---|---|
| [2] | Reddit | 2019 vs. 2020 | General Sentiment | 25% higher odds of negative sentiment in 2020; 16.20% | < 0.001 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | relative increase in negative sentiment. | |
| 8 | Reddit | 2019 vs. 2021, 2022 | General Sentiment | 7.3% higher odds negative sentiment in 2021; 6.3% higher odds negative sentiment in 2022 (partial recovery). | Significant at 5% level (multiplicity-adjusted) |
| 9 | Instagram | 2020 vs. 2024 | General Sentiment | Positive sentiment decreased from 38.35% to 28.69%; Neutral sentiment increased from 44.19% to 58.34%. | Not specified |
| 10 | Twitter | 40 days pre vs. 40 days post March 11, 2020 | Alcohol-related discourse | More positive emotion, authentic language, and affiliation-oriented language in post-declaration tweets. | Significant |
| 4 | Twitter | April 2020 - Dec 2021 | Online Learning | Prevalent negativity (37.19%) vs. 29.37% positive and 33.43% | Not specified |

| | | | | | |
|---|---|---|---|---|---|
| | | | | neutral. | |
| [2] | Reddit | In-person vs. Remote Learning (during 2020) | Learning Modalities | 48.3% higher odds of negative sentiments associated with in-person learning. | 0.029 |
| [11] | Reddit | Jan 2020 - Jun 2022 | COVID-19 mitigation measures (lockdowns, masks, vaccines) | Sentiment influenced by previous posts, leading to homophily and polarization. | Not specified |
| [1] | General Social Media | Pandemic period | News coverage (pandemic situations, anti-pandemic measures, government mentions) | Supportive actions reduced negative emotions; government mentions revealed more negativeness. | Not specified |

## 3. Methodological Approaches and Computational Tools

This section delves into the technical aspects of how these sentiment analyses were conducted, highlighting the evolution and sophistication of the tools used.

### 3.1. Social Media Platforms Utilized

Research studies analyzing COVID-19 sentiment widely leveraged major social media platforms, recognizing their extensive user-generated content as a rich data source. **Reddit** was frequently utilized due to its forum-like structure, which facilitates longer posts and more detailed discussions, particularly within university communities [2] and specific national contexts like Belgium.[11] Reddit data was also used in comparative studies alongside Twitter for discussions on topics such as loneliness.[7]

**Twitter** emerged as a prominent platform for capturing real-time, short-form communication. It was extensively used for analyzing general sentiment shifts [10], attitudes towards online learning [3], and broader public health discussions.[12] Twitter also featured in comparative studies with other platforms.[7] **Instagram** proved valuable for multilingual sentiment analysis over an extended period (2020-2024), demonstrating its utility for tracking evolving discourse across different linguistic communities.[9] In specific geopolitical contexts, such as China during the outbreak, **Sina Weibo** was a key platform for sentiment studies.[14] Other platforms like Facebook and Gab have also been explored in comparative studies, indicating a broader scope of platforms being considered for sentiment analysis.[6]

Data collection methods commonly involved utilizing platform-specific Application Programming Interfaces (APIs), such as the Twitter API for tweets and the Pushshift API for Reddit messages.[2] Researchers also leveraged existing datasets, like the GeoCoV19 dataset for Twitter, to obtain tweet IDs for analysis.[10]

### 3.2. Computational Tools and Methodologies for Sentiment Analysis

The field of social media sentiment analysis during the pandemic witnessed a rapid adoption of advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques to navigate the complexity of human language. **Transformer-based models**, particularly Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT pre-training approach (RoBERTa), were highlighted for their superior performance in sentiment classification.[2] RoBERTa, for example, was employed to learn text embeddings from Reddit messages, demonstrating its capability to handle internet slang and non-standard spellings effectively.[2] Fine-tuned BERT models have consistently shown high accuracy rates, with one study reporting 94.80% accuracy in sentiment classification.[12]

Beyond text content, **Graph Attention Networks (GATs)** were utilized in conjunction with RoBERTa to leverage relational information among posted messages, such as message-reply relationships, thereby improving sentiment classification accuracy.[2] This approach signifies a move towards incorporating structural relationships within social media discourse, not just the textual content itself. To further enhance robustness and mitigate inconsistencies from individual models, researchers employed **model stacking** or **ensemble learning** approaches. This involved combining prediction probabilities from models like RoBERTa and GAT using meta-models, often logistic regression, to yield a more reliable final sentiment classification.[2]

Traditional NLP tools also played a role. The **Linguistic Inquiry and Word Count (LIWC) software** was used for analyzing psychological, emotional, and social sentiment by categorizing language use into predefined linguistic categories. LIWC provides summary variables such as Authenticity, Clout, and Emotional Tone, offering insights into psychological states expressed in text.[10] **VADER (Valence Aware Dictionary for Sentiment Reasoning)**, a rule-based sentiment analysis tool from the Natural Language Toolkit (NLTK), specifically designed for social media text, was employed for identifying keywords and phrases related to specific emotions like loneliness.[7] Additionally, **TF-IDF (Term Frequency-Inverse Document Frequency)** was used for summarizing topics of posts, often in combination with sentiment analysis to provide thematic context.[14]

More recently, emerging research has demonstrated the effectiveness of **Large Language Models (LLMs)**, such as OPT and ChatGPT, in sentiment analysis of health-related survey data. ChatGPT, in particular, outperformed other sentiment analysis tools, exhibiting 6% higher accuracy and 4-7% higher F-measure.[6] This development points to a significant advancement in automated sentiment analysis, with the potential to reduce the need for extensive human labor in labeling tasks.[6] Finally, the outputs from these machine learning models were often subjected to rigorous **statistical modeling**. Generalized Linear Mixed-Effects Models (GLMM) were applied to estimate temporal trends in sentiment and assess the effects of various factors, such as the mode of teaching (in-person vs. remote), demonstrating the integration of advanced computational outputs with robust statistical inference.[2]

The progression from rule-based tools like VADER and dictionary-based methods such as LIWC to sophisticated deep learning models like BERT and RoBERTa, and now to Large Language Models, reflects a continuous drive for higher accuracy and a more nuanced understanding of social media sentiment, especially in complex and evolving contexts like a pandemic. Earlier sentiment analysis often relied on lexicons or simpler machine learning algorithms. As social media text became increasingly complex—incorporating slang, sarcasm, and rapidly evolving topics—these initial methods faced significant challenges. The advent of BERT and RoBERTa, with their ability to understand context and semantic relationships, marked a substantial leap forward, as evidenced by their high accuracy rates. The integration of Graph Attention Networks further refined this capability by capturing relational information within social media interactions. The most recent development, the application of LLMs, suggests a new frontier where models can perform effectively even with minimal or no prior examples (few-shot or zero-shot learning), thereby reducing the need for extensive human labeling. This methodological evolution is critical because it enables researchers to extract more reliable and granular insights from increasingly massive and complex datasets, moving beyond broad sentiment classification to discern subtle emotional shifts and their underlying drivers.

**Table 2: Summary of Sentiment Analysis Methodologies and Tools**

| Study ID | Social Media Platform(s) | Primary Sentiment Analysis Tool(s) / Model(s) | Key Methodological Features | Accuracy/Performance (if specified) |
|---|---|---|---|---|
| [2] | Reddit | RoBERTa, Graph Attention Network (GAT), Model Stacking | Binary (Negative vs. non-Negative), Generalized Linear Mixed-Effects Model (GLMM) | High accuracy (specific % not stated for stacked model), 84.04% human agreement for labeling |
| [6] | Health-related survey data (NIH, Stanford) | OPT, ChatGPT | Few-shot and zero-shot learning, Human raters for gold standard labels | ChatGPT outperformed OPT: 6% higher accuracy, 4-7% higher F-measure |
| [10] | Twitter | Linguistic Inquiry and Word Count (LIWC) | Categorizes language into ~80 linguistic categories, 2x2 mixed ANCOVA | Provides summary variables (Authenticity, Clout, Emotional Tone) |
| [7] | Twitter, Reddit | VADER (Valence Aware Dictionary for Sentiment Reasoning) | Rule-based, identifies keywords/phrases related to loneliness | Not specified |
| [14] | Sina Weibo | BERT, TF-IDF | Unsupervised BERT for | Fine-tuned BERT with |

| | | | | |
|---|---|---|---|---|
| | | | classification (positive, neutral, negative), TF-IDF for topic summarization | considerable accuracy |
| [12] | Twitter | BERT, BiLSTM, CNN, DistilBERT, XLNET, ALBERT | Multi-class (positive, negative, neutral), various feature sets and classifiers | BERT accuracy: 94.80% [12], Proposed approach: 96.66% [12] |

## 4. Implications for Public Health, Mental Health, and Policy

This section explores the real-world consequences and policy relevance of the observed sentiment changes.

### 4.1. Correlations with Mental Health Outcomes

The observed rise in negative sentiment on social media platforms directly corresponds with documented increases in mental health issues during the pandemic. The 25.7% higher odds of negative sentiments expressed on Reddit in 2020 compared to 2019 [2] aligns with systematic reviews that reported increased symptoms of anxiety, depression, and post-traumatic stress disorder (PTSD) within the general population during this period.[2] Furthermore, studies indicate that excessive social media use and the constant stream of pandemic-related news contributed to information overload, which in turn exacerbated stress, anxiety, depression, and social isolation.[5] A meta-analysis specifically highlighted a significant positive association between the amount of time individuals spent on social media platforms and the likelihood of experiencing heightened levels of anxiety and depressive symptoms.[5] Exposure to adverse news articles and postings on social media was also identified as a contributing factor to the risk of depression in certain individuals.[5]

The finding that in-person learning during the pandemic was associated with 48.3% higher odds of negative sentiments than remote learning [2] points to specific contexts that intensified mental distress. This suggests that the anxieties related to physical presence, potential exposure to the virus, and the disruptions inherent in structured in-person environments significantly impacted emotional well-being.

Despite these negative impacts, social media also served as an adaptive coping mechanism for some individuals, providing avenues for social support and potentially reducing stress and anxiety levels.[5] The observed increase in "affiliation-oriented language" on Twitter during the pandemic's onset [10] supports this, indicating a fundamental human need for connection during periods of isolation.

Social media sentiment analysis functions as a powerful, non-invasive epidemiological tool, offering real-time insights into the collective psychological burden of a crisis. This capability can precede or complement traditional mental health surveillance data. Conventional mental health data, such as clinical diagnoses or comprehensive surveys, often involve a time lag in collection and analysis. Social media, however, captures immediate, unfiltered public sentiment. The consistent rise in negative sentiment observed across platforms provides an early warning system for widespread mental distress, allowing public health officials to anticipate needs and allocate resources more proactively. This real-time monitoring is invaluable for understanding the "infodemic"—the rapid spread of both accurate and inaccurate information [5]—and its psychological ripple effects, acting as a direct reflection of the public's emotional state in response to unfolding events and information.

### 4.2. Influence on Public Health Behaviors and Policy Responses

Gauging public sentiment is valuable for disease modelers, epidemiologists, and policymakers.[11] Understanding public opinion on critical measures like lockdowns, mask mandates, and vaccination campaigns can directly inform strategies to improve public adherence and more accurately predict disease transmission dynamics.[11]

A significant challenge during the pandemic was the rapid spread of both true and false information, termed an "infodemic," which profoundly influenced public opinion and their willingness to adhere to public health measures.[5] Misinformation, conspiracy theories, and false claims amplified anxiety and stress among the population.[5] This underscores the critical need for public health authorities to manage information dissemination effectively and actively address misinformation to prevent its detrimental effects on public sentiment and behavior.[5]

Conversely, social media platforms also positively influenced public health protection by increasing public health awareness and promoting beneficial behavioral changes, such as improved hand washing, consistent mask-wearing, social distancing, and self-isolation.[16] These platforms provided direct and efficient channels for government agencies and health organizations to communicate with the public, disseminate vital information, and offer advice on infection prevention.[5] Furthermore, social media was instrumental in organizing community efforts, ranging from neighborhood support groups to initiatives aiding vulnerable populations, thereby fostering community resilience during the crisis.[5]

The interplay between social media sentiment, information dissemination (including misinformation), and public adherence to health measures creates a complex feedback loop that policymakers must actively monitor and influence to achieve effective public health outcomes. This means it is not merely that sentiment reflects the impact of policy; it also directly influences public behavior and compliance. If negative sentiment is primarily driven by misinformation, it can erode public trust and diminish compliance with public health directives. Conversely, positive sentiment linked to clear and effective communication can foster cooperation and adherence. This dynamic system suggests that public health messaging must be agile, responsive to real-time sentiment, and specifically designed to counter negative emotional contagion and the spread of misinformation. The "infodemic" is not merely a side effect of a crisis but a critical determinant of public health outcomes, making sentiment analysis an indispensable component of any comprehensive crisis communication strategy.

## 5. Challenges and Limitations in Social Media Sentiment Analysis

This section critically examines the inherent difficulties and biases that can affect the accuracy and generalizability of sentiment analysis studies using social media data, particularly in a crisis context.

### 5.1. Inherent Complexity of Human Emotion and Language

One of the most significant challenges in social media sentiment analysis lies in the inherent complexity and nuance of human emotion and language. Emotions are often expressed with sarcasm, idioms, cultural references, and subtle undertones that are exceedingly difficult for computational models to detect accurately.[17] This poses a substantial hurdle for creating universally effective sentiment models that can truly capture the full spectrum of human feeling.[17]

Furthermore, social media data, particularly from platforms like Twitter, is characterized by informal language, misspellings, and often careless grammar.[12] This "noisy" and heterogeneous data makes fundamental NLP tasks—such as sarcasm detection, aspect extraction, and subjectivity detection—more challenging, which in turn impacts the reliability of sentiment detection.[12] Identifying informative content from this vast and often messy dataset requires significant pre-processing and advanced techniques.[12] Additionally, some sentiment analysis studies explicitly limit their scope to classifying "sentiment" (e.g., positive, negative, neutral) and do not delve into the broader spectrum of specific "moods and emotions" (e.g., anger, fear, sadness).[12] This narrow focus can lead to an incomplete picture of the public's emotional state during a crisis, missing critical emotional nuances that could inform targeted interventions.

The persistent challenge of capturing the full spectrum and nuance of human emotion in social media text implies that even the most advanced sentiment analysis models provide a *proxy* for public sentiment rather than a perfect reflection. While models like BERT can achieve high accuracy, the inherent limitations of natural language, including sarcasm, idioms, and cultural context, mean that a purely computational approach will inevitably miss some subtleties. This suggests that sentiment analysis should be regarded as a powerful indicator or trend identifier, but its outputs must be interpreted with caution and ideally cross-referenced with qualitative methods or domain expertise. The "black box" nature of complex artificial intelligence models further complicates full interpretability, meaning that while researchers can identify "what" is happening in terms of sentiment shifts, the underlying "why" in nuanced emotional changes is not always fully transparent.

### 5.2. Biases and Limitations in Data and Models

Sentiment analysis models are inherently susceptible to biases introduced through imbalanced training datasets.[17] These biases can lead to skewed predictions, potentially favoring certain demographics, misclassifying sentiments in underrepresented groups, or amplifying existing stereotypes.[17] This issue is particularly problematic when applying sentiment analysis to sensitive areas like health or social issues, where biased results could have severe consequences.[17]

Moreover, moods, opinions, and attitudes towards a given topic are heavily influenced by culture.[13] Models trained predominantly on data from one cultural context may not accurately reflect sentiments in another, leading to errors and limiting the generalizability of findings.[13] A significant limitation in much of the reviewed research is its restriction to English-language text, which severely limits global applicability and prevents a comprehensive understanding of sentiment among non-English speaking populations.[13]

During a crisis, the proliferation of fake news and incorrect scientific results on social networks can significantly distort true public sentiment.[12] If sentiment analysis captures reactions to misinformation, it can lead to misleading conclusions about genuine public opinion or dissatisfaction with crisis responses.[13] Furthermore, the comparison of different sentiment analysis tools has revealed high variability and disagreement across evaluated tools, even when applied to similar health-related data.[6] This highlights a pressing need for comprehensive evaluation and standardization within the field to ensure consistent and reliable results.[6]

Some studies simplify sentiment classification into binary categories (positive/negative), which can oversimplify the emotional landscape and miss nuances that a multi-class approach (positive/negative/neutral) could reveal.[13] For instance, a shift from "neutral" to "negative" might be more informative than a change within a broad "negative" category. Additionally, many regional studies may overlook the influence of worldwide news and data, potentially skewing before/after comparisons if broader global events are not adequately accounted for.[13] Finally, much of the research was conducted over relatively short periods, which might miss long-term shifts, recovery patterns, or the evolving sentiment dynamics inherent in a prolonged crisis like a pandemic.[13]

The pervasive nature of bias—whether stemming from data, cultural context, or algorithmic design—and the widespread presence of misinformation in social media sentiment analysis mean that findings, especially before/after comparisons, must be interpreted with a critical awareness of their inherent limitations and potential for misrepresentation. If the training data used for models is biased, the model's output will inevitably reflect that bias, potentially reinforcing societal inequalities. If cultural nuances are overlooked, cross-cultural comparisons become invalid. If the input data is contaminated with fake news, the detected "sentiment" might be a reaction to falsehoods rather than reality. This implies that while sentiment analysis offers valuable insights, it cannot be treated as a definitive measure without rigorous validation, transparency in its methodology, and a deep understanding of both the data's origin and the model's limitations. This necessitates a concerted move towards "bias-aware" frameworks and a greater emphasis on ethical considerations throughout the deployment of these analytical tools.

**Table 3: Identified Challenges and Biases in Social Media Sentiment Analysis**

| Challenge/Bias Category | Specific Challenge/Bias | Explanation/Impact on Analysis | Relevant Snippet ID(s) |
|---|---|---|---|
| **Data Quality & Nuance** | Sarcasm, Idioms, Cultural References | Difficult for machines to detect, leading to misclassification of sentiment. | [17] |

| | | | |
|---|---|---|---|
| | Informal Language & Noisy Data | Requires significant pre-processing; challenging for NLP tasks like sarcasm detection, subjectivity detection. | 12 |
| | Sentiment vs. Mood/Emotions | Studies may not capture the full spectrum of human emotional responses (e.g., anger, fear, sadness). | 12 |
| **Model & Methodological Limitations** | Bias in Training Data | Leads to skewed predictions, favoring demographics, misclassifying underrepresented groups, amplifying stereotypes. | 17 |
| | Cultural Influence on Opinions | Models trained on one culture may not accurately reflect sentiments in another; limits generalizability. | 13 |
| | Tool Variability & Inconsistency | High disagreement across different sentiment analysis tools, highlighting need for standardization. | 6 |
| | Binary vs. Multi-Class Classification | Oversimplifies emotional landscape, misses nuances (e.g., shift from neutral to negative). | 13 |
| | Limited to | Restricts | 13 |

| | | | |
|---|---|---|---|
| | English-Language Text | generalizability and global applicability of findings. | |
| | Short Timelines of Research | May miss long-term shifts, recovery patterns, or evolving sentiment dynamics in prolonged crises. | 13 |
| **External Factors** | Misinformation & Fake News | Distorts true public sentiment, leading to misleading conclusions if reactions to falsehoods are analyzed. | 12 |
| | Ignoring Global News & Data Effects | Regional studies may not fully capture broader influences on public opinion, skewing comparisons. | 13 |

## 6. Conclusion and Future Directions

### 6.1. Synthesis of Overarching Findings

The COVID-19 pandemic exerted a profound and measurable impact on public sentiment, consistently reflected in social media discourse. Research demonstrates a significant initial increase in negative emotions across platforms during the pandemic's peak, followed by a partial, albeit incomplete, recovery. Social media platforms proved to be vital, near real-time barometers of public psychological states, revealing nuanced emotional responses to specific pandemic-related topics, policies, and even coping mechanisms. The evolution of sentiment analysis methodologies, from traditional NLP techniques to advanced deep learning models and emerging Large Language Models, has substantially enhanced the capacity to extract meaningful insights from the vast and complex social media datasets. However, the utility of social media sentiment analysis is inherently tempered by significant challenges, including the complexity of human emotion, the pervasive noise and biases within social media data, and the widespread impact of misinformation. These factors necessitate careful interpretation of findings and continuous methodological refinement.

### 6.2. Recommendations for Advancing Social Media Sentiment Analysis in Crisis Contexts

To enhance the reliability, depth, and applicability of social media sentiment analysis in future crisis contexts, several key recommendations emerge:

- **Enhance Methodological Robustness:** Future research should prioritize multi-modal and multi-platform analysis, integrating data from diverse social media platforms and potentially other data sources (e.g., traditional news media, public surveys) to provide a more holistic view of public sentiment and allow for cross-validation of findings. It is crucial to implement and further develop bias-aware sentiment analysis frameworks [17] to systematically address demographic, cultural, and linguistic biases, thereby ensuring more equitable and accurate predictions. This includes developing models capable of analyzing text in multiple languages, moving beyond the current predominant focus on English-language content.[13] Additionally, a shift from binary sentiment classification towards more nuanced multi-class or fine-grained emotion detection is necessary to capture the full spectrum of human emotional responses.[12] Finally, conducting longer-term longitudinal studies is essential to track sentiment evolution beyond the immediate crisis phase, capturing patterns of adaptation, recovery, and potential long-term psychological impacts.[13]
- **Address Information Quality:** Robust mechanisms for misinformation and fake news detection should be integrated directly into sentiment analysis pipelines to either filter out or appropriately contextualize sentiment driven by false narratives.[12] Furthermore, methods to weigh sentiment based on the credibility of the source or the verified spread of information could provide a more accurate reflection of informed public opinion.
- **Strengthen Interdisciplinary Collaboration:** Fostering deeper collaboration between computational social scientists, public health experts, psychologists, and policymakers is paramount. This ensures that research questions are policy-relevant, methodologies are robust and ethically sound, and findings are effectively translated into actionable insights for public health interventions and communication strategies.
- **Prioritize Ethical Considerations:** Ethical considerations must be paramount at all stages of research, encompassing data privacy, transparency in model architecture and training data, and accountability for the potential societal

impacts of sentiment analysis tools.[17] This includes ensuring that such tools do not inadvertently reinforce harmful stereotypes or disadvantage vulnerable groups.

### 6.3. Future Research Avenues

Several promising avenues for future research can further advance the field:

- **Causal Inference:** Developing and applying methodologies that can establish stronger causal links between specific pandemic events, policy changes, and observed sentiment shifts, moving beyond mere correlation, would provide more definitive guidance for crisis management.
- **Predictive Modeling:** The development of predictive models capable of forecasting shifts in public sentiment based on evolving crisis conditions could enable proactive public health interventions and more timely policy adjustments.
- **Personalized Interventions:** Investigating how sentiment analysis can inform personalized public health messaging or targeted mental health support, tailored to specific demographic groups or emotional states identified on social media, represents a significant step towards precision public health.
- **Cross-Cultural Comparative Studies:** Conducting more extensive comparative studies across diverse cultural contexts is crucial to understand how cultural norms, societal structures, and communication styles influence emotional expression and sentiment dynamics during global crises. This would enhance the generalizability and cultural sensitivity of sentiment analysis tools.

### Works cited

1. News Coverage of the COVID-19 Pandemic on Social Media and ..., accessed June 10, 2025, https://www.jmir.org/2024/1/e48491/
2. COVID-19 sentiment analysis using college subreddit data - PMC, accessed June 10, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9635711/
3. How Attitudes Toward Online Learning Changes During Pandemic, and Reasons Behind It, accessed June 10, 2025, https://www.ewadirect.com/proceedings/lnep/article/view/12228
4. Analysing Public Sentiment on Online Learning During COVID-19 - Academic Conferences International, accessed June 10, 2025, https://papers.academic-conferences.org/index.php/ecsm/article/download/2238/2053/8231
5. Public mental health through social media in the post ... - Frontiers, accessed June 10, 2025, https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1323922/full
6. A Comparison of ChatGPT and Fine-Tuned Open Pre-Trained Transformers (OPT)

Against Widely Used Sentiment Analysis Tools: Sentiment Analysis of COVID-19 Survey Data - JMIR Mental Health, accessed June 10, 2025, https://mental.jmir.org/2024/1/e50150
7. Understanding Loneliness Through Analysis of Twitter and Reddit Data: Comparative Study, accessed June 10, 2025, https://www.i-jmr.org/2025/1/e49464
8. Has sentiment returned to the pre-pandemic level? A sentiment ..., accessed June 10, 2025, https://pubmed.ncbi.nlm.nih.gov/38489275/
9. arxiv.org, accessed June 10, 2025, https://arxiv.org/abs/2410.03293
10. Tweeting about alcohol: Exploring differences in Twitter sentiment ..., accessed June 10, 2025, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0276863
11. Sentiment spreads, but topics do not, in COVID-19 discussions within the Belgian Reddit community - arXiv, accessed June 10, 2025, https://arxiv.org/html/2505.20185v1
12. COVID-19 Related Sentiment Analysis Using State-of-the ... - Frontiers, accessed June 10, 2025, https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2021.812735/full
13. BERT-Deep CNN: State-of-the-Art for Sentiment Analysis of ... - arXiv, accessed June 10, 2025, https://arxiv.org/pdf/2211.09733
14. COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model, accessed June 10, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8545339/
15. COVID-19 sentiment analysis using college subreddit data | PLOS ..., accessed June 10, 2025, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0275862
16. Influence of Social Media Platforms on Public Health Protection ..., accessed June 10, 2025, https://www.jmir.org/2020/8/e19996/
17. A Comprehensive Approach to Bias Mitigation for Sentiment ... - MDPI, accessed June 10, 2025, https://www.mdpi.com/2076-3417/14/23/11471