Identifying and Analyzing False Information Discourses: A Text Mining Study of COVID-19 Related Tweets in 2020

Mohamad Kaddoura

A Thesis

In The Department

of

Business Analytics and Technology Management

Presented in Partial Fulfilment of the Requirements

For the Degree of

Master of Science (Business Analytics and Technology Management)

at Concordia University

Montréal, Québec, Canada

April 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Mohamad Kaddoura

Entitled: Identifying and Analyzing False Information Discourses: A Text Mining Study of COVID-19 Related Tweets in 2020

And submitted in partial fulfilment of the requirements for the degree of

Master of Science (Business Analytics and Technology Management)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

____________________________________ Chair

Dr. Danielle Morin

____________________________________ Examiner

Dr. Danielle Morin

____________________________________ Examiner

Dr. Mahdi Mirhoseini

____________________________________ Supervisor

Dr. Mohsen Farhadloo

Approved by ____

Dr. Suchit Ahuja, Graduate Program Director

April 10^{th} 2025 ____

Dr. Anne-Marie Croteau, Dean of Faculty

Abstract

Identifying and Analyzing False Information Discourses: A Text Mining Study of COVID-19 Related Tweets in 2020

Mohamad Kaddoura

False information is an ongoing challenge to global health crises that influences public perception and weakens emergency response efforts. This study investigates how false information is structured and framed during pandemics by using COVID-19 pandemic as a case study. Focusing on COVID-19 related Twitter posts, our study filters tweets that possibly contain false information and groups them into conversations that are analyzed through two of the main topic modeling techniques: Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). The results of NMF, which was found to discover more coherent and interpretable topics, were considered for theme identification and framing analysis. We were able to identify eight dominant themes that shaped the entire COVID-19 narratives. These narratives were highly politicized, but also included themes like severity, virus origin, potential treatments, health measures and global responses. The framing analysis showed that linguistic characteristics across false information often included emotionally charged words and evolved through different political and social contexts. Frames like blaming, resistance and conspiracy were recurring across the identified themes, indicating mixed feelings that amplified the spread of false information and challenged public health efforts in combating the pandemic. By combining topic modeling with manual interpretation, this research presents a novel approach to understanding the context of false information and its dynamics during times of health crises. Our findings contribute to the studies of false information and crisis communication research by showing how narratives are framed and how they evolve over time.

Keywords: False information, topic modeling, framing, COVID-19

Acknowledgment

I would like to begin by sincerely thanking my supervisor, Dr. Mohsen Farhadloo, for his support, guidance, and encouragement throughout the development of this thesis. His expertise in the field and thoughtful advice were influential at every stage of this work, and I am truly grateful for his mentorship, patience, and belief in my potential.

I am also thankful to the members of my thesis committee, Dr. Danielle Morin and Dr. Mahdi Mirhoseini, for their time and insightful feedback. Their constructive comments and suggestions strengthened the overall quality of this research.

Finally, I am extremely indebted to my family and friends for their unconditional love and support. Your presence, even from afar, has been a source of strength. I am thankful to have you by my side, and I hope to make you proud.

Table of Contents

List of Figures

List of Tables

1. Introduction

In today's world, the rapid flow of information across social media platforms has reshaped how we consume information and interpret events. The digital age provides people with unlimited access to information, giving individuals the privilege to receive global news after seconds of their occurrences. However, this instant access to information on social platforms often comes with real challenges. One of the challenges is that individuals are permitted to share whatever information they would like to without actually caring about its validity. In such cases, the spread of accurate and inaccurate information poses a threat, as it makes it difficult and confusing for people to differentiate between reliable news and fake ones.

False information is categorized according to the intent of the author sharing it as either misinformation or disinformation (Kumar & Shah, 2020). Misinformation is defined as the spread of false information without the author's intent to deceive others. Since it is extremely challenging for researchers to understand the intent behind sharing a piece of false information, Wu et al. (2019) described misinformation as an "umbrella" that holds inaccurate and false information. Unintentionally spread misinformation could be about contagious diseases like HIV and Ebola, urban legends, unverified information, spam messages, rumors, and other fake news. Although there are numerous debates around the exact definition for misinformation, it is worth mentioning that misinformation was classified as a global risk by the World Economic Forum (Muhammed & Mathew, 2021). As for disinformation, it can be defined as the opposite of misinformation, where the intent of their author is to deceive readers and mislead them (Shu, 2022; Tandoc Jr et al., 2020; Aïmeur et al., 2023). According to Diaz Ruiz and Nilsson (2023), disinformation proposes more threats and is carried strategically to leverage emotional appeals and divisive topics to manipulate public opinion and reinforce biases. In this research, all types of misleading content will be referred to as false information, given the complexity of understanding the actual intent behind sharing it.

Social media platforms like Twitter have changed the way people share information and allowed rapid dissemination of content to users from different locations across the world. Although these platforms make it easier for users to share information, they also enable users to share false information widely and uncontrollably. To clarify this, algorithms of all social media platforms are designed to prioritize and personalize content according to the user behavior to maximize engagement (Dujeancourt & Garz, 2023). These algorithms, particularly on Twitter, rely on engagement metrics such as likes and replies to determine which content appears on the user's feed. In fact, technological advancements enabled developers of these conglomerates to suggest content based on the time spent by users on each post (Farsi, 2021). Twitter's ranking system still prioritizes tweets with higher engagement, meaning that tweets will become more visible as the number of likes and retweets increases. In the context of false information, this looks threatful as these kinds of narratives rely on emotions to deceive people.

Furthermore, echo chambers and filter bubbles on social media expose users to content that matches their existing beliefs and desires. Echo chambers, as defined by Song et al. (2024), are homogeneous clusters created in selective social media environments where users with matching interests receive reinforcing feedback to strengthen their beliefs and limit their exposure to different views. Similarly, filter bubbles are personalized online environments where content is automatically curated and shown for users with matching beliefs to strengthen their views and hide opposing opinions (Ross Arguedas et al., 2022). This selective exposure, although sometimes useful, becomes more dangerous under the scope of false information dissemination. Elon Musk, the CEO of X (previously Twitter), hinted about these new algorithmic advancements when he suggested that the use of hashtags is no longer needed (Smith, 2024).

During global health emergencies, managing the flow of information is as critical as managing the outbreak itself. Previous crises, like the outbreak of Ebola in West Africa and the H1N1 influenza pandemic, highlighted the threats of false information (Allgaier & Svalastog, 2015; Chowdhury, Khalid, & Turin, 2023). False information usually spreads faster in times of emergencies, as people spend more time online looking for major news and announcements. This widespread of false information, often created by confusion and fear, leads to behaviors that dented public health efforts in containing the virus (Rodríguez et al., 2020). The spread of different and inaccurate claims about the origin of the disease, potential treatments, and its severity has weakened public trust and driven harmful actions.

The challenges posed by the dissemination of false information have intensified as the world became more interconnected. This is where the term "infodemic" comes on-stream, a term created by the World Health Organization (WHO) in 2003 during the SARS outbreak. It describes the rapid spread of accurate and inaccurate information that possibly led to mistrust among the public during a disease outbreak (Pian, Chi & Ma, 2021). An infodemic worsens the impact of a health crisis, not necessarily through increasing fatality, by distorting public perceptions and obstructing public compliance with health measures (Chen & Fu, 2022).

Circulating false information on social media platforms imposes many problems (Marwick, 2018). It can create panic among the public and influence their perceptions, often leading to unsafe behaviors that might be dangerous to their health and safety. For example, false information about the efficacy of certain treatments or specific health measures has been widely discussed and believed by many people (Van der Meer & Jin, 2020). The COVID-19 pandemic illustrates how fast false information can circulate and how challenging it is to retrieve public trust after the spread of such narratives (Eysenbach, 2020). This phenomenon highlighted the need for extensive research that tackles false information during pandemics, its nature and its impact on public compliance.

During the COVID-19 outbreak, Twitter emerged as a key medium for public discussions about the virus, providing real-time updates and a space for users to share their beliefs, behaviors, and concerns. As WHO declared the outbreak as a global pandemic on March 11, 2020 (Cucinotta & Vanelli, 2020), people had already turned Twitter into a breeding ground of false information (Kouzy et al., 2020). They began spreading conspiracy theories and unverified health advice that contradicted public health efforts. This combination of legitimate and illegitimate information poses significant challenges to public health authorities whose main goal is to ensure public safety (Rathore & Farooq, 2020).

The spread of false information does not only lead to the dissemination of false facts but also creates new discourses that are promoted by these falsehoods and sometimes overlooked by public health agencies (Vasconcellos-Silva & Castiel, 2020). These discourses might introduce new ideas that affect public opinions and attitudes. In the context of false information, these discourses include conspiracy theories about the virus, its origin, and its severity to finally suggest that this pandemic is a “hoax” (Bolsover & Tizon, 2020). Such discourses target existing societal fears, political ideologies, and anxieties which make them hard to track and resistant to correct (Leão et al., 2021).

Interpreting and analyzing these discourses are essential for developing strategies and public policies to limit the spread of false information during public health crises. By identifying the most common narratives during pandemics, public health officials and policymakers can design better health initiatives and become more prepared in the future in cases of new outbreaks. This research will contribute to the ongoing research on crisis management by considering COVID-19 as a case study, analyzing related false information narratives on Twitter, and interpreting the most persistent discourses and frames.

The primary goal of this research is to identify and analyze the most common false information discourses related to health pandemics on Twitter. Using text mining techniques, the study aims to uncover the characteristics of key narratives containing false information. By analyzing these narratives, the research will provide insights about the themes that contain false information and how they evolved during the first months of the pandemic. Also, this research will explore how these narratives were structured and portrayed to understand the strategies taken by different people to amplify the false narrative and affect public trust.

The objective of this study is to provide a clearer picture of the main false information discourses during health pandemics, including recurrent patterns in dissemination and prevalent subtopics. This research will contribute to public policy by offering insights that can help create more effective strategies and interventions during future pandemics. The analysis of our results will provide policymakers and public health experts with a better understanding of the nature of false information, allow them to develop better interventions, and ensure public compliance with their measures. This research aims to aid public health in mitigating the spread of false information and its associated risks.

The significance of this research extends beyond the scope of COVID-19 pandemic to future health crises, when the spread of false information can again impose similar challenges that undermine public health efforts. This research has the potential to offer insights into how future pandemics might be affected by the spread of false information and how these narratives alternate and evolve in importance over time. Eventually, this would allow public health officials to counteract false information narratives early and promote accurate health information more effectively.

Our research is grounded in Framing Theory, which explains how information is designed and structured to impact public perception. Frames present information in a way that emphasizes certain aspects while downplaying others, thereby guiding how individuals interpret narratives and react to them. Applying this theory will enable us to better understand how false information narratives reflect and influence how people perceive the pandemic.

This study addresses the following research questions:

a. What are the most prevalent false information discourses related to COVID-19, and how can they be categorized using text mining techniques?

b. What are the linguistic differences used in false information narratives, and how are they related to framing strategies?

Having established the research objectives and the significance of this study, and to better understand the foundations of this research, the following section reviews recent literature related to false information, its dissemination, and its dynamics.

2. Literature Review

Numerous studies have examined the spread and impact of false information related to COVID-19 on social media platforms. These studies offered valuable insights into how these narratives are constructed, disseminated and received. The following literature review summarizes the key areas of research related to this topic, presenting findings into false information discourses, dissemination mechanisms, and detection methods.

To ground this study within a conceptual framework, this literature introduces the base theory of this research which is the Framing Theory. This theory will help us understand how false information is amplified and framed. This theoretical approach acts as a solid foundation to analyze how certain narratives influence public perception and why specific frames resonate more than others. The next section provides an overview of the framing theory, its key principles, and its relevance to false information research in social media.

Theoretical Framework: Framing Theory

Framing Theory, initially introduced by Erving Goffman and later expanded Robert Entman, explains how the framing or presentation of information shapes public behaviors (D'Angelo, 2017). In the context of health-related false information, frames are often created by hashtags, key words, and emotionally charged language for the purpose of evoking strong reactions. These frames shape how the public perceives false information and reacts to it.

This study aims to analyze the framing strategies followed by users to uncover the characteristics of false information disseminated during pandemics. For instance, false information related to unproven cures could emphasize false hope while those related to approved vaccines might be framed through certain keywords that invoke distrust and uncertainty of public health institutions. Understanding the role of these framing techniques is significant for identifying the most prominent false information narratives and their role in influencing public perceptions during health crises.

The ability to understand and address these frames is vital for ensuring that public health response during future pandemics is not weakened by false narratives. By understanding which of these frames resonate the most with the public and how they influence engagement, public health officials and policymakers can design their initiatives to counteract false information, potentially limiting their spread across people. This research contributes to building a clear and effective framework for combating health pandemics to ensure that public behaviors are established on accurate evidence-based information rather than misleading fake news.

Wang et al. (2024) highlighted how framing theory could be applied to analyze the dissemination on false information on social media platforms. They pointed out that false information often uses problem definitions, causal interpretations, moral evaluations, and treatment recommendations to influence public behavior. This signifies a growing trend in false information analysis that leverages the Framing Theory to understand how narratives are strategically constructed to amplify reach and shape public opinion.

Mohammadi et al. (2022) also studied how COVID-19 false information narratives evolved over time, focusing on how prevalent discourses shift in prominence during different stages of the pandemic. They utilized the framing analysis to uncover the thematic patterns and structural elements, eventually providing a framework to analyze the content of false information and the context in which it spreads. They explained how framing strategies leverage emotional triggers and persuasive messaging to sustain and capture audience attention.

Building on these two foundations, this research applies the Framing Theory to investigate how false information discourses were constructed and framed on Twitter. It examines the use of keywords and narrative structures as framing tools and analyzes their amplification. By identifying the most prominent themes and their evolution during the first months of COVID-19, this study aims to provide insights about the role of framing strategies in manipulating and shaping public perception during pandemics.

The following section provides a detailed review of the literature, highlighting the themes that were most common to false information discourses, their dissemination patterns, and detection methods to position this study within the broader academic context.

2.1. COVID-19 False information: Themes and Narratives

Recent literature has focused on identifying the nature of COVID-19 false information on social media platforms. Studies have identified different related information that is based on false claims, such as conspiracy theories, unproven remedies, and political agendas. For instance, Chen et al. (2020) tracked the discourses related to COVID-19 pandemic by analyzing the most frequent hashtags and terms appearing during times of major global announcements. Moreover, Balasubramaniam et al. (2021) employed spatio-temporal topic dynamics, which seemed very efficient in analyzing how false information narratives change across distinct stages and different geographical locations. By using matrix tensor factorization, they revealed how false information patterns shifted over time and space.

Conspiracy theories were among the most prevalent themes in COVID-19 false information research. Smith et al. (2021), Nguyen et al. (2021), and Van der Meer and Jin (2020) concluded that most of these theories were included in discussions about the origin of the virus, where people debated whether the virus was naturally spread or artificially created. Bolsover and Tizon (2020) found out that describing the pandemic as a “hoax” gained much attention on social media. These narratives persisted as users truly believed that credible media agencies abstained from sharing the real story due to political pressures. Ahmed et al. (2020) added to this by showing how false information narratives focused on unverified treatments like “hydroxychloroquine” and disinfectants that contradicted with public health recommendations.

2.2. Politically Motivated False Information

Wang et al. (2021) were able to detect politically motivated false information in the United States. They showed how political speeches often lead people to downplay the pandemic severity even when cases surged. They explored how these narratives were used to decrease compliance with public health advice and politicize discourses related to mask mandates and vaccines. Influential users were found to amplify the reach and spread of fake news, which stands as a real concern to public health organizations. Therefore, it was concluded that politically motivated false information played a critical role in undermining public trust in official policies and measures.

Nguyen et al. (2021) further highlighted that exposure to false information led to decreased trust in public health institutions and lower compliance with health measures. The results of their survey showed a strong direct link between online false information and offline behaviors. Similarly, Kim and Lee (2021) underlined the significant impact of politically motivated false information on public perception. They found out that these types of narratives contributed to increased polarization and impacted public support for health interventions.

2.3. Dissemination Dynamics of False Information

Lee et al. (2021) and Brown et al. (2020) investigated the dissemination patterns of COVID-19 false information on Twitter. They discovered that false information tends to spread through echo chambers, where users are entitled to engage with content that most aligns with their existing beliefs. Through network analysis, they revealed that false information clusters within certain communities, where fact-checking efforts are minimal.

Gisondi et al. (2022) analyzed the spread dynamics of COVID-19 false information by analyzing Twitter engagement metrics. They found out that false information narratives often received higher engagement, allowing it to gain more traction and spread faster than other posts. This difference in engagement poses another threat for public health officials, where controlling these narratives on social media platforms becomes hard to achieve. This puts more pressure on policymakers and urges them to design effective strategies to combat the spread of false information.

Kim et al. (2021) conducted a time analysis and discovered that high false information spread coincided with major public health announcements, meaning that users who intended to share false information to deceive public perception strategically chose times of high public uncertainty. They timed their activities during critical periods to maximize the reach of false information and complicate public health efforts.

2.4. Sentiment Analysis of False Information

Several studies have tackled the emotional polarity of false information discourses. Nemes and Kiss (2021) used deep learning models to analyze how sentiment shaped these narratives. Surprisingly, they found a predominance of positive sentiment despite the ongoing spread of false information. Their findings emphasize the emotional appeal of certain false information narratives and how it supported the spread of fake news among many users.

Liu et al. (2021) explored how certain false information narratives used fear, anxiety and anger to manipulate public emotions. For example, false information about COVID-19 treatments often included fear emotions about side effects or government overreach. This emotional framing aided the spread of false information as it tapped into public anxieties and doubts.

2.5. Detection and Classification of False Information

There has been significant research on detecting and classifying false information content on social media platforms. Alonso et al. (2021) categorized over 5 million tweets as credible or misleading using machine learning. They found out that supervised machine learning methods like Random Forest and Support Vector Machines are effective in detecting false information based on linguistic characteristics. Similarly, Zhao et al. (2021) applied a hybrid model of deep learning and unsupervised clustering to categorize false information into relevant themes like politics, medical issues and conspiracy theories.

Wong and Zhou (2020) used ensemble learning models that utilized both tweet content and network structure. They improved false information detection as their model outperformed traditional methods by identifying the influencers who played significant roles in spreading misleading content. Chen et al. (2020) identified the most recurring false information themes and analyzed how they target different population groups. They signified that different theme, such as conspiracy theories and vaccine reluctance target specific demographics.

2.6. Framing of False Information

The framing of false information related to health pandemics focuses on the critical role of social media in shaping public perceptions and amplifying false information. Tsao et al. (2021) found that infodemics is a central theme when it comes to global pandemics, emphasizing the spread of false information and conspiracy theories through platforms like Twitter and Facebook. They highlighted that false content could lead to widespread confusion and harm, especially when amplified by viral sharing. Malecki et al. (2021) added to this by grounding their discussion in risk communication theory, introducing the "hazard and outrage" framework. They explain how public responses are shaped by scientific facts and emotional reactions. They concluded that social media plays a dual role as both a tool for communication and a medium for false information that increases public outrage and undermines health initiatives.

Wicke and Bolognesi (2020) explored the metaphorical framing of COVID-19 false information on Twitter. They identified recurring rhetorical language like war, storm, and monster that described how the public perceived the pandemic. Their analysis showed that war metaphors were among the most used, especially in tweets regarding treatment and containment, highlighting a combative and urgent framing of the crisis. However, they claim that this framing may be harmful when discussing other aspects like social distancing or emotional well-being. Together, the articles demonstrated that framing has significant implications on public behavior and must be accompanied with strategic communication during crises.

The literature review has demonstrated that false information during health crises, particularly the COVID-19 pandemic, relies heavily on emotional appeals and narrative structures to shape public perception. Existing studies provided significant insights into how false information spreads by exploring themes like political motivations, sentiment analysis, dissemination patterns, and classification methods. However, there are still some gaps in exploring how framing strategies influence user behaviors and amplify false information.

The following Methodology section introduces the research design, data collection process, and topic modeling techniques used to address the study objectives and fill the specified gaps.

3. Methodology

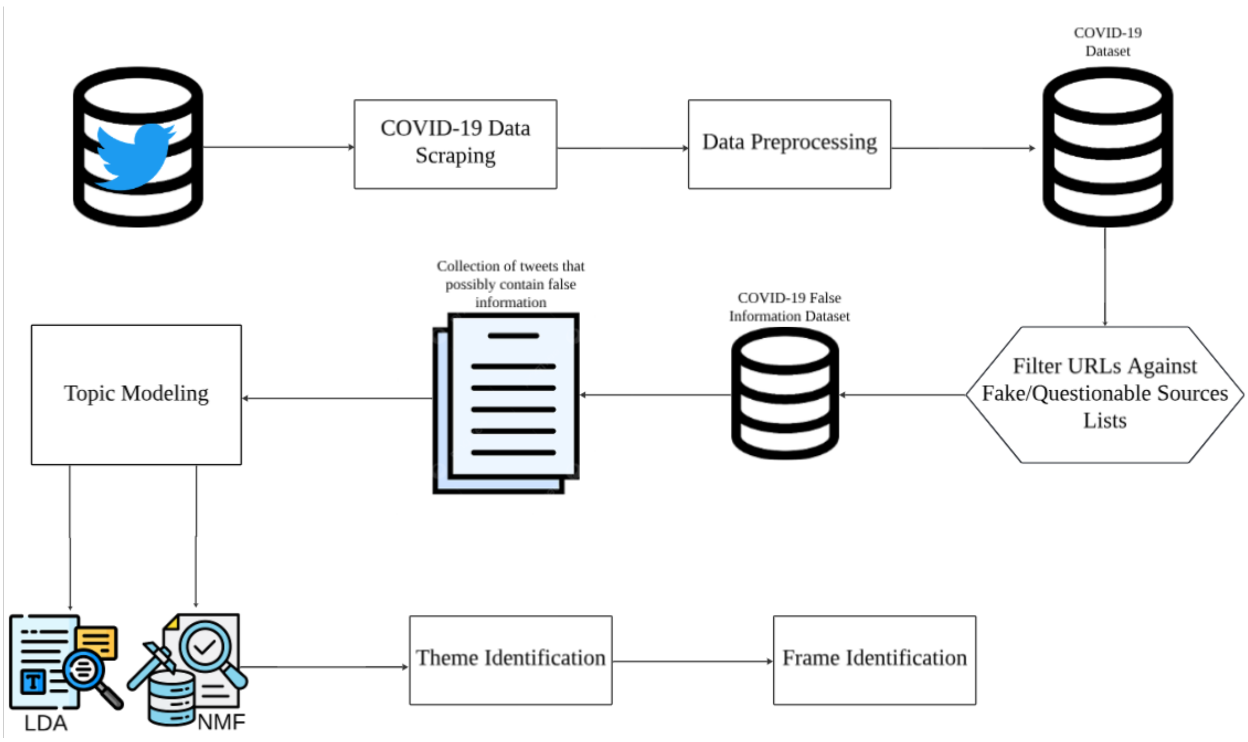


Figure 1. Research Workflow

3.1. Research Design

This study uses a cross-sectional research design to analyze COVID-19 false information on Twitter, focusing on tweets posted between March and July 2020, during the initial outbreak of COVID-19. The first objective of this study is to identify and analyze the most prominent discourses found in Twitter conversations that possibly contain false information. To identify possible false information, and as shown in Figure 1, we flagged tweets that include links from unreliable websites. As we were not able to obtain private access to fact-checking websites' lists, we relied on two publicly available lists of unreliable and questionable websites: False, Misleading, Clickbait-y, and/or Satirical News' Sources by Melissa Zimdars and MediaBias Fact Check list of questionable websites.

Figure 1 shows the overall workflow of our research. This study combines quantitative techniques with qualitative interpretations to analyze false information narratives related to COVID-19 on Twitter. First, we identified the prevalent discourses by applying Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF). Qualitative analysis was then applied to the results to categorize these narratives and discover the framing strategies used to amplify false information. This integrated approach provides a strong framework for studying the content and context of false information, providing interpretations about how framing techniques impact visibility and reach.

3.2. Data Collection

Twitter data was used to analyze public discourse during the early stages of the COVID-19 pandemic, covering the period from March till July 2020. This period was characterized by widespread uncertainty and rapid spread of accurate and inaccurate information. An existing dataset, the "COV19Tweets dataset" created by Lamsal (2020), collected tweet IDs related to COVID-19 since October 2019. We rehydrated over 182 million tweets using the twarc library and stored full tweet content and its characteristics in a MongoDB database. Data preprocessing involved removing retweets, identified by the "RT" tag, to create a master dataset that contains 70 million unique tweets. These tweets were compiled according to their publication month and by conversation ID to facilitate structured analysis. Other metrics were also extracted and they are presented in Table 1.

Metric	**Definition**
Tweet ID	Unique identifier for the tweet
Conversation ID	Unique identifier for the conversation thread
Text	The body of the tweet including all mentions and hashtags
Expanded URL	The original URL referred to in the tweet
Retweet count	The number of times the tweet was retweeted
Reply count	The number of replies the tweet has
Like count	The number of likes the tweet received
Quote count	The number of times the tweet was quoted
Impression count	The number of views the tweet received
Username	The name of the Twitter user
Followers count	The number of followers the user has
Tweet count	The number of tweets the user posted
Verified	The verification status of the user
Location	The self-reported location of the user
Country	The real-time GPS-based location based on latitude and longitude coordinates

Table 1. Attribute Definitions

3.3.Data Sources

The Zimdars list was prepared by Melissa Zimdars, a communications professor at Merrimack College. She created a list of websites that contain different types of questionable sources ranging from fake news websites to satirical ones (Zimdars, 2016). The list was first intended for educational purposes until it gained much attention and became a reference for fact-checkers. Although the list contains different kinds of news sources, we only considered the websites tagged as “fake news”. Despite some criticism on the inclusion of certain websites, Zimdars list still serves as a significant tool for false information researchers and journalists. Sharma et al. (2020) relied on this list to classify and analyze tweets that contain false information. It allowed them to analyze the spread of misinformation narratives related to COVID-19 on Twitter. Similarly, Ai, Liu, and Hirschberg (2023) used this list to filter tweets related to unreliable sources and flag them as misleading.

As for the MediaBias/Fact Check list of questionable websites, it is publicly available on their official website. This list includes websites that display extreme bias, publish fake news, and lack credible sourcing and transparency. It is also widely used by researchers to filter and analyze content that possibly contains false information. Chen et al. (2022) and Sharma et al. (2020) both used the Media Bias/Fact Check list to detect tweets containing links from questionable sources. They analyzed these tweets to discover the most prevalent false information narratives and analyze their engagement trends during the COVID-19 pandemic.

3.4. Data Preprocessing

The preprocessing phase was essential to clean and transform data to ensure it was suitable for quantitative analysis. We started by removing duplicate tweets to eliminate redundancy and standardizing text by converting it to lower case, removing special characters, and normalizing accented characters. We then expanded contractions to their original form and then removed hashtags, mentions, and URLs to prepare conversations for topic modeling. We also applied stopword removal and lemmatizations to delete all non-informative words and transform words into their base forms to enhance consistency. After that, texts were tokenized into individual words and n-grams were generated to capture common phrases related to false information narratives. The tokenization was extremely important to construct a document-term matrix through topic modeling.

3.5. Topic Modeling

Topic modeling is an unsupervised machine learning technique used in Natural Language Processing (NLP) to discover hidden thematic structures in large texts (George & Sumathy, 2023). It was first developed in the 1990’s after several advancements in probabilistic modeling that enabled researchers to extract meaningful patterns from unordered documents (Jelodar et al., 2019; Churchill & Singh, 2022). Over time, various topic modeling techniques have been developed to accommodate specific data relationships and structures (Lafferty & Blei, 2006; Yan et al., 2013; Li & McCallum, 2006). Topic modeling was found to be effective in analyzing social media posts as it helped in identifying recurrent narratives and categorizing them into meaningful themes (Ramage, Dumais, & Liebling, 2010).

The most common topic modeling techniques include LDA, which is a generative probabilistic model that discovers underlying topics in texts by considering each document as a collection of topics, where each topic is characterized by a distribution over words (Blei, Ng, & Jordan, 2003). In simpler words, it transforms documents into mixtures of topics. It uses the "bag of words" approach to generate topics that are represented as probability distributions over a number of words (Hong & Davison, 2010). Maier et al. (2021) describe how LDA models a corpus by decomposing it into two probability matrices:

1. Word-Topic assignment matrix ϕ (K×V): It represents the probability $\phi_{k,w}$ of each word w appearing in topic k to define topic characteristics.

2. Document-Topic assignment matrix θ (K×D): It represents the probability of topic k appearing in document d to classify documents according to dominant topics.

The second common topic modeling technique is NMF, which is a non-probabilistic model that uses matrix factorization (Egger & Yu, 2022). It relies on TF-IDF, which is a measure of evaluating the importance of a word in a collection of documents. NMF decomposes the term-document matrix (A) into a product of a terms-topics matrix (W) and topics-documents matrix (H) (Chen et al., 2019; Egger & Yu, 2022). Wang and Zhang (2013) explain how NMF models a corpus by factorizing the non-negative term-document matrix X into two lower-rank non-negative matrices:

1. Basis Matrix W (K×V): It represents the contribution of each word w to topic k to define topic characteristics.

2. Coefficient Matrix H (K×D): It represents the weight of topic k in document d to classify documents according to dominant topics.

Both LDA and NMF will be applied and tested to uncover the dominant themes and narratives found in the conversations. These models will enable us to uncover the hidden structures of false information discourses and categorize them without prior knowledge of the topics. Our analysis will reveal the most prevalent discourses within the dataset and how they change across conversations and months. Only conversations with more than 10 tweets are considered to sustain contextual richness, each containing at least one tweet that references an unreliable website. This approach allows us to correctly identify the discourses that contain false information, even if other tweets belonging to the same conversation do not explicitly include false information. Both models will be tested with 10, 15, 20, 25, 30, 35, 40, and 50 topics to determine the optimal number of topics (k) that best represent the false information themes across conversations.

Few configurations were tested for each number of topics to ensure the interpretability of these topics. Coherence and perplexity scores were used to evaluate the model's performance. According to Newman et al. (2010), coherence scores measure how interpretable and semantically related the keywords of a topic are. They evaluate whether these words have a meaningful theme that humans can label. Röder, Both & Hinneburg (2015) stated that they range between 0 to 1. A score of 1 indicates perfect coherence while 0 indicates no coherence between words. Moreover, Neishabouri, Desmarais and Montreal (2021) explained that perplexity scores measure how well a probabilistic model predicts unseen data. Researchers often study perplexity scores to evaluate different models and determine the optimal number of models by choosing the one with lower perplexity.

As for the NMF model, we calculated the coherence and other scores for each number of topics. The performance of the model was evaluated using coherence and reconstruction error. Reconstruction error, as described by Ghaly and Laksito (2023), measures how well the result of the factorization of the non-negative matrix into matrices W and H can reconstruct the original matrix. The error is usually between 0 and 1, and a lower score means that the model is better at data reconstruction (Saha & Sindhwani, 2012). Coherence score definition remains the same for this model.

In both models, we applied similar hyperparameter preprocessing techniques to ensure comparability and reliability. For NMF, we configured the TF-IDF with max_df=0.96 and min_df=20, aligning with the no_above and no_below thresholds used in the LDA dictionary filtering. This process was important to exclude extremely common and rare items from both models in order to improve topic quality. Additionally, for both models, we tested multiple numbers of topics (k), that range from 10 to 50, to evaluate model performance. These choices allowed us to make meaningful comparisons between LDA and NMF and choose the one with better results.

We will be looking at the average score for the previously defined metrics for each number of topics across months in order to choose the optimal number of topics for each model. Once this number was determined for each model, we explored the topics and their keywords to proceed with topic labeling. This manual process ensures that these topics contribute to meaningful narratives rather than being clusters of words (Ahammad, 2024). We examined the top ten keywords for each topic under each model. During this step, we noticed how LDA and NMF differ in allocating words to topics as LDA keywords have the highest probabilities in each topic and NMF keywords have the highest weights in the topic-word matrix. Topics that contained similar words and overlapping narratives were manually grouped under the same theme. After rounds of assignments, we compared and merged topics in different months to achieve consistency in theme categorization. By merging topics into fixed themes, we confirmed that these themes are representative of the false information narratives included in the dataset.

3.6. Manual Labeling and Validation

A random sample containing 5% of the total dataset was then considered for manual labeling, yielding 650 conversations out of the 13,038 conversations. To ensure a representative distribution of false information discourses over time, we chose an equal proportion of unlabeled conversations from each month to minimize biases associated with temporal variations in topic prevalence.

Each conversation was manually classified into two themes as twitter conversations often combine multiple narratives. Labeling was done based on the linguistic and contextual characteristics of conversations and without prior knowledge of the topic modeling results to make sure that model assignments did not bias the classification process. After an initial round of labeling, ambiguous conversations that were not assigned to a second theme were reviewed to ensure that each conversation manual assignments can be compared to the two themes with highest probabilities in LDA and the two themes with highest weights in NMF.

To systematically assess the performance of the topic models, we calculated the accuracy of topic models assignments. This accuracy computes the percentage of conversations where at least one of the two model themes matched a manually labeled theme. A higher accuracy score implies better thematic alignment between model-generated and human-labeled categorizations.

Accuracy Formula = $\frac{Number\ of\ Matching\ Conversations}{Total\ Number\ of\ Conversations}$

The results of this validation process directed the final model selection, ensuring that the most accurate and interpretable topic modeling approach was adopted for further analysis. This manual labeling and validation framework contributed to establishing a benchmark for false information discourse analysis in social media contexts and strengthened the credibility of the model-generated false information themes.

3.7. Conversation Assignments to Themes

To ensure a structured and meaningful analysis of the themes generated by both models, conversations were assigned to themes according to different criteria for LDA and NMF, considering the difference in which each model represents topic distributions. Since further analysis relies on dominant and clear themes per conversation, we applied threshold-based filtering for LDA and weight-based approach for NMF. In LDA, each conversation is represented as a probability distribution over multiple topics, meaning that one conversation will have different probabilities for multiple themes. To ensure that we only consider conversations with dominant themes for further analysis, we implemented a probability threshold of 40%. This means that conversations were assigned to themes that have a probability higher than 40%. This approach allowed us to filter out an acceptably low number of conversations that tackled several themes as they do not provide meaningful insights in later analyses. We hereby prioritized conversations that strongly represent the false information themes generated by LDA rather than including those that contain a mix of multiple topics. In NMF, each conversation is characterized by a weighted representation across multiple topics, where weights do not necessarily add up to 1 compared to the probabilities for themes in LDA. We assigned conversations to the theme with the highest cumulative weight. This again ensures that each conversation was assigned to the most dominant false information discourse.

By using different strategies for assigning conversations to themes in LDA and NMF, we confirmed that our final dataset includes only thematically strong and relevant conversations, enhancing the consistency of subsequent findings.

3.8. Temporal Analysis for Theme Evolution

To analyze the evolution of false information themes across months, we tracked the number of conversations assigned to each theme in each month. By looking into shifts in thematic prevalence, we will be able to reveal which false information discourses gained or lost attention as the pandemic progressed. Certain themes remained dominant across all months, indicating a strong presence of these false information narratives, while others only peaked in alignment with public health announcements.

As the number of conversations might not be the best indication of the theme presence, a theme strength score was computed which equals the sum of weight or probability assigned to each theme. Analyzing these trends provides insights into the continuity of certain false information narratives. While some themes became marginal as public interest changed, others increased in dominance in response to new developments. These patterns help us in recognizing how false information adapts to changing circumstances, what factors drive its amplification, and what frames enhance its influence over time. These temporal changes also offer insights into public health and policy as it enables them to predict the timing for interventions in future pandemics in order to limit the spread of false information.

3.9. Term Frequency Analysis

To understand the linguistic characteristics within each false information theme, term frequency analysis was conducted. Term Frequency (TF) refers to the number of times a specific word appears in a document (Azam & Yao, 2012). This analysis allowed us to look for the most occurring words in each theme in each month, providing insights into how each narrative was linguistically structured and how they evolved over time. In each month, we identified the top 50 most occurring words for each theme. This selection allows us to discover the recurrent terms shaping each false information theme and enables us to quantitatively examine the term or language trends.

Word clouds were then generated for each theme in each month to analyze the persistence and evolution of false information discourses. Certain words appeared consistently in each month, emphasizing the stability of some narratives, while others emerged or declined over time, signaling changes in public perceptions. These findings act as a foundation for further framing identification, where the repetitive use of specific linguistic choices implies the existence of unique frames that reinforce certain narratives.

3.10. Framing Strategies

To understand how false information is framed to influence public perception in each theme, we decided to analyze the most frequently occurring words within each theme. While some themes rely on emotionally charged terms to motivate distrust in public policies, others employ scientific skepticism and incorporate specific terms that dispute established medical practices. These differences in linguistic patterns allow us to identify strategies that are used to reinforce and legitimize false information narratives without need for sentiment or emotional analysis.

As specific terms continuously appear across months, we realize that some framing strategies persist during the pandemic while others only appear in response to global updates. Themes related to politics often used politically charged language to frame these narratives as ideological struggles rather than a health-related issue. By exploring these framing strategies, we will be able to understand how false information adapts to different contexts and how they become more difficult to counter.

4. Results

4.1. Topic Modeling Performance

At first, we compared the evaluation metrics inside each model to determine the optimal number of topics. Since the performance of the model should be consistent across all the five months, we decided to consider the average values of these metrics.

For LDA, the mean coherence and log perplexity scores were calculated and presented in the table below.

Number of topics	Mean Coherence	Mean Perplexity(log)
10	0.338	-7.043
15	0.348	-7.036
20	0.364	-7.035
25	0.361	-7.042
30	0.36	-7.04
35	0.36	-7.045
40	0.359	-7.049
50	0.359	-7.051

Table 2. LDA Evaluation Metrics

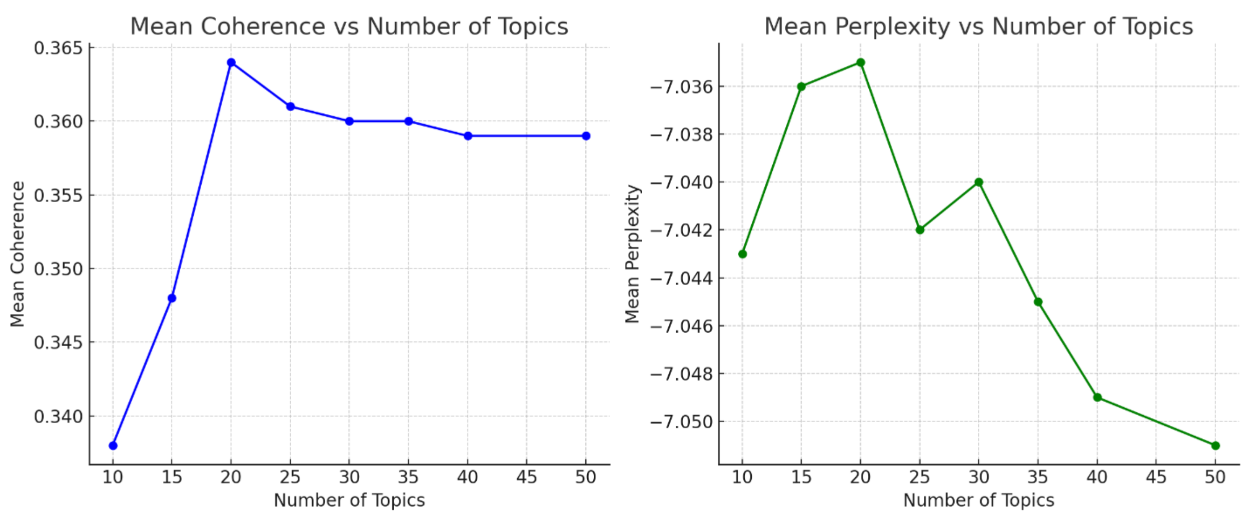


Figure 2. Comparison of LDA Coherence and log Perplexity Scores by Number of Topics

As can be seen from Figure 2, coherence score increased sharply from 10 topics to 20 topics, and the improvement peaked and stabilized at 20 topics (0.364), then it slightly decreased in performance at 25 topics, but the difference was minimal (-0.03). The score decreased after that as the number of topics increased. As for log perplexity, which favors lower values, showed continued improvement as the number of topics increased beyond 20 topics. At 25 topics, log perplexity reached -7.042, compared to -7.035 at 20 topics and -7.04 at 30 topics. This indicates that the model had a marginal yet noticeable improvement at 25 topics. Based on these observations, 25 topics were selected as the optimal number of topics for the LDA model, balancing model performance and complexity, and ensuring a delicate representation of the data.

As for NMF, the mean coherence scores and reconstruction errors were calculated and presented in Table 3.

Number of Topics	Mean Coherence	Mean Reconstruction Error
10	0.542	42.055
15	0.561	41.532
20	0.569	41.120
25	0.576	40.809
30	0.585	40.541
35	0.589	40.300
40	0.588	40.079
50	0.595	39.681

Table 3. NMF Evaluation Metrics

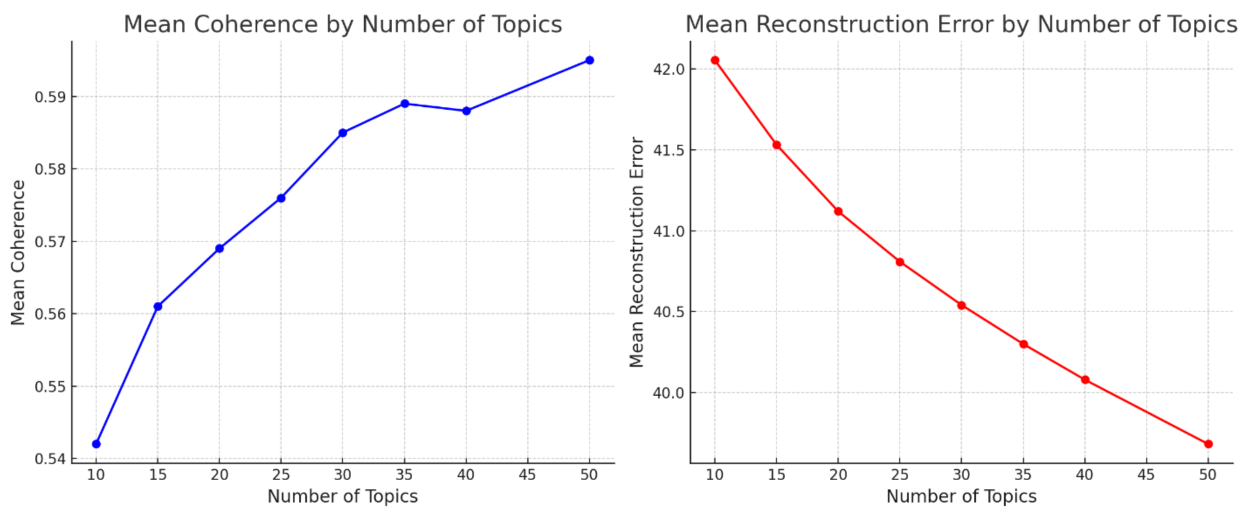


Figure 3. Comparison of NMF Coherence and Reconstruction errors by Number of Topics

As can be seen from Figure 3, coherence score showed a steady increase from 10 topics to 35 topics with a peak value of 0.589, right before it slightly decreased and then reached a maximum value of 0.595 at 50 topics. As for the reconstruction error, which favors lower values, improved consistently as the number of topics increased. The reconstruction error graph shows a linear relationship between reconstruction error and number of topics. The error started dropping from 42.055 at 10 topics, maintaining values between 41.0 and 40.0 at 25, 30, 35, and 40 topics, until reaching a minimum of 39.681 at 50 topics. Eventually, 35 topics were selected as the optimal number of topics in the NMF model as it maintains a good balance between coherence and reconstruction error, providing a well-structured representation of data and minimizing redundancy.

4.2. Topic Labeling

After selecting the optimal number of topics for each model, we proceeded by manually labeling each topic in each month and assigning it into a general theme. Aggregating subcomponents of larger narratives into specific themes allowed for a better interpretation of false information discourses. We examined the top 10 keywords in each topic to look for common patterns and recurring patterns between topics. Topics that included similar words were grouped under the same theme. For example, topics containing words like “Chinese”, “virus”, “originate”, “bat”, “lab”, “Wuhan” were all merged into the “COVID-19 Origin” themes. Similarly, words like “injection”, “bleach”, “vaccine”, “hydroxychloroquine”, “disinfectant” were all grouped under the “COVID-19 Treatments” theme.

Under both LDA and NMF, we identified 7 common themes, with common words and different numbers of topics. NMF produced one additional theme, which is present in 4 different months. The results are shown in Table 4.

Theme	March		April		May		June		July	
	LDA	NMF	LDA	NMF	LDA	NMF	LDA	NMF	LDA	NMF
Politics and Leadership	6	6	6	6	3	9	6	9	5	7
COVID-19 Severity	4	5	2	4	4	3	4	5	4	6
COVID-19 Origin	2	3	1	2	3	2	2	1	3	2
COVID-19 Treatments	1	2	4	3	1	2	1	2	1	2
COVID-19 Impact	3	4	5	6	3	3	5	3	4	7
COVID-19 Global Response	3	3	5	4	4	5	3	6	2	4
COVID-19 Health Measures	6	4	2	3	7	4	4	3	6	2
COVID-19 & Society	N/A	3	N/A	2	N/A	2	N/A	1	N/A	0
Total	25	30	25	30	25	30	25	30	25	30

Table 4. Distribution of topics under topics models

As shown in Table 4, across the 5-month period, the number of topics assigned to each theme varied in both LDA and NMF, reflecting shifts in false information focus. In LDA, the “Politics and Leadership” theme remained stable for 4 months (5-6 topics) but dropped to 3 in May. However, under NMF, the number of topics remained stable in the same 4 months but peaked in May to 9, suggesting differences in how each model handles linguistic characteristics and extract topics. “COVID-19 Severity” theme remained stable across all months in both LDA and NMF models. Discussions about “COVID-19 Origin” in both models were relatively brief as the theme had a low number of topics. As for the “COVID-19 Treatments” theme, discussions were consistent across all months and under both models, with a slight increase in April. “COVID-19 Impact” theme was mostly represented by more topics under NMF than LDA, except for the month of June, where the number of topics under the LDA model was 5 compared to 2 under NMF. For the “COVID-19 Global Response” and “COVID-19 Health Measures” themes, the results fluctuated under both models to reflect changes in extracting topics with different algorithms.

Finally, the NMF model captured a new theme “COVID-19 & Society”, which was not found by the LDA model. It showed a gradual decline from 3 topics in March to 0 topics in July, which suggests that discussions about societal behaviors were more prominent in the early months of the pandemic.

Each theme was represented by several discourses. These discourses reflect the different false information narratives that were circulated during the early stage of the pandemic. Table 5 shows the distribution of discourses among themes.

Theme	Discourse
Politics and Leadership	COVID-19 as a political hoax
	Political leaders' pandemic response
	Use of pandemic for political gain
	Political agendas
COVID-19 Severity	COVID-19 vs flu
	Cases/deaths/hospitalizations
	Asymptomatic transmission
	Herd immunity
COVID-19 Origin	Virus conspiracies
	Virus as a bioweapon
	Government cover-ups
COVID-19 Treatments	Potential cures/treatments
	Vaccine development and trials
	Vaccine side effects and microchipping
COVID-19 Impact	Economic harm
	Public closures
	Nursing homes
	Prisons
COVID-19 Global Response	International responses
	Governments' incompetence
	Religious groups' pandemic handling
COVID-19 Health Measures	Testing
	Masks
	Social distancing
COVID-19 & Society	COVID-19 as religious test
	Marginalized communities
	Government digital surveillance

Table 5. Distribution of discourses within themes

4.3. Manual Validation of Theme Assignments

The results of the manual validation process confirmed that NMF outperformed LDA in terms of accuracy, demonstrating a better alignment with human-labeled categorizations. With an accuracy of 89%, NMF consistently produced theme assignments that were strongly aligned with the manual annotations. Although LDA achieved a reasonable accuracy of 85%, this indicates that LDA had a lower precision regarding theme assignments and suggests that NMF was more effective in capturing dominant themes across conversations. The higher accuracy of NMF considers the model more suitable for extracting meaningful discourses.

Given these findings and the differences in mean coherence scores shown in Tables 2 and 3, NMF was selected as the primary model for further analysis. Its superior accuracy and stronger thematic assignments made it the most reliable choice for studying false information discourses. By using NMF, the following examination of false information trends and framing strategies is grounded on the most robust and interpretable theme assignments. This decision guarantees that the analysis captured well-defined false information narratives, enabling us to understand how these discourses evolved, became popular, and amplified over time.

4.4. Theme Distribution Across Conversations

According to Table 5, roughly 2% of the total conversations extracted from the dataset were considered for this study. This approach ensures that all conversations have more than 10 tweets and contributes to the overall goal of the analysis. Below is a summary of the number of conversations considered for each month, compared to the actual number of conversations.

Month	Total Number of Conversations with false information	Total Number of Conversations with more than 10 tweets	Percentage of conversations with more than 10 tweets	Percentage of conversations across months
March	238,838	4,767	2.00%	36.56%
April	73,284	1,039	1.42%	7.97%
May	192,990	3,625	1.88%	27.80%
June	115,119	2,567	2.23%	19.69%
July	31,561	1,040	3.30%	7.98%

Table 6. Conversation Statistics per month

The number of conversations assigned to false information themes fluctuated substantially across the five-month period, reflecting changes in public discourse and false information focus. It is also worth noting that the number of conversations across months changed, and the percentage of conversations with more than 10 tweets varied as well.

Theme	**March**	**April**	**May**	**June**	**July**
Politics and Leadership	1752	353	1375	1021	404
COVID-19 Severity	1138	109	492	394	203
COVID-19 Origin	632	51	188	63	33
COVID-19 Treatments	123	110	289	113	57
COVID-19 Impact	203	170	305	141	153
COVID-19 Global Response	273	193	628	479	117
COVID-19 Health Measures	286	38	272	299	73

COVID-19 & Society	360	15	76	57	0
Total	4767	1039	3625	2567	1040

Table 7. Conversation Statistics per theme

According to Table 6 and 7, March recorded the highest number of conversations (4,767) across all months, followed by May (3,625), and June (2,567). “Politics and Leadership” was the most dominant theme across all months, represented by almost 40% of conversations in June. These numbers indicate that false information narratives related to the COVID-19 pandemic were centered around political responses and governmental actions. “COVID-19 Severity” was also a prevalent discourse, especially in March where it peaked with around 24% of total conversations. This signifies that people were discussing the severity of the pandemic, either downplaying or amplifying the health effects of the virus. “COVID-19 Origin” was mostly discussed in March as well, possibly meaning that people were debating about the origin of the virus when COVID-19 was declared as a global pandemic by WHO. As for “COVID-19 Treatments”, there was a notable increase in the number of conversations during May, suggesting a rise in discussions about possible cures. “COVID-19 Impact” did not have a strong presence in March but fluctuated after to reach a maximum of 305 conversions in May, which shows an increase in the discussions around the impact of the pandemic on the economy. Similarly, “COVID-19 Global Response” peaked in May, suggesting a rise in the discussions regarding the various measures taken by different governments to combat the virus outbreak. As for the “COVID-19 Health Measures” theme, most of its discussions were in June, with fluctuating number of conversations across other months, which indicates varying public engagement with false information related to safety protocols. Finally, the “COVID-19 & Society” theme appeared mostly in March, then started to decline until it disappeared completely in July. This shows how specific false information narratives faded over time while others persisted.

4.5. Term Frequency Analysis

The top 50 frequent terms were extracted for each theme in each month and a summary of the common words were illustrated in Table 8.

Theme	March	April	May	June	July
Politics and Leadership	virus, trump, hoax, president, american, democrat, obama	covid, trump, american, president, country, world, hoax, pandemic, briefing	covid, trump, death, president, state, case, country, pandemic, response	covid, trump, pandemic, american, rally, president, state, economy	covid, trump, pandemic, american, country, president, state
COVID-19 Severity	virus, flu, people, death, rate, spread, test, number, symptom, kill	covid, death, hospital, patient, died, rate, many, data, heart, disease	covid, death, flu, number, case, hospital, patient, rate	covid, death, case, number, hospital, flu, died, many, new, positive	covid, death, case, pandemic, positive, flu, rate, died, testing, immunity, data
COVID-19 Origin	virus, chinese, wuhan, racist, name, originated, started, blame	china, death, wuhan, lab, january, first, human, make	covid, china, trump, chinese, wuhan, blame, evidence	covid, china, trump, angry, chinese, blame, ugly, racist, damage	covid, china, american, mask, angry, blame, stop, population, fault
COVID-19 Treatments	virus, vaccine, drug, patient, treatment, chloroquine, cure	covid, trump, disinfectant, cure, bleach, treatment, drug, vaccine, injecting	covid, drug, hydroxychloroquine, vaccine, treatment, malaria, cure, effective, risk	covid, vaccine, drug, treatment, hcq, trial, cure, life	covid, study, trump, drug, patient, hydroxychloroquine, treatment, vaccine, fda
COVID-19 Impact	virus, trump, wall, stop, border, market, mexico, money	covid, case, home, hospital, patinet, life, care	covid, home, child, nursing, symptom, parent, family, wife, elderly, kid	covid, pandemic, home, nursing, crime, cuomo, elderly	covid, people, stop, school, job, case, child, home, work, open
COVID-19 Global Response	virus, india, italy, china, spread, government, immunity, health, quarantine, stop	corona, india, death, country, fight, spread, doctor, patient, china	covid, death, home, lockdown, government, country, hospital, india	covid, pandemic, lockdown, government, country, india, world, test, rate	covid, lockdown, pandemic, country, case, government, social, mask, distancing
COVID-19 Health Measures	virus, quarantine, mask, need, test, ppe, dying, spread	covid, testing, mask, case, home, beach, hospital, open	covid, mask, wear, home, spread, case, testing, nursing	covid, mask, wear, social, distancing, spread, testing, positive, patient	covid, mask, wear, hospital, social, distancing, spread, stop, risk

COVID-19 & Society	virus, trump, god, need, stop, spread, help, prayer	covid, harvard, app, need, money, get, care, aid, help, fund, give	covid, church, trump, death, god, virus, worship, service, care, pray	covid, black, protest, death, racism, white, community, blame, social	N/A

Table 8. Monthly key term examples per theme

4.6. Frames Identification

The frequent terms extracted for each theme in each month highlight recurring language patterns used in false information narratives. In the "Politics and Leadership" theme, the terms reflected a strong focus on political figures and leadership decisions during the pandemic. Across all months, terms like "Trump", "american", "president", and "country" remained dominant, indicating that false information narratives around this theme are mainly discussing statements from political figures and national response to the pandemic. There was also a notable presence of words like "hoax", "state", and "economy". These narratives evolved over time and discussed political briefings and rallies in later months.

In the "COVID-19 Severity" theme, discussions included words that highlighted mortality rates and case numbers. Common words such as "virus", "death", "flu", and "spread" suggest that false information narratives around this theme focused on minimizing or overstating the severity of the virus. The presence of words like "testing", "positive", and "immunity" suggests that people might be questioning the accuracy of testing and the long-term effect of the virus on the human body. Over time, discussions shifted from general concerns about case numbers to debates about mortality rates and credibility of official data.

In the "COVID-19 Origin" theme, discussions were shaped by words that are related to geopolitical and conspiratorial ideologies. Discussions in early months were dominated by words like "Wuhan", "Chinese", "bat", and "lab", suggesting that people were questioning the origin of the virus as the first cases appeared in China. As the months progressed, new words started to appear as "blame", "racist", and "evidence", reflecting continuous discussions around political accusations about the origin of the virus.

In the "COVID-19 Treatments" theme, false information narratives were mainly about potential cures and alternative treatments. Common words like "vaccine", "drug", "cure", and "treatment" remained prominent across all months, which indicates persistent desire to find medical solutions. Unverified medications like "hydroxychloroquine" and treatments like "bleach" and "disinfectant" appeared in different months, reflecting false information trends motivated by public political statements. In later months, discussions shifted from speculative treatments to vagueness about vaccine development.

In the "COVID-19 Impact" theme, false information mainly focused on the social and economic consequences of the pandemic. Early discussions included terms like "market", "border", and "money", which signals concerns about financial stability and restrictions. In the following months, there was an increase in words like "home", "school", "job", and "work" to highlight the false information about lockdowns, school closures, and economic disruptions. The presence of emotionally charged words like “stop” and “crime” implies that some narratives accused the pandemic as an attack on personal freedoms.

In the “COVID-19 Global Response”, false information suggests a focus on global efforts to handle the pandemic and the role of different governments and organizations in mitigating the virus spread. Common terms such as “government”, “country”, and “lockdown” highlight discussions about global policy decisions, while words like “India”, China”, and “Italy” indicate region-specific false information narratives. In later months, false information advanced to include concerns about vaccine distribution, testing policies, and the impact of international organizations which created public suspicion toward global health initiatives.

In the “COVID-19 Health Measures” theme, false information focused on the impact of protective actions such as mask wearing, social distancing, and testing. Common words like “mask”, “quarantine”, “spread”, and “distancing” indicate that false information narratives were mainly about the efficacy of these measures. The presence of words such as “need”, “open”, and “risk” suggests that public debates on mandates and restrictions were overly stated in false information discussions. Over time, the appearance of words like “hospital” and “patient” signals that false information about the necessity of health measures extended into skepticism about hospital protocols and patient care.

In the “COVID-19 & Society” theme, false information was mainly related to religion, community impact, and social movements. Early months featured words like “prayer”, “help”, and “god” indicating that some false information narratives framed the pandemic within religious contexts. Later months saw a shift toward protests, racism, and social justice, with words such as “black”, “white”, “protest”, and “community” emerging more frequently. The disappearance of this theme by July indicates that false information under this theme was either merged with other narratives or lost attention as the pandemic progressed.

5. Analysis

In the following analysis, we start by presenting the eight discourses separately. We then look at the changes of themes’ presence across the five months to identify key false information trends and explore how this generalizes on future pandemics.

5.1. Themes

1. Politics and Leadership: This theme included various discussions about politics all over the world with a notable concentration on U.S. politics. It reflects how political figures and policies became central topics in false information narratives. In the early months, these narratives focused on the reactions of political leaders with accusations and critiques against Trump, Biden and Pelosi. People were mainly debating how Trump was handling the pandemic and describing it as a “hoax”. Later, the focus shifted towards election-related narratives, with increased discussions about mail-in voting and Biden vs Trump. Discussions also expanded to include racial justice protests and their connections to COVID-19, along with relief funding and public policies. In June and July, false information narratives linked COVID-19 with

conspiracy theories (e.g. Hillary Clinton, Ghislaine Maxwell) and continued to focus on Trump's rallies, leadership failures and political consequences.

Example tweets: *"The WH has legal authority to prevent a private citizen from speaking about the spread of a contagious, life-threatening virus???l didn't realize we were officially governed under a fascist regime yet."*

"Trump doesn't want to tell Americans the truth about coronavirus. The man is afraid of losing elections and put thousands of Americans at risk, with his disinformation campaign!! #PenceIsAntiScience #COVID2019 #TrumpVirus #TrumpLiesMatter #LetCDCHandleCoronavirusCrisis."

"My grandma died recently of the wuhan flu and I can't believe these dems and their tricks. If it wasn't for Trump shutting down all flights and getting everyone tested, we wouldn't have known that the Dems had killed her".

2. COVID-19 Severity: This theme included narratives about the severity of the virus focusing on the reported positive cases and death rates. People were comparing the new virus to the flu and other common lethal infections. In March, they were downplaying the risks associated with the virus and questioning the mortality rates. Later on, there was an increased skepticism about inflated hospitalization cases. It was believed that hospitals were increasing cases on purpose and forging death certificates. In June and July, people discussed asymptomatic spread, herd immunity, and antibody testing, with many claiming that immunity has already developed.

   Example Tweets: *"I know. Its mostly the hysteria that is causing this. Seriously hundreds more are killed by household accidents then this damn flu. People are just cowards."*

   "It's fake, there is a virus but not as bad as they are saying. They are putting deaths down as covid when some people haven't even been tested. It's a plan for control and mass vaccination"

   "And try going online to find flu stats for this last season. Darn near impossible. Yet they can tell us what the COVID case and death counts are at any moment of the day."

3. COVID-19 Origin: This theme included narratives about the origin of the virus and scientific justifications. In March, there was heavy blame on China and the Chinese Communist Party (CCP), with racist accusations targeting the Chinese and Asian communities. In April and May, the discourse focused on lab-leak theories, with considerable mentions of Wuhan, bats, and research labs. By June, these discussions persisted with added anti-China sentiment and considerable mentions of the U.S., retaining a strong geopolitical tone. In July, these narratives remained prominent but incorporated more scientific language.

Example Tweets: *“In #China, a #Wuhan resident claims that "critically ill #coronavirus patients [are] sealed up in body bags while 😱alive😱, then sent for cremation.”*

“The problem is... Covid-19 was detected in the Seattle area prior to October of 2019. So it was HERE months before the theoretical "patient zero" in Wuhan China. It would be just as accurate to theorize the virus originated here. Research Dr. Helen Chu.”

“Now they’re cooking up some new pig virus. Like the “accidentally “ escaped Corona from Wuhan wasn’t enough. CHINA CN The country of the most virulent strains of the flu.”

4. COVID-19 Treatments: This theme included narratives about the possible cures of the virus and the need for vaccine development. In March, discussions were filled with optimism around developing a vaccine and growing attention to hydroxychloroquine as a potential cure. In April, false information escalated with dangerous suggestions like injecting disinfectants, which was mentioned by Donald Trump, and continued promotion of hydroxychloroquine. In May, these treatments remained present but tied to doctors and lupus patients, while vaccine conspiracies increased in volume. June saw a persistent focus on hydroxychloroquine trials and an increase in anti-vaccine narratives linking Bill Gates and Dr. Anthony Fauci. By July, people kept discussing possible treatments and vaccine development, while focusing on international actors like India’s ICMR and Modi.

   Example Tweets: *“Dr. Vladimir Zelenko, a board-certified family practitioner in NY, has now treated 699 Covid-19 patients with 100% success using Hydroxychloroquine Sulfate, Zinc and Z-Pak. All symptoms of shortness of breath resolved within 4-6 hr”*

   “My chlorox wipes kill HIV my Lysol kills human coronavirus. That's where I put my money in the markets. Spray the chlorox wipe with some Lysol & viola.”

   “Covid vaccine is worse than we thought. A GlaxoSmithKline whistleblower says Covid vaccine contains anti-fertility drugs! Tests on 63 women caused 61 to become infertile. Men affected too. DNA in sperm is killed. They want the end of humanity.”

5. COVID-19 Impact: This theme included narratives about the economic and institutional impact of COVID-19. In March, discussions were focusing on the economic disruptions caused by the pandemic like job losses, business closures, school shutdowns, and market crashes. In April, this discourse expanded to include nursing home deaths, protests against lockdowns, and the risks of reopening decisions. In May, the bad situation in nursing homes gained attention alongside continued concerns about the children’s health and education. In June, global economic narratives were discussed in addition to school safety and nursing home accountability. By July, this discourse expanded to include reopening controversies, unemployment, and travel resumption, while referring to the U.S. governors’ decisions.

Example tweets: *"@ewarren Shut up you crazy woman , you want to open borders to let virus in and put us all in danger because you put illegals over Americans We have had it with you American hating dems who will do anything to help illegals. The wall keeps us safe from crimes, drugs, and diseases."*

"Trying to save lives??? Trump built Cuomo a facility JUST FOR COVID cases & sent a navy hospital ship. No cases were sent there, instead NY's top idiot > @NYGovCuomo forced senior centers to admit COVID infected patients. THIS is why NY has such a high death rate."

"UNDER DRUM AMERICA, WE HAVE THE MOST CORONAVIRUS DEATHS IN THE WORLD, 130,000+3MILLIOM INFECTIONS AND THE MOST DEATHS FROM GUNS & FIREARMS, INCLUDING MASS SHOOTINGS, TENS OF THOUSANDS EACH YEAR, & MILLIONS UNEMPLOYED WITH RECESSION & MILLIONS UNINSURED WITH NO HEALTHCARE"

6. COVID-19 Global Response: This theme included discussions about the international efforts taken by different countries to combat the pandemic. In March, narratives focused on U.K.'s herd immunity strategies, India's lockdown measures, and Italy's healthcare crisis. In April, false information linked the virus to religion, with narratives targeting Hindu and Muslim communities in India, while the U.K.'s management of PPE shortages and government decisions were criticized. In May, Sweden's no-lockdown policy gained more attention as people were debating about freedom and restrictions. In June, discussions remained focused on India's healthcare response, Sweden's policies, and political tensions in the U.K., along with regional responses in Scotland and Leicester. By July, people were still analyzing broader global pandemic responses, with continued criticism of government policies.

   Example tweets: *"Herd immunity is useless without a vaccine. Govt strategy is criminally reckless and against the expert advice of the W.H.O. We're the only country in the world proposing this nonsense."*

   *"@guyverhofstadt Tell that to Italy, whom the EU couldn't even send medical equipment to. We all know that the only countries the EU has ever really cared about are Germany and France. And f**k everyone else. Especially the Italians and Greeks."*

   "@DGCAIndia Dear Hardeep, Can u please tell us how Banning International Flights has helped fight COVID? After banning international flights Covid number has only gone up in India!! Therefore this is pure greed on Indian Gvt part!! BJP Gvt is giving more grief than help!"

7. COVID-19 Health Measures: This theme included narratives about the different health measures taken globally like testing, quarantine, and mask wearing, especially in the US. In March, people were discussing the efficiency of these measures in stopping the virus spread.

In April, discussions focused on testing rates by population, mask wearing in public spaces, and reopening parks and beaches. By May, false information regarding the mask mandate continued, along with discussions about antibody testing and prison releases. In June, all these narratives persisted as governments started reopening public spaces. By July, people were criticizing public health guidelines during political events and rallies.

Example tweets: *"Getting tested won't do anything to treat the virus. If he becomes symptomatic, then test. Otherwise, practice social distance and good hygiene."*

"Supposedly to protect someone from spreading it. however most masks, particularly cloth ones, are not dense enough to block the tiny microscopic covid molecules."

"Covid was found in fecal matter, so farts can also spread it. Wear your diaper too!"

8. COVID-19 & Society: This theme included discussions about religion, media influence, and general societal issues. In March, people discussed religious protections from the virus and the role of prayer in limiting the virus spread. In April, false information narratives included discussions about privacy concerns over contact tracing applications. By May, people were criticizing religious gatherings, church reopening, and government digital surveillance. In June, these narratives shifted towards racial justice movements like BLM, alongside other general discussions about public violence and racism.

Example tweets: *"… these all the sign of judgement day ,the Quran already shows that the disease will spread critically but out technology and scientist and politicians can't do anything against the nature"*

"In the UK, nearly two-thirds of people polled are in favour of using mobile data to track coronavirus sufferers and those they come into contact with—a reflection, perhaps, of the public's desperation to see lockdown rules lifted."

"Corona virus is so intelligent it knows to attack only small businesses, church, and white people.... 🙄"

5.2. Theme Evolution

Figure 4 shows the proportional strength of false information themes from March to July 2020. This visualization accounts for varying conversation volumes and highlights how each theme evolved over time.

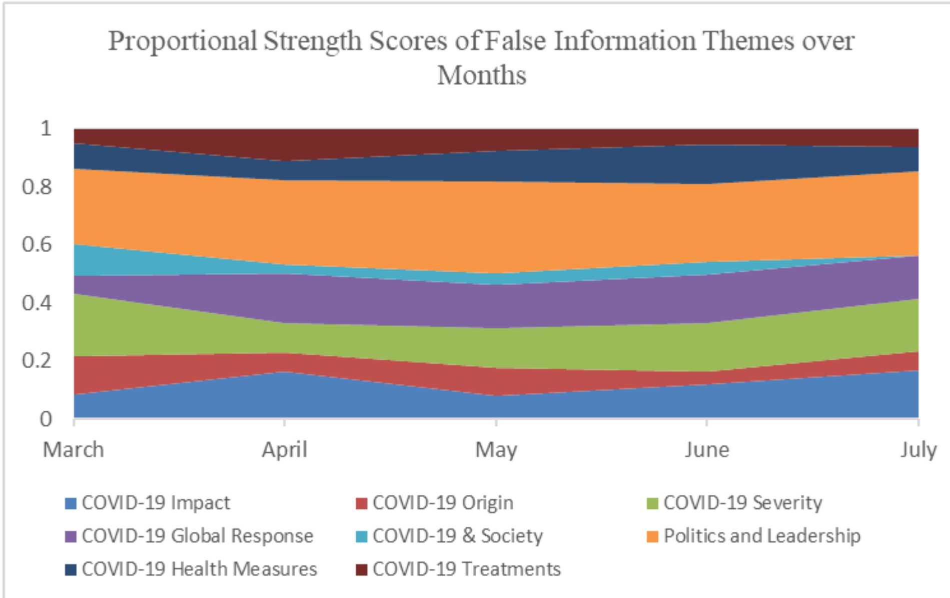


Figure 4. Proportional Strength Scores of False Information Themes over Months

Figure 4 shows that Politics and Leadership theme was the most salient theme across all months, peaking in May, when political narratives regarding presidential responses and election-related issues gained momentum. This theme alone accounted for around 32% of all false information discourses that month. This reflects users' high engagement with political actors and decisions during the initial outbreak period of the pandemic.

COVID-19 Severity theme was also leading in March. It represented around 21% of all narratives as the world was trying to understand more about the virus. After decreasing in April, it increased again in June and July. This shows that the public focus on health effects was renewed as the cases started increasing again in many areas.

Another key theme, which is COVID-19 Origin was dominant in March as it accounted for more than 13% of the monthly discourse. Discussions were mainly concentrated on China and the first identified cases in Wuhan, and their relationship to the virus origin. However, this theme decreased with time and reached below 5% in June and remained marginal in July. This indicates that the initial hypothesis shifted toward more important concerns.

According to Figure 4, the COVID-19 Global Response theme observed a significant trajectory. It was very prominent between March and April, which reflected comparisons between the measures used by different governments throughout the world. It stayed stable until June and July which suggested public interest in how governments handled the pandemic.

COVID-19 Impact and COVID-19 Health Measures themes became more present as the crisis unfolded. For example, there was a steady growth in May and June in the COVID-19 Health Measures theme which aligns with the widespread public debates about preventive behaviors. Also, in April and June, the COVID-19 Impact theme observed an expected increase which reflected a rise in discussions around the impact of public closures and reopening on the economy.

Covid-19 Treatments theme remained minor and marginal throughout all months. It observed an increase in April when the attention of the public was shifted towards potential cures for the virus. Its low proportional presence shows that these hypotheses and speculations were purely extracted from the political reactions with limited information around treatments during the first COVID-19 wave.

Finally, the COVID-19 & Society theme appeared prominently in March but rapidly decreased over time and then disappeared in July. This shows that although the societal and cultural narratives were part of the discourse, they became less prominent when the public discussions moved towards policy, politics, and pandemic management.

These alternating proportions show what the public was most worried about at each stage. This allows for a more thorough analysis of how false narratives might shift during future public health crises.

Brennen et al. (2020) and Motta et al. (2020) found out that political figures and policy makers are frequent targets of false information campaigns. This aligns with the findings of our study, where the dominance of the "Politics and Leadership" themes across all months show that false information during health pandemics is highly politicized. People were overreacting to how political leaders are handling the pandemic and questioning the measures they put into practice. Themes such as "COVID-19 Severity" and "COVID-19 Origin" were widely known during the early stages of the pandemic which shows that initial uncertainty fuels narratives questioning the threat level and the origin of the virus. This agrees with the research done by Swire-Thompson and Lazer (2020), who concluded that misinformation prospers in periods of low trust and ambiguity.

As the pandemic progressed, other themes like "COVID-19 Global Response", "COVID-19 Impact", and "COVID-19 Health Measures" became more noticeable, indicating a shift from uncertainty to criticism of institutional responses, public health protocols, and economic consequences, which is a pattern concluded from recent studies on the lifecycle of misinformation (Islam et al., 2020). Opposingly, "COVID-19 Treatments" theme only gained brief presence during specific moments when public figures promoted unproven remedies. This brief prominence highlights that specific narratives are event-driven rather sustained. As for the "COVID-19 & Society" theme, its early presence followed by marginalization suggests that social and moral narratives may only gain attention when tied with broader cultural movements.

These findings imply that false information during future health crises may follow a similar trajectory, starting with doubts about origin and severity, escalating into political polarization and criticism of global responses, and finally disbanding across themes related to policy, societal disruptions, and treatments. Understanding these changes is essential for intervening in the spread of false information, especially during the early stages when public trust needs to be established. It also reinforces the importance of targeted communication strategies that evolve as the crisis unfolds and directly addresses the false information narratives right after they emerge.

In the next paragraphs, we analyze the word clouds for each theme across all months to discover the false information frames used by individuals.

5.3. False Information Frames

Politics and Leadership

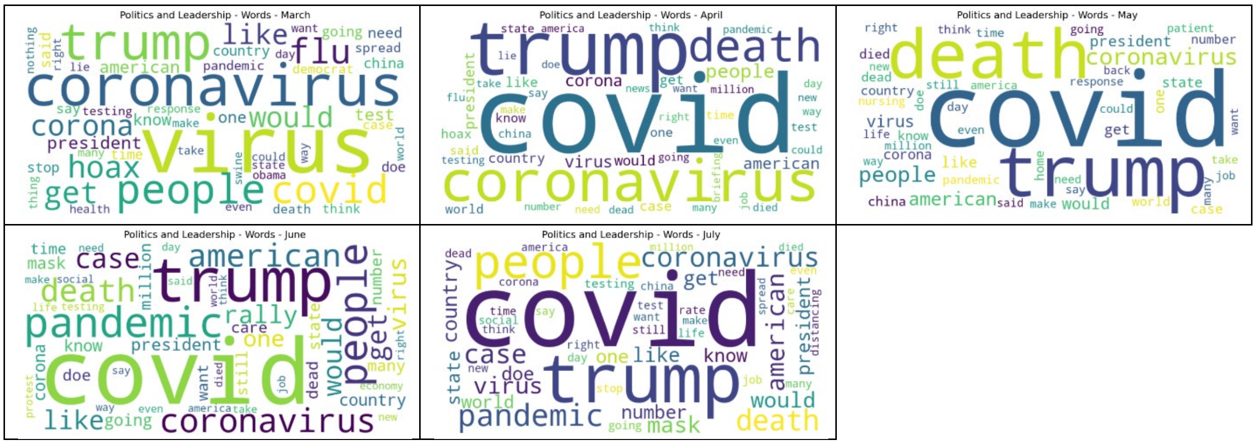


Figure 5. "Politics and Leadership" Theme Word Clouds

1. Minimization Frame

This frame was mostly prominent in March, when COVID-19 was downplayed as a mild illness and compared to less-severe diseases. As shown in Figure 5, words such as "hoax", "flu", "like", and "get" suggest an ongoing effort to minimize the severity of the virus. These terms often coincided with political figures' names like "Trump" and "Pelosi", highlighting that minimization was performed through political messaging. This frame functioned to reduce public fear and delay governmental responses.

Example tweet: *"Trumps response to the Coronavirus saved lives. Unlike Obama's 6 month delay response to the Swine flu. After thousands of deaths, Obama finally called a "state of emergency". So grateful that Trump and his team are handling this and not some reckless Democrat."*

2. Blaming Frame

This frame spans the entire timeline and highlights how the pandemic became a tool for political conflicts. In Figure 4, key words like "Trump", "president", "democrats", "Biden", "Pelosi", and "Obama" often came along emotionally charged words like "lie" and "hoax" to suggest that both sides accused each other of incompetence or using the pandemic for political gains. Blaming was directed toward either party leaders depending on the individual's political stance.

Example tweet: *"Joe Biden helped rescue our economy from the Great Recession that the last republican president drove our economy into. Trump golfing, holding parties, holding rallies, and doing nothing about covid for two months caused our economy to tank. Republicans always tank the economy."*

Example tweet: *"IF Trump had spent all his golf time this year working on solving the pandemic, reform for police (including the unions which both bills left out), and banning assault weapons, well he could have done some great things -- but he didn't and his time is up. We need Joe Biden change"*

COVID-19 Severity

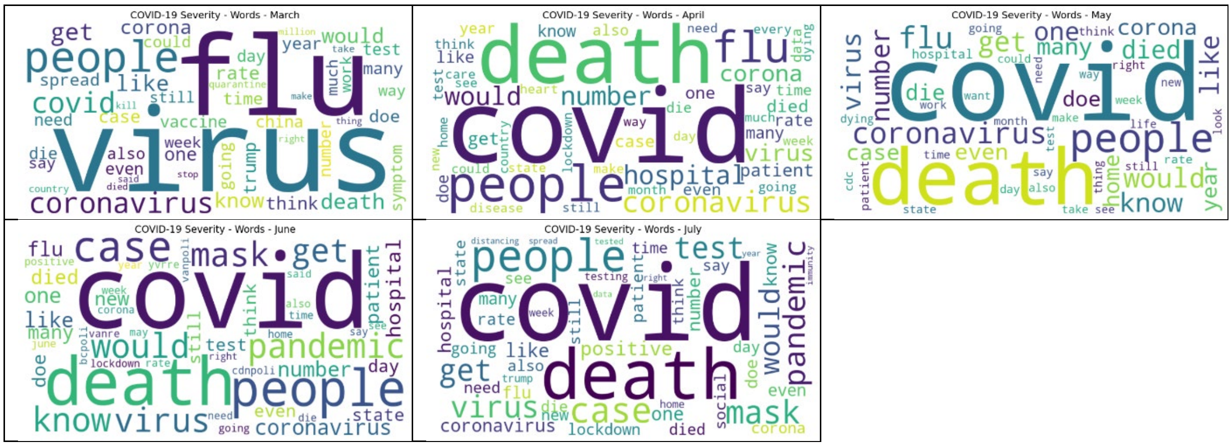


Figure 6. "COVID-19 Severity" Theme Word Clouds

1. Amplification Frame

This frame increased in prominence from April through July as the death counts surged. It highlights the severity and lethality nature of the virus. As shown in Figure 6, key terms include "death", "rate", "high", "hospital", and "die". It often expressed urgency in dealing with the outbreak, warning of overwhelmed hospitals and insufficient healthcare infrastructure. This frame was used manipulatively, sometimes promoting safety while other times intensifying fear.

Example tweet: *"How do deaths go down? The Lazarus effect? Nearly 125 k Americans are dead. Cases are rising toward all time highs. The fatality rate may be lower but no one has risen from the dead. Shame on you for minimizing the importance of this pandemic."*

COVID-19 Origin

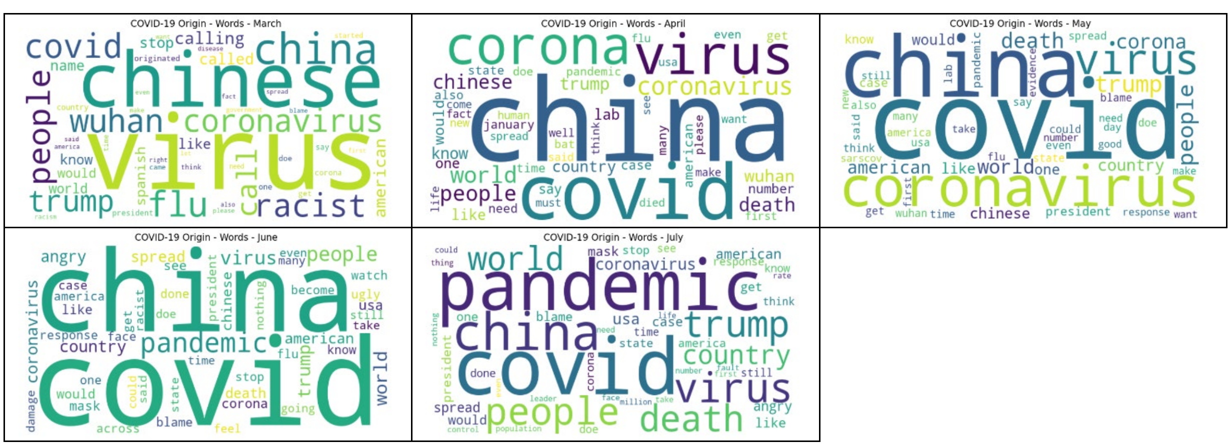


Figure 7. "COVID-19 Origin" Theme Word Clouds

1. Conspiracy Frame

This frame was mostly present in March and April and persisted through May and June, suggesting that the virus originated from a laboratory in Wuhan, China. As shown in Figure 7, repeated words like “lab”, “bat”, “Wuhan”, and “China” suggests that there is a conspiracy speculation regarding the real origin of the virus. People considered COVID-19 a human made or manipulated virus, which aligned with narratives pushing geopolitical blame.

Example tweet: *“Unsanitary conditions in Chinese wet markets (Wuhan Providence) is believed to have caused it. Bats are the likely source of animal to human transmission.”*

2. Blaming Frame

This theme persisted from March till May and targets China as solely responsible for the virus outbreak. As shown in Figure 6, words like “china”, “wuhan”, “government”, and “blame” dominate this narrative. The frame highlights the need for political accountability, describing China as deceitful or negligent. It is often accompanied with xenophobic language, framing the pandemic as a reason to blame and punish specific countries.

Example tweet: *“China lied, and now thousands of innocent souls have been lost, economies are at risk of collapsing... The world is at stand still... Let them be accountable for everything concerning the covid-19”*

COVID-19 Treatments

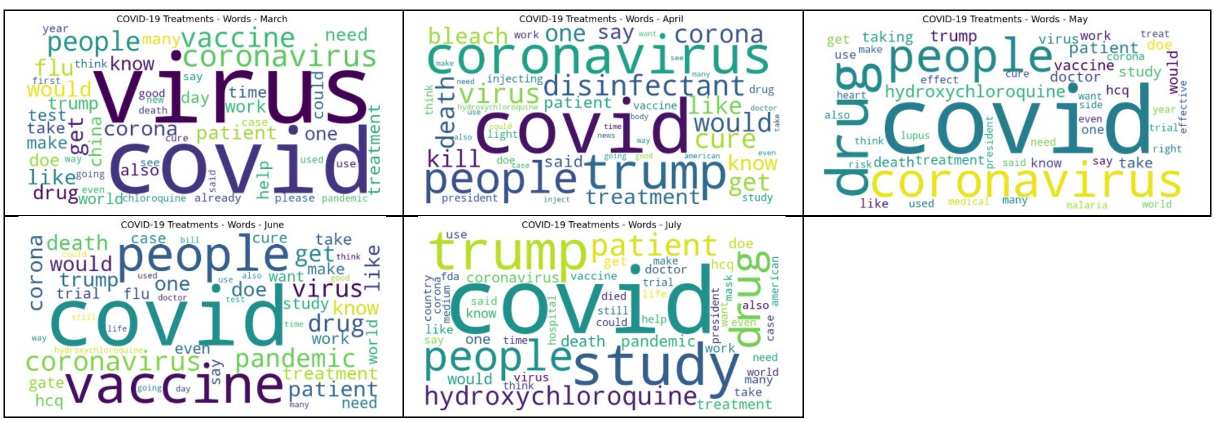


Figure 8. “COVID-19 Treatments” Theme Word Clouds

1. Hope Frame

This frame introduces unproven treatments as potential cures that could ease the symptoms of the virus. It was mostly present in March till May. As shown in Figure 8, it included terms like “hydroxychloroquine”, “treatment”, “cure”, “drug”, “good”, and “patient”. It was usually linked to endorsements from public figures and appealed to public optimism by suggesting that existing remedies could end the pandemic. It offered emotional reassurance but lacked scientific backing.

Example tweet: *"My brother got CoVid. MDs gave him Hydroxychloroquine. It worked. We don't need or want your vaccine! We already have a cure and it's cheap, but you want to make your $"*

2. Conspiracy Frame

This frame criticizes treatments and vaccines not just as medical issues but also as tools of control or political manipulation. It appeared in June and included terms like "bill", "gates", "trial", and "vaccine". These words appeared in narratives suggesting that the pandemic was directed to push pharmaceutical or geopolitical agendas. This frame extends into blaming scientific institutions and global leadership.

Example tweet: *"That's the vaccine 💉 Bill's conglomerate has developed to control Muslims. Once infected, Bill uses his personally developed 'X-Box Æ 12' to control Muslim populations. NOTE: There is a bug in the current code, once fixed, will stop spreading #COVID19"*

COVID-19 Impact

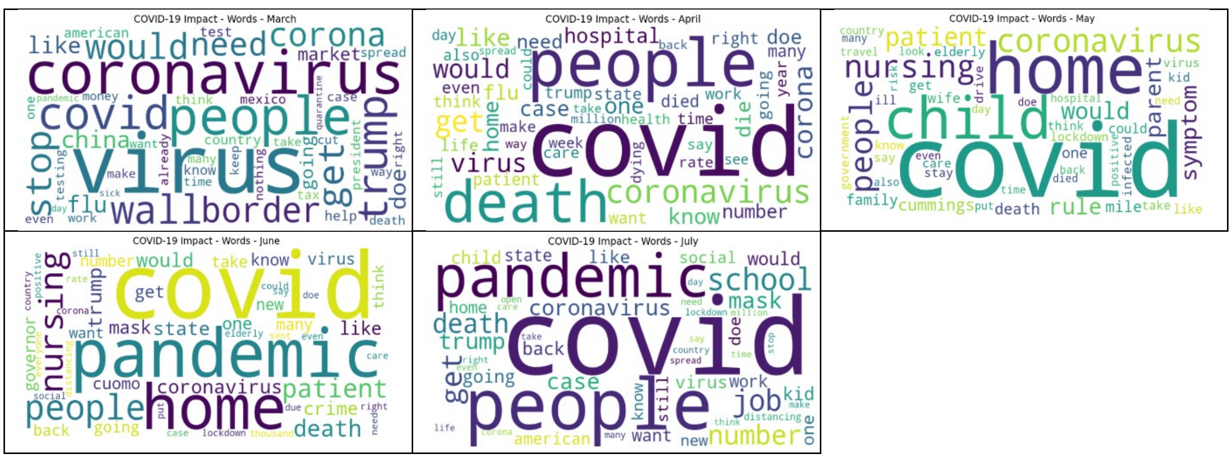


Figure 9. "COVID-19 Impact" Theme Word Clouds

1. Criticism Frame

This frame emphasizes policy-level failures to decrease the effect of the pandemic on vulnerable populations which include the elderly and low-income workers. It emerged in April and continued till July, and the repeated words used include "trump", "nursing", "home", "elderly", "risk", and "care". The tweets within this frame blamed political figures for the unsafe conditions in nursing homes, ineffective support for essential services or prolonged school closures.

Example tweet: *"Especially when Cuomo sends COVID positive people to nursing homes where others caught it & died. It's like he *wants* to get rid of old people. Interesting as he's making cuts to Medicaid at the same time."*

2. Resistance Frame

This frame appeared when lockdown fatigue and anti-measures protests became more popular in July. As shown in Figure 9, words like "open", "people", "back", "school," "work", and "stop" appeared frequently. Tweets within this frame described the pandemic as limiting personal freedom. This frame included tweets that challenged lockdown orders or school closures and considered them as overreach rather than protection.

Example tweet: *"-Can't go to church. -Can't go fishing *by yourself* in your motor boat. -And now you can't buy seeds to garden at your house? In some states, the #coronavirus attack on freedom is out of control. Americans are right to demand their civil liberties back."*

COVID-19 Global Response

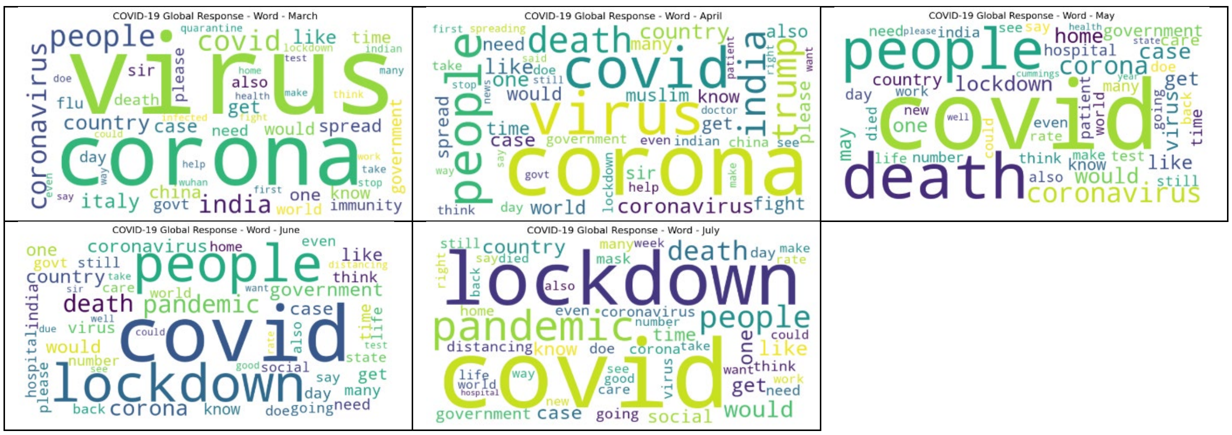


Figure 10. "COVID-19 Global Response" Theme Word Clouds

1. Comparative Frame

This frame persisted across all months, and it focuses on how different countries handled the pandemic. It was often used to praise or criticize certain approaches and it included words like "India", "Sweden", "lockdown", "government", "Germany", "country" and "response". These comparisons were sometimes neutral but usually carried judgement to consider some nations as models of success and others as models of failure.

Example tweet: *"No-lockdown in Sweden is having no economic benefits, but also over five times more deaths per capita vs. Finland."*

2. Blaming Frame

This frame attacks specific populations or religious groups and blames them for worsening the pandemic through their non-compliance. As shown in Figure 10, words like "muslim", "spit", "religion", "tablighi", "freedom", "iran", and "uae" appear frequently in this frame. Tweets within this frame focused on isolated incidents or culturally charged narratives to accuse certain nations and spread false information. This frame contains xenophobic language, mainly about hygiene or public behavior.

Example tweet: *"What about those Muslims who spits on corona warriors? What about those Muslims who don't follow the lockdown and promotes religious gatherings?"*

COVID-19 Health Measures

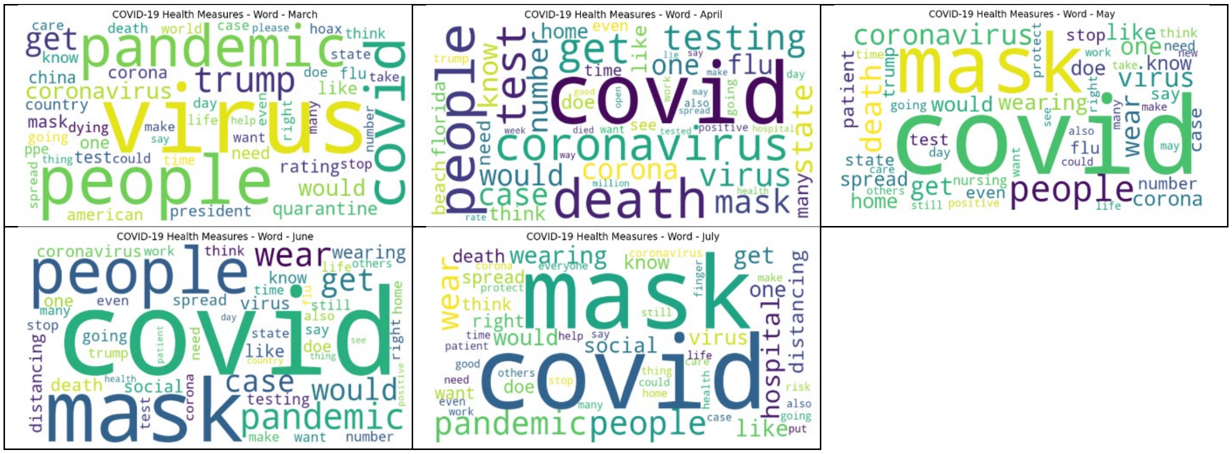


Figure 11. "COVID-19 Health Measures" Theme Word Clouds

1. Preventive Frame

This frame supports protective health measures like mask wearing, testing, and social distancing. It appeared consistently across all months and included words like "mask", "wear", "protect", "spread", "face", "stop", "please", and "testing", as shown in Figure 11. Tweets within this frame often encourage compliance with public health guidelines and shared factual updates on safety protocols.

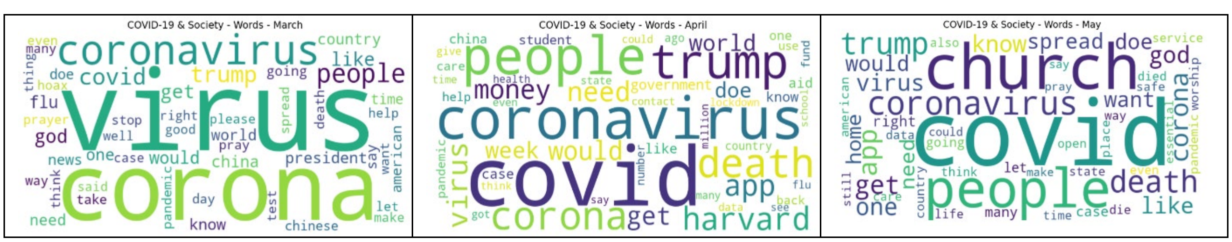
Example Tweet: *"You need to close the whole STATE AND ENFORCE A LOCKDOWN! If you don't lives will continue to be lost. Lives & economy. It's like the government is completely ignoring the obvious solutions. I'm tired of seeing people in critical conditions!"*

2. Resistance Frame

This frame opposes the previous frame and focuses on resisting health mandates. It reflects frustration and criticism of the mandates perceived as overreach. It appeared from May and it included words like "distance", "social", "people", "event", "gathering", "rally", "risk", and "work". Tweets within this frame criticized the prolonged lockdowns and mask rules as they neglect personal freedom.

Example tweet: *"Try to pay attention this time: Both the New England Journal of Medicine and the WHO say there is NO need and NO value in wearing a mask anywhere outside a health care facility unless you have Covid or are caring for someone with Covid. Period. End."*

COVID-19 & Society



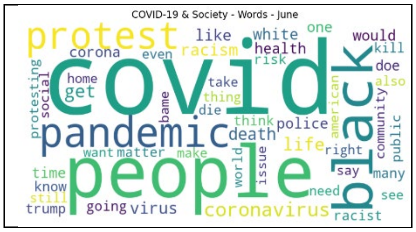


Figure 12. "COVID-19 & Society" Theme Word Clouds

1. Religious Frame

This frame interprets the pandemic from a religious perspective, and it mostly shows up during the first month. "God," "pray", "prayer", "thank", "bless", "jesus", "worship", and "church" are the common words that were used as we look into Figure 12. In this frame, the tweets interpreted Covid-19 as a test of faith or a call for prayer. Also, the religious frame included narratives about spiritual protection and reopening places of worship.

Example Tweet: *"No fear. God has a plan for us. We just have to have faith. I believe God is using COVID-19 to remind us to love each other."*

"His prayers aren't going to do anything more to stop COVID-19 then he was able to do with his requests that Catholic priests stop molesting your children. If you really want to do something, release some of the money the church has been hoarding and help people battling C-19."

6. Discussion and Limitations

This study addressed two main research questions: identifying the most prominent false information narratives related to health pandemic with COVID-19 as the case study and investigating the framing strategies employed to structure these narratives. In response to RQ1, various major themes related to COVID-19 were identified. They include themes related to politics, severity, origin, treatments, measures and responses. These themes varied in prominence during the five-month period as they often aligned with major worldwide events. However, the Politics and Leadership theme remained the most present across all months. In response to RQ2, the framing analysis explored recurring strategies across themes. Frames like blaming, resistance, and conspiracy were identified in multiple themes. Additional frames like hope, minimization, and criticism were only present in specific themes. These findings demonstrate that false information is framed to align with different political and social contexts.

The application of topic modeling on conversations, rather than tweets, offered more coherent textual insights. Findings of recent literature align with the outcomes of our study, where NMF outperformed LDA in extracting more coherent and interpretable topics, especially when thematic clarity mattered (George & Vasudevan, 2020; Latif, Shafait & Latif, 2021). We then performed manual validation on a random sample of conversations to compare the results of NMF and LDA and to enhance the reliability of this research. This was also done by Maier et al. (2018), who showed that human labeling should be compared to machine learning outputs to measure their efficiency. By that, we concluded that the topics generated by NMF were more meaningful than those of LDA and aligned with the context of false information narratives.

The temporal evolution of false information narratives witnessed in this research reflects different stages of health pandemics and how they associate with established crisis communication models. During the early months (March and April), narratives about COVID-19 Severity and COVID-19 Origin were more prominent, as the public sought explanations and clarity in the initial stage of the pandemic. This aligns with the findings of Ophir et al. (2021) who concluded that false information narratives in the early stages of health pandemics often included confusion, comparisons to previous outbreaks and diseases, and questions about the origin. As the pandemic progressed, focus was shifted to global responses and institutional interventions, themes like Politics and Leadership, COVID-19 Treatments and COVID-19 Impact became more dominant. This supports the findings of Tsao et al. (2021) and Wicke and Bolognesi (2020), which suggested that false information shifted toward blame, institutional criticism, and politicized narratives. Later on, false information around global governance and public health interventions became more embedded. This, again, aligns with the findings of Malecki et al. (2021), who explained how long-term crises are dominated by polarizing narratives and fatigue-induced resistance. These patterns justify how false information changed over time, either in response to major events or to match people's feelings and demands during distinct stages of the crisis.

As for the framing strategies presented in this research, the most dominant ones were blaming, resistance, and conspiracy. First, the blaming frame within health-related false information aligns with the findings of Chevalier et al. (2024) and Chang et al. (2020), who concluded that false information during the early stages of the COVID-19 pandemic targeted external actors like China and WHO before shifting to local governments and specific politicians. They targeted specific groups or authorities to trigger emotional responses. Second, the resistance frame aligns with the findings of Walter et al. (2023) and Xu et al. (2022), who found out that opposition to public interventions like lockdowns and mask mandates were regarded as defense of personal liberty and against government overreach. Users actively encouraged non-compliance with these safety protocols and health guidelines by showing distrust in science and government. Third, the conspiracy frame agrees with the findings of Shahsavari et al. (2020), Jiang et al. (2021), and Walter et al. (2023) and supports the research conclusion that false information was systematically framed to appeal to ideological identities and motivate distrust in institutions.

Overall, the findings of this study offer significant insights to academic researchers, policymakers, public health agencies, and social media platforms aiming to better understand the dynamics of false information during health pandemics. By identifying the most prominent false information discourses, this research offers insights that guide future efforts to detect and respond to false and misleading content in a way that is both timely and sensitive to the surrounding context.

From a theoretical perspective, this study contributes to the growing and ongoing body of research on false information in the context of global health crises. The identification of recurrent frames across themes supports the idea that false information is usually structured in ways that resonate with emotional rather than rational reasoning. The accuracy of content indeed matters; however, it is important to analyze the linguistic patterns that were used in order to amplify certain narratives and affect public perceptions. Additionally, by combining unsupervised topic modeling with frame analysis, this study contributes to communication research by bridging computational and interpretive approaches. Moreover, this study highlights the need for hybrid approaches that balance algorithmic efficiency with human interpretations when analyzing complex social media posts.

On a practical level, the findings of this study offer valuable insights for public health analysts and policymakers. By understanding the recurring frames used in spreading false information, institutions become better prepared to counter these claims and design targeted messaging strategies that address unverified information. In addition to that, this study contributes to the continuing efforts of social media platforms to limit the spread of fake news. Building on generic fact-checking tools, social media developers could also incorporate the linguistic characteristics and framing tools found in false information to alert users exposed to such misleading content while maintaining user autonomy and freedom of speech. Furthermore, by mapping the evolution of themes over time, communicators can predict which types of narratives are most likely to appear and intensify in future health crises, permitting for more proactive content moderation and campaigns tailored to different public concerns.

While this study offers valuable insights into the framing and thematic structure of false information narratives related to health pandemics, few limitations should be acknowledged. First, the identification of tweets possibly containing false information relied on publicly available lists of unreliable and questionable websites. While the lists from Zimdars and Media/Bias Fact Check are used in prior research, other fact checking databases like Poynter, PolitiFact, and Snopes were not utilized as we were not able to obtain access to them. This approach remained systematic and reliable in extracting false information content; however, we may have excluded tweets that included links not covered by these lists or tweets that did not contain any URL. Consequently, we ignored user-generated false claims. Moreover, the data was collected exclusively from Twitter, which limits the generalizability of the findings across the broader social media landscape.

Second, although our method of analyzing full conversations added contextual richness, the framing analysis employed interpretive subjectivity. This means that frame identification was done qualitatively by interpreting word clouds and reading conversations. Different researchers might classify narratives under different frames as multiple framing strategies might be extracted from a single conversation. Similarly, NMF produced interpretable topics but it remains sensitive to preprocessing decisions and different parameter settings. This means that multiple topic model runs could yield to slightly different themes. Also, the study also lacked sentiment analysis which might discover additional emotions that are helpful in detecting frames across the themes.

Finally, our analysis was limited to English-language conversations that occurred between March and July 2020, which only captures the early stages of the pandemic. As a result, the identified themes and frames may reflect only the public reactions specific to that phase. Later developments in the pandemic are outside the temporal scope of this study. Similarly, the linguistic and cultural dynamics of false information were not analyzed for non-English contexts, which limits the generalizability of this study on other communities. Moreover, thematic labelling and manual validation of theme assignments was based on manual interpretation of keywords generated by the topic models and conversations, which introduces a degree of subjectivity. This means that the absence of multiple human coders may limit the replicability of theme assignments.

7. Conclusion

This research examined how health-related false information can be identified and understood on social media using the COVID-19 pandemic as a case study. By applying two different topic modeling techniques to conversation-level data, interpreting the results of the NMF-generated themes as they showed better coherence and accuracy with human-labeled conversations, and exploring the common linguistic patterns across discourses, our analysis identified the key false information themes and the framing techniques used to affect public perception.

Our study introduces a novel evaluation approach for topic modeling techniques by combining coherence-based comparison with manual validation. Rather than directly choosing the topic model with the higher coherence, we manually annotated a sample of conversations to compute classification accuracy. This approach enhances the interpretability and reliability of topic modeling in false information research and offers a replicable method for validating discourse analysis.

This study also extends the themes and discourses of health-related false information. Instead of using the model generated topics for further analysis, we decided to assign topics to general themes to avoid having redundant topics. We were able to identify 8 dominant themes and 27 discourses across all conversations. Although some of the themes were common among prior research, such as politics, treatments, origin, and impact, our study introduced a new unusual theme which we named "COVID-19 & Society". This new theme mainly includes false narratives about religious coping and cultural practices during health pandemics.

While our analysis offered key insights into how false information narratives were constructed and circulated, supporting public health agencies and policymakers, the framing of false information remains a complex subject influenced by different broader social, political and cultural contexts. The way individuals interpret and spread false information depends on the content and the other factors like political ideology, cultural values, and trust level in institutions. However, we strongly believe that this study contributes to future research that seeks to understand the persuasive strategies embedded in false information and the role of narrative framing in shaping public perception in future healthcare crises. Even though this study contains few limitations, it offers a structured and theoretically grounded approach to understand how false information narratives are constructed and potentially develop effective campaigns to mitigate their spread.

Future research could expand the scope of this research by incorporating more data sources to capture a broader range of false information discourses. Further studies might explore how frames differ across user types or how audiences from different communities react to false information over time. Applying sentiment analysis into the framing process could offer more insights about the emotional tone associated with each frame, which enhances our understanding of how emotional appeal contributes to the spread of false information. Additionally, more advanced methods, such as dynamic topic modeling and network analysis, could also help discover how themes and frames interact and shift over time.

References

Ahammad, T. (2024). Identifying hidden patterns of fake COVID-19 news: An in-depth sentiment analysis and topic modeling approach. *Natural Language Processing Journal*, *6*, 100053.

Ai, L., Liu, Z., & Hirschberg, J. (2022). Combating the COVID-19 Infodemic: Untrustworthy Tweet Classification using Heterogeneous Graph Transformer.

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, *13*(1), 30.

Allgaier, J., & Svalastog, A. L. (2015). The communication aspects of the Ebola virus disease outbreak in Western Africa–do we need to counter one, two, or many epidemics?. *Croatian medical journal*, *56*(5), 496.

Al-Rakhami, M. S., & Al-Amri, A. M. (2020). Lies kill facts save: Detecting COVID-19 misinformation in Twitter, IEEE Access 8 (2020) 155961–155970.

Azam, N., & Yao, J. (2012). Comparison of term frequency and document frequency based feature selection metrics in text categorization. *Expert Systems with Applications*, *39*(5), 4760-4768.

Balasubramaniam, A., Balasubramaniam, T., Jeyaraj, R., Paul, A., & Nayak, R. (2021). Nonnegative matrix factorization to understand spatio-temporal traffic pattern variations during COVID-19: a case study. In *Data Mining: 19th Australasian Conference on Data Mining, AusDM 2021, Brisbane, QLD, Australia, December 14-15, 2021, Proceedings 19* (pp. 223-234). Springer Singapore.

Blei, D. M., & Lafferty, J. D. (2006, June). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning* (pp. 113-120).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Bolsover, G., & Tizon, J. T. (2020). Social media and health misinformation during the US COVID crisis. *arXiv preprint arXiv:2008.05271*.

Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation.

Burel, G., Farrell, T., & Alani, H. (2021). Demographics and topics impact on the co-spread of COVID-19 misinformation and fact-checks on Twitter. *Information Processing & Management*, *58*(6), 102732.

Ceron, W., de-Lima-Santos, M. F., & Quiles, M. G. (2021). Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content. *Online social networks and media*, *21*, 100116.

Chang, A., Schulz, P. J., Tu, S., & Liu, M. T. (2020). Communicative blame in online communication of the COVID-19 pandemic: Computational approach of stigmatizing cues and negative sentiment gauged with automated analytic techniques. Journal of Medical Internet Research, 22(11), e21504.

Chen, E., Jiang, J., Chang, H. C. H., Muric, G., & Ferrara, E. (2022). Charting the information and misinformation landscape to characterize misinfodemics on social media: COVID-19 infodemiology study at a planetary scale. *Jmir Infodemiology*, *2*(1), e32378.

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR public health and surveillance*, *6*(2), e19273.

Chen, L., & Fu, L. (2022). Let's fight the infodemic: the third-person effect process of misinformation during public health emergencies. *Internet Research*, *32*(4), 1357-1377.

Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*, *163*, 1-13.

Chevalier, M., de la Sablonnière, R., Harel, S. O., Ratté, S., Pelletier-Dumas, M., Dorfman, A., ... & Lacourse, É. (2024). Who's To Blame for the COVID-19 pandemic? Perceptions of responsibility during the crisis using text mining and latent Dirichlet allocation. Social Sciences & Humanities Open, 9, 100825.

Chowdhury, N., Khalid, A., & Turin, T. C. (2023). Understanding misinformation infodemic during public health emergencies due to large-scale disease outbreaks: a rapid review. *Journal of Public Health*, *31*(4), 553-573.

Churchill, R., & Singh, L. (2022). The evolution of topic modeling. *ACM Computing Surveys*, *54*(10s), 1-35.

Cucinotta, D., & Vanelli, M. (2020). WHO Declares COVID-19 a Pandemic. *Acta bio-medica : Atenei Parmensis*, *91*(1), 157–160.

D'Angelo, P. (2017). Framing: media frames. *The international encyclopedia of media effects*, *2*, 634-644.

Diaz Ruiz, C., & Nilsson, T. (2023). Disinformation and echo chambers: how disinformation circulates on social media through identity-driven controversies. *Journal of public policy & marketing*, *42*(1), 18-35.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439-453.

Dujeancourt, E., & Garz, M. (2023). The effects of algorithmic content selection on user engagement with news on Twitter. *The Information Society*, *39*(5), 263-281.

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, *7*, 886498.

Eysenbach, G. (2020). How to fight an infodemic: the four pillars of infodemic management. *Journal of medical Internet research*, *22*(6), e21820.

Farsi, D. (2021). Social media and health care, part I: literature review of social media use by health care providers. *Journal of medical internet research*, *23*(4), e23205.

George, L., & Sumathy, P. (2023). An integrated clustering and BERT framework for improved topic modeling. *International Journal of Information Technology*, *15*(4), 2187-2195.

George, S., & Vasudevan, S. (2020). Comparison of LDA and NMF topic modeling techniques for restaurant reviews. *Indian J. Nat. Sci*, *10*(62), 28210-28216.

Ghaly, M. Z., & Laksito, A. D. (2023, December). Topic Modeling of Natural Disaster in Indonesia Using NMF. In *2023 Eighth International Conference on Informatics and Computing (ICIC)* (pp. 1-6). IEEE.

Gisondi, M. A., Barber, R., Faust, J. S., Raja, A., Strehlow, M. C., Westafer, L. M., & Gottlieb, M. (2022). A deadly infodemic: social media and the power of COVID-19 misinformation. Journal of medical Internet research, 24(2), e35552.

Hong, L., & Davison, B. D. (2010, July). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88).

Islam, M. S., Kamal, A. H. M., Kabir, A., Southern, D. L., Khan, S. H., Hasan, S. M., ... & Seale, H. (2021). COVID-19 vaccine rumors and conspiracy theories: The need for cognitive inoculation against misinformation to improve vaccine adherence. *PloS one*, *16*(5), e0251605.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, *78*, 15169-15211.

Jiang, L. C., Chu, T. H., & Sun, M. Characterization of vaccine tweets during the early stage of the COVID-19 outbreak in the United States: topic modeling analysis. JMIR Infodemiology. 2021 Sep 14; 1 (1): e25636. 10.2196/25636.

Kalantari, N., Liao, D., & Motti, V. G. (2021, December). Characterizing the online discourse in Twitter: Users' reaction to misinformation around COVID-19 in Twitter. In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 4371-4380). IEEE.

Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., ... & Baddour, K. (2020). Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. *Cureus*, *12*(3).

Latif, S., Shafait, F., & Latif, R. (2021). Analyzing LDA and NMF topic models for Urdu tweets via automatic labeling. *IEEE Access*, *9*, 127531-127547.

Leão, T., Amorim, M., Fraga, S., & Barros, H. (2021). What doubts, concerns and fears about COVID-19 emerged during the first wave of the pandemic?. *Patient education and counseling*, *104*(2), 235-241.

Li, D., & Santos Jr, E. (2020). Discriminating deception from truth and misinformation: an intent-level approach. *Journal of Experimental & Theoretical Artificial Intelligence*, *32*(3), 373-407.

Li, W., & McCallum, A. (2006, June). Pachinko allocation: DAG-structured mixture models of topic correlations. In *Proceedings of the 23rd international conference on Machine learning* (pp. 577-584).

Madani, Y., Erritali, M., & Bouikhalene, B. (2021). Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets. *Results in Physics*, *25*, 104266.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Adam, S. (2021). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In *Computational methods for communication science* (pp. 13-38). Routledge.

Malecki, K. M., Keating, J. A., & Safdar, N. (2021). Crisis communication and public perception of COVID-19 risk in the era of social media. *Clinical infectious diseases*, *72*(4), 697-702.

Marwick, A. E. (2018). Why do people share fake news? A sociotechnical model of media effects. *Georgetown law technology review*, *2*(2), 474-512.

Memon, S. A., & Carley, K. M. (2020). Characterizing covid-19 misinformation communities using a novel twitter dataset. *arXiv preprint arXiv:2008.00791*.

Motta, M., Stecula, D., & Farhart, C. (2020). How right-leaning media coverage of COVID-19 facilitated the spread of misinformation in the early stages of the pandemic in the US. *Canadian Journal of Political Science/Revue canadienne de science politique*, *53*(2), 335-342.

Muhammed T, S., & Mathew, S. K. (2022). The disaster of misinformation: a review of research in social media. *International journal of data science and analytics*, *13*(4), 271-285.

Neishabouri, A., Desmarais, M. C., & Montreal, P. (2020, May). Reliability of Perplexity to Find Number of Latent Topics. In *FLAIRS* (pp. 246-251).

Nemes, L., & Kiss, A. (2021). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, *5*(1), 1-15.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010, June). Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics* (pp. 100-108).

Ophir, Y., Walter, D., Arnon, D., Lokmanoglu, A., Tizzoni, M., Carota, J., ... & Nicastro, E. (2021). The framing of COVID-19 in Italian media and its relationship with community mobility: a mixed-method approach. *Journal of Health Communication*, *26*(3), 161-173.

Pian, W., Chi, J., & Ma, F. (2021). The causes, impacts and countermeasures of COVID-19 "Infodemic": A systematic review using narrative synthesis. *Information processing & management*, *58*(6), 102713.

Ramage, D., Dumais, S., & Liebling, D. (2010, May). Characterizing microblogs with topic models. In *Proceedings of the international AAAI conference on web and social media* (Vol. 4, No. 1, pp. 130-137).

Rathore, F. A., & Farooq, F. (2020). Information overload and infodemic in the COVID-19 pandemic. *J Pak Med Assoc*, *70*(5), S162-S165.

Ravichandran, B. D., & Keikhosrokiani, P. (2023). Classification of Covid-19 misinformation on social media based on neuro-fuzzy and neural network: A systematic review. *Neural Computing and Applications*, *35*(1), 699-717.

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).

Rodríguez, C. P., Carballido, B. V., Redondo-Sama, G., Guo, M., Ramis, M., & Flecha, R. (2020). False news around COVID-19 circulated less on Sina Weibo than on Twitter. How to overcome false information?. International and Multidisciplinary Journal of Social Sciences, 9(2), 107-128.

Ross Arguedas, A., Robertson, C., Fletcher, R., & Nielsen, R. (2022). Echo chambers, filter bubbles, and polarisation: A literature review

Saha, A., & Sindhwani, V. (2012, February). Learning evolving and emerging topics in social media: a dynamic nmf approach with temporal regularization. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 693-702).

Shahi, G., Dirkson, A., & Majchrzak, T. (2020). An exploratory study of COVID-19 misinformation on Twitter. Online Social Networks and Media, 22, 100104.

Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R., & Roychowdhury, V. (2020). Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. Journal of computational social science, 3(2), 279-317.

Sharma, K., Seo, S., Meng, C., Rambhatla, S., & Liu, Y. (2020). Covid-19 on social media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*.

Shu, K. (2022). Combating disinformation on social media: A computational perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *2*(1), 100035.

Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., ... & Wang, Y. (2020). A first look at COVID-19 information and misinformation sharing on Twitter. arXiv 2020. *arXiv preprint arXiv:2003.13907*.

Smith, J. (2024). Are hashtags dead? Elon Musk says' please stop'using them on X. Fox5dc.

Song, X., Petrak, J., Jiang, Y., Singh, I., Maynard, D., & Bontcheva, K. (2021). Classification aware neural topic model for COVID-19 disinformation categorisation. *PloS one*, *16*(2), e0247086.

Swire-Thompson, B., & Lazer, D. (2022). Reducing health misinformation in science: A call to arms. *The Annals of the American Academy of Political and Social Science*, *700*(1), 124-135.

Tandoc Jr, E. C., Lim, D., & Ling, R. (2020). Diffusion of disinformation: How social media users respond to fake news and why. *Journalism*, *21*(3), 381-398.

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

Van der Meer, T. G., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health communication*, *35*(5), 560-575.

Vasconcellos-Silva, P. R., & Castiel, L. D. (2020). COVID-19, fake news, and the sleep of communicative reason producing monsters: the narrative of risks and the risks of narratives. *Cadernos de Saúde Pública*, *36*, e00101920.

Walter, D., Ophir, Y., & Ye, H. (2023). Conspiracies, misinformation and resistance to public health measures during COVID-19 in white nationalist online communication. Vaccine, 41(17), 2868-2877.

Wang, Y. X., & Zhang, Y. J. (2012). Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on knowledge and data engineering*, *25*(6), 1336-1353.

Wicke, P., & Bolognesi, M. M. (2020). Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. *PloS one*, *15*(9), e0240010.

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD explorations newsletter*, *21*(2), 80-90.

Xu, W. W., Tshimula, J. M., Dubé, È., Graham, J. E., Greyson, D., MacDonald, N. E., & Meyer, S. B. (2022). Unmasking the Twitter discourses on masks during the COVID-19 pandemic: User cluster–based BERT topic modeling approach. Jmir Infodemiology, 2(2), e41198.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013, May). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1445-1456).

Zimdars, M. (2016). My "fake news list" went viral. But made-up stories are only part of the problem. *Washington Post*.