

Received December 23, 2021, accepted January 27, 2022, date of publication February 9, 2022, date of current version February 16, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3150329

Investigating the Emotional Response to COVID-19 News on Twitter: A Topic Modeling and Emotion Classification Approach

FRANCISCO BRÁULIO OLIVEIRA¹, AMANUL HAQUE², DAVOUD MOUGOUEI³,
SIMON EVANS⁴, JAIME SIMÃO SICHMAN¹, AND MUNINDAR P. SINGH², (Fellow, IEEE)

¹Department of Computer Engineering and Digital Systems, University of São Paulo, São Paulo 05508070, Brazil

²Department of Computer Science, North Carolina State University, Raleigh, NC 27695, USA

³School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia

⁴School of Psychology, University of Surrey, Guildford, Surrey GU2 7XH, U.K.

Corresponding author: Francisco Bráulio Oliveira (fcoabraulio@usp.br)

This work was supported in part by the University Global Partnership Network (UGPN) as well as the Australian Academy of Science; and in part by the Department of Industry, Science, Energy and Resources.

ABSTRACT Media has played an important role in public information on COVID-19. But distressing news, e.g., COVID-19 death tolls, may trigger negative emotions in public, discouraging them from following the news, which, in turn, can limit the effectiveness of the media. To understand people's emotional response to the COVID-19 news, we have investigated the prevalence of basic human emotions in around 19 million user responses to 1.7 million COVID-19 news posts on Twitter from (English-speaking) media across 12 countries from January 2020 to April 2021. We have used Latent Dirichlet Allocation (LDA) to identify news themes on Twitter. Also, the Robustly Optimized BERT Pretraining Approach (RoBERTa) model was used to identify emotions in the tweets. Our analysis of the Twitter data revealed that anger was the most prevalent emotion in user responses to the news coverage of COVID-19. That was followed by sadness, optimism, and joy, steadily over the period of the study. The prevalence of anger (in user responses) was higher for the news about authorities and politics while optimism and joy were more prevalent for the news about vaccination and educational impacts of COVID-19 respectively. The prevalence of sadness in user responses, however, was the highest for the news about COVID-19 cases and deaths and the impacts on the families, mental health, jails, and nursing homes. We also observed a higher level of anger in the user responses to the (COVID-19) news posted by the USA media accounts (e.g., CNN Politics, Fox News, MSNBC). Optimism, on the other hand, was found to be the highest for Filipino media accounts.

INDEX TERMS COVID-19, media, news, emotion, Twitter, RoBERTa model, topic modeling, NLP.

I. INTRODUCTION

Media has played a pivotal role in containing the COVID-19 pandemic through public information [1]. But COVID-19 news often contains distressing contents, e.g., death tolls, which may contribute to negative feelings in the audience [2], [3], discouraging them from following the news. Due to the multifaceted nature of this problem, it would be hard to hypothesize that COVID-19 news has been the primary factor in people experiencing certain emotions. But understanding the emotions of the news audience in relation to the published

news is a step toward developing a theoretical framework for examining such hypotheses.

There have been a few attempts to study public sentiments on social media in the context of the COVID-19 pandemic, but they have not focused on the COVID-19 news [4], [5]. Others have merely studied the emotions embedded in the COVID-19 news headlines without considering the emotions of the news audience [6], [7]. To overcome these limitations, we have investigated the emotional responses to the media coverage of the COVID-19 pandemic on Twitter from January 2020 to April 2021 across popular media outlets of 12 countries; the following research questions were formulated:

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

- (RQ1):** What are the COVID-19 news themes on Twitter?
(RQ2): What is the longitudinal distribution of the emotions and themes in the user responses?
(RQ3): What is the latitudinal distribution of the emotions and themes in the user responses?
(RQ4): What is the distribution of the emotions and themes in the user responses across different media outlets?

To answer these questions, we collected around 19 million user responses to 1.7 million COVID-19 news posts published on Twitter by 276 (English) news accounts from the popular media outlets across 12 countries. Our methodology adopted an emotion classification approach to detect Ekman's basic emotions [8] (joy, optimism, anger, and sadness) in the user responses to the news posted on Twitter. In doing so, we adopted a Robustly Optimized BERT Pretraining Approach (RoBERTa) model [9], retrained on Twitter data, fine-tuned for emotion detection and evaluated on a popular Twitter classification benchmark [10]. To understand the emotional responses to the COVID-19 news about different topics (e.g., Vaccination), we identified the news themes using an unsupervised topic modeling approach referred to as Latent Dirichlet Allocation (LDA) [11]. The longitudinal distributions of the emotional responses and news themes were generated through a weekly grouping of the Twitter posts; while the latitudinal distribution was generated from grouping the themes and emotions in the news and user responses by geographical locations.

Our analysis of the Twitter data identified anger as the most prevalent emotion in user responses to the news coverage of COVID-19. That was followed by sadness, optimism, and joy, steadily over the period of the study. Our results also showed that, among all news themes, the prevalence of anger was highest in user responses to the news about "Authorities & Politics" (health authorities, political authorities, and elections). Also, the highest level of optimism and joy were observed in user responses to the news about "COVID-19 Vaccination" and "Education" respectively. Sadness, however, was mainly associated with the COVID-19 news about "Cases and Deaths" and the "Family Stories", "Mental Health", "Jails", and "Nursing Homes". Our results also revealed that 4 of the top 5 news accounts with the highest prevalence of anger in their user responses were from the USA news media (CNN Politics, Fox News, MSNBC, Breitbart News). Sadness and joy, were both more prevalent in user responses to the news posted by the Indian media. The news accounts from the Filipino media, on the other hand, demonstrated the highest prevalence of optimism in their user responses.

The remainder of this paper is organized as follows. Section II gives some background on this research and discusses the related work. Then, Section III describes our methodology for investigating the emotional responses to the COVID-19 news. Our findings are presented in Section IV, followed by a discussion to contextualize those findings in the literature (V). The limitations of the work are listed in Section VI and the paper is concluded in Section VII.

II. BACKGROUND AND RELATED WORK

A. COVID-19 NEWS AND EMOTIONS

Journalism is becoming more interactive, interconnected, participatory, and global, giving rise to a networked journalism, producing a constant stream of data, and comments [12], [13]. While this expands journalists' reach and influence, it also increases their accountability for responsible journalism.

Emotion-evoking features such as images in news reports can increase the perceived severity of the event among its readers [14]. Lack of effective public communication has resulted in a misconception of risks and irrational fears that have led to surge in demand and shortage of flu shots and vaccines [15] and contributed to stigmatization of victims and health workers, as happened in the case Ebola outbreak [16]

Additionally, there is evidence that exposure to social media enhances the negative mental effects caused by intense contact with media content [17], [18]. Prior research on identifying immediate priorities for research on COVID-19 with a focus on mental health and well-being recognized an urgent need for collecting high-quality data on the mental health and psychological effects of the COVID-19 pandemic across different populations and vulnerable groups [19].

As distressing news have become an integral part of COVID-19 information, the media outlets need to take a more responsible approach towards developing effective news policies. Policies that mitigate the adverse impacts of negative news on people's mental health while communicating the essential information for the containment of the pandemic. This is particularly important to protect the mental well-being of the vulnerable groups (e.g., people with preexisting mental conditions).

Beckett *et al.* [20] highlighted how quality news reporting and editing has always had emotion at its core, and identified three major factors currently driving journalists toward using emotion as a tool, namely, *economic* (competition has never been more), *technological* (greater reach and engagement) and *behavioral* (people respond to emotions more than ideas or facts). In this work, we focused on technological (use of social media) and behavioral (people's reactions to posts from news sources) aspects of emotions associated with news coverage.

B. EMOTION DETECTION FROM TEXT

Identifying emotions from unstructured text has been a subject of prior research [21]–[23]. While sentiment analysis provides a general polarity (positive or negative) of sentiments in a text, emotion recognition gives a more fine-grained analysis of the emotional (i.e., affectual) state of the author of the text.

Emotion detection can broadly be classified into *discrete* and *dimensional* models of emotions. Some of the popular discrete emotion models include the following. Ekman's model [24] distinguishes emotions based on six basic emotions (happiness, sadness, anger, disgust, surprise, and fear). Plutchik's model [25] postulates that basic emotions occur

in opposite pairs and complex emotions are produced by their combinations naming eight of such fundamental emotions (adding trust and anticipation to the six basic emotions posited by Ekman). Orthony, Clore, and Collins’s (OCC) model [26] discretizes emotions into 22, adding 16 emotions to the emotions Ekman posited as basic.

Dimensional emotion models presuppose that emotions are not independent and that there exists a relation between them. Some of the popular dimensional emotion models include Plutchik’s 2-dimensional wheel of emotions [25] and Russell’s circumplex [27] of affect model. Both models distinguish emotions based on arousal (activation vs deactivation) and valence (pleasantness vs unpleasantness).

Colnerič *et al.* [28] proposed a model that can recognize emotions in tweets based on Ekman’s emotion classes [24], Plutchik’s emotion wheel [25] and POMS (Profile of Mood States) [29]. The model trains for all three classification tasks with shared parameters. Training dataset was curated using distant supervision by leveraging hashtags in tweets as labels. Aslam *et al.* [7] used the National Research Council (NRC) word-emotion lexicon [30] to calculate the presence of sentiments and emotions in COVID-19 related news headlines.

Recent methods on emotion detection from text include deep attentive RNNs and transfer learning [31], Bidirectional Long Short Term Memory (Bi-LSTM) with attention [32], combination of word and document level embedding along with a set of psycho linguistic features [33], skip thought vector and sentiment neurons with an ensemble of multiple predictive models [34], lexicon based emotion extraction using a generative Unigram Mixture Model [35].

We adopted the Robustly Optimized BERT Pretraining Approach (RoBERTa) model retrained on ≈58M tweets and fine-tuned for emotion detection to detect four basic emotions in tweets {joy, optimism, anger, and sadness} inspired by Ekman’s basic emotion classes [24]. The emotion detection model achieved state-of-the-art results for emotion detection from tweets with a weighted average F1-score of 79.8% on TweetEval benchmark [10], a unified benchmark and comparative evaluation for Tweet classification. RoBERTa model has also been used to analyze the emotional experiences in society at large [36], emotion detection from tweets [37], identifying worry [38] and mental illness [39] on social media, identifying misinformation regarding Covid-19 [40], and towards unsupervised bias reduction for emotion and sentiment classification [41].

C. RELATED WORK

COVID-19 has impacted populations around the world and has been the focus of many recent research studies. Recent research reveals a range of emotional reactions (e.g., anguish, stress, anger) and behavior changes (e.g., panic buying) in public across different societies [50]. Meanwhile, it has been found that positive emotion and resilience were significantly higher in individuals exposed to news having positive content

TABLE 1. Overview of mentioned publications in relation to this work.

Publications	Modeling		Timing		Other		
	Sentiments	Emotions	Topics	Events	Longitudinal	News Replies	Multi-Country
Gaspar[42], Kim[43], Aslam[44]		✓		✓			
Cheong[45], Barnaghi[46]	✓			✓			
Thelwall[47], de Melo[48]		✓	✓		✓		
Xue[49], Xue[4]		✓	✓	✓			
Ghasiya[6]	✓		✓	✓			✓
This work		✓	✓		✓	✓	✓

about the COVID-19 than those who were exposed to negative content about the same [51].

The microblogging platform Twitter has been demonstrated to be a useful environment for the study of major events. Being able to generate relevant information for public policy decision-making by anticipating the occurrence of extreme situations and demonstrating how population groups react to them while it unfolds [45], [47]. Prior research on social media content analyses had used Natural Language Processing (NLP) techniques to extract sentiments from the text [45], [46], [52], [53]. However, it has been shown that sentiment analysis alone might not be enough to characterize a group response without the context [42].

In the context of the COVID-19 pandemic, Ghasiya and Okamura [6] investigated more than 100 thousand news headlines and articles from four countries. They used semantic search based model to identify topics within the data set, and fine-tuned the RoBERTa-Base model for sentiment analysis. They found the UK as the country with the highest percentage of negative sentiments. Xue *et al.* [4], [49] focused on examining COVID-19-related discussions and sentiments using tweets posted by Twitter users. They used Latent Dirichlet Allocation (LDA) to identify topics and the NRC Emotion Lexicon for emotion classification and found that anticipation is the dominant emotion and fear is relevant when the tweet is related to reports of new COVID-19 cases. Kim, Cho and LoCascio [43] used a nationally representative survey of South Koreans adults and found that media exposure has impact on adoption of preventive measures regarding the pandemic and on negative emotions.

De Melo and Figueiredo [48] investigated both news articles and user tweets to compare the context in which the COVID-19 discussion is realized in Brazil over time. They used LDA for topic modeling along with lexicons and rule-based sentiment analysis tool called VADER for sentiment analysis, finding more negativity related to political themes in both media. Aslam *et al.* [44] analyzed sentiments and emotions in news headlines on COVID-19 from 25 top global English news sources and found that a majority (52%) of the them evoked negative sentiments, and only 30% evoked positive sentiment among its readers while 18% showed



FIGURE 1. Different parts of a news tweet from “The Times of India.” A media outlet may own multiple news accounts on Twitter.

neutral emotion. COVID-19 news coverage has also attracted a fair amount of work on addressing the misinformation spread and fake news [54]–[57].

Ayyoubzadeh *et al.* [58] used google trends to predict new cases of COVID-19 in Iran. They look for a set of keywords, including hand sanitizers, antiseptics, COVID-19 cases and so on, in the google trends data that may be indicative of increasing COVID-19 cases. Naseem *et al.* [5] created a dataset of 90,000 COVID-19 related tweets labelled for sentiments (positive, negative and neutral). They also present a word cloud to visualize common keywords and graphs to demonstrate dominant topics (set of words). However, their ground truth is based on predictions from TextBlob and may not be very reliable and their model shows a high error rate on test data.

Though the COVID-19 pandemic has been the target of extensive research, the public emotional response to the media coverage has not been adequately characterized. In this work, we analyzed data from users in a broad set of English-speaking countries; used an LDA model to identify topics from news tweets; exploited a large pretrained language model (RoBERTa) to classify the emotions aroused by media coverage of the pandemic on user responses; and reported this distribution of emotions in settings that include changes over time, location, news tweets themes and emotions. Table 1 shows a comparison between our work and related prior works.

III. METHODOLOGY

This section presents our methodology for collecting and analyzing Twitter data to investigate the prevalence of emotions in user responses to the COVID-19 news published on Twitter by popular media (Figure 1).

For the purpose of topic modeling we used unsupervised machine learning, which is an approach used to observe patterns in unlabeled data. It can be used to identify clusters of instances that are semantically related and sufficiently unrelated with other clusters in large unstructured text corpus. Advanced Natural Language Processing (NLP) deep learning techniques are already used to this intent [59], but we chose to use a statistical model, since the results are better interpretable and we could set the desired number of topics. We used MALLET (Machine Learning for Language Toolkit [60]) implementation of Latent Dirichlet Allocation (LDA) [11].

LDA is a generative probabilistic model of a corpus (i.e. a data set of texts), where each tweet content is represented as random mixtures over latent topics, and each topic is characterized by a distribution over words. It is a method that has been successfully used in a diversity of domains, such as research paper, clinical data, Twitter data for health care, and so on. [61] The MALLET implementation uses an optimized Gibbs sampling algorithm for LDA. Table 3 presents examples of news tweets labeled with some of the themes and sub-themes that we defined.

To identify emotions in tweets, we adopted a Robustly Optimized BERT Pretraining Approach (RoBERTa) model [9] retrained on ≈ 58 million tweets to capture the Twitter language specifics and fine-tuned for emotion detection on SemEval “Affects in Tweets” dataset [62]. RoBERTa is based on Bidirectional Encoder Representations from Transformers (BERT), a transformer-based deep learning language representation model. Compared to BERT, RoBERTa is trained on more data, with longer sequences, bigger batches, for a longer time and does not include the next sentence prediction objective during pretraining. While BERT advanced the state-of-the-art for eleven benchmark NLP tasks, RoBERTa achieved further improvements on GLUE and SQuAD benchmarks.

The “Affects in Tweets” dataset contains tweets labeled for multi-label classification. Each tweet is annotated for eleven emotions to capture the affectual state of the author of the tweet. The dataset was re-purposed for a multi-class classification problem by keeping only tweets with a single emotion. Emotions with less than 300 tweets were discarded, reducing the number of labels to four basic emotions {joy, optimism, anger, and sadness}. The fine-tuned model is evaluated on TweetEval [10], an evaluation benchmark for Twitter-specific classification tasks, and achieved state-of-the-art performance of 79.8% macro-averaged F1-score on the task of emotion recognition.

In summary, we started with identifying verified Twitter accounts from popular English news publishers (Figure 2) to collect the news tweets about COVID-19. We collected user responses to those news tweets and classified them based on their embedded emotions using a pretrained RoBERTa model and use Latent Dirichlet Allocation (LDA) to identify latent topics in news tweets.

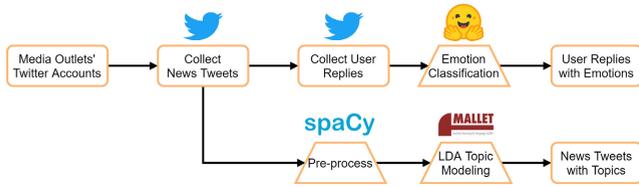


FIGURE 2. Method for selecting, collecting and processing content from social networks applied in this paper.



FIGURE 3. Word cloud of media outlets with text size proportional to the number of user responses.

A. DATA ACQUISITION

We first selected 282 English news accounts based on their popularity and user engagement, out of which six accounts were removed as they did not publish any news on COVID-19. A list of keywords¹ was used to collect news tweets and the responses to those tweets from 276 news accounts across 12 countries and 5 continents. For each news account, we collected the archived Twitter data using SnsCrape² for January 1st 2020–April 30th 2021.

As a result, 1,705,830 news tweets were collected containing at least one of the keywords and identified as written in English by Twitter metadata. Some news tweets just included a URL link to their news articles, without any text in the tweet content that we could search a keyword on. In such cases we used the URL to fetch the title of the news and included only the tweets with links to news articles whose headlines contained a COVID-19-related keyword.

From each of the identified news tweets, we collected all available Twitter replies (a response to another person’s tweet), excluding those which were (i) replies of replies, (ii) deleted by the author or by the platform, (iii) not identified as written in English in Twitter metadata, or (iv) non-public tweets from private Twitter accounts. This resulted in 18,882,812 user responses, from 3,038,495 unique Twitter accounts. Each collected tweet contains the following fields: id, conversation_id, author_id, created_at, text, lang, source, reply_count, retweet_count, like_count, quote_count, coordinates and place.

¹keywords: wuhan, ncov, coronavirus, covid, sars-cov-2, pandemic, lockdown, quarantine, social distancing, wearing masks, vaccination, vaccine, outbreak, panic buying, remote working, homeschooling.

²github.com/JustAnotherArchivist/snsCrape

B. TOPIC MODELING

The input to an LDA model is a text corpus, a list where each element represents an instance of text and indicates which words from a dictionary (i.e., a finite set of words) it contains. Nevertheless, the raw text of a tweet has multiple issues for the generation of a text corpus, such as different inflections of the same word (e.g., run, running), emojis, mentions to other Twitter users and web links. To address those issues we proceeded with pre-processing the tweet’s textual content. The pre-processing step was only applied for the topic modeling task.

We started by applying a regular expression to keep words and spaces, while removing numbers and the underscore symbols (regex code: `[^\w\s][0-9][_]`). Then, we converted each news tweet into a list of tokens (i.e. a sequence of characters that is a useful semantic unit for processing) in the lowercase lemma format (i.e. the base or dictionary form of a word) and removed all tokens with length equal to one. We also removed tokens that start with ‘@’ symbol, to avoid references to other Twitter accounts.

Tweets from the same location tend to share similar text features (such as similar entity names) which may bias our topic modeling towards grouping tweets originating from a geographical place together. To avoid this, we removed tokens classified as geographical entities (i.e. countries, cities, states names) so the news tweets are better reflective of the different themes related to the COVID-19 pandemic. We also removed tokens that are URLs, punctuation symbols, stop words, or spaces.

Next, we created bigrams and trigrams, frequent sequences of two and three tokens, and we removed tokens that occur in less than 0.05% or more than 33% of the documents, hence prohibiting rare and highly common tokens to bias the topic modeling. Finally, we obtained a dictionary with 1,966 tokens and a corpus indicating which of these words are contained in each news tweet.

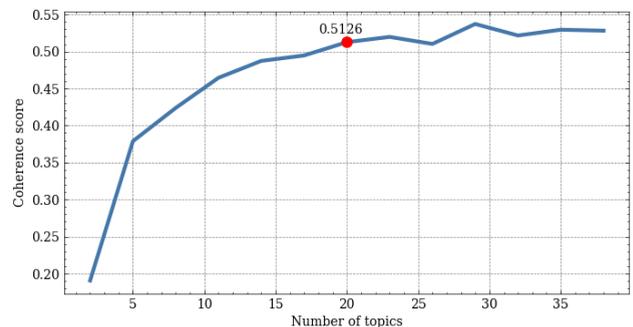


FIGURE 4. Coherence score over number of topics from LDA Mallet modeling.

Another required input for an LDA model is the number of topics we expected to extract from the corpus. We defined the optimal number of topics by analyzing the plot shown in Figure 4, where each point represents the coherence score of the LDA Mallet topic modeling with a varying number of topics, ranging from 2 to 40 with steps of 3. The coherence of

TABLE 2. Topic, theme and sub-theme description. This table presents the automatically generated topics at the left, its coherence and prevalence, and the themes and sub-themes we manually derived from them.

Topic top keywords	Topic Coherence	News tweets prevalence	Themes	Theme Coherence	Subthemes
day, surge, continue, high, country, hit, week, infection, wave, datum	49.34%	6.01%	Cases and deaths	54.99%	(1) Cases and deaths increasing (2) Cases and deaths decreasing
case, report, rise, record, numb, death, infection, daily, confirm, spike	64.86%	5.64%			
case, report, death, total, confirm, hour, recovery, breaking, death_toll, record	72.332%	3.68%			
pandemic, time, face, world, change, year, write, long, life, work	50.76%	5.16%	Economic impact	36.63%	(1) Economic impact over national economies and big companies (2) Economic plans to support jobs and food programs
pandemic, business, economy, impact, year, company, market, demand, employee, support	55.25%	3.84%			
health, vaccination, work, provide, service, community, fight, question, launch, worker	42.98%	2.09%			
amid, pandemic, school, close, student, child, open, fear, concern, shut	47.09%	2.75%	Educational impact	44.92%	None
patient, hospital, die, test_positive, doctor, family, man, woman, contract, treatment	52.80%	5.89%	People stories	50.79%	(1) Nursing homes and elderly victims (2) Mental health impact (3) Jail system and cops (4) Family stories
government, country, read, disease, national, announce, area, list, official, include	45.41%	7.30%	Authorities & Politics	29.25%	(1) Health authorities (2) Political authorities (3) Elections
amp, watch, pm, today, fight, post, meet, live, pandemic, speak	50.64%	3.29%			
news, update, late, live, city, follow, resident, outbreak, breaking, local	38.98%	2.82%			
response, pandemic, trump, crisis, american, call, president, plan, president_trump, deal	50.58%	6.53%			
state, test, positive, govt, centre, maharashtra, coronavirusoutbreak, live_update, district, conduct	52.99%	2.72%	Preventive measures	32.67%	(1) Mobility restrictions (2) Travelling protocols (3) People against restrictions or breaking rules (4) Non-pharmacological measures (5) Festivities and holidays (6) Easing restrictions
week, start, month, plan, begin, end, set, announce, march, early	51.05%	6.71%			
quarantine, test_positive, home, day, travel, return, leave, send, stay, passenger	49.47%	8.02%			
people, order, public, urge, issue, social_distance, police, follow, guideline, place	51.59%	3.98%			
lockdown, restriction, rule, extend, ease, measure, level, break, lift, impose	51.03%	4.66%			
outbreak, virus, spread, warn, global, official, country, world, emergency, vanguardnews	37.59%	5.53%			
people, find, expert, virus, risk, study, infect, variant, claim, spread	48.63%	2.86%			
vaccine, dose, receive, astrazeneca, give, trial, supply, ready, jab, shoot	61.93%	10.50%	Vaccines and vaccination	60.39%	None

a topic, used as a proxy for topic quality, is based on the distributional hypothesis that words with similar meaning tend to co-occur within a similar context [63], while the coherence score of a model with multiple topics is the average of the coherence score of all topics. Fang [64] found in several data sets that a few dozen topics modelled using LDA usually yield a coherence score between 0.4 and 0.5. We used 20 topics, that yield a coherence score of 0.51, and it is the highest number of topics that increases the coherence score more than 1% in relation to the previous number of topics. These topics are shown in Table 2. We then used both the dictionary and the number of topics to model the MALLET implementation of LDA.

C. EMOTION ANALYSIS

The model uses the text content from each tweet as input and recognizes one of the four emotion states {joy, optimism, anger, sadness} that the tweet is representative of. However, a tweet may contain other emotions that our model can't recognize. To address it, we used the softmax score generated by the model we adopted (for each emotion). Softmax is a normalized exponential function that normalizes the output of a model to a probability distribution over predicted classes, we used it as confidence score for each emotion. A tweet containing emotions other than the ones our model can recognize usually get a more homogeneous distribution of confidence scores across the four emotions it can detect.

TABLE 3. Examples of news tweets and its classification to theme and sub-theme.

Tweet content	Country	Theme	Sub-theme
"How to hold a kid's birthday party while social distancing"	Australia	Preventive Measures	Non-pharmacological measures
"... vaccines are fast approaching, but a second pandemic might impede efforts to recover from the first."	India	Vaccines & Vaccination	None
"Covid-infected Donald Trump instantly refuses to take part in virtual debate with Joe Biden"	UK	Authorities & Politics	Political authorities
"Which masks protect those around you best? Researchers weigh in"	US	Virus Spreading	Research & Treatment

TABLE 4. Examples of the emotional classification of the user replies to the news tweets.

Tweet content	Country	Emotion	Score
"... He doesn't respect the reporters neither the people that are at the frontline fighting for this virus ..."	United States	anger	0.937
"... so many doctors in Europe who have to decide who gets a respirator and who has to die ..."	United States	sadness	0.899
"... I do hope you and the rest of the 1743 family are well ..."	United Kingdom	optimism	0.556
"... I heard this as well via a reputable news source. You're not imagining it. :) ..."	United Kingdom	joy	0.581

We choose the emotion with the highest softmax score as representative of the emotion in a tweet with a threshold of 0.50. In case of mixed emotions when none of the emotion softmax score exceeds 0.50, we label such tweets as having an "undefined" emotion. Table 4 shows examples of tweets and their corresponding corresponding emotion scores. The news tweets, their user responses, and the emotion scores are available in our published dataset named REACT³ (Responses and Emotions from the Audience of COVID-19 news Tweets).

IV. RESULTS

Section III explained our employed topic modeling and emotion classification approach to understanding the emotional response to the COVID-19 news on Twitter. This section presents our findings and discusses them to answer the research questions listed below.

- (RQ1):** What are the COVID-19 news themes on Twitter?
(RQ2): What is the longitudinal distribution of the emotions and themes in the user responses?
(RQ3): What is the latitudinal distribution of the emotions and themes in the user responses?
(RQ4): What is the distribution of the emotions and themes in the user responses across different media outlets?

A. THE COVID-19 NEWS THEMES, ANSWERING (RQ1)

The topic modeling process described in Section III identified 20 different topics, with a mean coherence score of 51.26%. These topics were, manually, grouped into themes

based on the frequent words used to characterise the topics. For instance, for a topic T1 with {case, infection, rise} as its frequent words and topic T2 with {case, report, death} as its frequent words, we grouped T1 and T2 under "Cases and Deaths". This resulted in eight main themes, with a mean coherence of 43.01%. Each resulting theme was represented by the union of the words from the topics that generated it. Next, we extracted sub-themes from the generated themes, when appropriate. For example, the theme "Cases and Deaths" includes keywords such as {rise, increase, high, low, drop, fall}, that can be grouped in two different sub-themes representing news reporting on *increase* and *fall* in COVID-19 cases and deaths, respectively. The resulting topic, themes and sub-themes are listed in Table 2.

B. THE LONGITUDINAL DISTRIBUTION OF THE EMOTIONS AND THEMES, ANSWERING (RQ2)

The weekly distribution of the COVID-19 news tweets and their corresponding user responses can be seen in Figure 5. Media coverage on the pandemic peaked globally around April 2020, when the weekly number of the news tweets reached $\approx 100k$, followed by a steady pace of $\approx 30k$ news tweets per week until the end of the studied period. Most of the early tweets did not contain "covid"; they, rather, used alternative keywords such as "outbreak" and "coronavirus".

Keyword "covid" became common after WHO coined the term on February 11, 2020 [65]; "pandemic" became prevalent in news tweets after WHO declared COVID-19 a pandemic, on March 11, 2020 [66]. Although "lockdown" was highly prevalent from April to June of 2020, it became less prevalent as the restrictions eased. On the contrary, the prevalence of keyword "vaccine" increased with advent of the early instances of the COVID-19 vaccine; it was already a popular term when the first person was vaccinated on December 8, 2020.

Our results further demonstrated (Fig 5) that although sadness was the most prevalent emotion in the early COVID-19 news posted on Twitter, its prevalence declined steadily over time. The prevalence of optimism and joy in the news, on the other hand, demonstrated a steady increase after hitting the lowest early. Anger in the news did not show much variability.

A longitudinal analysis of the data also revealed that, unlike news tweets, the fluctuation of the emotions in the user responses to the news tweets was not notable over time: anger remained, by far, the most prevalent emotion during the course of the study, followed by sadness, optimism, and joy.

C. THE LATITUDINAL DISTRIBUTION OF THE EMOTIONS AND THEMES, ANSWERING (RQ3)

Figure 6a shows the distribution of the news themes across different countries. We also observed that the news media in the United States have a high prevalence of COVID-19-related news tweets about "Economic Impact." Also, the media outlets from Australia and the Philippines were associated with a higher prevalence of COVID-19 related news about "Authorities & Politics" while news about

³<https://globalaffects.org/covid-news>

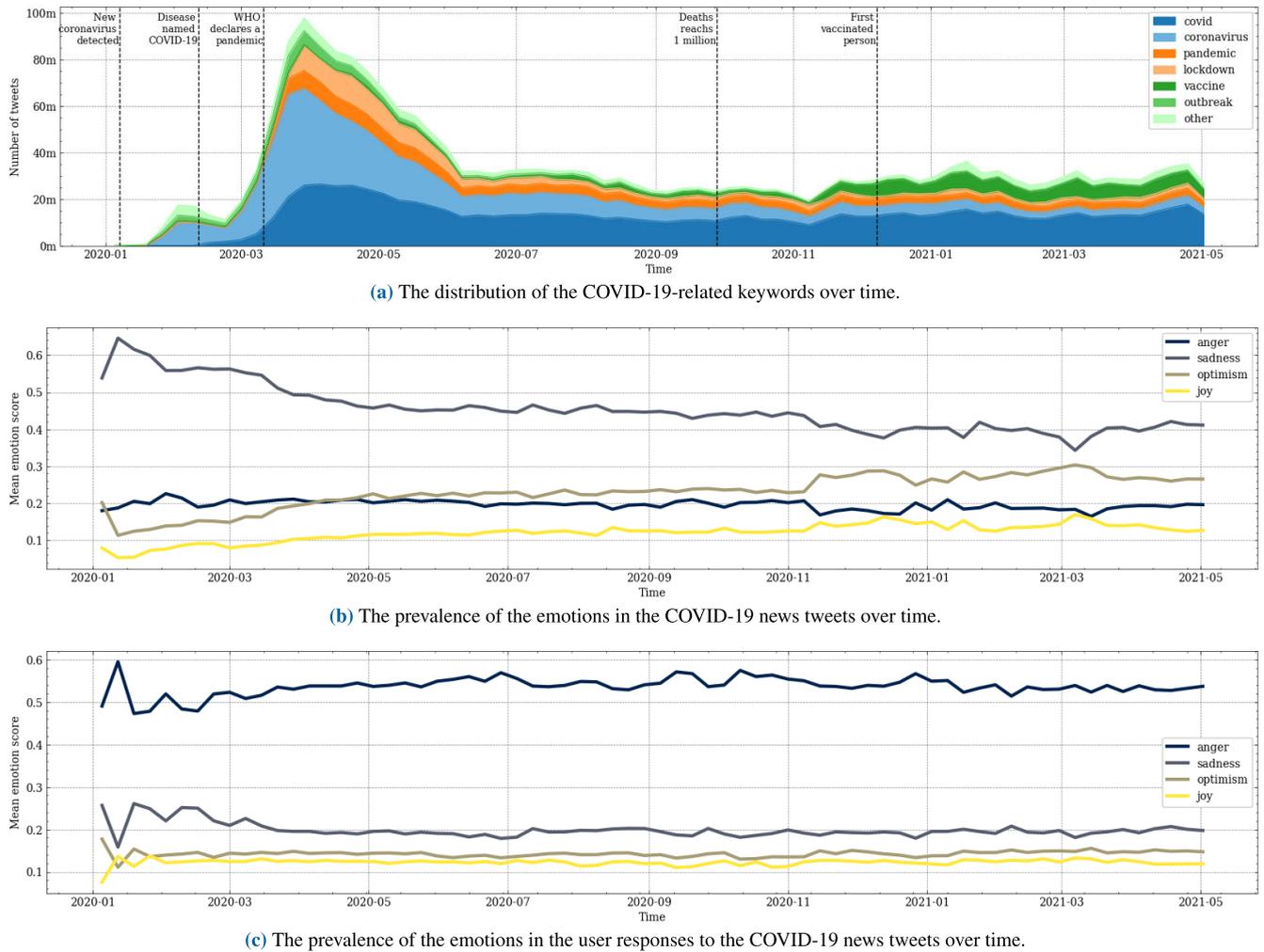


FIGURE 5. A longitudinal analysis of the collected tweets: (a) shows the number of tweets collected per week, the colors correspond to the keywords used for collecting the tweets; “other” denotes less frequent keywords. Also, (b) shows the weekly mean score of each emotion over the news tweets, and (c) shows the weekly mean score of each emotion over the user responses.

“Preventive Measures” were more prevalent in New Zealand and the UK.

Figure 6b helps understand the distribution of the emotions in the user responses to the COVID-19 news published by the media outlets from different countries: anger was the most prevalent emotion in the user responses across all countries; the media outlets from United States, Australia, and Canada constituted above 55% of user responses with anger being prevalent in them, this is about 35% for the Philippines. Also, in 20% of the user responses from Philippines, optimism was the most prevalent emotion.

Since the emotion distribution is similar across different countries, we used the *emotion lift* to illustrate regional differences. The emotion lift is calculated as the ratio of the emotion mean score of a region to the global mean score of that emotion, in an aggregated time interval (e.g., a month).

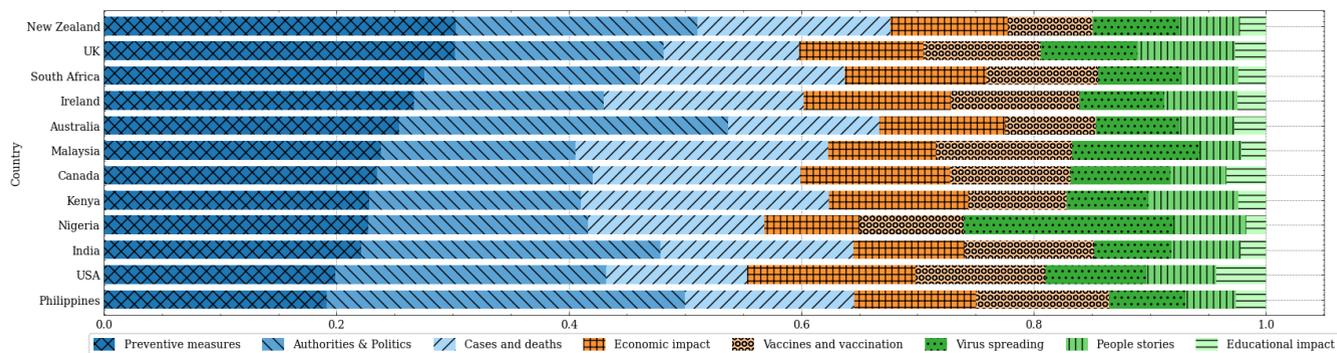
Figure 7 presents the monthly emotion lifts grouped by continent from January 2020 until April 2021, where a dashed black line represents a global average, and each continent is represented by the countries that are located in that continent

and their media outlets that have been studied in this paper. The results showed that Asia had the highest divergence from the global average: the emotion lifts of optimism and joy in user responses were larger than the global average, while anger remained lower than the global average.

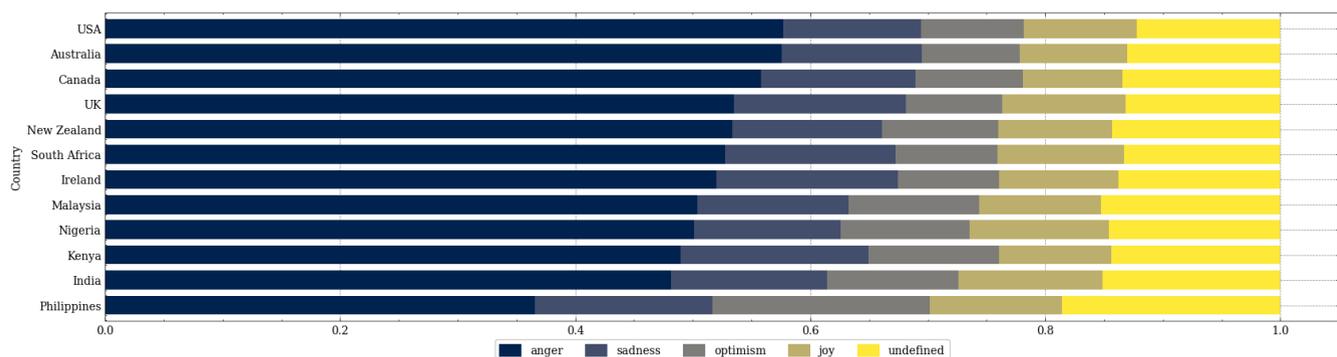
Oceania and North America demonstrated similar distributions, with emotion lifts closer to the average for all emotions, with the exception of anger, whole emotion lifts were slightly above the global average. In Europe, the highest emotion lift belonged to sadness, while joy was more prevalent in the user responses in Africa. The mean scores for anger were below average for both Africa and Europe.

D. THE DISTRIBUTION OF THE EMOTIONS AND THEMES ACROSS DIFFERENT MEDIA OUTLETS, ANSWERING (RQ4)

Table 5 shows the distribution of the emotions identified in the user responses to the COVID-19 news themes across different news accounts - the Twitter accounts associated with different media outlets. Table 5 has listed the top five news accounts



(a) The distribution of the emotions in the news tweets per country.



(b) The distribution of the emotions in the user responses to the news tweets per country.

FIGURE 6. The distribution of (a) news themes per country, and (b) emotions per country.

TABLE 5. The top 5 news accounts with the highest emotion scores for their user responses, and the distribution of the theme of their news tweets related to COVID-19. We only consider news accounts that have on average one or more COVID-19 related news post and user response per day on Twitter.

Emotion	News Accounts	Country	Followers (Millions)	Emotion Score	Authorities	Cases & deaths	Economic impact	Educational impact	People stories	Preventive measures	Vaccines & vaccination	Virus spreading
Anger	CNN Politics	US	4.14	61.8	51.0	7.1	7.6	2.2	2.5	17.1	8.4	4.2
	Fox News	US	20.19	61.8	38.1	8.3	5.3	1.7	5.7	19.7	15.9	5.3
	Breitbart News	US	1.47	61.1	32.7	9.6	10.0	3.0	5.3	22.0	5.7	11.7
	MSNBC	US	4.22	60.5	36.7	11.3	13.4	2.5	5.0	13.8	10.8	6.4
	Herald Sun	AU	0.33	60.3	24.4	14.1	14.9	3.6	4.6	25.6	6.3	6.5
Anger Top 5 News Accounts Average					37.6	10.5	11.7	2.6	4.4	17.8	9.0	6.4
Sadness	Firstpost	IN	2.09	34.0	23.4	15.1	16.0	2.3	4.8	23.6	8.1	6.7
	The New Indian Express	IN	0.47	32.5	22.1	10.4	13.5	3.4	8.0	27.1	7.0	8.5
	Free Press Journal	IN	0.02	27.8	26.7	17.2	7.4	2.5	7.7	24.8	8.9	4.8
	Oneindia News	IN	0.06	27.5	27.2	16.5	5.9	1.8	6.1	25.1	11.0	6.4
	The Hindu-Hyderabad	IN	0.06	26.1	27.9	18.6	6.1	2.0	10.5	24.0	6.3	4.7
Sadness Top 5 News Accounts Average					25.8	15.8	8.9	2.4	6.8	25.0	9.2	6.1
Optimism	Business Mirror	PH	0.04	39.6	37.5	7.8	19.9	2.9	1.9	14.5	10.7	4.7
	The Manila Times	PH	0.10	34.5	42.3	9.1	13.6	2.7	3.0	16.2	6.7	6.4
	Phil News Agency	PH	0.03	32.8	42.7	10.7	11.1	2.2	2.7	16.7	10.9	3.1
	Business Line	IN	0.08	30.8	23.5	15.2	17.4	1.5	3.3	15.1	13.5	10.5
	Manila Bulletin News	PH	0.86	26.9	28.0	17.0	8.2	2.0	5.7	20.0	13.9	5.3
Optimism Top 5 News Accounts Average					33.3	13.4	12.2	2.1	3.9	17.4	11.9	5.7
Joy	BBC Radio 4	UK	0.53	38.2	25.8	6.8	19.7	2.7	8.0	21.8	9.6	5.6
	Mid Day	IN	0.67	33.7	25.5	11.4	11.2	2.6	7.3	29.3	7.4	5.3
	BellaNaija	NG	1.79	27.1	29.4	7.0	22.3	2.7	3.2	17.7	5.2	12.6
	DD News	IN	3.23	23.3	30.7	20.7	8.2	1.2	4.2	19.6	10.3	5.0
	All India Radio News	IN	2.74	22.2	30.4	24.2	7.0	1.0	4.3	19.3	8.8	5.1
Joy Top 5 News Accounts Average					29.3	19.8	9.0	1.5	5.0	21.4	8.6	5.4
Global News Accounts Average					22.7	15.3	11.1	2.8	5.9	23.3	10.5	8.4

with the highest average emotion scores in user responses to different news themes. For instance, BBC Radio 4 on Twitter is associated with the highest joyful user responses, with a average emotion score of 38.2% (i.e., the mean emotion score for joy across all user responses has been 38.2%).

Table 6 shows the distribution of the emotional responses across different news sub-themes. For example, “Family Stories” has a value of 1.27 for joy, which means 1.27% of all user responses to the news tweets with joy being prevalent in them correspond to the news about “Family Stories.” Also,

TABLE 6. The distribution of the emotions in user responses to the news sub-themes. Column “Global” represents the overall distribution of the sub-themes. News tweets that were not associated to any of the predefined sub-themes were labelled with a generic tag “(default)”.

Theme Description	Subtheme Description	Anger	Joy	Optimism	Sadness	Global
Cases and deaths	Cases and deaths (default)	4.49	13.95	4.58	9.01	10.26
	Cases and deaths increasing	1.61	6.14	1.72	4.32	4.39
	Cases and deaths decreasing	0.12	1.25	0.21	0.11	0.66
Economic impact	Economic impact (default)	4.71	6.53	8.44	7.59	6.39
	Economic impact over national economies and bi...	1.67	2.65	4.03	1.47	2.48
	Economic plans to support jobs and food programs	1.21	2.61	3.43	1.10	2.24
Educational impact	Educational impact (default)	1.85	3.71	2.02	3.23	2.75
People stories	People stories (default)	3.55	6.77	2.54	4.51	4.91
	Nursing homes and elderly victims	0.06	0.23	0.03	0.06	0.13
	Mental health impact	NaN	0.01	0.00	NaN	0.00
	Jail system and cops	0.19	0.06	0.01	0.02	0.06
	Family stories	0.37	1.27	0.18	1.16	0.78
Authorities & Politics	Authorities & Politics (default)	16.89	14.21	17.96	19.07	16.31
	Health authorities	0.69	0.53	1.05	0.43	0.73
	Political authorities	10.41	3.21	7.46	3.01	5.37
	Elections	0.38	0.20	0.30	0.09	0.25
Preventive measures	Preventive measures (default)	11.11	12.02	9.87	15.64	12.11
	Mobility restrictions	21.09	7.38	4.08	6.35	8.61
	Travelling protocols	2.02	1.37	1.51	1.22	1.49
	People against restrictions or breaking rules	1.59	0.14	0.10	0.19	0.29
	Non-pharmacological measures	0.97	0.47	0.56	0.62	0.57
	Festivities and holidays	0.10	0.17	0.11	0.47	0.16
	Easing restrictions	0.04	0.08	0.22	0.08	0.12
Virus spreading	Virus spreading (default)	4.91	5.24	4.66	2.91	4.89
	Outbreak and pandemic origins	1.58	2.68	1.11	0.57	1.92
	Research and treatment	0.44	1.16	2.21	0.40	1.19
	New waves and variants	0.20	0.33	0.23	0.13	0.31
	Health workers	0.09	0.14	0.10	0.04	0.10
Vaccines and vaccination	Vaccines and vaccination (default)	7.68	5.50	21.28	16.20	10.52

TABLE 7. Engagement metrics for the COVID-19 news tweets per prevalent emotion. Reply, Retweet, Like, and Quote denote the average number of the user replies, tweet sharing, tweet appreciations, and comments on tweets, respectively: there are (on average) 21.6 replies to the news with anger as the prevalent emotion.

Prevalent Emotion	Reply	Retweet	Like	Quote
Anger	21.6	37.4	115.2	11.6
Joy	8.5	23.0	99.8	7.8
Optimism	11.5	22.9	88.9	7.0
Sadness	10.3	25.6	68.6	7.0
Undefined	11.2	22.8	74.3	7.5

0.78% of all user responses correspond to the news tweets about “Family Stories.”

Table 6 also shows that 21.09% of all tweets with anger as a prevalent emotion (i.e., score greater than 50%) are on the sub-theme of “Mobility Restrictions”, while this sub-theme appears in only 8.61% of all tweets related news from COVID-19. Similarly, 10.41% of the angry tweets were on the sub-theme of “Political Authorities”, while this sub-theme is present at 5.37% of the news tweets related to COVID-19. This shows that these sub-themes are strongly associated with a angry audience response.

TABLE 8. Emotion lift of the user responses to news tweets, grouped by theme.

Theme	Emotion Lift			
	Anger	Sadness	Optimism	Joy
Cases and deaths	0.95	1.16	1.01	0.95
Economic impact	0.97	1.02	1.06	1.06
Educational impact	0.94	1.06	1.03	1.14
People stories	0.92	1.27	0.94	1.01
Authorities & Politics	1.06	0.88	0.96	0.95
Preventive measures	1.01	0.98	0.97	1.03
Virus spreading	0.99	1.04	1.00	0.97
Vaccines and vaccination	0.96	0.96	1.16	1.06

Four of these accounts are from the United States media outlets and one is from Australia. They are also among the Twitter accounts observed in this paper with the highest number of followers.

Table 7 shows that news tweets associated with anger as their prevalent emotion had the highest level of engagement for all metrics. News tweets with the prevalence of joy have the smallest average number of user responses (i.e., Reply), although they had the second-highest number of likes. Also, the news tweets associated with sadness have second lowest

average number of user responses and likes, but have the second highest number of retweets.

The themes associated with the news tweets seem to bear some correlation with the emotions it triggers among its readers. Table 8 shows that the emotion lift of anger is higher for user responses to the news about “Authorities & Politics” compared to other themes. Table 5 shows that “Authorities & Politics” corresponds to 22.66% of all news tweets on COVID-19, while it is the theme of 37.6% of the news tweets from the top 5 news accounts with the highest prevalence of anger in their user responses, on average.

V. DISCUSSION

The impact of news on human emotions has long been recognized; such impacts can be significant in the presence of global crises such as the COVID-19 pandemic, due the far-reaching impacts of such phenomena on various aspects of human lives. Although this paper has studied the emotional expressions of the news readers (users) in response to the COVID-19 news on Twitter, our findings must be interpreted with certain considerations. First and foremost, human emotions are complex and various factors might have influenced the expression of certain emotions in the user responses to COVID-19 news. Our results thus should be interpreted with the following caveat. While the prevalence of anger, for instance, in user responses to a news post imply a (potential) relationship between the content of that news post and the expression of anger, the nature of that relationship cannot be established based on our current analysis, meaning we can not conclude that this particular news post has been the primary factor for such the expression of anger in the users response. It is particularly important to avoid overgeneralizing our results (Section IV).

Having said that, we had interesting observations which can inform research to further investigate the impacts of the news (framing) on people’s emotions amid the COVID-19 pandemic. As one example, the mean score of joy demonstrated to be the highest in user responses to the news posted by BBC Radio 4, described as “Your friendly lockdown companion,” on Twitter. The account adjusted the COVID-19 news to share less about Cases and Deaths and more about People Stories (Table 5). That, arguably, instigated positive feelings in the audience, resulting in the highest prevalence of joy in the user responses, compared to the other news accounts. This can be further investigated to identify possible relationships between different ways of framing COVID-19 news and the emotional responses to the news.

Another example is a COVID-19 news post on Twitter that was perceived as inaccurate and politically loaded by the audience.⁴ Published by a popular Nigerian media account (Punch Newspaper) the news post, arguably, instigated a high level of anger in the user responses to the posted news. This was also observed in our collected data in October 2020,

⁴<https://punchng.com/hoodlums-attack-warehouse-loot-covid-19-palliatives-in-lagos>

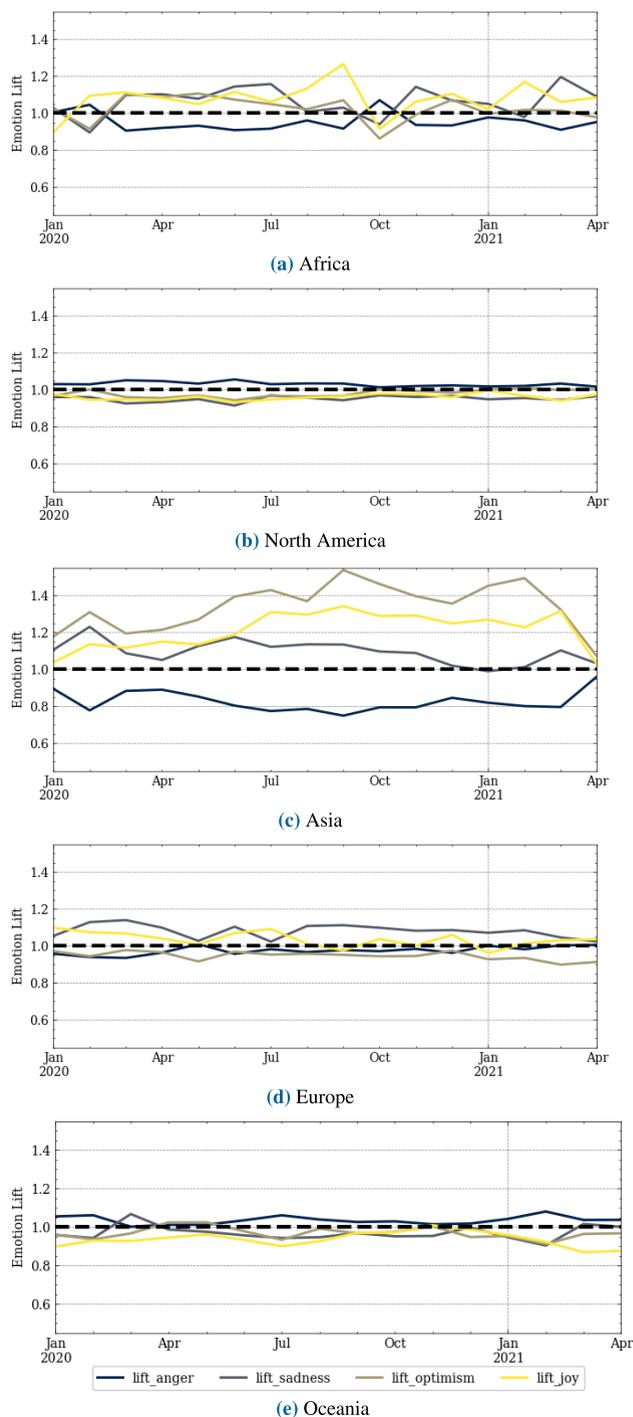


FIGURE 7. Monthly lift of emotions over the user responses grouped by continent. The lift shows the mean score of an emotion in a month in relation to the global mean score of that emotion in the same month. A lift greater than one indicates that the continent has had a mean score that is higher than global average.

where anger was found to have gained the highest emotion lift in the timeline (Figure 7a). This aligns with the research findings [67] that suggest “emotion-laden” news can impact the perceived severity of the incidents, and individuals may devalue the quality of such news. It also demonstrates how news sensationalization can influence the type and magnitude

of the public's emotional response to the news. This is also consistent with the research findings showing media outlets perceived as less fair in their published news witness a fall in public trust [68].

Another important finding of this research (Sub-section IV-D) was that the emotional responses to the COVID-19 news varied across the media outlets from different geographical locations. For instance, the news accounts from the USA (CNN Politics, Fox News, MSNBC, Breitbart News) constituted 4 out of the top 5 accounts with the highest average score of anger in their user responses; the other news account belonged to an Australian news agency (Herald Sun). Optimism, on the other hand, was the most prevalent emotion in user responses to the Filipino news accounts on Twitter. Also, all five news accounts with the highest mean score of sadness in their user responses were from India. While it may not be accurate to generalize these to the population of the aforementioned countries, significant variances in the emotional expressions across different regions may raise the question: "to what extent can cultural and socioeconomic factors influence the emotional responses to the COVID-19 news?"

While different emotions were prevalent in the contents of the COVID-19 news tweets, our study revealed that the level of engagement varied for the news with different prevalent emotions (Table 7). For instance, news with higher presence of anger in their contents had the highest average number of user replies, retweets, likes and quotes on Twitter. This further supports research that suggest the contents associated with anger often become more viral on social media [69] and that emotionally charged Twitter messages tend to be retweeted more frequently [70].

VI. LIMITATIONS AND THREATS TO VALIDITY

This section discusses the limitations of the paper to confine it to its theoretical and practical borders.

A. INTERNAL VALIDITY

We have used emotion classification to characterize human emotions in the COVID-19 news replies posted on Twitter. The identified emotions might have been triggered by the news contents per se or the way the news has been presented (i.e., framed) to the audience. The methodology used in this paper and, consequently, our findings do not distinguish between these two; complementary research is needed to separate the impact of the news content and news framing on emotions.

Moreover, although we have identified the emotions directly from the user responses to the news posted on Twitter, we cannot rule out the possibility that other factors beyond news itself, e.g., events such as elections and economic issues, might have triggered the users to express certain emotions in their replies. Identifying such factors and controlling them, however, is not trivial and goes beyond the scope of this research. As such, we have only reported the emotions

identified in the news replies without establishing any causal relationship between the emotional responses and the news.

B. CONSTRUCT VALIDITY

For countries with a multilingual population, such as India, the Philippines, and Singapore, some tweets may contain non-English words or phrases. We rely on identifying the language used in the tweets based on the Twitter metadata that recognizes the language a tweet is written in. This may limit the accuracy in recognizing language in multilingual tweets common from these geolocations. This may have an impact on the accuracy of detecting emotions in tweets from these countries where both native languages and English are used to communicate on Twitter.

Lastly, we have used a broad set of keywords to identify news pertaining to COVID-19, but the set is not comprehensive to cover all COVID-19 news topics; some news tweets related to the COVID-19 pandemic may not have been captured.

C. EXTERNAL VALIDITY

We have used Twitter messages to analyze the emotional responses to the COVID-19 news. While a sample of a few dozen million tweets has been used across the countries of interest, the authors avoid generalizing the results to the population of those countries, as the samples may not necessarily represent the attitude of the society in general. Moreover, the results of our study are limited to the COVID-19 news posts as well as the replies to those posts in English only; non-English posts have not been analyzed due to the shortcomings of the current NLP techniques [71].

VII. CONCLUSION AND FUTURE WORK

In this paper, we investigated the emotional response to the COVID-19 news to develop a theoretical framework for analyzing the impact of the COVID-19 news on people's emotions. In doing so, we used the Robustly Optimized BERT Pretraining Approach (RoBERTa) model to identify the prevalence of basic human emotions in about 19 million user responses corresponding to 1.7 million COVID-19-related tweets from official twitter accounts of popular news media from 12 countries between January 2020 to April 2021. A topic modeling (LDA) approach was used to identify the news themes.

Our analysis of the Twitter data identified anger as the most prevalent emotion in user responses to the news coverage of COVID-19. That was followed by sadness, optimism, and joy, steadily over the period of the study. We also observed a higher prevalence of optimism and joy in user responses to the news on vaccination and educational impacts of COVID-19 respectively. However, news about authorities and politics was associated with the highest prevalence of anger in user responses. Also, sadness in user responses, was mainly linked with the news about COVID-19 cases and deaths and the impacts on the families, mental health, jails, and nursing homes. Our findings further demonstrated that anger was

more prevalent in the user responses to the (COVID-19) news posted by the USA media accounts (e.g., CNN Politics, Fox News, MSNBC). Optimism in user responses, however, was the highest in Filipino media.

This work can be extended in several directions. First, correlation analysis can be used to identify the links between news framing and the emotional responses to the news. Second, our findings can be looked at from a psychology perspective to understand the potential impacts of the COVID-19 news on mental health. Third, the scope of the research can be extended to a longer period of the pandemic with more geographical locations. Finally, our findings can be studied in the light of the major events during the pandemic, e.g., elections, that might have affected user's emotional responses to the COVID-19 news.

RESOURCES

Our source code and dataset (REACT Data Set) can be accessed from <https://globalaffects.org/covid-news>.

REFERENCES

- [1] N. Mheidly and J. Fares, "Leveraging media and health communication strategies to overcome the COVID-19 infodemic," *J. Public Health Policy*, vol. 41, no. 4, pp. 410–420, Dec. 2020.
- [2] T. J. VanderWeele and A. C. Brooks, "A public health approach to negative news media: The 3-to-1 solution," *Amer. J. Health Promotion*, May 2020, Art. no. 089011712091422, doi: [10.1177/0890117120914227](https://doi.org/10.1177/0890117120914227).
- [3] R. Jones, D. Mougouei, and S. L. Evans, "Understanding the emotional response to COVID-19 information in news and social media: A mental health perspective," *Hum. Behav. Emerg. Technol.*, vol. 3, no. 5, pp. 832–842, 2021, doi: [10.1002/hbe2.304](https://doi.org/10.1002/hbe2.304).
- [4] J. Xue, J. Chen, R. Hu, C. Chen, C. Zheng, Y. Su, and T. Zhu, "Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach," *J. Med. Internet Res.*, vol. 22, no. 11, Nov. 2020, Art. no. e20550.
- [5] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, "COVID-Senti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 4, pp. 1003–1015, Aug. 2021.
- [6] P. Ghasiya and K. Okamura, "Investigating COVID-19 news across four nations: A topic modeling and sentiment analysis approach," *IEEE Access*, vol. 9, pp. 36645–36656, 2021.
- [7] F. Aslam, T. M. Awan, J. H. Syed, A. Kashif, and M. Parveen, "Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak," *Humanities Social Sci. Commun.*, vol. 7, no. 1, pp. 1–9, Dec. 2020.
- [8] P. Ekman, "Basic emotions," *Handbook Cognition Emotion*, vol. 98, nos. 45–60, p. 16, 1999.
- [9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.
- [10] F. Barbieri, J. Camacho-Collados, L. Espinosa Anke, and L. Neves, "TweetEval: Unified benchmark and comparative evaluation for tweet classification," in *Proc. Findings Assoc. Comput. Linguistics, (EMNLP)*, 2020, pp. 1644–1650. [Online]. Available: <https://aclanthology.org/2020.findings-emnlp.148>
- [11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
- [12] C. Beckett, *SuperMedia: Saving Journalism so it Can Save the World*. Hoboken, NJ, USA: Wiley, 2011.
- [13] C. Beckett, *The Value Networked Journalism*. London, U.K.: POLIS, London School of Economics and Political Science, 2010.
- [14] D. Zillmann, R. Gibson, and S. L. Sargent, "Effects of photographs in news-magazine reports on issue perception," *Media Psychol.*, vol. 1, no. 3, pp. 207–228, Sep. 1999.
- [15] T. May, "Public communication, risk perception, and the viability of preventive vaccination against communicable diseases," *Bioethics*, vol. 19, no. 4, pp. 407–421, Aug. 2005.
- [16] I. Y. S. Yahaya, "Role of media in portraying Ebola in and outside Africa," *J. Tropical Diseases*, vol. 3, no. 1, 2015, Art. no. 100152.
- [17] M. Y. Ni, L. Yang, C. M. C. Leung, N. Li, X. I. Yao, Y. Wang, G. M. Leung, B. J. Cowling, and Q. Liao, "Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: Cross-sectional survey," *JMIR Mental Health*, vol. 7, no. 5, May 2020, Art. no. e19009.
- [18] J. Gao, P. Zheng, Y. Jia, H. Chen, Y. Mao, S. Chen, Y. Wang, H. Fu, and J. Dai, "Mental health problems and social media exposure during COVID-19 outbreak," *PLoS ONE*, vol. 15, no. 4, Apr. 2020, Art. no. e0231924.
- [19] E. A. Holmes, R. C. O'Connor, V. H. Perry, I. Tracey, S. Wessely, L. Arseneault, C. Ballard, H. Christensen, R. C. Silver, I. Everall, and T. Ford, "Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science," *Lancet Psychiatry*, vol. 7, no. 6, pp. 547–560, Jun. 2020.
- [20] C. Beckett and M. Deuze, "On the role of emotion in the future of journalism," *Social Media + Soc.*, vol. 2, no. 3, 2016, Art. no. 2056305116662395, doi: [10.1177/2056305116662395](https://doi.org/10.1177/2056305116662395).
- [21] N. M. Shelke, "Approaches of emotion detection from text," *Int. J. Comput. Sci. Inf. Technol.*, vol. 2, no. 2, pp. 123–128, 2014.
- [22] L. Canales and P. Martínez-Barco, "Emotion detection from text: A survey," in *Proc. Workshop Natural Lang. Process. 5th Inf. Syst. Res. Work. Days (JISIC)*, 2014, pp. 37–43.
- [23] H. Binali, C. Wu, and V. Potdar, "Computational approaches for emotion detection in text," in *Proc. 4th IEEE Int. Conf. Digit. Ecosyst. Technol.*, Apr. 2010, pp. 172–177.
- [24] P. Ekman, "An argument for basic emotions," *Cognit. Emotion*, vol. 6, nos. 3–4, pp. 169–200, 1992.
- [25] R. Plutchik, "Emotions: A general psychoevolutionary theory," *Approaches Emotion*, vol. 1984, pp. 197–219, Jan. 1984.
- [26] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
- [27] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, p. 1161, Dec. 1980.
- [28] N. Colneric and J. Demsar, "Emotion recognition on Twitter: Comparative study and training a unison model," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 433–446, Jul. 2020.
- [29] J. C. Norcross, E. Guadagnoli, and J. O. Prochaska, "Factor structure of the profile of mood states (POMS): Two partial replications," *J. Clin. Psychol.*, vol. 40, no. 5, pp. 1270–1277, Sep. 1984.
- [30] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, Aug. 2013.
- [31] C. Baziotis, A. Nikolaos, A. Chronopoulou, A. Kolovou, G. Paraskevopoulos, N. Ellinas, S. Narayanan, and A. Potamianos, "NTUA-SLP at SemEval-2018 task 1: Predicting affective content in tweets with deep attentive RNNs and transfer learning," in *Proc. 12th Int. Workshop Semantic Eval.*, Jun. 2018, pp. 245–255. [Online]. Available: <https://aclanthology.org/S18-1037>
- [32] J. H. Park, P. Xu, and P. Fung, "PlusEmo2 Vec at SemEval-2018 task 1: Exploiting emotion knowledge from emoji and #hashtags," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 264–272. [Online]. Available: <https://aclanthology.org/S18-1039>
- [33] M. Abdullah and S. Shaikh, "TeamUNCC at SemEval-2018 task 1: Emotion detection in English and Arabic tweets using deep learning," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 350–357.
- [34] V. Duppada, R. Jain, and S. Hiray, "SeerNet at SemEval-2018 task 1: Domain adaptation for affect in tweets," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 18–23.
- [35] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, "Lexicon based feature extraction for emotion text classification," *Pattern Recognit. Lett.*, vol. 93, pp. 133–142, Jul. 2017.
- [36] D. Garcia, M. Pellert, J. Lasser, and H. Metzler, "Social media emotion macroscopes reflect emotional experiences in society at large," 2021, *arXiv:2107.13236*.
- [37] O. Kaminska, C. Cornelis, and V. Hoste, "Nearest neighbour approaches for emotion detection in tweets," in *Proc. 11th Workshop Comput. Approaches to Subjectivity, Sentiment Social Media Anal.*, Apr. 2021, pp. 203–212. [Online]. Available: <https://aclanthology.org/2021.wassa-1.22>

- [38] R. Verma, C. von der Weth, J. Vachery, and M. Kankanhalli, "Identifying worry in Twitter: Beyond emotion analysis," in *Proc. 4th Workshop Natural Lang. Process. Comput. Social Sci.*, Nov. 2020, pp. 72–82. [Online]. Available: <https://aclanthology.org/2020.nlpssc-1.9>
- [39] A. Murarka, B. Radhakrishnan, and S. Ravichandran, "Classification of mental illnesses on social media using RoBERTa," in *Proc. 12th Int. Workshop Health Text Mining Inf. Anal.*, Apr. 2021, pp. 59–68. [Online]. Available: <https://aclanthology.org/2021.louhi-1.7>
- [40] G. Tziafas, K. Kogkalidis, and T. Caselli, "Fighting the COVID-19 infodemic with a holistic BERT ensemble," 2021, *arXiv:2104.05745*.
- [41] J. Mowery, "Contrastive clustering: Toward unsupervised bias reduction for emotion and sentiment classification," 2021, *arXiv:2111.07448*.
- [42] R. Gaspar, C. Pedro, P. Panagiotopoulos, and B. Seibt, "Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events," *Comput. Hum. Behav.*, vol. 56, pp. 179–191, Mar. 2016.
- [43] S. Kim, S. K. Cho, and S. P. Locascio, "The role of media use and emotions in risk perception and preventive behaviors related to COVID-19 in south Korea," *Asian J. Public Opinion Res.*, vol. 8, pp. 297–323, Aug. 2020.
- [44] F. Aslam, T. M. Awan, J. H. Syed, A. Kashif, and M. Parveen, "Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak," *Humanities Social Sci. Commun.*, vol. 7, no. 1, pp. 1–9, Dec. 2020.
- [45] M. Cheong and V. C. S. Lee, "A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter," *Inf. Syst. Frontiers*, vol. 13, no. 1, pp. 45–59, Mar. 2011.
- [46] P. Barnaghi, P. Ghaffari, and J. G. Breslin, "Opinion mining and sentiment polarity on Twitter and correlation between events and sentiment," in *Proc. IEEE 2nd Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Mar. 2016, pp. 52–57.
- [47] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in Twitter events," *J. Amer. Soc. Inf. Technol.*, vol. 62, no. 2, pp. 406–418, Feb. 2011.
- [48] T. de Melo and C. M. S. Figueiredo, "Comparing news articles and tweets about COVID-19 in Brazil: Sentiment analysis and topic modeling approach," *JMIR Public Health Surveill.*, vol. 7, no. 2, Feb. 2021, Art. no. e24585.
- [49] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, "Public discourse and sentiment during the COVID 19 pandemic: Using latent Dirichlet allocation for topic modeling on Twitter," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0239441.
- [50] M. Pierce, H. Hope, T. Ford, S. Hatch, M. Hotopf, A. John, E. Kontopantelis, R. Webb, S. Wessely, S. Mcmanus, and K. M. Abel, "Mental health before and during the COVID-19 pandemic: A longitudinal probability sample survey of the U.K. population," *Lancet Psychiatry*, vol. 7, no. 10, pp. 883–892, Oct. 2020.
- [51] S. P. Giri and A. K. Maurya, "A neglected reality of mass media during COVID-19: Effect of pandemic news on individual's positive and negative emotion and psychological resilience," *Personality Individual Differences*, vol. 180, Oct. 2021, Art. no. 110962. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0191886921003378>
- [52] S. Boon-Itt and Y. Skunkan, "Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study," *JMIR Public Health Surveill.*, vol. 6, no. 4, Nov. 2020, Art. no. e21978.
- [53] C. Crocamo, M. Viviani, L. Famigliini, F. Bartoli, G. Pasi, and G. Carrà, "Surveilling COVID-19 emotional contagion on Twitter by sentiment analysis," *Eur. Psychiatry*, vol. 64, no. 1, p. e17, 2021.
- [54] J. S. Brennen, F. Simon, P. N. Howard, and R. K. Nielsen, "Types, sources, and claims of covid-19 misinformation," *Reuters Inst.*, vol. 7, no. 3, p. 1, 2020.
- [55] A. Bridgman, E. Merkley, P. J. Loewen, T. Owen, D. Ruths, L. Teichmann, and O. Zhilin, "The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media," *Harvard Kennedy School Misinformation Rev.*, vol. 1, no. 3, Jun. 2020, doi: [10.37016/mr-2020-028](https://doi.org/10.37016/mr-2020-028).
- [56] S. van der Linden, J. Roozenbeek, and J. Compton, "Inoculating against fake news about COVID-19," *Frontiers Psychol.*, vol. 11, p. 2928, Oct. 2020.
- [57] B. O. Ahinkorah, E. K. Ameyaw, J. E. Hagan, A.-A. Seidu, and T. Schack, "Rising above misinformation or fake news in Africa: Another strategy to control COVID-19 spread," *Frontiers Commun.*, vol. 5, p. 45, Jun. 2020.
- [58] S. M. Ayyoubzadeh, S. M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, and S. R. N. Kalhori, "Predicting COVID-19 incidence through analysis of Google trends data in Iran: Data mining and deep learning pilot study," *JMIR Public Health Surveill.*, vol. 6, no. 2, Apr. 2020, Art. no. e18828.
- [59] D. Angelov, "Top2 Vec: Distributed representations of topics," 2020, *arXiv:2008.09470*.
- [60] A. K. McCallum. (2002). *Mallet: A Machine Learning for Language Toolkit*. [Online]. Available: <http://mallet.cs.umass.edu>
- [61] U. Chauhan and A. Shah, "Topic modeling using latent Dirichlet allocation: A survey," *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–35, Sep. 2022.
- [62] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "SemEval-2018 task 1: Affect in tweets," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 1–17. [Online]. Available: <https://aclanthology.org/S18-1001>
- [63] S. Syed and M. Spruit, "Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Oct. 2017, pp. 165–174.
- [64] A. Fang, "Analysing political events on twitter: Topic modelling and user community classification," in *ACM SIGIR Forum*, vol. 53, no. 1, New York, NY, USA: ACM, 2021, pp. 38–39.
- [65] WHO. (2020). *Novel Coronavirus(2019-Ncov) Situation Report 22*. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/%20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2
- [66] WHO. (2020). *Novel Coronavirus(2019-Ncov) Situation Report 51*. [Online]. Available: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/%20200311-sitrep-51-covid-19.pdf>
- [67] C. Klemm, T. Hartmann, and E. Das, "Fear-mongering or fact-driven? Illuminating the interplay of objective risk and emotion-evoking form in the response to epidemic news," *Health Commun.*, vol. 34, no. 1, pp. 74–83, Jan. 2019, doi: [10.1080/10410236.2017.1384429](https://doi.org/10.1080/10410236.2017.1384429).
- [68] N. Newman, R. Fletcher, A. Schulz, S. Andi, C. T. Robertson, and R. K. Nielsen, "Reuters institute digital news report 2021," Reuters Inst. Study Journalism, Oxford, U.K.
- [69] M. Buckley and K. L. Milkman, "What makes online content viral?" *J. Marketing Res.*, vol. 49, no. 2, pp. 192–205, 2012. [Online]. Available: www.marketingpower.com/jmr_
- [70] S. Stieglitz and L. Dang-Xuan, "Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior," *J. Manage. Inf. Syst.*, vol. 29, no. 4, pp. 217–248, Apr. 2013, doi: [10.2753/MIS0742-1222290408](https://doi.org/10.2753/MIS0742-1222290408).
- [71] M. Kaity and V. Balakrishnan, "Sentiment lexicons and non-english languages: A survey," *Knowl. Inf. Syst.*, vol. 62, no. 12, pp. 4445–4480, Dec. 2020, doi: [10.1007/s10115-020-01497-6](https://doi.org/10.1007/s10115-020-01497-6).



FRANCISCO BRÁULIO OLIVEIRA is currently pursuing the master's degree with the University of São Paulo under the supervision of Dr. Jaime Sichman. He is an Electrical Engineer and a Specialist in data science, with more than five years of experience in the retail, banking, energy and mobile industries, where he applied machine learning models to client clusterization, demand forecasting, fraud prevention, and others. His main domain of knowledge is in artificial intelligence products and NLP solutions. His research interests include creating models to better understand human behavior and optimize decision-making.



AMANUL HAQUE received the master's degree in computer science from North Carolina State University under the supervision of Dr. Collin F. Lynch, where he is currently pursuing the Ph.D. degree with the Department of Computer Science. He has two years of industry experience working at Oracle on Software Deployment Infrastructure (SDI) for Oracle Public Cloud (OPC) and smaller stints of working on machine learning projects at Lenovo and Seagate.

His research interests include social aspects of artificial intelligence and focuses on using traditional and deep learning approaches for information extraction and inference from the unstructured text; sentiment and emotion identification and text summarization.



DAVOUD MOUGOUEI is currently a Lecturer at the University of Southern Queensland. His research interests include the intersection of computing (software engineering, AI, and data science) and social sciences. He was a recipient of the Regional Collaborations Program Covid-19 Digital Grants (Australian Academy of Science) and the University Global Partnership Network Fund (UGPN) for his interdisciplinary research on the emotional impacts of Covid-19 news on people.



JAIME SIMÃO SICHMAN is currently a Full Professor with the Computer Engineering and Digital Systems Department (PCS), Escola Politécnica (EP), Universidade de São Paulo (USP), Brazil. His main research interests include multi-agent systems, more particularly social and organizational reasoning, multi-agent-based simulation, reputation and trust, and interoperability in agent systems. He is a member of the Editorial Board of several journals, such as the *Knowledge Engineering Review* and *Autonomous Agents and Multi-Agent Systems*. He is a member of the Board of Directors of the International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS). He was the AAMAS Tutorial Chair (2007), the Program Co-Chair (2009), the Local Chair (2017), and the Workshop Co-Chair (2020). He was indicated as a Distinguished Speaker of the Association of Computing Machinery (ACM), in 2012.



SIMON EVANS received the Ph.D. degree from the UCL Institute of Neurology investigating social decision-making processes and their neural correlates. He then conducted postdoctoral work at the University of Sussex, applying MRI techniques to explore how genetic factors affect brain activity patterns and cognitive performance. In January 2017, he joined the University of Surrey, as a Lecturer in neuroscience. His research interests include the use of advanced statistical and brain imaging techniques to investigate factors affecting mental health and cognition across the lifespan; with the aim of informing interventions.



MUNINDAR P. SINGH (Fellow, IEEE) is currently an Alumni Distinguished Graduate Professor with the Department of Computer Science, North Carolina State University. He has graduated 29 Ph.D. and 39 M.S. research students. His research interests include AI and sociotechnical systems. He is a fellow of the Association for the Advancement of Artificial Intelligence (AAAI) and the American Association for the Advancement of Science (AAAS). He is a member of Academia Europaea. He has won the ACM/SIGAI Autonomous Agents Research Award, the IEEE TCSVC Research Innovation Award, and the IFAAMAS Influential Paper Award. He has been the Editor-in-Chief of the *ACM Transactions on Internet Technology* and the IEEE INTERNET COMPUTING.

...