Learning from Tweets: Designing an Interactive Visualization Interface Leveraging Interactions of Long-Hauler Communities on Twitter

ANONYMOUS AUTHOR(S)*

COVID-19 has affected many minority communities throughout the pandemic. In this paper, we focused on one such community: COVID-19 long-hauler community. Long-hauler communities consist of COVID patients who experience new, recurring, and ongoing conditions for a long time. The concerns of this community were initially ignored by healthcare providers primarily because of limited information. By analyzing their Twitter interactions, we found several insights such as a wide range of unusual symptoms experienced by long-hauler patients and emotional support that they shared while almost no institutional support was available. Based on these insights, we built LongCov, an interactive visualization tool that allows both long-haulers and other stake holders to explore, analyze, and summarize information related to Long-COVID conditions seamlessly using a single platform. A user-study showed that LongCov not only helps new members to get familiar with the community norms but also inspires existing members to explore data more analytically.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: long-hauler, gaslighting, interactive visualization, social media communities

ACM Reference Format:
Anonymous Author(s). 2018. Learning from Tweets: Designing an Interactive Visualization Interface Leveraging Interactions of Long-Hauler Communities on Twitter. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 28 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The COVID-19 pandemic, a global health crisis, has caused deaths to more than 4 million people worldwide so far. In the USA only, more than 34 million people got infected by the virus and the numbers are still counting. Since the pandemic has continued to spread for more than a year now, it has no longer remained a health crisis; rather it has become an economic and social crisis. Many communities (such as minority communities) suffered a disproportionate amount of difficulties throughout the pandemic. In this paper, we would like to focus on one such community: the COVID-19 long-haulers' community.

Who belongs to the COVID-19 long-haulers community? Typically, mild or moderate COVID-19 symptoms last about three to six weeks for most people. But, a recent study found that 10-30% individuals who had COVID-19 experience new, recurring, and ongoing conditions even when they have recovered from the acute phase of the illness [27]. For those patients, there is no longer a live Coronavirus running amok in their body but they still suffer from prolonged symptoms for months. People living in such conditions are known as "COVID-19 long haulers", "long-COVID", or "post-acute sequelae of COVID-19" as referred by the National Institutes of Health (NIH). Long-hauler patients have faced a wide range of symptoms which made it extremely difficult for them to go back to their usual routine. However, physical symptoms were not the only challenges that they faced during this pandemic. They experienced a great deal of mental pain and anxiety since healthcare professionals initially did not believe this condition was a real one. In many cases, these patients were referred for psychiatric evaluations, although their conditions were not remotely related to their mental health.

The concept of downplaying physiological condition as a psychological problem such as stress, anxiety, or psychosomatic symptom disorder (a mental condition when someone has excessive and unrealistic worries about their health) is not new. This concept is called "medical gaslighting". Women experience medical gaslighting the most because of the knowledge and trust gap of healthcare providers [47]. In the recent past, the concept received a lot of media attention because of the book titled "Doing Harm" [38] published by Maya Dusenbery where she shared her own traumatic experience of medical gaslighting, discussed more largely gender bias in medicine, and how it hurts women. Long-haulers have made this condition worse for everyone irrespective of their gender or age.

This ignorance motivated Long-hauler patients to seek help from each other. At the time of the pandemic, when social gatherings were almost impossible to arrange, these individuals started communicating through social media such as Twitter. They shared their experiences with others who were also experiencing similar symptoms or some unexplained symptoms that their healthcare providers could no address. In the last year, a substantial number of research projects have addressed various socio-economical challenges that people are still facing due to the pandemic. Unfortunately, to our knowledge, no prior work has focused their attention on long-hauler patients, their challenges, and initiatives. Because of growing interests and recent actions taken for addressing the conditions of Long-hauler patients, it is critical to understand how these individuals utilized social media to cope up with difficulties and challenges as a community when their complaints were mostly ignored by everyone else. In this paper, we primarily focused on communities on Twitter that used the following two hashtags: 1) #longhauler and 2) #LongCovid. We chose these two hashtags because these are the only two hashtags that were exclusively dedicated to discussing issues related to long-hauler patients. In the context of this long-hauler social media community, we primarily asked the following research question:

RQ1: What did the long-hauler communities discuss on social media groups and communities? What can we learn from their discourse that may have long-term implications on long-hauler treatment facilities, rehabilitation programs, and most importantly public health care policies?

To address this research agenda, we extracted all tweets that used #longhauler and #LongCovid hashtags from November 3rd, 2020 till August 15th, 2021. Our dataset contains 124,543 tweets in total. To address our research question, we performed LDA, an unsupervised topic modeling algorithm, on the entire dataset to identify the naturally occurring topics of discussion in this group. We identified three primary topics: 1) reporting long-covid symptoms and urging for taking action, 2) sharing information from external sources, and 3) showing gratitude. We developed a group of machine learning classifiers that used features based on word embeddings, psycholinguistic attributes, open-vocabulary-based n-grams, and sentiments to identify these three types of tweets from the corpus. After demonstrating the best performing classifier to provide robust and stable performance with an AUC of 0.80, we machine label our entire dataset. To summarize, our findings show how social media groups might organically develop a support system when there is a scarcity of regular resources and infrastructure.

Machine learning classifiers helped us identify three types of information that can be a great resource for long-hauler communities and anyone who is interested in this issue. However, during our analysis we realized that consuming this information directly from Twitter may not be a practical approach as in each category we have thousands of data points, and accessing them without any organization and suitable data structure would be hard to manage. Thus, we asked the following question:

RQ2: How can we design an interactive visualization interface that can assist long-haulers patients and others access reported symptoms and other related information (shared on social media) effectively and seamlessly?

To this end, we designed and built a functional prototype of LongCov, an interactive web-based visualization interface that can assist individuals to access systematically organized long-covid symptoms and other relevant news or articles (shared on social media) seamlessly and effectively from a single platform. To evaluate the efficacy of the LongCov tool, we conducted a remote user study with 28 participants. The primary objective of our evaluation was the following: 1) whether the information presented through LongCov is useful, 2) whether that information can be accessed and interpreted seamlessly and with minimum effort, and 3) what other information can be presented using LongCov that will be useful for long-hauler patients and others interested in this topics. Our results show that individuals can effectively access information regarding long-hauler communities on social media using LongCov and all participants found that information useful and thought-provoking. We also discovered several areas to improve the design of LongCov for making the tool more easily accessible and for adding more relevant information that can be useful for long-hauler communities. We believe these discoveries lead to future opportunities for developing tools for long-hauler patients. As more patients are getting diagnosed with long-hauler symptoms, our contribution should be considered as an early effort towards building more effective solutions for serving long-hauler communities.

Ethics and Disclosure. Because we used social media data, we had got this study approved by the relevant institutional review board. In addition, we took great care in the way data and analyses are presented in the paper, for instance, by avoiding any personally identifiable information. We intentionally avoided including any quote in the paper because of the growing concern of using publicly available social media data without users' consent. We recognize and acknowledge the limitations of our methodological approach and our position as researchers and outsiders to this particular online community. We describe our limitations and ethical considerations further in the discussion section.

2 RELATED WORK

2.1 Socio-economical Challenges during COVID

On March 11, 2020, World Health Organization (WHO) declared COVID-19 as a global pandemic. At that time, only 118,000 confirmed cases were reported which has increased to 187,519,798 so far and we are still counting. Because of this rapid increase in the number of confirmed cases, COVID-19 did not remain only as a regular infectious disease; rather it has impacted almost every aspect of our society. For instance, when counseling services to mental health patients were provided through an online program, counselors found it challenging to build rapport in the online environment. On the other hand, mental health patients faced hardship related to finance, housing, and distance learning due to the pandemic which often resulted in an increased level of anxiety, stress, addiction, depression, or psychosis [43, 62].

Getting accustomed to online programs and the working environment turned out to be challenging not only for mental health patients. Older adults also faced numerous challenges because of this change of norms. According to a 2017 Pew Research study, three-quarters of those older than 65 said they needed someone else to set up their electronic devices [3]. A third also said they were only a little or not at all confident in their ability to use electronics and to navigate the web [15]. This problem became much worse during the pandemic when older adults had to isolate themselves completely as they were facing a high risk of getting infected by the COVID-19 virus [44]. Similar to older adults, getting used to the online environment introduced many challenges for children as well. An initial survey has found that due to lockdown, children could rarely interact face-to-face with their teachers. The problem was more grave for below-average income families as children from those families felt less strongly about their own capacities to cope with online learning activities than other children [64].

Below-average income families also experienced a great deal of unemployment and lay off during this pandemic which made it even harder for their families to survive. Unemployment was found to be at the core of many socio-economical challenges such as lower standard of living [7], domestic violence [37], and mental health issues [35]. All these socio-economical challenges discussed so far had hit even harder to the members of the minority community such as Black communities, Latinos, immigrants and so on [68]. In this paper, we focused on one such minority community — the COVID-19 long-hauler community. The long-hauler community was created as a direct impact of the pandemic. Until recently, even medical professionals knew very little about the conditions and challenges faced by the long-hauler community. Thus, the long-hauler community tried to comfort and assure each other by forming social media groups. In this paper, we focused our attention on such communities that were formed on Twitter to emotionally support COVID-19 long-hauler patients. We aimed to investigate from their discussions how they extended support to each other when only a few people even believed in their concerns.

2.2 Minority Groups on Social Media

In the last few years, many research studies have found a close connection between minority groups and social media. For instance, Muller et al. have shown that since the 2016 presidential primaries and President Donald Trump's political rise, one standard deviation increase of Twitter usage increases anti-Muslim hate crimes by 32% [45]. In fact, Trump's tweets about Islam-related topics increased xenophobic tweets by his followers, cable news attention paid to Muslims, and hate crimes on the following days. During the global pandemic, another form of hate message that spread across all major social media platforms was Anti-Asian racial messages. About 17% of Asian Americans said in January 2021 they experienced severe online harassment compared with 11% during the same period last year [28]. Online hate and harassment aren't unique to Asians. For many years, social media users who identify as Black, Jewish, transgender, or as part of other marginalized groups have also complained that Facebook and Twitter aren't doing enough to stamp out hate speech, despite having rules against that type of behavior [67].

Despite the threat of spreading hate messages, misinformation, disinformation, and rumors through social media, minority communities often found social media as the only place where they could express their opinion freely and fearlessly. A 2011 national survey of LGBTQ youth reported that this population spent more time online, and were more likely to have close online friends, compared to non-LGBTQ youth [25]. Social media platforms enable LGBTQ people to seek and find health information [30, 39], yet this practice can sometimes be invalidating when one's specific identity or health concern is not represented online [39]. Tumblr has often been recognized as particularly LGBTQ friendly [12, 13, 21, 48]. Some of Reddit's features, such as anonymous and pseudonymous identities, enable LGBTQ communities to form and thrivedarwin2017doing, farber2017transing. Overall, studies found that social media can be an important place for online LGBTQ presentation due to the ability to maintain boundaries between different identities and networks, thus enabling a relatively safe space for identity exploration and transition [10, 29].

In addition to identity exploration, minority communities often relied on social media to protest discrimination against the communities. In 2014, following the non-indictments of officers in the murders of Michael Brown and Eric Garner, the youth of color used hashtags such as "#AllLivesMatter" and "#BlackLivesMatter" to shape the national discourse about race in the wake of high-profile tragedies [11]. Black Lives Matter (BLM) group frequently used social media for building connections, mobilizing participants and tangible resources, coalition building, and amplifying alternative narratives [33, 46].

Similar to the BLM group and LGBTQ community, the long-hauler community also utilized Twitter to get united and to find strength and solidarity within the group. However, unlike other minority groups, we know very little so far about the long-hauler community. Our work aims to fill this existing gap by examining this community through their discussion on Twitter.

2.3 Social Discrimination against Minority Communities

Virtually all countries in the world have national or ethnic, linguistic, and religious minorities within their populations. Many violations of civil, political, economical, social, and cultural rights have a basis in discrimination, racism, and exclusion on the grounds of the ethnic, religious, national, or racial characteristics of the victim group. Prior work has shown that individual and institutional measures, responsible for racial discrimination, were associated with the poor health status of Asian Americans [24]. In addition to the physiological health condition, in the USA, interpersonal discrimination has also been associated with increased rates of hypertension, depression, and stress; poorer self-rated health; and more reported days spent unwell in bed [5, 36].

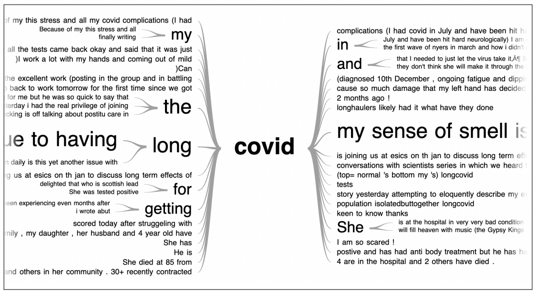
Other than the health sector, discrimination against minority communities have also been observed in the job sector. Imana et al. [32] found that job ads delivered through Facebook were skewed for gender minorities and it could not be justified by differences in qualifications. Discrimination was also observed in the job sector when promotions and increases in salary were considered [16]. Personal prejudice is a critical issue that makes minority communities experience discrimination in many areas. For instance, careful observation revealed that rental housing discrimination still exists in several important types of housing agent behavior, and in most of the cases, the discrimination was caused by agents' own prejudice [14] and/or by the centrality of powerful institutional (i.e., banks, realtors, and insurance companies) [51].

In our paper, we focused on COVID-19 long-hauler community, a minority community that primarily experienced discrimination because no one knew how to explain such concerns. We believe our findings will add value to the existing literature on minority discrimination, especially to understand the characteristics of ad-hoc minority communities who may appear during any type of natural or man-made crisis.

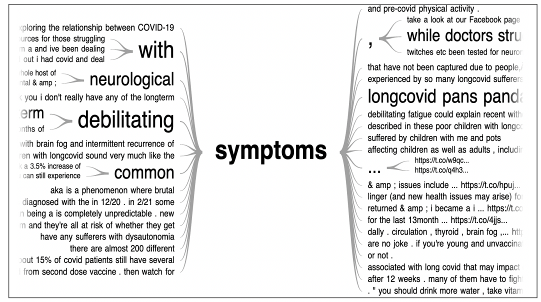
2.4 Interactive Visualization and its applications

Interactive interface and visualizations have been used for a wide range of applications. Visualizations presented through an interactive interface often enable us to present complex or abstract data in a more human-interpretable way. For example, applications where users' privacy is of key importance, Wang et al. [65] designed and implemented an interactive interface, called VailMe, that helped users to understand, configure, and maintain the sharing preferences of their personality data in their workplace. Motahhare et al. [22] used this same medium of interactive visual representation to convey abstract information to laymen users. They built FeedVis, an interactive application to reveal the difference between the algorithmically curated and an unadulterated News Feed on social media to users. In recent work, Dey et el. [20] adopted this theme for building VidLyz, an interactive tool, that allows novice users to explore the implications of audience engagement through persuasion factors in the context of making appealing ads for crowdfunding campaigns.

Interactive visualization can also be a powerful tool when observing the trend of a large dataset. In law enforcement and intelligence, an investigative analyst often starts by gathering data on individual incidents, locations, and persons. However, to extract meaningful and valuable insight from those data chunks, analysts often need to read thousands of articles, documents, evidence files, and so on. Stasko et al. [58] proposed Jigsaw, a tool that represents documents and their entities visually in order to help analysts examine them more efficiently and develop theories about potential



(a)



(b)

Fig. 1. Example word-trees built around "covid" and "symptoms" keywords on Tweets in our dataset. The font sizes are proportionate with relative occurrence.

actions more quickly. In this paper, we took inspiration from this line of work and designed LongCov for visualizing a large dataset on social media conversations of long-hauler communities. Our goal is to present this data in a way so that users can effectively infer useful information and critical trends by using our interactive elements and visual representations that will contribute to the broader field of sense-making [52].

3 METHOD

3.1 Collecting Social Media Data on COVID Long Haulers' Community

The first step in this phase was to collect data from social media communities dedicated to long-hauler patients. We started searching with the keyword long-hauler and identified three hashtags that were exclusively related to long-hauler condition: #longhauler, #LongCovid, and #LongCovidKids. We chose not to consider #Covid or #COVID19 hashtags for our dataset since those hashtags considered a wide range of topics related to the pandemic which were not only limited to long-hauler conditions. Specifically, we accessed data through the Twitter streaming API, continuously from November 2020 to August 2021 and collected a total of 98,678 tweets. We further collected all tweets that replied to them. In total, our dataset contains 124,543 tweets and their responses. There were 37,829 unique users in our dataset who either posted using the above-mentioned hashtags or responded to a tweet that used one of those hashtags. To give the reader a broad sense of our dataset, we present example word-trees in Figure 1 which show examples of post snippets of how people express about "covid" and "symptoms" within the long-hauler communities on Twitter.

3.2 Adopting the Latent Dirichlet Allocation (LDA) technique to identify Main Discussion topics of the Group

We started addressing our first research question where our main goal was to identify the main discussion topics of long-hauler communities on social media. To this end, we began by applying standard text-processing steps such as removing special characters, hyperlinks, punctuation, digits, stop words, and lowercasing all characters. We adopted the Latent Dirichlet Allocation (LDA) technique [6] to extract the range of discussion topics contributed by the long-haulers' communities. LDA is a widely used unsupervised statistical model to discover hidden topics by analyzing the semantic structure of the documents. Each topic consists of a set of keywords that define it, and text tokens are distributed over latent topics throughout each document. We treated each tweet as a document and applied LDA on all of them.

Table 1. Top 15 topics from LDA model with example Tweets. The first column shows the percentage of Tweets in each topic along with the thematic name of the topic. Column 3 lists representative terms from each topic.

Topic description	Category	Sample keywords	One example sentence per topic
Showing gratitude/thankfulness (10%)	thankfulness	nice, better, ty, gratitude, like	So greatful to have a supportive team behind me while I'm battling long covid! Huge thanks the GOSH Digital Learning Team!!
Research findings related to ME/CFS and PEM and long-hauler symptoms (8%)	news	mecfs, PEM, finding, longhaulers, research	Why are recoveries from Covid so different? What makes someone develop long Covid? Researchers are honing in on that question. This week's "Your Health" column: LongCovid https://t.co/t1Azg8Ay7X
Cases of death among long-covid and hospitalized patients (5%)	news	infection, deaths, reinfection, morbidity, breathing	Disgusting disregard for deaths aside, he also conveniently never addresses LongCovid, its human and health care costs, which can arise in even mild cases.
Continuation of symptoms for months (10%)	action/symptom	long, weeks, months, last, hundreds	While most recover from COVID19 within a few weeks, we're now hearing about individuals who experience lingering symptoms months after diagnosis. Join us tomorrow at 2pm ET for an expert discussion on LongCOVID ; how it's impacting your workforce. https://t.co/ThyAfeHGJi
News reporting uncertainty of survival of long-covid patients (4%)	news	uncertain, report, unclear, worrisome, damage	People with no underlying conditions end up with longcovid. They face chronic ill-health, disability and uncertain prospects.
News reporting kids experiencing long-covid symptoms (4%)	news	children, symptoms, young, news, longcovid	Some children have been taken into intensive care months after inital infection, with new symptoms. #LongCovid
Asking for research initiatives on herd immunity (6%)	news	initiative, herd, research, shielding	They are pushing herd immunity before a vaccine is ready, forcing people to work, overloading hospitals, increasing death rates as well as LongCovid.
Symptoms experienced for more than a year (5%)	action/symptom	years, yearsold, long, still, months	I've heard people saying they've had chronic fatigue syndrome for years and now that LongCovid has come along people are taking post-viral syndromes seriously, as if it's a bad thing. Bad they've been ignored for so long but great if that changes surely?
Symptoms related to loss of taste/smell and joint and chest pain (16%)	action/symptom	smells, smoke, taste, pain, chest	If you have had C19 at any point, what symptoms or effects do you still have? For me, I still can't smell bad smells (all other taste and smell has returned) and I find high levels of noise almost unbearable (eg. the background noise of a supermarket)
Cases of fatigue and brain fog (9%)	action/symptom	fatigue, brain, fog, brain-mental, scalp	Today I realized I'm terrified of getting brain fog or cognitive slowdown or losing my words in front of anyone who is not my wife or sister. And! Zoom makes that fear much worse because IRL my close friends know me well enough they could easily fill in my words for me!
News on reinfection (6%)	news	news, reinfection, virus, symptoms	We are certainly hearing of many reinfections - uncertain at this stage how it impacts LongCovid symptoms and whether it might lengthen or shorten the illness https://t.co/LCHG7BO5mF
Efficacy of Vaccines and plasma treatments (2%)	others	plasma, pulse, vaccine, convalescent	Cummings: testing of his eyesight opened the door for others to break the rules. My partner: having survived Covid, donated his plasma to save at least 21 lives, has weakened immunity and now partially lost his sight. #oneruleforthem #LongCovid #StayAtHome"
Asking for help for racial and ethnic minority groups (4%)	action/symptom	community, support, help, groups	People with longcovid, we've interviewed 100 of you, but not enough from *non-White ethnic groups*. If your symptoms went on for more than 3 weeks, and especially if for more than 12 weeks, please consider doing an interview! Not compulsory but if interested, reply to this tweet and I'll DM you.
Lack of facilities for treating long-covid patients (5%)	others	struggling, need, inability, facilities	TheView: all the selfish, ignorant and disengaged monsters unmasked, un-socialdistanced who don't give a damn about the dying alone the suffering longhaulers the aged in facilities see families behind window https://t.co/3M3DeJbXaS
Asking for prayers and good thoughts (6%)	thankfulness	pray, please, support, good, prayers	So sorry to hear this Maneesh. Praying for complete recovery for you and other friends with LongCovid

The performance of the LDA model depends on the choice of hyperparameters α and β and the number of topics (k). Here, α controls the sparsity of document-topic distribution and β determines the sparsity of topic-word distribution. A low value of α is preferred (less than 1), because it produces a sparse distribution, leading to very few topic assignments per comment. This intuitively makes sense, because it is almost unlikely to mention a large number of topics in a single tweet. Similarly, lower values of β favor fewer words per topic. To tune the value of the hyperparameters, we followed the similar procedure that was proposed by Pathik et al. [6]. We considered k = 20 topics as a seed value and ran the LDA model for a range of values of α and β. We considered all values in the range of [0.01 0.99] at regular intervals of 0.05. For each unique combination of α and β, we ran the model and recorded the coherence score. Thus, we chose α = 0.01 and β = 0.11 as the best-fitting hyperparameters for our dataset since the coherence score of the model was the highest for this combination.

Once the values of α and β were identified, we followed the same procedure to tune the value of the number of topics (k). With α = 0.01 and β = 0.11, we ran the model for all values in the range of [5 50] at regular intervals of 5. We observed the highest coherence score at k = 15 and the score did not increase significantly after that. We also investigated the topics themselves and increasing the value of k beyond 15 resulted in repeated appearances of the same keywords in multiple topics which were not intended in our context. Finally, we decided to run the model for α = 0.01, β = 0.1, and k = 15 and generate topics for further analysis. Once the topics were identified, two human coders familiar with the concept of social media group for COVID long haulers independently reviewed those 15 topics and the top words in each topic. Following an inductive open coding method, they individually identified the non-overlapping themes from those topics. In the process, they merged two or more topics when they were thematically overlapping with each other. Finally, they resolved disagreements through discussions. We identified three main themes and one other group that contained the remaining topics.

Table 1 lists the 15 topics presenting the main discussions in the long-haulers' Twitter communities. The table shows the percentage of posts (column 1) on each topic along with the top five keywords. Primarily, the topics generated by LDA can be divided into three major categories: 1) the topics that expressed gratitude and thankfulness to other members of the community for showing support and keeping them in their prayers at difficult times, 2) topics discussing a specific symptom or a group of symptoms and urging the community to take action, and 3) topics sharing news articles and blog posts regarding long-hauler conditions, findings from new research projects, and available treatments.

A small percentage of tweets (7%) did not belong to any of these three categories. These tweets mostly highlighted discussions on vaccination and its efficacy and new facilities established especially equipped to treat long-hauler patients. Although we observed a lot of discussions related to COVID vaccination on Twitter, our dataset could not capture them as a major discussion topic. One possible explanation can be the hashtags that we chose for this analysis. Individuals who discussed topics related to COVID vaccines did not use #longhauler or #LongCovid hashtags. #Moderna, #Pfizer, #Vaccine are some notable example hashtags that are used predominantly for COVID vaccine-related discussions. Another topic that we considered in the fourth category is the discussion of not having facilities specialized for treating long-hauler patients. Since facilities especially equipped to treat long-hauler patients are still at a very early stage, we found only a few tweets discussing that topic in our dataset. We believe with more and more facilities getting established, this would be a major topic of discussion in this community in a few months.

Overall, the main discussion topics identified by LDA show that long-hauler communities actively participated on social media not only to share their sufferings and ailments but also utilized social media to spread news articles and blog posts on continuous development made in research studies related to the long-covid condition. At the time of lock-down and social distancing, social media has become the most accessible alternative for communication among minority groups such as COVID long-haulers. Our findings are consistent with media reports that showed that long-hauler patients used various forms of online media such as Instagram and Slack for running their own research studies in a quest to find out the root cause of their condition when doctors dismissed their concerns as psychosomatic conditions [4]. While we observed a large number of discussions on the role of social media in spreading misinformation and disinformation on COVID-19, our analysis revealed how marginalized, close-knit communities can use social media for building awareness, continuing communication, and even for running their own research studies. Based on the findings from the LDA topic modeling algorithm, we applied machine learning algorithms to automatically classify all tweets) in our dataset as to one of these four classes.

3.3 Building A Machine Learning Classifier for Identifying Categories of the Tweets related to Long-Hauler Communities

Our next goal centers around identifying topics discussing COVID-19 long-hauler conditions at scale from our Twitter dataset. We draw on natural language analysis techniques to build machine learning classifiers on the annotated dataset. We describe our approach, features, and models below.

3.3.1 Machine Learning Features. Inspired from several prior work on analyzing text data collected from social media [34, 54, 55], our work uses four kinds of features:

Latent Semantics (Word Embedding). To capture the semantics of language beyond raw keywords, we use word embeddings, which are essentially vector representations of words in latent semantic dimensions. Several studies have revealed the potential of word embeddings in improving natural language analysis and classification problems [41]. In particular, we use pre-trained word embeddings (GloVe) in 50-dimensions that are trained on word-word co-occurrences in a Wikipedia corpus of 6B tokens [50].

Psycholinguistic Attributes (LIWC). Prior literature in the space of social media and psychological wellbeing has established the potential of using psycholinguistic attributes in building predictive models [18, 54] We use the Linguistic Inquiry and Word Count (LIWC) lexicon to extract a variety of psycholinguistic categories (51 in total). These categories consist of words related to affective process, cognitive and perceptual process, informal language, time orientations, linguistic dimensions, biological process, social process, and personal concerns [63].

Open Vocabulary (n-grams). Drawing on prior work where open-vocabulary-based approaches have been extensively used to infer psychological attributes of individuals [56, 57] we also extracted the top 500 n-grams (n = 1, 2, 3) from our dataset as features.

Sentiment. An important dimension in social media language is the tone or sentiment of a tweet. Sentiment has been used in prior work to understand psychological constructs and shifts in the mood of individuals [26, 53]. We use Stanford CoreNLP's deep-learning-based sentiment analysis tool [40] to identify the sentiment of a tweet among positive, negative, and neutral sentiment labels.

Modeling Approach. We used 1000 manually annotated tweets from the previous section to build a machine learning classifier with a total of 619 features. The process followed for qualitative annotation is explained in Appendix A. We considered and evaluated multiple classifiers, including Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Random Forest, AdaBoost, and Convolutional Neural Network (CNN) algorithms. We use stratified k-fold cross-validation (k = 5) to parameter tune our classifiers. Table 2 summarizes the performance metrics of these models. All of these classifiers performed better than the baseline accuracy of 25% on our dataset (based on a chance model). We found that the Boosting classifier outperforms all with a median AUC of 0.80, median precision of 0.77, and median recall of 0.73. Table 3 summarizes the performance metrics of this classifier, where we find that the classifier is reasonably

Table 2. Median metrics in k-fold (k=5) cross-validation.

Model	Accuracy	Precision	Recall	F1	AUC
Naïve Bayes	0.683	0.467	0.660	0.525	0.693
Linear Regression	0.706	0.702	0.706	0.680	0.752
SVM	0.679	0.580	0.679	0.577	0.762
Random Forest	0.716	0.750	0.706	0.637	0.781
XGBoost	0.753	0.777	0.739	0.703	0.804
CNN	0.585	0.483	0.626	0.505	0.601

Table 3. Detailed accuracy metrics in k-fold (k=5) cross-validation in the COVID-19 long-haulers' Tweeter conversation Classifier (AdaBoost).

Metric	Min.	Max.	Mean	Stdev.
Accuracy	0.688	0.753	0.703	0.015
Precision	0.589	0.777	0.639	0.069
Recall	0.688	0.739	0.703	0.069
F1	0.589	0.703	0.627	0.024
AUC	0.698	0.855	0.787	0.020

Table 4. Incremental accuracy metrics of adding features in the COVID-19 long-haulers' Tweeter conversation Classifier (AdaBoost).

Model	Accuracy	Precision	Recall	F1	AUC
N-grams	0.697	0.598	0.688	0.587	0.728
.+ Word Embeddings	0.711	0.609	0.711	0.646	0.740
.+.+Sentiments	0.730	0.759	0.728	0.657	0.713
.+.+.+LIWC	0.753	0.777	0.739	0.703	0.855

stable (STDEV = 0.02) across the five-folds, and Table 4 summarizes the step-wise improvement with the addition of each kind of feature in the AdaBoost model. For the rest of the paper, we would use the AdaBoost as the classifier for classifying all Tweets in our dataset. We found that 16% tweets were classified as the first category, i.e., expressing gratitude and thankfulness to other members of the community, 44% of them were classified as the second category, i.e., discussing symptoms and urging for taking action, and 33% in the third category, i.e., sharing news articles and blog posts. The remaining 7% were classified as the fourth category which is a miscellaneous category for the remaining tweets.

We used K-best univariate statistical scoring model using mutual information to obtain the relative importance among features. We established their statistical significance using ANOVA to obtain the top features of the Twitter conversation of long-hauler communities, which are reported in Table 5. This table only includes the "interpretable features", and excludes word-embedding dimensions. We found that the features obtained from LIWC and n-gram vectors equally contributed to the desired model. Many of the psycholinguistic (LIWC) categories are significant, which also align with the health crisis and its long-term effect on people's work and family. For instance, the social process categories, such as family and friends show high relative importance as many individuals in long-hauler communities showed concerns regarding their family members and closely-connected friends.

We also found personal concerns, such as work and death, biological processes such as body, and cognitive processes such as tentative and causation show high relative importance — these keywords are known to be associated with

Table 5. Top 22 Features in the long-haulers' Twitter conversation classifier. p-values reported after Bonferroni correction following ANOVA (*** p<0.0001, ** p<0.001, * p<0.01).

Feature	Score	Feature	Score
n-gram: fever	0.17 ***	n-gram: chest	0.11 ***
LIWC: family	0.17 ***	LIWC: body	0.10 ***
n-gram: sleep	0.16 ***	n-gram: taste	0.08 ***
LIWC: verb	0.16 ***	LIWC: death	0.08 **
LIWC: feel	0.14 ***	LIWC: tentative	0.08 ***
n-gram: blood	0.14 ***	n-gram: headache	0.08 ***
LIWC: work	0.14 ***	LIWC: health	0.08 ***
LIWC: social	0.13 ***	LIWC: friends	0.08 ***
n-gram: body	0.12 ***	n-gram: pain	0.07 ***
LIWC: causation	0.11 ***	n-gram: smell	0.06**

first-hand accounts of the real world happenings, events, and experiences [8, 9]. Boals and Klein found that individuals describing stressful events use words that belong to cognitive processes which help them actively find the meaning of their condition [8]. N-gram features that have shown high significance mostly highlighted various types of symptoms or body parts where long-hauler patients experienced those symptoms. Some notable examples are fever, blood, taste, headaches, and pain which are all related to some ailments that long-hauler patients frequently experience.

The machine learning classification process provided us with a list of Tweets that belong to four different categories. However, the Tweets in the second (Tweets related to symptoms experienced by long-hauler patients) and third (Tweets regarding news articles and blog posts on long-covid related topics) categories contained critical information which was hard to convey to a broader audience when they are simply stored in a list format. To make them easily accessible and hence useful for long-hauler communities, we designed and built an initial prototype of an interactive interface that can systematically present Tweets in these two categories to the target audience. In the next section, we first explained the design of the tool followed by the feedback that we received on this initial prototype from online long-hauler communities.

4 LONGCOV: A INTERACTIVE INTERFACE FOR PRESENTING SOCIAL MEDIA DATA TO LONG-HAULER COMMUNITIES

In the second research question (RQ2), we aimed to explore some form of an interactive visualization interface that can assist long-hauler members to access important information extracted from social media conversation. To this end, we designed and built LongCov. The primary objective of the LongCov interface is to access social media data discussed among long-hauler communities effectively and with the least amount of effect. We hypothesize that this interface will be a complementary tool for the long-hauler communities. Individuals, finding it difficult to access and follow critical information directly from social media, will find this tool useful and thought-provoking. Moreover, anyone interested in long-covid will find this tool as a convenient medium to get familiar with the discussion topics that are prevailing among long-hauler communities. Here, we first explained the steps that we followed to design the interface. We will follow it up by discussing the feedback that we received on the interface from interview participants.

4.1 Interface Design

We built the LongCov interface using the Dash platform provided by Plotly that uses only Python for creating advanced, interactive interfaces for the web [2]. We used Jupyter notebook for developing our interactive elements [1]. Finally, we deployed the interface in Heroku server. The interface can be accessed using all major web browsers including Chrome, Firefox, Edge, and Safari.

We used the same dataset that we used for topic modeling and ML classification for designing the LongCov interface. We conducted two rounds of online participatory design sessions with ten members of long-hauler communities from Twitter. Each session continued for 30 minutes and five participants joined each session through a Zoom call. Participants were recruited using an open invitation posted on Twitter that asked people to contribute to designing an interactive interface for long-hauler communities. The participation was voluntary and participants did not receive any payment for their contribution.

To explain the design objective of the LongCov interface, we presented the following two hypothetical scenarios. In scenario 1, let us consider Maria as an existing member of long-hauler communities on social media. Maria recently experienced chest pain symptom. She remembers that other members of the community mentioned this symptom earlier on Twitter but she could not find those discussions readily. She wants to check what members of long-hauler communities have discussed earlier on chest pain symptom. She also wants to know how common this symptom is and what other symptoms may appear with this one.

In scenario 2, let us consider Jun. Her brother was diagnosed as COVID-19 positive two months ago but he has not completely recovered yet. He is still experiencing many symptoms almost every day and has not felt fit enough to go back to work. Jun is suspecting that her brother may be suffering from long-COVID symptoms but she does not know much about the long-COVID condition. She has found several social media communities for long-hauler patients. However, she wants to know a little more about these communities and their topics of discussion before she feels comfortable communicating with other members of the community. She tried to read Tweets posted by these communities but there were so many Tweets available that it was practically impossible for her to go through all of them. She wants a summarized view of the discussions going on in these communities so that she can feel comfortable and confident to speak with other members of the community.

Here we explain the functionality of the LongCov interface that we designed to meet the requirements of Maria, Jun, and of other long-hauler patients similar to them. Based on the outcomes of our LDA analysis, we decided to design the LongCov interface for presenting two sets of information: 1) long-COVID symptoms and 2) news articles, shared by the members of long-hauler communities on Twitter. The interface has two parallel views: 1) symptom tracker view and 2) news organizer view.

4.1.1 Symptom-Tracker View. The symptom-tracker view supports exploration through accessing a wide range of symptoms that long-hauler communities discussed on Twitter. We divided this view into three panels: 1) the left panel — control area (A), 2) the middle panel — explore symptoms (B), and 3) the right panel — supporting information (C).

Control Panel (Left) (Section A). The control area (Section A in Fig 2) shows the list of all symptoms discussed so far in these communities. To extract the list of symptoms from our data corpus, we applied parts-of-speech tagging followed by an open vocabulary (n-grams) feature extraction technique. The process of symptom extraction is explained in Appendix B. The full list of symptoms was sorted based on their frequencies and was presented at the top of the control panel in a drop-down list. Imagine that if Maria is interested to know more about ongoing discussions on "chest

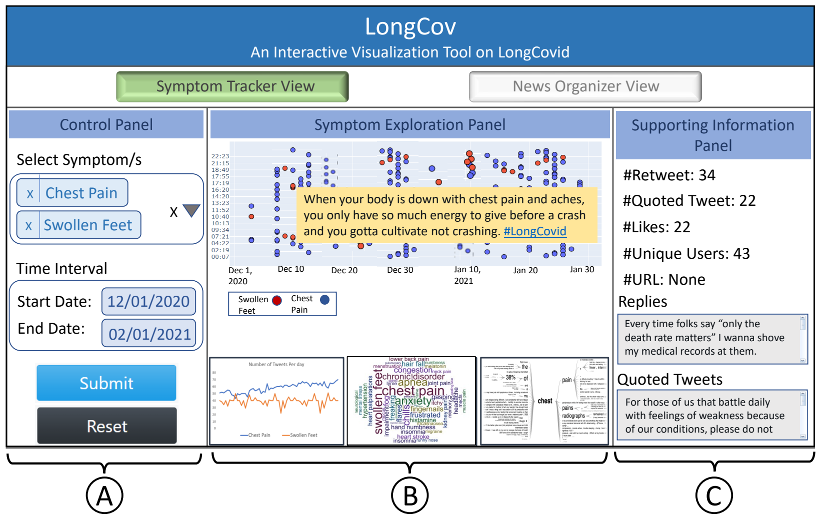


Fig. 2. A screenshot of LongCov, an interactive visualization interface for accessing and examining Tweeter conversation of long-hauler communities. Here, the figure shows the symptom tracker view. The left panel (A) is the control panel that defines the type of data that the user wants to access using LongCov. The middle panel (B) uses different types of graphs and charts to present the desired data to users. Finally, the right panel (C) shows additional information on selected tweets.

pain" symptom, she will have to choose that symptom from the drop-down list and click the search button at the bottom of the control panel. This action will find all Tweets that have mentioned the symptom "chest pain" in some context and will present them using a scatter plot in the middle panel (explained in more detail in the next paragraph). Maria may also choose more than one symptoms (such as "chest pain" and "swollen feet" together) at a time from the drop-down list. In that scenario, all tweets that mentioned either of those two symptoms or both symptoms together will be presented using the scatter plot. Maria can further curate the data by choosing the time interval (start date and end date) available in the control panel. If a time interval is chosen, the data extraction engine will only consider those tweets that were posted in that time interval.

Symptom Exploration Panel (Middle) (Section B). The middle panel (Section B in Fig 2) is designed to show Tweets based on the criteria mentioned in the control panel (A). For instance, if Maria chooses "chest pain" and "swollen feet" in the control panel and sets the time interval from December 1st, 2020 till February 1st, 2021, the data extraction engine will pick all Tweets that satisfy those criteria and will present them on the scatter plot. Each dot on the plot will represent one unique Tweet. Each symptom will be represented with a unique color. For example, in this scenario, all Tweets that contain only "chest pain" symptom will be presented by blue dots and all Tweets that contain only "swollen feet" symptom will be presented by red dots. However, a Tweet containing both of these symptoms may be presented with another unique color, such as a green dot (not shown here). The X-axis of the plot presents the time interval (dates) and the y-axis presents the time of the day (time of the post made on Twitter).

Maria can click on any dot and the corresponding Tweet will be presented on a pop-up text box (shown using the yellow pop-up box). Supporting information related to that Tweet will be shown on the right panel (explained in the next paragraph). At the bottom part of the right panel, there will three more plots. These plots will encourage more analytical exploration of the Tweets presented on the scatter plot.

The first plot will show the number of Tweets on each day for each symptom using a line graph. In this case, there will be two lines (one for chest pain, and one for swollen feet). A third line for the combination of chest pain and swollen feet can be added if required. The colors of the lines will match with the corresponding colors on the scatter plot. This line chart will help Maria understand how the frequency of Tweets on a specific symptom has changed for a certain time interval. This piece of information may be useful for the user to understand whether that symptom is a commonly experienced symptom among the long-hauler community or not.

The second plot will show a word cloud that will present all other symptoms that have co-appeared in our dataset along with "chest pain" and "swollen feet". The size of the words on the word cloud will be equivalent to the number of times they have been identified co-appearing with "chest pain" and "swollen feet". This word cloud will help Maria to know more about other relevant symptoms that long-hauler patients generally experience along with "chest pain" and "swollen feet".

The third plot will show one word-tree per symptom of interest. In this example, since Maria is interested in "chest pain" and "swollen feet", it will generate two word-trees, one for each symptom. The word-trees are hard to read when presented in a small box. Therefore, when Maria will click on the plot, it will open a large pop-up box and that will show one word-tree at a time. Maria can click the left or right arrow to switch between different word-tress. A click outside of the pop-up box will close the word-tree plot. The word-trees will show words most often follow or precede the name of a specific symptom. We hypothesized that this visualization will present in a more human-interpretable way what and how Twitter users talked about a certain symptom in long-hauler communities.

Supporting Information Panel (Right) (Section C). The panel in the right (Section C in Fig 2) is the panel for showing supporting information regarding a Tweet selected by the user. Imagine that Maria wants to read a few tweets on "chest pain" symptom. She will have to click on any tweet from the scatter plot (here she will pick any blue dot from the scatter plot). This action will open a pop-up box to show the content of the original Tweet. Moreover, in the right information panel, this will load all additional relevant information regarding the Tweet. For example, the right panel will show all responses and quoted tweets of that original tweet. Oftentimes, members of the long-hauler community suggest some home remedies for certain symptoms. They also discuss different variations of the same symptom. These conversations will help Maria to know more about the symptom of her interest. The panel will also show the original posting date, the number of "like" and the number of "Retweet" for the original tweet. In addition, the panel will show how many unique users interacted with the original tweet (by responding, liking, or retwritting) which will indirectly tell her how relatable that post was to the members of the long-hauler communities. Finally, if any Tweet contains an URL to some external resources, the information panel will also show that which Maria will be able to access just by a click.

4.1.2 News-Organizer View. The news-organizer view is parallel to the symptom-tracker view. On various stages of Covid-19, members of long-hauler communities shared news articles on a wide range of topics. For example, in early 2020, most news articles shared by long-hauler members were related to research articles that confirmed that the long-hauler symptoms are not some form of psychosomatic condition; rather a certain percentage of COVID patients are experiencing these lingering conditions for a long time that have no known cure yet. However, the trend of sharing news articles has changed during the last few months of 2020 when most of the discussions were focused on vaccinations and whether long-hauler patients should take the vaccine or not. The purpose of this view is to facilitate users to explore news articles and blog posts shared by long-hauler communities on Twitter. Similar to the previous one, the news-organizer view also uses a three-panel structure: 1) the left-panel — control area (A), 2) the middle panel — new articles exploration panel (B), and 3) the right panel — supporting information (C).

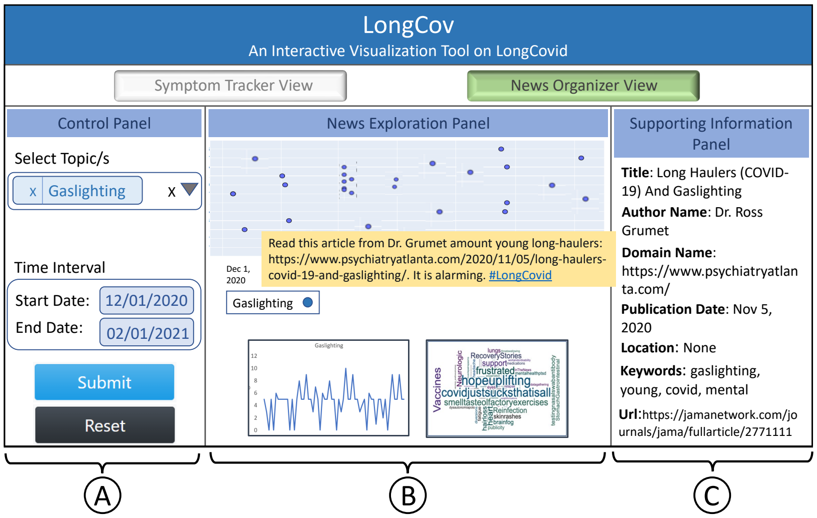


Fig. 3. A screenshot of LongCov's news organizer view. Again, the left panel (A) is the control panel that defines the type of data that the user wants to access using LongCov. The middle panel (B) uses graphs and charts to present the desired data to users. Finally, the right panel (C) shows additional information on selected news articles.

Control Panel (Left)(Section A). The control panel (Section A in Fig 3) in this view is very similar to the control panel of the symptom tracker view. The only difference is that here, instead of the list of symptoms, the drop-down list shows the list of all keywords relevant to the articles and blog posts shared by long-hauler communities on Twitter. To prepare this list of relevant keywords, we extracted all URLs from our dataset. For each URL, we automatically extracted all information such as title, author's name, source domain, publication date, all text-based content (body of the article), and publication location (if available). We considered the title and the body of each article to identify the relevant keywords for that specific article. We applied the same parts-of-speech tagging followed by an open vocabulary (n-grams) feature extraction technique as we did for extracting the symptoms for the previous view. Once all keywords of an article were identified, we only stored the first 10 most probable keywords for the corresponding article in our database as the representative keywords. Later, we took all these keywords to populate the drop-down list in the control panel of the news-organizer view. We arranged these keywords chronologically where the most frequent keyword appeared at the top. Imagine that Jun is interested to know more about the topic of "gaslighting" that long-hauler patients experienced throughout the pandemic. In this scenario, she will have to choose the topic "gaslighting" from the drop-down list and our search engine will pick up all URLs for which we stored "gaslighting" as a representative keyword and will show them on the scatter plot (explain in the next section).

News Exploration Panel (Middle)(Section B). Similar to the symptom tracker view, the top part of the middle panel (Section B in Fig 3) utilizes a scatter plot to present all URLs that satisfy the criteria mentioned in the control panel. For example, in this current scenario, the data extraction engine will present all URLs that are stored in our database with the representative keyword "gaslighting" on the scatter plot. If multiple keywords are chosen, the dots on the scatter plot will be color-coded accordingly. Jun can click on any dot and the pop-up window will appear to ask for her permission to open the URL in a new tab. All meta-data corresponding to that URL will be presented in the right panel (explained in the next section). At the bottom part of the right panel, there will be two more plots. These plots will encourage more analytical exploration of the external URLs presented on the scatter plot.

The first plot, similar to the symptom tracker view, will show the number of URLs on each day for each keyword using a color-coded line graph. This line chart will help Maria understand how the frequency of the shared URLs on a specific topic has changed for a certain time interval.

The second plot will show a word cloud that will present all keywords that have co-appeared with "gaslighting" in our database. The size of the words on the word cloud will be equivalent to the number of URLs containing that keyword. This word cloud will help Maria to know more about other relevant topics that are frequently mentioned on news articles in relation to the keyword "gaslighting" which can potentially inspire her to investigate more on this topic.

Supporting Information Panel (Right)(Section C). The information panel (Section C in Fig 3) at the right side is very similar to the information panel of the symptom tracker view. If Jun clicks on a dot on the scatter plot which is representing a specific news article or blog post, the information panel will show her all meta-data of the corresponding article/blog. For example, the right panel will show the title of the article, the author's name, the domain name of the article, the publication date, and location (if available). The panel will also show the representative keywords (stored in our database) for the article. These keywords will let Jun know a little more details of the content of the article even before she decides to read it. In addition, the panel will also show links of three other news articles similar to the original one from our database as suggestions. To calculate the similarity of two news articles, we have considered multiple criteria such as the source of the articles, title of the articles, and representative keywords. We hypothesized that these suggested URLs will be helpful if Jun wants to examine more on this topic.

5 EVALUATION

Since the functionality and objective of the LongCov interface are new to the best of our knowledge and long-hauler communities are still going through a large number of physical and mental symptoms, we designed the evaluation plan to gauge the usefulness of the tool and to gain insights into how we can improve the LongCov tool in the future. We put our best effort to design the evaluation process in a way so that participants could feel comfortable and relaxed during the session. Hence, we decided to conduct semi-structured interviews with participants rather than conducting task-based evaluation sessions.

5.1 Participants

We recruited 28 participants for our user study. To recruit these participants, we posted a Tweet from the personal accounts of the authors of the paper with hashtags "#LongCovid", "#Longhauler", and "#LongCovidKids". The post briefly explained the purpose of the LongCov tool and requested the members of long-hauler communities to contact us through Twitter if they are interested to participate in a 45-minute online interview session. Anyone, above 18, who had some experiences with long-hauler patients was welcome to participate in the study. The participation was voluntary and we did not offer any monetary compensation for their participation. The participants were allowed to leave the interview at any time. Among these 28 participants, 15 of them were long-hauler patients at the time of the interview, 8 of them were close relatives of at least one long-hauler patient (either parents, spouse, children, or siblings), and 5 of them had at least one acquaintance who was a long-hauler patient. Participants' average age was 42.19 (SD = 8.32), and 56% were females. On average, each interview session lasted for approximately 43 minutes.

Table 6. Responses of the participants for the first three questions that helped us understand the overall impression of the participants of the LongCov Interface. We used five-point likert scale for all these questions where 5 means "definitely" or "extremely likely".

Questions	Mean	SD
How easy and effortless was to use the LongCov interface?	3.74	0.67
Would you like to use this interface in the future?	4.39	0.31
Would you recommend this interface to your friends, family, and acquaintances?	4.46	0.29

5.2 Procedure

We conducted semi-structured interviews with participants in this user study. The interview session was divided into three stages. In the first stage, the interviewer explained the functionality and possible operations of the tool to each participant. The interviewer took approximately 7 minutes to complete this stage. The next stage is the exploration stage where we asked participants to explore the tool online by themselves for 20 minutes. During this stage, the interviewer was readily available to answer any questions; however, the interviewer kept her camera off so that participants would feel comfortable and at ease during the exploration stage. The last stage is the feedback stage where the interviewer asked for the feedback of participants on the tool which lasted for around 15 minutes. In the final feedback stage, we asked participants primarily the following six questions:

- How easy and effortless was it to use the LongCov interface?
- Would you like to use this interface in the future?
- Would you recommend this interface to your friends, family, and acquaintances?
- What elements or functionality of the interface did you like the most?
- What were the elements or functionality of the interface that you would not like to see in the revised version of the interface?
- What additional elements or functionality would not like to see in the revised version of the interface?

We included the first three questions to understand the overall impression (positive vs negative) of the participants of the LongCov interface whereas we anticipated that the last three questions will help us gather some feedback that can be useful to revise the interface before releasing it for public use. For the first three questions, we asked participants to use a five-point Likert scale to answer where 1 means either "definitely not" and 5 means "definitely". The full scale can be found in Appendix C. Participants shared their responses verbally with the interviewer. For the last three qualitative questions, we recorded the interview sessions and later transcribed them for qualitative analysis. Two authors of the paper independently performed iterative open-ended coding [60, 61] for all interview sessions followed by thematic analysis [19, 42, 49]. In the end, they discussed together to resolve any disagreement. Cohen's kappa test showed a substantial agreement between the two coders (K = 0.89). We have summarized the main themes identified by the coders in the next section.

5.3 Results

5.3.1 Responses from the first three overall impression questions. Table reftab:my-table shows the ratings of the participants that represented their overall impression of the LongCov interface. Most of the participants mentioned that they would most likely like to use the LongCov interface in the future (4.39 out of 5). We observed the same trend when we asked them about recommending the interface to their friends, family, and acquaintances (4.46 out of 5). High ratings in those two questions show that participants liked the purpose of the interface and found that the interface will be useful for the long-hauler community. However, when we asked them about the easiness of use of the interface, they gave a rating of 3.74 out of 5 on average. Although this is better than neutral (3 out of 5), the lower rating in this question (in comparison to the other two ratings) shows that there were some elements in this interface that were not easy for them to interpret in the first place. To understand the overall usability of the interface better, we analyzed their qualitative answers.

5.3.2 What participants liked about the LongCov Interface. During their feedback stage, most of the participants (N = 25) started their conversation by explaining the elements that they liked about the interface the most. In this section, we summarized the strengths of the LongCov interface as explained by the participants of our user study.

A convenient tool for referring a specific information. A majority of the participants (N = 17) found that LongCov can be a great resource for referring a piece of specific information to someone which is often extremely difficult to do if that information is only available on Twitter. Participants in this community often gather virtually to discuss issues related to long-hauler communities. Some notable issues are gathering support for long-hauler patients from public health sectors, raising awareness through social media movement, and taking action for building a support system for long-hauler patients at every possible location outside of their homes. During such discussion, members of the community go through many supporting material, research articles, and reports which they have encountered on social media (Twitter) Finding a certain document from Twitter can be tricky. Participants felt that a tool like LongCov can be handy in those contexts because of its functional and systematic way of organizing and retrieving information.

A great way to finding assurance from the community. Participants during our interviews discussed several instances when they experienced some painful symptoms and they had no place to go. This is because finding professional help with minimum delay for long-hauler conditions during this pandemic has been extremely difficult. In those situations, many participants (N= 19) relied on their social media community. Although such communities had been a big support for each other, in some scenarios waiting for responses on social media can be frustrating and depressing for long-hauler patients. This is particularly true during holiday seasons when response rates in these communities naturally go down. The LongCov interface can be of great assistance to long-hauler patients. Conversations presented through this interface can be reassuring in many scenarios where real-time responses are not readily delivered.

A helpful resource for public sector policy makers. Participants also mentioned the lack of sources that have logged the concerns of long-hauler patients systematically over the entire duration of the pandemic. Many research groups have started investigating this topic recently, but their analysis is not complete yet. Participants of our user study (N = 11) believed that the public health sector has not yet taken this issue as seriously as they should because they do not have access to a sufficient amount of information that is critical for taking action and changing policy in public sectors. Participants mentioned that if LongCov is publicly available and regularly updated with new information from Twitter and other similar resources, this will be a valuable resource for public health sectors and even for people who are running social movements to push public health sectors for taking effective initiatives for long-hauler communities.

5.3.3 What participants did not like about the LongCov Interface. Participants highlighted several areas that they did not like in the current version of the interface and wanted us to redesign the revised version.

Complexity of the visualization. Most of the participants found the scatter plot intuitive and easy to interpret (N = 24). However, participants (N = 16) initially found that the word-clouds and the word-trees are not easy to interpret if they are not explained first. Once explained, participants found those visualizations interesting and thought-provoking. They also expressed that those plots might motivate users to further investigate this topic and that can ultimately help them gather more knowledge on this issue. Despite that, participants felt that new users may find these visualizations hard to interpret, and thus, they may assume that the interface is hard to use. Participants suggested that we should consider adding some text cues at the front end instead of showing the visualizations directly. Anyone interested in those cues would click on them and that action will open the corresponding visualization. Thus, users will understand the purpose of that visualization by themselves without any external help.

Disconnection between symptom tracker and news organizer view. Some participants (N = 6) found the two views of the LongCov interface seemed disconnected and they wanted those two views to be joined together in the future version. Participants felt that the news articles shared on Twitter had many references on long-covid symptoms. Thus, they did not find the design of arranging them in two separate views appealing. They also mentioned that someone new to this interface may completely miss the view which is not loaded as the default view.

5.3.4 *What additional elements participants want to add to improve the LongCov Interface.*

Participants suggested many additional elements that may improve the functionality of the LongCov interface significantly. Here we have discussed them starting from the most popular one.

A direct extension of Twitter. Most of the participants (N = 21) suggested that the LongCov interface should directly be accessible from Twitter. Participants felt that Tweets using #longCovid or #longhauler hashtags should have a direct link to the LongCov interface so that the long-hauler community can easily get access to such an analytical interface. They felt that this direct access will be critical for building awareness about the concerns that long-hauler communities are struggling with for a long time among the people who do not know yet a lot about these communities. They suggested that not being able to reach the broader community will be a big failure for the LongCov interface as this interface has the potentials to create an impact for long-hauler communities.

A link between a user and their symptoms. Participants in the user study found that the symptom tracker view is a great tool for investigating a specific symptom. However, some participants (N = 8) commented that the more they spent time reading about different symptoms and discussion around them on Twitter, the more they became curious about those people who experienced such symptoms. Did they have any pre-conditions? How was their condition when they were infected by the COVID-19 virus? Did they have to go through hospitalization? How old are they? Did they eventually recover from their condition? In summary, many participants mentioned that linking social media conversation with real people and their unique journey through the long-covid condition, would be a great extension for the LongCov tool.

Adding pictures for making these conversations more appealing. Some participants (N = 9) noted that the LongCov interface only focused on the text conversation on Twitter. However, the interface completely ignored all forms of visual media such as images, infographics, memes, graphs, and videos in some scenarios. This choice of ignoring all visual mediums made the interface less dynamic. In the future version, participants suggested that a small window for visual content may make the overall interface more appealing to a broader audience.

More advanced classification of the list of symptoms. When we asked for suggestions from participants on how to improve the interface in the future revised version, some participants (N = 7) suggested including a hierarchical structure for the list of long-hauler symptoms rather than adding them all in a flat drop-down list. These participants anticipated that when someone would take an extra effort to visit the LongCov interface, they would like to see some advanced features that are not often be seen on Twitter or any other similar social media platform. One suggestion in this direction was to categorize symptoms in some broader categories such as pain, cognitive, and breathing so that users can access them more conveniently. Another suggestion in this direction was to add interactive actions for word-cloud and word-tree visualizations. Some participants suggested that making them interactive will be a great way to inspire more users to access those graphs.

Incorporating data from various social media platforms. Many participants agreed that collecting data from Twitter for designing the LongCov interface is a good choice since data collected from government resources might not represent the full picture of the long-covid condition. However, some participants (N = 9) argued that although Twitter is one of the major social media platforms at this time, only capturing conversation from this one social media platform can make our system more biased toward a certain group of people who are regular to Twitter and might completely miss the community who prefer to use other mainstream social media platforms such as Facebook and Instagram. They suggested that in order to represent the majority of members of Long-hauler communities, we should also incorporate data from other mainstream social media platforms such as Facebook, Instagram, and Reddit.

6 DISCUSSION

6.1 Theoretical Implications

This study assesses long haulers' community and their harrowing journey through difficulties via an inexpensive and unobtrusive data source, social media data. Long hauler patients received denials and ignorance throughout the pandemic. Only recently, initiatives are being taken to establish facilities that would be equipped to assist conditions related to long-hauler patients. However, gathering information from existing long hauler patients and taking action as per the requirements would most likely take a substantial amount of time and effort. Since social media data consists of long hauler patients' self-initiated and candid opinions and experiences, this data provides us with a rich and accessible lens to examine this community, their requirements, and complaints. This approach has clear advantages beyond traditional mechanisms — such as surveying a large group of COVID-19 patients and closely monitoring their symptoms and lingering conditions. The purpose of this study is to not replace the traditional approach; rather establish a complementary strategy so that new infrastructures getting built to assist long haulers can make the best use of social media data. The large-scale availability of social media data provides opportunities to understand the breadth of mental, physiological, socio-economical issues that the long-haulers' communities are facing throughout this global pandemic.

Summarily, this study opens up interest in conceptualizing long hauler patients specifically (their unique conditions) without confusing them with regular COVID patients or people who faced extreme socio-economical hurdles due to the pandemic. In particular, the LongCov interface allowed us to observe the discussion that revealed not only a long list of lingering symptoms (symptom tracker view) but also mental and social issues which are often missed (news organizer view). The word-tree and word-cloud visualizations as part of the symptom tracker view illustrate the importance of studying these conditions as a group of co-related symptoms rather than any separate ailment.

During the pandemic, social media was frequently accused of spreading fake news, misinformation, and disinformation on mask usage, the safety of getting vaccinated, COVID-19 prevention measures and treatments. In recent work, Su et al. [69] found that social media news use was associated with higher conspiracy beliefs. Individuals who trust social media news more are more likely to believe in conspiracy theories. Our findings present an opposite side of social media where a minority community (COVID long haulers) gathered together and found assurance, hope, and support from each other.

6.2 Practical Implications

6.2.1 Monitoring the existing symptoms and their continuous changes over time. The challenges and discrimination against minority groups such as COVID long haulers' communities are often difficult to assess at a finer granularity. Although long hauler patients did not experience denials and negligence for a long time (as other minority communities such as the LGBTQ+ community), many of them had to go through physiological and mental challenges that they had no prior experience with. The level of anxiety was even higher for them because of the uncertainty and vulnerability that we are still experiencing because of this global pandemic. They had no way to prepare for these challenges beforehand. Existing health care facilities often did not have the resources to assist long-hauler patients. Realizing these challenges, a Facebook group called "Survivor CORPS" has recently created a live guideline on how to establish and operate a multi-disciplinary Post-COVID Care Center [17] based on published practices of established centers from around the world. Our findings can provide valuable insights for such organizations, allowing a richer and nuanced understanding of long haulers' physiological and mental conditions. An interactive visualization interface like LongCov can help organizations build tools that can monitor long-hauler communities on a continuous and real-time basis to establish their initial base and later to evolve with the changing need and situation. LongCov can provide continuous updates based on social media data that can be an excellent resource for keeping track of conditions and symptoms of a diverse group of long-hauler patients who are hard to keep track of from any physical location.

6.2.2 Developing Infrastructure for Long-Hauler Rehabilitation. The symptoms experienced by long hauler patients often continue for a few weeks to several months. Our results have shown that lingering symptoms not only impact the physiological and mental conditions of the patients but also affect their overall lifestyle. Individuals going through such conditions often become temporarily incapable of continuing their existing jobs. Members from the long-haulers' community also shared incidents where they had to quit their job because of some chronic condition such as ME/CFS (Myalgic encephalomyelitis/chronic fatigue syndrome) which can develop as an aftereffect of COVID-19 infection [31]. Since COVID long-hauler research is still at a very early stage, it is hard to predict how fast long-hauler patients would recover from this condition. In fact, in the last few years, the logic behind traditional physiological and cognitive therapy for ME/CFS has also received massive criticism. We believe the information presented using the news organizer view will provide initial insights for establishing rehabilitation facilities appropriate for long-hauler patients. Future work could adapt our approach to looking at more nuanced conceptualizations of long-term socio-economical challenges of long hauler patients that can be used to develop a concrete structure of rehabilitation program for them.

6.3 Policy and Social Implications

Pandemic generally occurs infrequently. The last major pandemic in the United States was H1N1 flu or commonly known as swine flu. It was detected in the spring of 2009 and 12,469 individuals died from the infection. In comparison to swine flu, we have seen 622,825 deaths so far in the USA because of COVID-19 which is approximately 50 times higher than swine flu. Preparing for such a major pandemic is always a challenging task but not quite impossible. Some countries that have set notable examples are Singapore, Japan, Hong Kong, Taiwan, and South Korea. These countries suffered severely from SARS-COV-2 in 2003 and swine flu in 2009. However, they applied lessons learned from those outbreaks to revise their public health systems to such an extent that they could handle COVID-19 much better than the United States, some European countries such as Italy, and India [66].

Some of these statistics mentioned above might sound demoralizing but we can still learn from this pandemic. World Health Organization (WHO) has shown that infectious diseases outbreaks are emerging alarmingly regularly over the last 30 years [59] and we cannot deal with them just by luck. In the USA, we need to reorganize our public health care system significantly to be better prepared for the next possible pandemic. Our findings reveal that medical gaslighting might not happen only to women, black people, and Latinos. Without careful monitoring and a well-structured public health care system, this kind of discrimination can happen to anybody. A pandemic similar to COVID-19 can make such a condition significantly worse especially when the health care system is not prepared in advance. We believe our work will provide data-driven insights for informed policy decisions and aid in building backup facilities even for people whose conditions are not initially well-defined by medical science. Layered, people-oriented health care divisions might become a more effective solution in these scenarios. Our findings have shown that automatic tools based on social media data can be developed to monitor these issues closely. Early detection of such conditions might be easier through regular monitoring of social media data where people from diverse communities are organically gathering together to share their experiences and an interface like LongCov can become a great tool to go ahead with this vision. To this end, we recognize that some of the alarms raised on social media can turn out to be false alarms and prompt actions against those alarms can be wasteful. We, therefore, suggest carefully constructed layered infrastructure which can filter false alarms raised on social media (if any) and pursue further actions only on legitimate cases.

From a methodological perspective, we further recognize the ethical complexities associated with automatically monitoring people's social media data for taking critical decisions in the public health care sector. There has been a growing concern on this issue. A recent survey that many Twitter users do not know that their Tweets are publicly available and a majority of them believed that even researchers should ask for their consents before using their data for further analysis [23]. The factors that motivate users to express their opinion on social media and enable their candid self-disclosure may be confounded with their perceptions of being monitored. Moreover, people may have reservations as to who uses the results of such analysis — they may not be comfortable having government officials assess their social media data, as it can raise questions surrounding the privacy of social media data. If such a monitoring system is set without careful consideration of the privacy of the social media data, users of such groups might get discouraged and uncomfortable sharing their personal, sensitive information in the group.

The above factors and their potential risks and benefits need to be carefully evaluated before establishing infrastructure for continuous monitoring of online social media data.

7 LIMITATIONS AND FUTURE DIRECTIONS

We acknowledge that our work has limitations, many of which suggest interesting directions for future research. We do not make any population-centric assessments because the Twitter communities considered in our work cannot be considered wholesome of online discussions of the long-hauler community. Rather, our work should be seen as a proof-of-concept study to examine the long-hauler community on social media. Future work that makes population-centric assessments associated with long-hauler patients should consider the caveats concerning missingness and quality of social media datasets.

Our work inherently suffers from self-selection biases, that it only works on the language of the individuals who self-selected to express themselves on online communities, particularly those that did not feel shy to share their experiences on social media. Relatedly, we only study the language on social media. Incorporating other behavioral and communicative signals like frequency of posting, the topic of interest, and the support-seeking or support-giving nature of posting can help us to comprehensively understand the long-hauler community on social media. Future work can further investigate this community across other mediums and social media platforms.

The machine learning classifiers can be further improved with more sophisticated models of machine learning and natural language processing. This can be tuned with respect to the objective of the problem, where our objective was to balance between predictability and interpretability — i.e., to not only build a stable model that reveals the potential in machine learning to scalably infer the language of Covid-19 long-haulers, but also to help us understand the linguistic nuances in expressing minority community on social media.

The design of the LongCov interface highlights only the key concepts that we envisioned would be appealing to long-hauler communities. However, the design needs to be revised to a great extent before making it available for public use. We are fortunate that we have received many suggestions through our user study on how to improve the interface for better accessibility and effective utilization. Our next goal is to focus on those suggestions and revise the design of the interface so that we can incorporate them without losing the primary functionality of the tool. Here, we need to keep in mind that this feedback was collected from a group of people who were members of the long-hauler communities on Twitter and self-selected themselves to participate in the study. A more thorough user study with participants recruited from different sources and diverse backgrounds may generate more insights on how to improve the design of the LongCov interface.

8 CONCLUSION

This paper studied the long-hauler community from their discussion on social media. Adopting a combined qualitative and computational approach, this paper examined the language on the online community specifically created for COVID-19 long-haulers, and makes three primary contributions. First, we identified three primary themes that can broadly summarize all the discussion topics of the COVID-19 long-haulers' Twitter communities. Second, a machine learning classifier to identify social media posts discussing various topics related to long-hauler communities at scale. The classifier used a variety of features, ranging across word embeddings, psycholinguistic attributes, sentiment, and open-vocabulary based n-grams. We achieved a mean AUC of 0.80. Finally, we designed and built LongCov, an interactive visualization interface that can present social media discussion of long-hauler communities in a systematically organized and meaningful way to assist anyone interested in this topic to learn more about it directly from long-hauler patients, their friends, families, and acquaintances. We believe our work bears the potential to better understand the long-hauler community from an honest point of view especially when the community was ignored initially by health care providers. Our work also supports tailored public health intervention and policy change to better address the requirements of minority groups during critical times such as a global pandemic.

REFERENCES

[1] 2021. Jupiter. https://jupyter.org/.
[2] 2021. Plotly. https://plotly.com/about-us/.
[3] Monica Anderson and Andrew Perrin. 2017. Barriers to adoption and attitudes towards technology. https://www.pewresearch.org/internet/2017/05/17/barriers-to-adoption-and-attitudes-towards-technology/.
[4] Tanya Basu. 2020. Covid-19 "long haulers" are organizing online to study themselves. https://www.technologyreview.com/2020/08/12/1006602/covid-19-long-haulers-are-organizing-online-to-study-themselves/.
[5] Lisa F Berkman, Ichirō Kawachi, and M Maria Glymour. 2014. *Social epidemiology.* Oxford University Press.
[6] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[7] David L Blustein, Ryan Duffy, Joaquim A Ferreira, Valerie Cohen-Scali, Rachel Gali Cinamon, and Blake A Allan. 2020. Unemployment in the time of COVID-19: A research agenda.

[8] Adriel Boals and Kitty Klein. 2005. Word use in emotional narratives about failed romantic relationships and subsequent mental health. *Journal of Language and Social Psychology* 24, 3 (2005), 252–268.

[9] Jed Brubaker, Funda Kivran-Swaine, Lee Taber, and Gillian Hayes. 2012. Grief-stricken in a crowd: The language of bereavement and distress in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 6.

[10] Yuliya Cannon, Stacy Speedlin, Joe Avera, Derek Robertson, Mercedes Ingram, and Ashely Prado. 2017. Transition, connection, disconnection, and social media: Examining the digital lived experiences of transgender individuals. *Journal of LGBT Issues in Counseling* 11, 2 (2017), 68–87.

[11] Nikita Carney. 2016. All lives matter, but so does race: Black lives matter and the evolving role of social media. *Humanity & Society* 40, 2 (2016), 180–199.

[12] Andre Cavalcante. 2019. Tumbling into queer utopias and vortexes: Experiences of LGBTQ social media users on Tumblr. *Journal of Homosexuality* 66, 12 (2019), 1715–1735.

[13] Alexander Cho. 2018. Default publicness: Queer youth of color, social media, and being outed by the machine. *New Media & Society* 20, 9 (2018), 3183–3200.

[14] Seok Joon Choi, Jan Ondrich, and John Yinger. 2005. Do rental agents discriminate against minority customers? Evidence from the 2000 Housing Discrimination Study. *Journal of Housing Economics* 14, 1 (2005), 1–26.

[15] Kate Conger and Erin Griffith. 2020. As Life Moves Online, an Older Generation Faces a Digital Divide. https://www.nytimes.com/2020/03/27/technology/virus-older-generation-digital-divide.html?action=click&module=RelatedLinks&pgtype=Article.

[16] Courtney Connley. 2021. Why Black workers still face a promotion and wage gap that's costing the economy trillions. https://www.cnbc.com/2021/04/16/black-workers-face-promotion-and-wage-gaps-that-cost-the-economy-trillions.html.

[17] Survivor CORPS. 2021. Post-COVID Care: Guidelines for Multidisciplinary Care Centers. https://static1.squarespace.com/static/5e8b5f63562c031c16e36a93/t/605a8a3262f0191b99584df0/1616546355297/PCCC+Standard+of+Practice+3_23.pdf.

[18] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.

[19] Norman K Denzin and Yvonna S Lincoln. 2011. *The Sage handbook of qualitative research.* sage.

[20] Sanorita Dey, Brittany RL Duff, Wai-Tat Fu, and Karrie Karahalios. 2019. VidLyz: An Interactive Approach to Assist Novice Entrepreneurs in Making Persuasive Campaign Videos. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.

[21] Stefanie Duguay. 2016. "He has a way gayer Facebook than I do": Investigating sexual identity disclosure and context collapse on a social networking site. *New media & society* 18, 6 (2016), 891–907.

[22] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. " I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 153–162.

[23] Casey Fiesler and Nicholas Proferes. 2018. "Participant" perceptions of Twitter research ethics. *Social Media+ Society* 4, 1 (2018), 2056305118763366.

[24] Gilbert C Gee. 2008. A multilevel analysis of the relationship between institutional and individual racial discrimination and health status. *American journal of public health* 98, Supplement_1 (2008), S48–S56.

[25] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel. 1968. The discovery of grounded theory; strategies for qualitative research. *Nursing research* 17, 4 (1968), 364.

[26] Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333, 6051 (2011), 1878–1881.

[27] Trisha Greenhalgh, Matthew Knight, Maria Buxton, Laiba Husain, et al. 2020. Management of post-acute covid-19 in primary care. *bmj* 370 (2020).

[28] Jessica Guynn and Aleszu Bajak. 2021. Asian Americans report biggest increase in serious incidents of online hate and harassment during COVID-19 pandemic. https://www.usatoday.com/story/tech/2021/03/24/asian-american-hate-crimes-covid-harassment-atlanta-google-facebook-youtube/6973659002/.

[29] Oliver Haimson. 2018. Social media as social transition machinery. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–21.

[30] Blake Hawkins and Jen Jack Gieseking. 2017. Seeking ways to our transgender bodies, by ourselves: Rationalizing transgender-specific health information behaviors. *Proceedings of the Association for Information Science and Technology* 54, 1 (2017), 702–704.

[31] HealthAffairs. 2021. Paradigm Lost: Lessons For Long COVID-19 From A Changing Approach To Chronic Fatigue Syndrome. https://www.healthaffairs.org/do/10.1377/hblog20210514.425704/full/.

[32] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *Proceedings of the Web Conference 2021*. 3767–3778.

[33] Jelani Ince, Fabio Rojas, and Clayton A Davis. 2017. The social media response to Black Lives Matter: How Twitter users interact with Black Lives Matter through hashtag use. *Ethnic and racial studies* 40, 11 (2017), 1814–1830.

[34] Kokil Jaidka, Salvatore Giorgi, H Andrew Schwartz, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2020. Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences* 117, 19 (2020), 10165–10171.

[35] Wolfram Kawohl and Carlos Nordt. 2020. COVID-19, unemployment, and suicide. *The Lancet Psychiatry* 7, 5 (2020), 389–390.

[36] Nancy Krieger. 2001. A glossary for social epidemiology. *Journal of Epidemiology & Community Health* 55, 10 (2001), 693–700.

[37] Emily Leslie and Riley Wilson. 2020. Sheltering in place and domestic violence: Evidence from calls for service during COVID-19. *Journal of Public Economics* 189 (2020), 104241.

[38] Casey Littlejohn. 2019. Doing Harm: The Truth About How Bad Medicine and Lazy Science leave Women Dismissed, Misdiagnosed, and Sick, and editor of Feministing. com, explains the evolution of the current medical system in America; starting in the 18th century the existing medical community was overrun by European ideas, and female caretakers were eradicated (5-11). Without women working in the medical field, they were underrepresented and lacked a voice in their treatment. (2019).

[39] Joshua C Magee, Louisa Bigelow, Samantha DeHaan, and Brian S Mustanski. 2012. Sexual health information seeking online: a mixed-methods study among lesbian, gay, bisexual, and transgender young people. *Health Education & Behavior* 39, 3 (2012), 276–289.

[40] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*. 55–60.

[41] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[42] Matthew B Miles and A Michael Huberman. 1994. *Qualitative data analysis: An expanded sourcebook.* sage.

[43] Carmen Moreno, Til Wykes, Silvana Galderisi, Merete Nordentoft, Nicolas Crossley, Nev Jones, Mary Cannon, Christoph U Correll, Louise Byrne, Sarah Carr, et al. 2020. How mental health care should change as a consequence of the COVID-19 pandemic. *The Lancet Psychiatry* (2020).

[44] Nancy Morrow-Howell, Natalie Galucia, and Emma Swinford. 2020. Recovering from the COVID-19 pandemic: a focus on older adults. *Journal of aging & social policy* 32, 4-5 (2020), 526–535.

[45] Karsten Müller and Carlo Schwarz. 2020. From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103* (2020).

[46] Marcia Mundt, Karen Ross, and Charla M Burnett. 2018. Scaling social movements through social media: The case of Black Lives Matter. *Social Media+ Society* 4, 4 (2018), 2056305118807911.

[47] Adrienne Nolan-Smith. 2021. MAYA DUSENBERY ON HOW MEDICAL GASLIGHTING SIDELINES WOMEN'S HEALTH PROBLEMS. https://getwellbe.com/medical-gaslighting-maya-dusenbery/.

[48] Abigail Oakley. 2016. Disturbing hegemonic discourse: Nonbinary gender and sexual orientation labeling on Tumblr. *Social Media+ Society* 2, 3 (2016), 2056305116664217.

[49] Michael Quinn Patton. 2002. Qualitative research and evaluation methods. Thousand Oaks. *Cal.: Sage Publications* 4 (2002).

[50] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[51] Vincent J Roscigno, Diana L Karafin, and Griff Tester. 2009. The complexities and processes of racial housing discrimination. *Social Problems* 56, 1 (2009), 49–69.

[52] Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 269–276.

[53] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–27.

[54] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–27.

[55] Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. 2019. The language of LGBTQ+ minority stress experiences on social media. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–22.

[56] Koustuv Saha, Ingmar Weber, and Munmun De Choudhury. 2018. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[57] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.

[58] John Stasko, Carsten Görg, and Zhicheng Liu. 2008. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization* 7, 2 (2008), 118–132.

[59] STAT. 2021. Luck is not a strategy: The world needs to start preparing now for the next pandemic. https://www.statnews.com/2021/05/18/luck-is-not-a-strategy-the-world-needs-to-start-preparing-now-for-the-next-pandemic/.

[60] Anselm Strauss and Juliet Corbin. 1998. *Basics of qualitative research techniques*. Citeseer.

[61] Anselm L Strauss. 1987. *Qualitative analysis for social scientists*. Cambridge university press.

[62] Zsófia Szlamka, Márta Kiss, Sámuel Bernáth, Péter Kámán, Amina Lubani, Orsolya Karner, and Zsolt Demetrovics. 2021. Mental health support in the time of crisis: are we prepared? Experiences with the CoViD-19 counselling programme in Hungary. *Frontiers in psychiatry* 12 (2021), 792.

[63] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[64] Riina Vuorikari, Anca Velicu, Stephane Chaudron, Romina Cachia, and Rosanna Di Gioia. 2020. *How Families Handled Emergency Remote Schooling during the COVID-19 Lockdown in Spring 2020: Summary of Key Findings from Families with Children in 11 European Countries. JRC Science for Policy Report.* ERIC.

[65] Yang Wang, Liang Gou, Anbang Xu, Michelle X Zhou, Huahai Yang, and Hernan Badenes. 2015. Veilme: An interactive visualization tool for privacy configuration of using personality traits. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 817–826.

[66] WIRED. 2020. Singapore Was Ready for Covid-19—Other Countries, Take Note. https://www.wired.com/story/singapore-was-ready-for-covid-19-other-countries-take-note/.

[67] Queenie Wong. 2021. Twitter, Facebook and others are failing to stop anti-Asian hate. https://www.cnet.com/news/twitter-facebook-and-others-are-failing-to-stop-anti-asian-hate/.

[68] Daniel Wood and Maria Godoy. 2020. What Do Coronavirus Racial Disparities Look Like State By State? https://www.npr.org/sections/health-shots/2020/05/30/865413079/what-do-coronavirus-racial-disparities-look-like-state-by-state.

[69] Xizhu Xiao, Porismita Borah, and Yan Su. 2021. The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs. *Public Understanding of Science* (2021), 0963662521998025.

A MANUAL ANNOTATION PROCESS

To make sure that we did not miss any critical discussion topics (other than those that were identified by the LDA technique), we performed manual annotation of a subset of our dataset. Two human coders (one of them was the first author of this paper) who were familiar with the COVID long hauler community and their activities on social media examined a random sample of 1000 Tweets from our dataset. In the absence of labeled ground-truth data, they adhered to an open inductive coding approach [25]. During this coding process, we organized three brainstorming sessions where both coders discussed their preliminary thoughts with each other. We followed an iterative process and after multiple iterations, we identified 19 initial themes.

Next, to avoid any bias imposed by the first author of this paper and to make this annotation process more applicable in general, we invited five undergraduate students, all with backgrounds in social media data analysis, social science, and behavioral psychology, to examine another random sample of 500 Tweets. This new set did not contain any comments from the previous set. To provide background on the annotation process, we conducted an hour-long information session that involved discussing themes identified earlier along with specific example comments. Following this discussion, all coders independently coded the new set of 500 Tweets. They could either apply any theme from the existing pool of 19 themes (if applicable) or create a new theme for each comment based on their judgment. Finally, we discussed their coding experiences and received feedback about potentially ambiguous, misrepresented themes, and possible new themes.

Based on the discussion with undergrad coders, we modified, removed, and added a few themes. Next, to assess the effect of the changes, the first author and a social science expert coded another random sample of 600 Tweets (did not include Tweets from any previous set). The disagreements in annotations were resolved through discussion until consensus was reached. We also combined multiple initially identified themes that were closely related to each other. Finally, we achieved a substantial agreement based on Cohen's kappa test (K = 0.87). Combined efforts in the three stages resulted in the same three major topics that were identified from the LDA algorithm (refer section 3.2).

B LIST OF SYMPTOMS EXTRACTION FORM THE DATASET

To identify all unique keywords from our Twitter data corpus, we used a pattern/rule-based strategy to extract causal links applying an NLP framework that could use syntactic information. The first step was to pre-process the raw data to eliminate punctuations, stop words, articles, and offensive words. and extract only the useful data. Once the noise had been removed, the next step was Part of Speech (PoS) tagging. A PoS tagger's purpose is to assign a linguistic tag or information to tokens. We filtered in only such tagged terms because we were only interested in retrieving the symptom element of the text, which was typically forms of nouns or adverbs (primarily).

Our process became more efficient step by step. We used a dictionary describing the symptom-related keywords, words were chosen from that, which were all nouns. The primary focus was on unigrams. As a result, removing all Noun-Noun (NN) words would direct us to the synonyms that were needed. The PoS tagger NN was implemented. It was not, however, 100% accurate. To increase the overall performance, we included bigrams and further n-grams because the symptoms can be a group of words. Bigrams or further n-grams could also be used to express adjectives for a specific symptom, such as chest pain or awful taste. This highlighted the importance of a model that could handle n-grams larger than unigrams.

In addition, noise removal was ineffective in retaining only the noun parts of the input data. By using n-grams we built a pattern to identify the symptom-related keywords which were termed as cause-effect relation extraction. Cause and effect relationships could be expressed in a variety of ways, including verb phrases and noun phrases. The cause-effect relations were determined based on trigger terms. For example, consider a sentence: "I have lost taste and smell", we can see the words "taste", "smell" are nouns that are combined by conjunction constitutes the symptom dictionary. We needed to build a pattern-based method that extracts based on the part of speech of a particular word in the sentence. By using the PoS tagger and the trigram we successfully extracted the list of symptoms that contained 281 entries. One human coder manually checked the entire list and discarded 17 keywords. The final list contains 264 symptoms in total.

C QUESTIONS USED FOR SEMI-STRUCTURED INTERVIEWS DURING USER-STUDY

Here we have included all questions that we used during the semi-structured interview.

Question 1: How easy and effortless was to use the LongCov interface?

(1) Not at all easy
(2) Slightly difficult
(3) Neither easy nor difficult
(4) To some extent easy
(5) Extremely easy

Question 2: Would you like to use this interface in the future?

(1) Never
(2) Rarely
(3) Sometimes
(4) Often
(5) Always

Question 3: Would you recommend this interface to your friends, family, and acquaintances?

(1) Extremely unlikely
(2) Unlikely
(3) Neutral
(4) Likely
(5) Extremely likely

Question 4: What elements or functionality of the interface did you like the most?

Question 5: What were the elements or functionality of the interface that you would not like to see in the revised version of the interface?

Question 6: What additional elements or functionality would not like to see in the revised version of the interface?

We used a 5-point Likert scale for recording the answers to the first three questions. For the last three questions, we asked users to answer them in their own words. Later, we transcribed those answers and manually coded them to extract the main themes of their responses. The interviewer verbally asked all six questions to each participant. No online form was used for them. For the Likert scale answers, the interviewer manually took note of them in an excel file for further processing.