COVID-19 Misinformation: The Evolution of Tweets Over Time on the Pandemic

Roxy Velazquez, Eury Sohn, Jerusha Seesala

ABSTRACT

The prevalence of COVID-19 misinformation has existed since the onset of the pandemic. And while much research has been conducted in relation to analyses of misinformation justification and its harms, scant research exists to observe the changes of such misinformation over time. Using the Elaboration Likelihood Model (ELM), we aim to look at individual media viewers' perceptions of misinformation and inspect the factors that may have influenced the change of these perceptions over the course of the pandemic. To do so, we look at datasets of tweets made by users in 2020 and 2021 and introduce our data exploration and analysis to contextualize it to the framework of ELM. The results of our analysis aligned with previous ELM misinformation research after finding a 10% increase in negative tweets and significantly more authority figures were mentioned in 2021, which are both heuristic cues that allude to peripheral processing. The implications this paper holds for future works includes providing a framework for time-based considerations of misinformation and contextual factors of media viewership and perception.

KEYWORDS

Misinformation, Elaboration Likelihood Model (ELM), Cognitive Reflection Test (CRT), central cues, fake news, Dual Process Theory (DPT), Natural Language Processing (NLP)

1 INTRODUCTION

Having access to accurate information is imperative to living a healthy full life. And access to correct, complete information is crucial for individual safety, health, and making informed political decisions. Unfortunately, not every piece of information we are exposed to is truthful. Misinformation has been defined by the dictionary as, "false information that is spread, regardless of intent to mislead" (Dictionary.com, 2022, para. 4). For the context of this study, all dissemination of false information will be lumped under the category of "misinformation."

Misinformation has existed well before modern technology. An example to contextualize the prevalence of misinformation prior to modern times was during the Roman Republic era around 2,000 years ago. Octavian slandered Mark Anthony with deceitful, unfounded faux commentary about his substance use, values, and infidelity with Cleopatra or in the 15th century when there were lies spread about the health of King George II using the printing press to spread information quickly to encourage a rebellion (BBC News, 2020). Clearly misinformation is not a new concept and will continue to spread and infect society, but using modern technology in conjunction with academic advances, is there a way to halt or reduce the impact? The goal of this paper is to explore datasets to attempt to discern this.

While misinformation spread was impactful in the past, in today's society, misinformation has evolved to spread to millions of people in a millisecond using the "share" button on any social media platform. Like a contagious virus, misinformation has become an entity that makes us question everything we read in a sea of information overload where every direction you turn, there are different viewpoints, stories, and topics being constantly shared. Misinformation can turn deadly when you apply it to the context of health information such as for the worldwide pandemic, COVID-19, especially considering that 87% of people have been exposed to some form of misinformation on social media (Zhao et al, 2022).

This paper sets out to analyze two main topics: why does misinformation continue to spread in a society that has all the informational resources at their fingertips and what causes misinformation to be believable or spread faster and wider than factual information (Ross et al, 2021). Using Dual Process Theory, a social psychological framework for how information we receive is processed, we hope to shed light on why and how misinformation is still such a prevalent topic. Thus, with these considerations in mind, we introduce our two primary research questions:

RQ1: How do misinformation tweets regarding Covid-19 differ between 2020 and 2021?

RQ2: Using the Elaboration Likelihood Model, do misinformation tweets from 2021 reference more cues related to peripheral processing, like emotion and authority, compared to 2020?

1.1 Dual Process Theory

Dual Process Theory is a theory rooted in communication science that demonstrates how the way information people come across is processed varies upon *how* the information is perceived. There are two different paths in which the information can be processed: Type I, which relies on intuition, or Type II, which involves reflective reasoning (Kahneman, 2011). The Elaboration Likelihood Model (ELM) is a model that uses Dual Process Theory as its framework to demonstrate how this information processing looks in real life. In ELM, the Type I model is referred to as "peripheral pathway", which is the pathway which relies on heuristics and little elaboration goes into processing the information. An example of this could be seen in a list of persuasive arguments for why someone should eat healthy. If this is perceived as common sense to someone, they may only look at the title and the first bullet point and not process all the actual information presented. This differs from the Type II model, or the "central pathway", which processes and evaluates the information based on merits and takes time to think through what is being presented (Booth-Butterfield & Welbourne, 2002). Overall, the central pathway tends to be more resistant to persuasive arguments, stable, and predictive of future behaviors.

Using the framework of the ELM, it can be hypothesized that misinformation shared on social media may follow either pathway depending on the person's motivational factors, beliefs, and values. COVID-19 misinformation has been on the internet since day 1 of the pandemic. What we are more interested in is what pathway the misinformation tends to gravitate towards. If misinformation is geared towards the peripheral pathway, it is likely to contain emotionally charged, time-sensitive, and eye-catching titles and photos. If misinformation is geared toward the central pathway, it is going to take more deliberation to determine what is correct. Examples of this could include non-doctors dressing up in lab coats to talk about COVID-19 treatments or a legitimate looking academic paper that has not been peer reviewed, but rather published on a blog or in a predatory journal. Using this information, our prediction is that at the start of COVID-19, the verifiable misinformation presented appealed to both the central and the peripheral pathways, but as COVID-19 continued to persist and factual information was published by governmental organizations, such as the Center for Disease Control (CDC), the misinformation being spread a year or more after the initial outbreak of the pandemic, will heavily skew to appeal toward the peripheral pathway.

2 RELATED WORKS

Researchers have previously applied the ELM to observe the spread of misinformation. Previous academic research has shown that "fake news" is not as prevalent as we may believe given how much society talks about it, but one worrying statistic shown is that "fake news" gets shared just as much if not more than factual news (Ross et al, 2021). We have thoroughly reviewed the related literature to provide empirical evidence of this phenomenon and reveal the current gaps that exist in this domain. One common argument surrounding the sharing of COVID-19 includes partisan ideology, and we will show that this has little impact on the misinformation.

In terms of partisanship, much of the COVID-19 misinformation material that is shared tends to be associated with the Conservative or Republican members of American society. This stems in part from the fact that Mr. Trump served as president during the start of the pandemic and would endorse misinformation as valid treatment options, such as when he suggested the population should just inject themselves with bleach/cleaner. The research, however, does not show this same sentiment. In an academic study of Americans completed by Ross et al (2021), they focused their research solely on information with a hyper partisan bias, either heavily Republican or Democratic. They provided true information and misinformation surrounding partisan current events and had participants determine which were true. The results of this study showed that political ideology did not play an impact on determination of misinformation, but rather their personal degree of analytical reasoning did.

One common thread that accompanies misinformation and ELM research is that overall cognitive deliberation on material, decreases the belief in misinformation. This is to say, that when people are fully utilizing their central pathway, they are more likely to make an accurate decision about whether the material they are presented is factual. Many studies utilize a self-reported Cognitive Reflection Test (CRT) measurement to look at cognitive deliberation and in all those studies people who score higher on CRT are objectively better at determining what is misinformation and what is not [(Ross et al, 2021), (Bago et al, 2020), (Borukhson et al, 2022)]. This provides evidence that the central pathway processing leads to a decrease in "fake news" believability.

In a different study completed by Bago et al (2020), looked at the idea of deliberation with both factual and fake headlines. They had people make a gut reaction about whether a headline was fact or fiction. After that, participants were able to deliberate about their decision for a short period of time and think it through thoroughly. They concluded that people were significantly less likely to believe false headlines and make fewer mistakes regarding the nature of the headlines when they deliberated about it, which again was not impacted by any partisan bias. They indicated that their results demonstrated that fast, intuitive processing (peripheral pathway) was involved in believing misinformation. Furthermore, they hypothesized that this may be due to the quick scrolling done on social media by users and thus misinformation needs to be emotional and extreme in terms of time or action.

Zhao et al (2022) went into depth about the language and dissemination of misinformation on a social media platform called Weibo. They determined that there were seven common indicators of misinformation that are meant to activate the peripheral pathway: authority, commitment, contrast, liking, reciprocity, scarcity, and social proof. They found that unverified accounts were more likely to produce or share misinformation, but those who have a higher following or status, ended up getting more followers to share the misinformation. Interestingly, they found a connection between the eliciting of positive emotion for misinformation surrounding unproven treatment methods and negative emotions for spreading of the epidemic or lockdowns. In terms of information dissemination, 97% of misinformation is disseminated using a "radiation level sharing," which means that the first level of sharing misinformation from the original source is the largest. This is different from "viral level sharing" where each wave of sharing can continue to grow.

Emotion also plays a large role in how we process information. Negative emotions, such as sadness, have been linked to a decreased reliance on the peripheral pathway as this correlates with an increase in skepticism. Positive emotions increase gullibility and anxiety creates a generalized feeling of doubt, both of which causes a higher belief in misinformation (Martel et al, 2020; Lee, 2020).

One final aspect to consider in previous research is that of individual user perceptions when sharing social media posts. In considering fact checks, or annotations or suggestions that refer users who come across a specific post, towards factual information web pages or urge them to do so. The findings that Chen et al, 2021, discovered are paramount to understanding how both peripheral and central cues are involved in such fact checks. The main conclusion of interest is that central processing, the pathway related to critical thinking, is vital to an increase in fact check sharing.

Overall, many features of social psychological research have uncovered the elements relating to perceptions of social media's misinformation on COVID-19. What the current literature does not include is an analysis of how such perceptions have changed over time, and what factors contribute to such changes. Our research aims to address such gaps and provide insights for future prevention of mass misinformation sharing, also coined the "infodemic," parallel to the COVID-19 pandemic (Chen et al, 2021).

3 DATA ANALYSIS INTRODUCTION

3.1 Dataset Information

The data used for our research analysis is sourced from three datasets holding tweets from the social media platform, Twitter. There has been an influx of COVID-19 datasets in a short amount of time, yet these datasets lack specific labels. For the misinformation analysis, our dataset would need to contain tweets between the years 2020-2021 and include annotations for misinformation.

For this purpose, the three chosen datasets include the following:

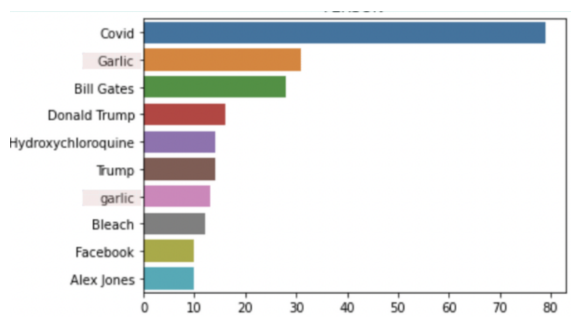
1. CMU-MisCov19 Dataset: 4573 tweets & 17 categories (Shuhan et al, 2020)
2. Misinformation Twitter Dataset: 1500 tweets & 5 categories (Zen, 2020)
3. ANTiVAX Dataset: 15000 tweets & 2 categories (Shuhan et al, 2020)

3.2 Dataset Cleaning

The corpus of the datasets underwent a cleaning process on the elements included in each. Tweets categorized under the annotation of "satire" were excluded from our data alongside those deemed "irrelevant".

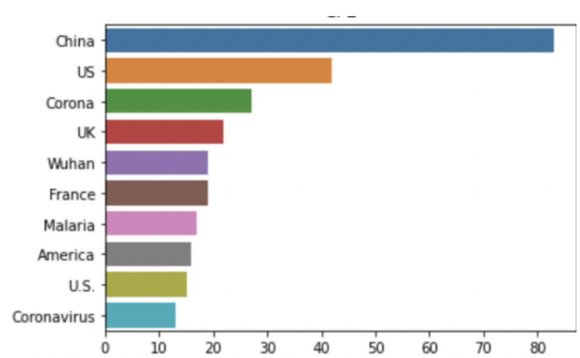
Pre-processing the tweets was done after ensuring the context of the keywords. As Figure 1 shows, the terms garlic and Garlic are some of the most common words, yet "garlic" and "Garlic," was treated separately as one referred to the misconception that "garlic" could cure Covid-19 whereas "Garlic" referred to a brand. Moreover, terms that were used interchangeably such as USA, U.S., US, and America were classified under the same title of "US".

| Figure 3.2.1 |



Histogram depicting the common nouns that occurred in the tweets corpus.

| Figure 3.2.2 |



Histogram depicting the most common countries which occur in the tweet corpus.

To fully ascertain the misinformation of Covid-19 over time our data set of tweets contains Tweet ID, the date (written as 2020 or 2021), the misinformation label, polarity, subjectivity, and sentiment. In our study, we will present this information alongside a sentiment value which is to judge the tweet's emotional severity. This feature will help focus the research on its influence on the peripheral pathway.

3.3 Dataset Preprocessing

The corpus was preprocessed using the Python NLP (natural language learning) toolkit (NLP, n.d.). This kit included the removal of punctuation, lemmatization, removing stop-words, and capital letters. However, as the lemmatization of the Python library often turns negative sentiment contraction words to positive sentiments, such as the word don't into do, this was done manually with a custom dictionary. Words with equivocated meanings, such as country names or nicknames, were also replaced using a custom dictionary. These words include USA, America, and US, all being deemed the same.

3.4 Sentiment & Polarity

Twitter sentiment analysis is a prominent topic in research of linguistics and their various methods of approaching the matter. These analyses include detecting polarity in text for the tweets in either a two-labeled method (negative, positive) or a three-labeled method (negative, positive, and neutral). The tweets in this study were analyzed with the help of TextBlob library and its inbuilt feature for polarity analysis (TextBlob, n.d.). Each tweet was given a value from -1 to 1 where -1 has the most negative sentiment and 1 as the most positive sentiment.

4 METHODOLOGY

4.1 Topic Extraction via PyLDAvis

Latent Dirichlet Allocation Visualization (LDAvis) was used to classify and explain interview topics through text mining (Formoso, 2022). LDA is a Python library that visually expresses the learning outcomes of models used in topic modeling. This technique also considers many terms to convey content to general audiences without prior knowledge and demonstrates the distinctions of derived topics. This model learns the subject vector from a given text, where the subject vector is P(w|t) consisting of probabilities. The process that takes place here is a high-dimensional vector, so it is difficult to grasp the relationship between each other. PyLDAvis helps to easily connect and understand the relationship of keywords using dimensionality reduction and extraction

By extracting the topic via pyLDAvis, the topic can be identified by a topic ellipse that is divided into x- and y-axes for visualization of a set of finding similar words. Also, the panel on the right-hand side shows the corpus and the selected words based on the topic ellipses.

pyLDAvis:
The representation includes the distribution of themes in two-dimensional space and these themes are represented in the form of bubbles.

- The size of the ellipse is proportional to the frequency of the subject.
- The distance between subjects approximates the semantic relationship between them.
- Topics that share common words overlap by comparing them to non-overlapping topics.

Right-hand panel:

- The blue bar graph showing the frequency distribution in the document
- The red shaded bar describes the frequency of each word for a given subject

4.2 Sentiment Analysis

Textblob is a lexicon-based Python library model widely used for sentiment analysis as it offers itself to streamline accessible text processing. It offers NLP activities such as translation, Parts of Speech tagging, tokenization, classification, and sentiment analysis which we have utilized in our analysis. Its default library allows for training and recognizing sentiment in each text, assigning values to each (TextBlob, n.d.).

After assigning each tweet with the sentiment, the term frequency was set using a count vectorizer and fitting it into a sparse matrix. This was essential as the same words may have different sentiments depending on each tweet; hence the context of the tweet determines the sentiment value of a word.

The term frequency was modeled as a scatterplot, making sure to use a harmonic mean to ensure that outliers would not influence the "trend" of the sentiment (Broomberg, 2022). The harmonic mean H of the positive real number xn is defined as:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

4.3 Authority

NLP allows users to assign POS to each word with the ability to extract different parts of speech and pronouns (nltk.org). With this feature, we were able to delineate the authority figures which occur in the dataset. Each of the tokens was looked through manually to ensure no equivocation of the same name was separate tokens. "President of the United States", "The Don", "Trump", and "Donald Trump" were all assigned the same name as "Donald Trump" for accurate frequency counting.

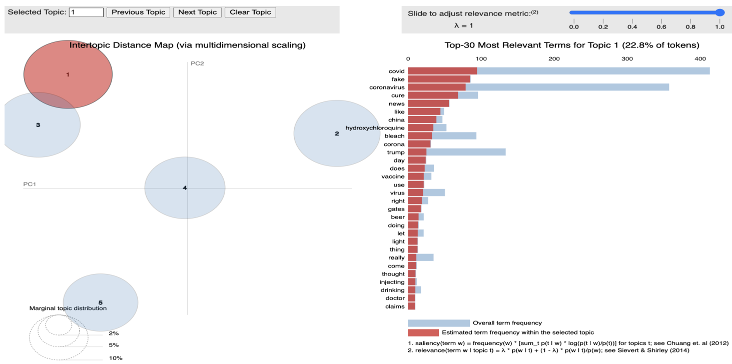
5 RESULTS

5.1 pyLDAvis

The pyLDAvis graphs were created for the following years: 2020 and 2021. Thematic qualitative analysis was performed on both sets of topics to determine overarching themes for each topic. Additional information related to analysis can also be found in Appendix A.

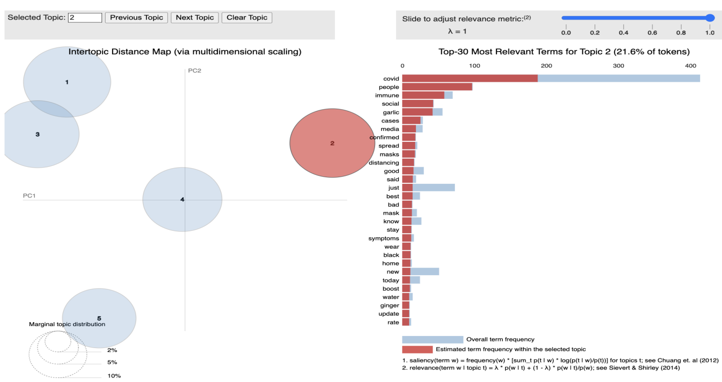
2020 Topic LDA Analysis

| Figure 5.1.1 |



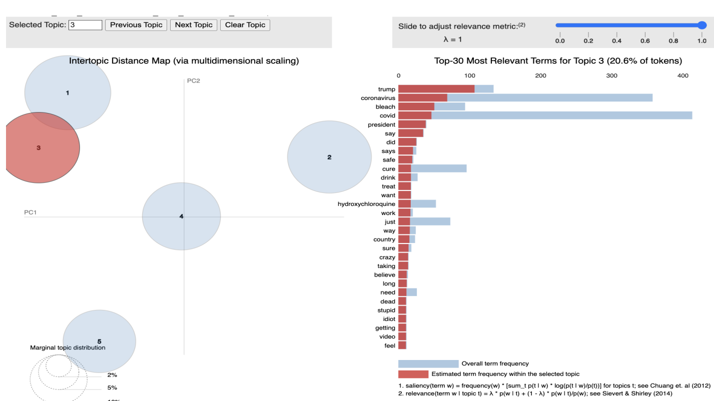
Topic 1 indicates the subject matter of fake news, claims and politics as the tweets center around words that stipulate the reality of Covid-19.

| Figure 5.1.2 |



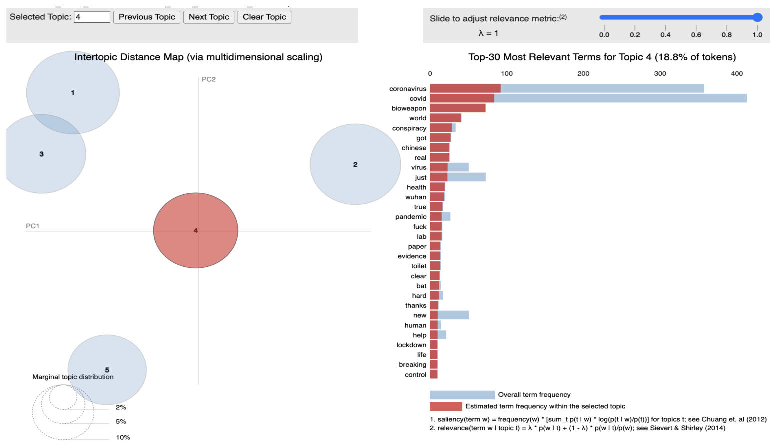
Topic 2 indicates a "naturalistic" approach to Covid-19 Treatment

| Figure 5.1.3 |



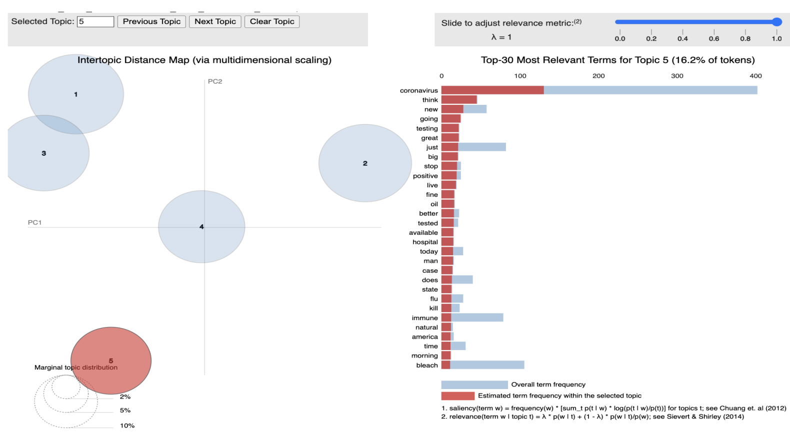
Topic 3 indicates a Trumpian rhetoric toward Covid-19 at the beginning of the pandemic

| Figure 5.1.4 |



Topic 4 indicates the subject matter of Covid-19's origin conspiracies

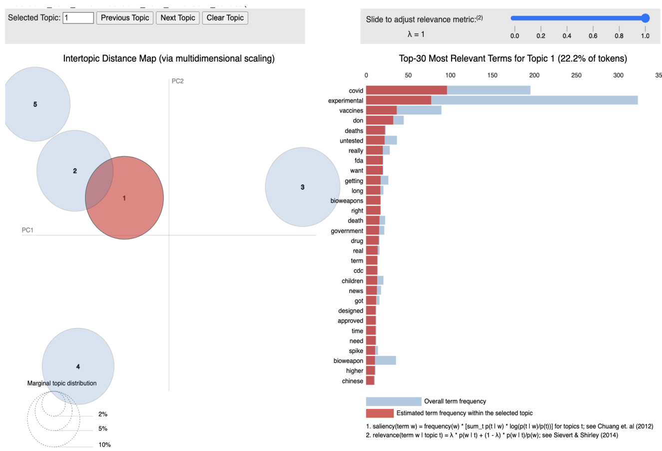
| Figure 5.1.5 |



Topic 5 indicates the subject matter of stipulations about the lethality of Covid-19.

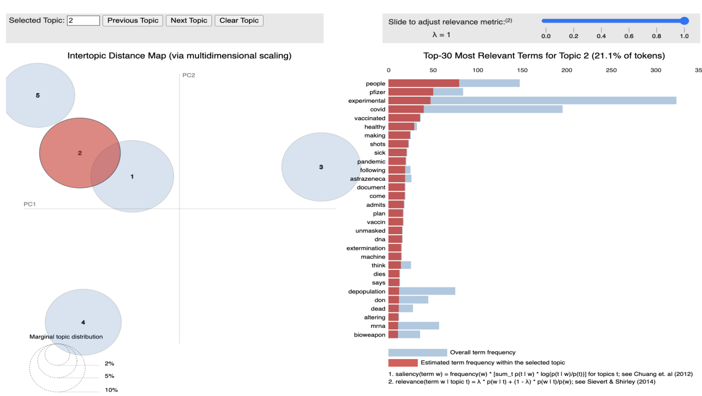
2021 Topic LDA Analysis

| Figure 5.1.6 |



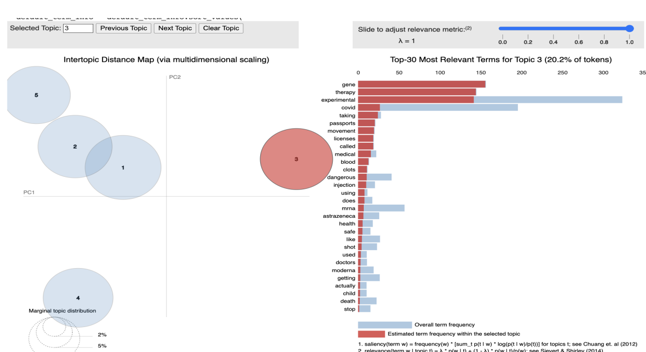
Topic 1 indicates the subject matter of distrust in governmental organizations' handling of Covid-19.

| Figure 5.1.7 |



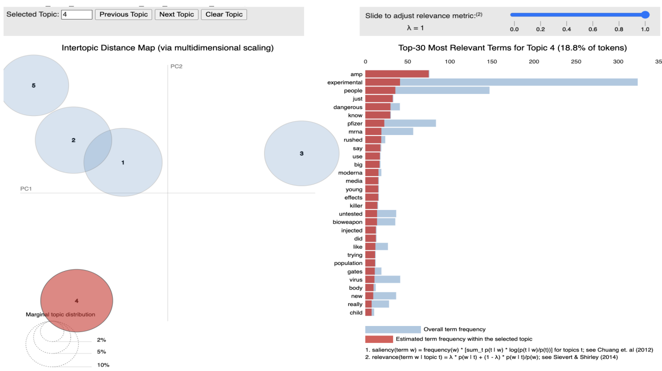
Topic 2 indicates the subject matter of the conspiracy that Covid-19 is used as a depopulation tool.

| Figure 5.1.8 |



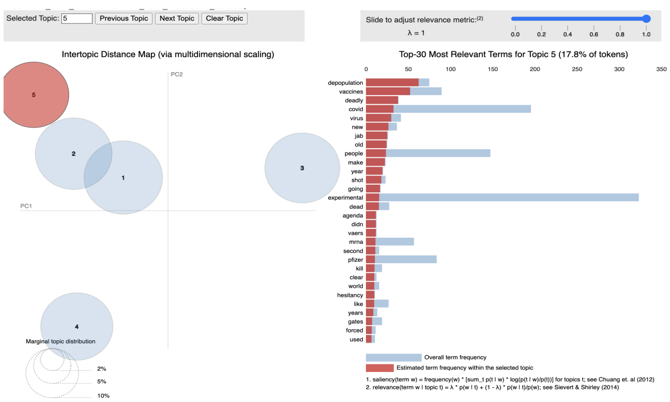
Topic 3 indicates the subject matter of the government's role in Covid-19.

| Figure 5.1.9 |



Topic 4 indicates the subject matter of the distrust of the vaccine, whether it's rushed or untested.

| Figure 5.1.10 |

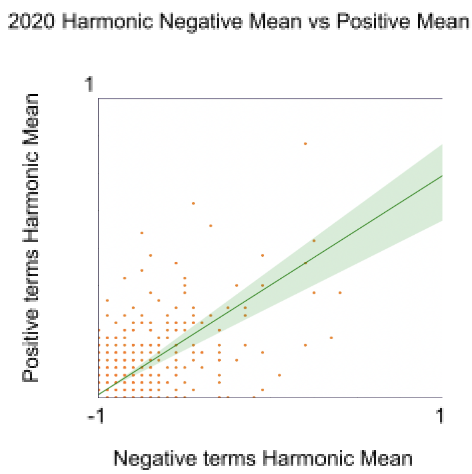


Topic 5 indicates the subject matter of the government policies regarding the vaccines and the trust in the matter.

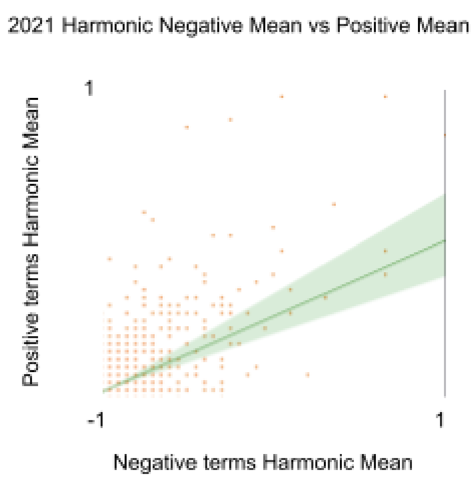
5.2 Sentiment Analysis

The cumulated harmonic sentiment for terms 2020 and 2021 with the line of best fit demonstrates the general skew of the sentiment of the following years.

| Figure 5.2.1: 2020 Sentiment Frequency |



| Figure 5.2.2: 2021 Sentiment Frequency |

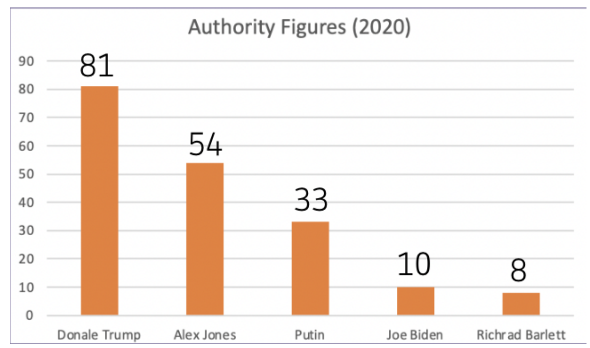


Comparing 2020's sentiment frequency to 2021's sentiment frequency, we can see that line skewness differs in its values as 2021's line is skewed towards the negative terms. This suggests that the 2021 corpus offers more negative sentiment tokens and that tweets themselves were more negative in 2021 compared to 2020.

5.3 Authority

The 2020 and 2021 datasets are split and show the frequency of which authority figures appear in each corpus.

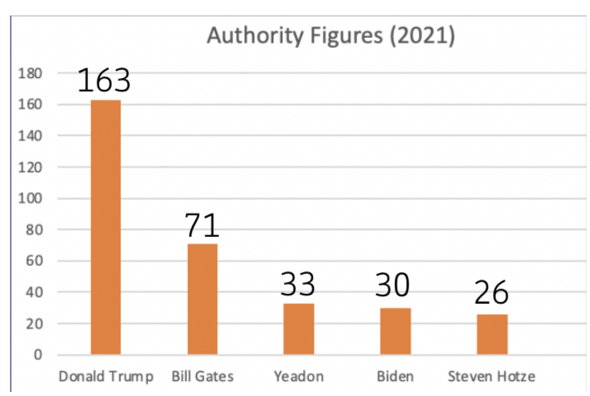
| Figure 5.3.1 |



In 2020, the five most mentioned authorities are "Donald Trump" (81 times), "Alex Jones" (54 times), "Putin" (33 times), "Joe Biden" (10 times), and "Richard Barrett" (8 times).

In total, authorities were mentioned 186 times in the corpus.

| Figure 5.3.2 |



In 2021, the five most mentioned authorities are "Donald Trump" (163 times), "Bill Gates" (71 times), "Yeadon" (33 times), "Biden" (30 times), and "Steven Hotze" (26 times).

In total, authorities were mentioned 323 times in the corpus.

6 DISCUSSION

While misinformation is a common topic in academic research due to its ubiquity in digital media, this is the first paper to our knowledge that examines misinformation dissemination on a specific topic longitudinally. Based on prior research, the single best predictor of whether someone believes misinformation is high performance on a Cognitive Reflective Test (CRT), which is used as a measure for determining level of central pathway processing (Pennycook & Rand, 2021). Misinformation tends to be anecdotally associated with political polarization, but even in studies where partisan misinformation is used, central pathway processing prevails as the lone significant result for misinformation belief (Ross et al, 2021). Our results add to the field of misinformation research and reaffirm previous study results that misinformation sharing is likely indeed correlated to peripheral processing, given the significant results in emotion and authority figures, which are two important cues used in peripheral processing (Zhao et al, 2022). Additionally, our results show the trajectory of misinformation from when Covid-19 was discovered and the misinformation a year later, which is vastly different.

6.1 Sentiment

Previous studies have shown the influence that emotions can have on misinformation belief and the sheer cognitive load of certain emotions, especially anxiety or anger, can result in peripheral pathway processing (Martel et al, 2020). For the purposes of this discussion section, sentiment refers to the attitude or opinions surrounding Covid-19 and the resulting emotions evoked by a piece of misinformation. While we cannot definitively measure how the users feel when sharing the misinformation, the language used can be a good indicator of emotion and tone. For example, words like "good" or "cure" evoke an inherently different set of emotions than words like "horrible" or "death." To analyze our results regarding emotion, we performed sentiment analysis using code and a thematic qualitative analysis using topics and tokens extracted using LDA, which can be found in Appendix A.

Vosoughi et al (2018) found that false stories tend to evoke feelings of fear, disgust, and surprise which is different from factual stories which tend to cause feelings of anticipation, sadness, joy, and trust. While factual news can elicit sadness, an inherently negative emotion, the relevant tokens generated are skewed toward fear and disgust. The code we used determined that there was a 10% increase in negative sentiment misinformation tweets between 2020 and 2021. The overarching theme of fear and disgust related to negativity is also reflected in our qualitative thematic analysis.

Using LDA, five topics were generated for 2020 and 2021, which can be seen in Appendix A. The tweets from each year respectively were grouped into topics based on similarity. It is important to note there were no topics that were heavily skewed to include a very large or very small percentage of the tweets, and most were between 22-18%. Each topic generated thirty of the most relevant tokens, in our case words. Using the tokens, themes were determined and tokens that influenced the determination of the topics are italicized. In 2020, the five topics included: dangerous Covid-19 treatment endorsed by Trump, naturalistic cures for Covid-19, a Trumpian rhetoric of the attitudes surrounding Covid-19 in early to mid 2020, conspiracy theories about virus origin, and those who downplayed Covid-19 as a "flu." In 2021, we see more homogeneity in topics with the five topics being: distrust in US Governmental organizations, vaccines as a depopulation tool, fear of governmental control with mentions of vaccine passports, distrust in vaccines due to it being developed too fast or an experimental treatment and forced vaccination agendas. In 2021, we see two major themes of distrust in the government and distrust in the vaccine, which can evoke fear and/or disgust in individuals. In 2020, we observe about five distinct topics although Topic 1 and 3 both relate heavily to Trump.

The five topics generated from 2021 misinformation tweets along with the 10% increase in negative sentiment tweets showcase the impact of peripheral processing cues in misinformation, but we went a step further and performed a thematic analysis on the actual tokens. When referring to the LDA table located in Appendix A, the bolded words are considered words that inherently evoke negative emotions, such as "death", "fake", or "depopulation." Out of the 150 tokens generated for all five topics, 2020 had 13 negative tokens compared to 2021 which had 36 negative tokens, which is a 2.77 times increase again showcasing the impact of peripheral processing cues.

6.2 Authority

A heuristic cue often used with peripheral processing is believing information based on the person who is sharing or saying it, especially if that person is a figure of authority (Zhao et al, 2022). Generally, if a person is in a position of authority, especially the government, people will expect they are telling the truth and have their best interest at heart. Another complex factor related to Covid-19 was that there was limited information in general on the virus in early 2020. In social psychology, the "illusory truth effect" is the concept that when an individual receives information on a topic they are unfamiliar with, they trust the person providing the information is telling the truth (Martel et al, 2020). If this person is in a position of authority, this would likely strengthen this feeling.

Our results indicate that in 2020, there were 186 instances of an authority figure being mentioned in misinformation tweets compared to 323 in 2021. This is alarming as 2021 had significantly less tweets than 2020, with 2021 having 951 tweets versus 2020 having 1,372 tweets after data cleaning. Trump accounted for the most appearances for both years: 81 (43.5%) of the occurrences in 2020 and 163 (50.5%) of the occurrences in 2021, which makes sense given he was president at the beginning of the pandemic and was a prime source of misinformation dissemination. In fact, he was banned from Twitter due to his repeated endorsement of pandemic related misinformation.

On April 23, 2020, the president of the United States, Donald Trump, "encouraged his top health officials to study the injection of bleach into the human body as a means of fighting Covid" (McGraw & Stein, 2021, para 1). The American Association of Poison Control Centers (AAPCC) reported a 121% increase in bleach poisoning in April 2020 compared to April 2019 (Kluger, 2020). This dangerous rhetoric of injecting or drinking bleach can also be seen throughout the 2020 relevant tokens. This is just one example of how the illusory truth effect and implicit trust in authority can impact individuals and cause serious or deadly impacts.

In 2020, other authority figures mentioned also included Alex Jones, a notorious right wing conspiracy theorist who was very vocal about Covid-19 misinformation and anti-vaccine, Vladimir Putin, the Russian president who Trump has praised heavily, Joe Biden, the current US president and winner of the 2020 election, and Richard Barrett, a doctor who endorsed alternative Covid-19 treatments such as Ivermectin.

In 2021, other authority figures mentioned included Bill Gates, as many conspiracy theorists thought he was responsible for inserting “microchips” in vaccine to surveil citizens, Michael Yeadon, an ex-Pfizer scientist who alleged the Covid-19 vaccine was dangerous for children and became an important figure for anti-vaxxers, Joe Biden, and Steven Hotze, a right wing conspiracy theorist that is anti-vaccine and repeatedly mentioned a “forced vaccination agenda” by the government.

7 CONCLUSION

Through examination of our combined dataset, we were able to retrieve valuable insights from the evolution of misinformation tweets throughout the COVID-19 pandemic. With thorough data analysis and consideration of the Elaboration Likelihood Model on tweets made on Twitter, we arrived at the conclusion that tweets have indeed changed over time. By method of sentiment and thematic qualitative analysis, the language that Twitter users held in 2020 and 2021 were observed and the following results were found:

- Trends of more peripheral, or emotion-related, cues were found to increase by 10% from 2020 to 2021.
- Additionally, more references to authority figures were made by Twitter users in 2021 than 2020 when comparing both years.

It is also important to recognize the trends of media viewership over time and its impact on misinformation believability. From the increase in peripheral cue references, it may suggest that viewers were more likely to believe misinformation in 2021 than in 2020. Our data analysis and research along with previous literature helps inform viewers of the types of cues related to peripheral processing and brings awareness to the factors that make misinformation more believable on social media, such that misinformation is made less trustworthy.

The implications of our research further emphasize its value. Our data analysis provides a framework to apply ELM to pandemic misinformation for future use by data scientists and social psychology researchers. It also offers a timespan data grouping on these tweets to be used for future related research interests. Lastly, it offers insights that will aid viewers with their interactions with social media, which may help decrease misinformation believability.

Overall, we have performed a data driven examination of COVID-19 misinformation tweets and analyzed trends from the year 2020 and 2021, while focusing on the Elaboration Likelihood Model as part of our consideration, leading to impactful findings.

References

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

BBC News. (2020, December 4). *A brief history of fake news*. https://www.bbc.co.uk/bitesize/articles/zwcgn9q

Booth-Butterfield, S., & Welbourne, J. (2002). The elaboration likelihood model: Its impact on persuasion theory and research. In J. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 155-173). Thousand Oaks, CA: Sage.

Borukhson, D., Lorenz-Spreen, P., & Ragni, M. (2022). When does an individual accept misinformation? an extended investigation through cognitive modeling. *Computational Brain & Behavior*, *5*(2), 244–260. https://doi.org/10.1007/s42113-022-00136-3

Bromberg, M. (2022, November 3). *Harmonic mean definition: Formula and examples*. Investopedia. https://www.investopedia.com/terms/h/harmonicaverage.asp#:~:text=The%20harmonic%20mean%20is%20calculated,of%20numbers%20in%20that%20series.

Chen, Q., Zhang, Y., Evans, R., & Min, C. (2021). Why do citizens share covid-19 fact-checks posted by Chinese government social media accounts? the elaboration likelihood model. *International Journal of Environmental Research and Public Health*, *18*(19), 10058. https://doi.org/10.3390/ijerph181910058

Dictionary.com. (2022, August 15). *"Misinformation" vs. "Disinformation": Get informed on the difference*. https://www.dictionary.com/e/misinformation-vs-disinformation-get-informed-on-the-difference/

Formoso, J. B. (2022, April 8). *4 steps of using latent Dirichlet allocation (LDA) for topic modeling in NLP*. Omdena. https://omdena.com/blog/latent-dirichlet-allocation/

Kahneman D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kluger, J. (2020, May 12). *Accidental poisonings rise after Trump disinfectant comments*. Time. https://time.com/5835244/accidental-poisonings-trump/

Lee, J. (2020). The effect of web add-on correction and narrative correction on belief in misinformation depending on motivations for using social media. *Behaviour & Information Technology*, *41*(3), 629–643. https://doi.org/10.1080/0144929x.2020.1829708

Natural language processing (NLP): What it is and why it matters. SAS. (n.d.). https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, *5*(1). https://doi.org/10.1186/s41235-020-00252-3

McGraw, M., & Stein, S. (2021, April 23). *It's been exactly one year since Trump suggested injecting bleach. We've never been the same.* Politico . https://www.politico.com/news/2021/04/23/trump-bleach-one-year-484399

Pakgohar, A., Habibirad, A., & Yousefzadeh, F. (2019). Lin–Wong divergence and relations on type I censored data. *Communications in Statistics - Theory and Methods*, *48*(19), 4804–4819. https://doi.org/10.1080/03610926.2018.1494839

Parker, A., Buckley, S., & Dagnall, N. (2009). Reduced misinformation effects following saccadic bilateral eye movements. *Brain and Cognition*, *69*(1), 89–97. https://doi.org/10.1016/j.bandc.2008.05.009

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Rabb, N., Cowen, L., de Ruiter, J. P., & Scheutz, M. (2022). Cognitive cascades: How to model (and potentially counter) the spread of fake news. *PLOS ONE*, *17*(1). https://doi.org/10.1371/journal.pone.0261811

Ross, R. M., Rand, D. G., & Pennycook , G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision Making*, *16*(2), 484–504.

Shahan A. M. & Carley, K. M.. Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset, *In Proceedings of The 5th International Workshop on Mining Actionable Insights from Social Networks* (MAISoN 2020), co-located with CIKM, virtual event due to COVID-19, 2020

TextBlob: Simplified text processing. TextBlob. (n.d.). https://textblob.readthedocs.io/en/dev/

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vraga, E. K., Kim, S. C., Cook, J., & Bode, L. (2020). Testing the effectiveness of correction placement and type on Instagram. *The International Journal of Press/Politics*, *25*(4), 632–652. https://doi.org/10.1177/1940161220919082

Zhao, Y., Zhu, S., Wan, Q., Li, T., Zou, C., Wang, H., & Deng, S. (2022). Understanding how and by whom covid-19 misinformation is spread on social media: Coding and network analyses. *Journal of Medical Internet Research*, *24*(6). https://doi.org/10.2196/37623

Appendix A: LDA Graphs

Figure 1: 2020 relevant tokens, topics, and inherently negative terms

2020 LDA Analysis

LDA	*30 most relevant tokens*	*Thematic Topic*	*Negative terms*
Topic 1 (22.8%)	covid, ***fake***, coronavirus, cure, news, like, *china*, *hydroxychloroquine*, *bleach*, corona, *trump*, day, does, vaccine, use, virus, right, gates, beer, doing, let, light, thing, really, come, thought, *injecting*, *drinking,* doctor, claims	Treatment misinformation pushed or repeated by Trump	1
Topic 2 (21.6%)	covid, people, *immune*, social, *garlic*, cases, media, confirmed, spread, *masks*, distancing, good, said, just, best, **bad**, *mask*, know, stay, symptoms, wear, black, *home*, new, today, boost, *water*, *ginger*, update, rate	Naturalistic treatment remedies for Covid-19	1
Topic 3 (20.6%)	*trump*, *coronavirus, bleach,* covid, *president*, say, did, says, cure, *drink*, treat, want, *hydroxychloroquine,* work, just, way, *country,* sure, ***crazy,*** talking, believe, long, need, **dead**, ***stupid, idiot***, getting, video, feel	Trumpian rhetoric of Covid-19 early in the pandemic	4
Topic 4 (18.8%)	coronavirus, covid, ***bioweapon***, world, ***conspiracy***, got, *chinese*, real, *virus*, just, health, *wuhan*, true, pandemic, **fuck**, *lab*, paper, *evidence*, toilet, clear, *bat*, hard, thanks, new, human, help, ***lockdown***, life, breaking, ***control***	Conspiratorial theories regarding virus origin	5
Topic 5 (16.2%)	coronavirus, think, new, going, testing, great, just, big, ***stop***, positive, live, fine, oil, better, tested, available, hospital, today, man, case, does, state, *flu,* **kill**, *immune*, *natural*, america, time, morning, bleach	People who downplayed Covid-19	2

Anything in italics relates to the Topic determined by Thematic Qualitative Analysis
Anything in bold relates to an inherently negative evoking word

Figure 2: 2021 relevant tokens, topics, and inherently negative words

2021 LDA Analysis

LDA	*30 most relevant tokens*	*Thematic Topic*	*Negative Terms*
Topic 1 (22.2%)	covid, ***experimental,*** *vaccines*, don, ***deaths, untested***, really, *fda*, want, getting, long, ***bioweapons***, right, ***death,*** *government,* ***drug***, real, term, *cdc*, children, news, got, designed, approved, time, need, spike, **bioweapon**, higher, chinese	Distrust in US Government Organizations	7
Topic 2 (21.1%)	people, *pfizer*, **experimental**, covid, *vaccinated*, healthy, making, *shots,* ***sick***, pandemic, following, *astrazeneca*, document, come, admits, *plan,* *vaccin* [SIC], unmasked, *dna,* ***extermination***, machine, think, ***dies,*** says, ***depopulation***, don, ***dead***, ***altering***, *mrna,* **bioweapon**	Vaccines as a depopulation tool	8
Topic 3 (20.2%)	gene, therapy, **experimental**, covid, taking, *passports*, *movement,* *licenses*, called, medical, blood, **clots, dangerous**, injection, using, does, mrna, astrazeneca, health, safe, like, shot, used, doctors, moderna, getting, actually, child, **death,** ***stop***	Governmental control	5
Topic 4 (18.8%)	amp, ***experimental,*** people, just, ***dangerous,*** know, *pfizer, mrna,* ***rushed***, say, use, big, *moderna,* media, young, effects, ***killer, untested, bioweapon***, *injected,* did, like, trying, population, *gates*, virus, body, new, really, child	Distrust in vaccine due to it being rushed / experimental / untested	6
Topic 5 (17.8%)	***depopulation, vaccines, deadly,*** covid, virus, new, *jab,* old, people, make, year, shot, going, ***experimental, dead, agenda,*** didn, ***vaers***, *mrna,* second, *pfizer,* ***kill,*** clear, world, ***hesitancy***, like, years, *gates*, ***forced,*** *used*	Forced vaccination agendas / distrust in vaccine	10

Anything in italics relates to the Topic determined by Thematic Qualitative Analysis
Anything in bold relates to an inherently negative evoking word

Appendix B: Data Exploration & Analysis References

Data Cleaning & Exploratory Data Analysis (EDA):
https://colab.research.google.com/drive/1MQtg9gEF7Ji7NixdH5DBL2UCDS82MdK6?usp=sharing

Sentiment Analysis:
https://colab.research.google.com/drive/1Z_F3seTYVcLqDNntpqN0alRZBsEGKnda?usp=sharing

LDAvis:
https://colab.research.google.com/drive/1ZaSGCyycY8vwRBq96N-5iPZkv2jZnXAx?usp=sharing

Keyword Extraction:
https://colab.research.google.com/drive/1VGo7PMn_PFYq7rFgBoRgQEKZgbCn1qMZ?usp=sharing