

Original Paper

# Detecting Tweets Containing Cannabidiol-Related COVID-19 Misinformation Using Transformer Language Models and Warning Letters From Food and Drug Administration: Content Analysis and Identification

Jason Turner<sup>1\*</sup>, MS; Mehmed Kantardzic<sup>1\*</sup>, PhD; Rachel Vickers-Smith<sup>2\*</sup>, MPH, PhD; Andrew G Brown<sup>3\*</sup>, PhD

<sup>1</sup>Data Mining Lab, Department of Computer Science and Engineering, J B Speed School of Engineering, University of Louisville, Louisville, KY, United States

<sup>2</sup>Department of Epidemiology and Environmental Health, College of Public Health, University of Kentucky, Lexington, KY, United States

<sup>3</sup>Department of Criminology and Criminal Justice, Northern Arizona University, Tempe, AZ, United States

\* all authors contributed equally

**Corresponding Author:**

Jason Turner, MS

Data Mining Lab

Department of Computer Science and Engineering

Speed School of Engineering, University of Louisville

Louisville, KY, 40292

United States

Phone: 1 859 302 0189

Email: [jason.turner@louisville.edu](mailto:jason.turner@louisville.edu)

## Abstract

**Background:** COVID-19 has introduced yet another opportunity to web-based sellers of loosely regulated substances, such as cannabidiol (CBD), to promote sales under false pretenses of curing the disease. Therefore, it has become necessary to innovate ways to identify such instances of misinformation.

**Objective:** We sought to identify COVID-19 misinformation as it relates to the sales or promotion of CBD and used transformer-based language models to identify tweets semantically similar to quotes taken from known instances of misinformation. In this case, the known misinformation was the publicly available Warning Letters from Food and Drug Administration (FDA).

**Methods:** We collected tweets using CBD- and COVID-19-related terms. Using a previously trained model, we extracted the tweets indicating commercialization and sales of CBD and annotated those containing COVID-19 misinformation according to the FDA definitions. We encoded the collection of tweets and misinformation quotes into sentence vectors and then calculated the cosine similarity between each quote and each tweet. This allowed us to establish a threshold to identify tweets that were making false claims regarding CBD and COVID-19 while minimizing the instances of false positives.

**Results:** We demonstrated that by using quotes taken from Warning Letters issued by FDA to perpetrators of similar misinformation, we can identify semantically similar tweets that also contain misinformation. This was accomplished by identifying a cosine distance threshold between the sentence vectors of the Warning Letters and tweets.

**Conclusions:** This research shows that commercial CBD or COVID-19 misinformation can potentially be identified and curbed using transformer-based language models and known prior instances of misinformation. Our approach functions without the need for labeled data, potentially reducing the time at which misinformation can be identified. Our approach shows promise in that it is easily adapted to identify other forms of misinformation related to loosely regulated substances.

(*JMIR Infodemiology* 2023;3:e38390) doi: [10.2196/38390](https://doi.org/10.2196/38390)

**KEYWORDS**

transformer; misinformation; deep learning; COVID-19; infodemic; pandemic; language model; health information; social media; Twitter; content analysis; cannabidiol; sentence vector; infodemiology

## Introduction

### Background

The Food and Drug Administration (FDA) defines medical “misinformation” as the information that misleadingly represents a product as able to mitigate, prevent, treat, diagnose, or cure a condition or disease, such as COVID-19 [1]. Misinformation is prevalent on social networking platforms and is often seen in product advertising. Since early 2020, the COVID-19 pandemic has offered a new medical condition for advertisers and sellers to exploit, who seek to profit from the crisis at the expense of public health [2].

Although some misleading posts regarding COVID-19 have been about the virus’s origins or the effectiveness and safety of masks and vaccines [3-5], others have focused on false information about alternative products in treating or preventing COVID-19. For example, since the onset of the pandemic, some cannabidiol (CBD) sellers have claimed that CBD can prevent

or cure COVID-19 (Figure 1). Preclinical studies have suggested that CBD could be effective in treating respiratory conditions and might confer cardioprotective, nephroprotective, hepatoprotective, neuroprotective, and anticonvulsant benefits [6]. Meanwhile, other evidence suggests that CBD could decrease one’s ability to fight infections, countering its potential clinical use as an anti-inflammatory agent [7]. Because CBD has not been thoroughly tested for efficacy and safety in treating COVID-19, its benefits are mostly considered unsubstantiated [8]. However, there is plentiful web-based marketing for CBD, touting these unverified benefits. This is not unique to CBD and COVID-19 and occurs with other loosely regulated substances (eg, kratom and herbal supplements) as well as other medical conditions such as Alzheimer disease, cancer, and autism. Although the consumption of CBD is typically well tolerated, misinformation about its effectiveness in treating COVID-19 has become widespread and poses a danger to public health [9,10]. Therefore, it is essential that this type of content be efficiently identified so that its potential harmful effects on public health can be minimized.

**Figure 1.** Example of a tweet promoting cannabidiol (CBD) as a prevention or treatment for COVID-19.



Misinformation has been shown to spread faster and further than accurate information on social media, and Twitter serves as an example of a social media site where misinformation about smoking products, drugs, vaccines, and diseases is abundant [11,12]. Specifically regarding COVID-19, for example, in March 2020, a daily average of 46,000 misleading or inaccurate news posts appeared on Twitter in Italy alone [13]. In Iran, a rumor originating on social networks claimed that ingesting neat alcohol could cure COVID-19, resulting in hundreds of alcohol poisoning deaths [14]. Furthermore, influential Twitter users such as former President Donald J Trump and celebrity Joe Rogan suggested taking the antimalarial medications chloroquine or hydroxychloroquine to treat COVID-19, even

though there has never been scientific evidence to support these claims [15-17]. Although the United States FDA warned against improper consumption of these substances in July 2020, there were still dozens of documented deaths and poisonings associated with their use, including at least one person who ingested fish tank cleaner containing chloroquine after being influenced by misinformation on Twitter [18-20]. Not only has misinformation dissuaded consumers from seeking effective treatments but also it has encouraged the use of dangerous and unfounded *treatments*.

Prior research has shown some success in using supervised and unsupervised machine learning techniques to detect and explore COVID-19 misinformation on social media platforms such as

Twitter [21–24] and YouTube [25]. Although some studies used an annotated data approach [21,24,26], it was noted that the rigor and costs of this process could preclude its widespread utility [21]. One study used an unsupervised topic model to examine inaccurate information about vaping CBD and COVID-19 [27], which is a more cost-effective technique than annotation, but is a less explicit way to identify misinformation than supervised approaches. Deep learning [23,28] and transformer language models [22,23,25,29] are the most commonly used techniques, likely because of their efficiency and highly advanced ability to interpret and understand natural language, which is key to examining content on social media.

This study leverages the expertise of the FDA, the regulatory body of the safety and efficacy of numerous products intended for human or animal use in the United States, to define misinformation regarding CBD and COVID-19 using Warning Letters as a gold standard. When the FDA becomes aware that a company has violated FDA regulations, they often issue Warning Letters to the company that outline the nature of the violation (eg, problems with claims about a product and incorrect directions for use), corrective action, and a timeframe [30]. The FDA then follows up to verify whether the company has completed the corrective action, and if not, the FDA may enact regulatory actions such as seizure or civil penalties [31]. These Warning Letters are made available to the public on the FDA's website [32]. By using verbatim quotes taken from relevant FDA Warning Letters and transformer language models, we propose an efficient and consistent approach to identify tweets that contain false claims regarding CBD and COVID-19.

## Objectives

This study has two primary objectives: (1) to examine misinformation concerning CBD and COVID-19 disseminated by Twitter advertisers and (2) to propose a framework that helps identify CBD- and COVID-19-related misinformation in tweets more efficiently and can also be easily modified to detect misinformation in advertisements about other substances and conditions.

Traditional text classification requires labeled data for model training and testing, which can be both time consuming and expensive. Transformer-based language models have shown promising results in various semantic textual similarity tasks, including in the medical domain [33].

## Methods

### Ethics Approval

This study leverages publicly available data and was registered as approved by the University of Louisville Institutional Review Board (approval protocol 20.1122).

### Collecting and Annotating Tweets

We used the Snsrape Python package to collect English language tweets from the Twitter website from January 1, 2020, to April 28, 2021, by searching for tweets using the following keywords: *CBD or cannabidiol*, and *COVID-19, COVID, corona* [34]. Although this method does not provide full access to Twitter's past data, it does afford the ability to collect thousands of historical tweets retrospectively after an unexpected event has occurred, such as the pandemic. Using this method of historical tweet searching, we were able to collect 37,526 tweets over a 484-day period.

To extract the commercial CBD tweets referencing COVID-19, we used a model previously developed by our group [35]. The model was trained on an earlier collection of CBD tweets to identify those that reflected commercial sales, promotion, and marketing of CBD. Applying the commercial CBD tweet classifier to the historical tweet collection resulted in 4937 tweets that were classified as commercial CBD tweets referencing COVID-19. We evaluated the performance of this model that identified commercial CBD tweets from the CBD or COVID-19 collection of tweets by annotating a random sample of classified 250 commercial CBD or COVID-19 tweets and 250 noncommercial CBD or COVID-19 tweets. We observed an improvement in the commercial tweet classification performance, with precision, recall, accuracy, and  $F_1$ -scores of 0.95 (Table 1).

**Table 1.** Performance of 2019 commercial cannabidiol (CBD) classifier in differentiating tweets from 2020 to 2021 (N=500).<sup>a</sup>

	Precision	Recall	$F_1$ -score	Random sample
Noncommercial CBD	0.95	0.95	0.95	249
Commercial CBD	0.95	0.95	0.95	251

<sup>a</sup>Accuracy of the model: 0.95.

The United States FDA has warned CBD advertisers and sellers about promoting CBD as a treatment or cure for several conditions, including COVID-19. These Warning Letters inform the advertiser that CBD is not an approved treatment or prevention for the condition that the advertiser mentions, warn the advertiser that further action will be taken if this style of misleading advertising continues, and include the exact misleading quote taken from the advertisement. This list of letters continues to grow as the FDA becomes aware of more advertisers making false claims about the medical benefits of

CBD. We drew directly from the FDA Warning Letters about CBD and COVID-19 to understand the types of statements that the FDA had flagged as misinformation.

We used the statements from these FDA Warning Letters to identify and label the CBD tweets containing COVID-19-related misinformation. Instead of using our annotated set of tweets to train and test a tweet classification model, we used them to establish the relationship of text similarity between the misinformation tweets and the statements in the FDA Warning Letters via the transformer language model. The transformer

architecture was introduced to address some of the shortcomings of the recurrent neural networks in tasks such as language translation [36].

Because the FDA’s CBD or COVID-19 Warning Letters did not include CBD-infused hand sanitizers, we included the Warning Letters sent to advertisers of nonalcohol and essential oil sanitizer products that had made false claims as guidance in our misinformation annotation process [37-43]. Figure 2 is an example of one of the Warning Letters issued by the FDA about misinformation surrounding the use of CBD for treating COVID-19.

The extracted commercial CBD or COVID-19 tweets were annotated for misinformation (yes or no), according to the FDA’s definition of misinformation, by 3 university-trained, nonmedical, professional annotators; discrepancies were decided by a majority vote to determine the overall label of the tweet.

Before labeling the tweets, we reviewed several FDA Warning Letters with the annotators so that the annotators were familiar with what the FDA considered a misleading statement. Because tweets are relatively short in length, if a tweet contained any misinformation related to CBD and COVID-19, the entire tweet was considered positive for misinformation. There were no noticeable discrepancies among the annotators, and they had an intercoder agreement of 91%.

The Warning Letters on the FDA website provided some of the quotes that sellers used in their advertisements that the FDA deemed misleading. Some examples of these letters are listed in Textbox 1. Along with the FDA’s definition of misinformation, these quotes provided guidelines for annotating the collection of tweets as to whether they contained misinformation. Textbox 2 displays selected examples of the nonmisinformation and misinformation tweets encountered during the annotation process.

Figure 2. Example Warning Letter taken from the Food and Drug Administration (FDA) website.

The image shows a screenshot of an FDA Warning Letter for 'For Our Vets LLC dba Patriot Supreme', dated October 16, 2020. The letter is addressed to Justin Elenburg at 1910 S. Stapley Dr. #221, Mesa, AZ 85204. The issuing office is the Center for Drug Evaluation and Research (CDER), United States, Federal Trade Commission. The letter discusses unapproved and misbranded products related to COVID-19. To the right of the letter, several red boxes highlight specific claims from the company's website that the FDA found misleading. These include: 'CBD AND COVID-19 - CAN CBD HELP COVID-19 LUNG INFLAMMATION?', 'CBD May Play a Role in Helping Lung Symptoms', 'Using CBD to Alleviate Inflammation', 'Acute respiratory distress syndrome (ARDS) is a type of respiratory failure characterized by rapid onset of widespread inflammation in the lungs, rapid breathing and the inability to sustain adequate oxygen levels to the body and brain. Shortness of breath or difficulty breathing are some of the early signs of COVID-19...', 'CBD and Covid-19 / Why Do Researchers Believe CBD Can Help?', 'CBD INFLAMMATION COVID / A PLAUSIBLE METHOD?', 'CBD Inflammation Covid', 'How is CBD Inflammation Covid a plausible method?', 'ARDS, which is "Acute Respiratory Distress Syndrome" is one of the severe symptoms of the covid-19. In this syndrome, the corona virus causes the inflammation of the lungs, which in turn may lead to severe difficulty in breathing, pain or perhaps even coughing.', and 'How is Inflammation cured?'. The highlighted text includes various claims about CBD's effectiveness in treating lung inflammation, its role in helping lung symptoms, its use in alleviating inflammation, and its potential to cure COVID-19 related lung conditions.

**Textbox 1.** Five example quotes from Warning Letters issued by Food and Drug Administration (FDA) related to cannabidiol (CBD) and COVID-19 misinformation.

- “Firstly, the research performed to date has shown that CBD can reduce a number of pro-inflammatory cytokines (numerous different types of substances, such as interferon, interleukin, and growth factors, which are secreted by certain cells of the immune system and have an effect on other cells) including IL-6, the one reduced by other drugs being studied for COVID-19. CBD was also shown to reduce interleukin (IL)-2, IL-1 $\alpha$  and  $\beta$ , interferon gamma, inducible protein-10, monocyte chemoattractant protein-1, macrophage inflammatory protein-1 $\alpha$ , and tumor necrosis factor- $\alpha$  – all of which are associated with the pathology of severe cases of COVID-19. In addition to reducing these pro-inflammatory cytokines, CBD has also been shown to increase the production of interferons, a type of signaling protein that activates immune cells and prevents viruses from replicating”
- “There has been an increased interest in CBD and Covid-19 to treat lung problems and symptoms (mental or physical) associated with the coronavirus”
- “CBD oil may help to prevent getting infected by strengthening your immune system. It has also been proven to offer relief to some of the symptoms”
- “By using CBD oil, you can keep inflammation at bay, retain a healthy or even higher than average white blood cell count, stay calm and relaxed (which is best for a strong immune system), and prevent catching a virus or infection beforehand”
- “Is CBD an Anti-Viral Agent for Coronavirus, Influenza, MERS, and Sars Plus Key Antiviral Supplements?”

**Textbox 2.** Examples of annotated commercial cannabidiol (CBD) or COVID-19 tweets.

- Misinformation
  - “#CBD is readily available for anyone who want to build-up their immune to help guard against the #coronavirus. It’s your responsibility to protect & take care of yourself, not the government. Order Now”
  - “I’ve ordered a 4 month supply of #CBD to help fortify my immune system & guard against the #CoronaVirusUpdates #COVID19. What have you done to protect yourself? Order Now”
  - “#COVID19 attacks the inside of body/lungs which are internal so topical solutions will not mitigate what's happening inside your body/lungs. Ask me about #CBD. #cbdoil <MASKED-URL>”
  - “#CBDL Could Double On Product Launch News. CBD Hand Sanitizer Could Help Stop Spread Of Covid-19. [Read Now] LINK #USA #Stocks #Bonds #Equities #Gold #Silver #Bitcoin #Cryptocurrency #Investing #Trading #Options #1Author #USStocks #StockMarket”
- Nonmisinformation
  - “Could CBD offer treatment options for COVID-19? Read more at <MASKED-URL> Global Go does not endorse the use of any product for medicinal purposes. Please consult with a physician before using any such products. #CBDtrials #hempresearch #hempnews #covid19”
  - “First day of the week... First day of the month of June! Would you like to try something new? #Covid19 #SanFrancisco #SanFranciscobayArea” #helpingthecommunity #Realestate #HartFordproperties #CBD Source:
  - “Online sales for cannabidiol (#CBD) products continue to #flourish despite in-store slowdowns amid the COVID-19 #pandemic. #CSPDailyNews”
  - “New post (Iowa Down To One Medical CBD Manufacturer Due To COVID-19 Pandemic) has been published on Buy Premium CBD and CBG Products | 100% Natural Cannabinol Store | Buy CBD Oils, Gummies, Topicals, Pet CBD and more.”

## Misinformation Search

The transformer follows an encoder-decoder structure wherein the encoder converts the text input into a vector representation. There is one vector per word in the sequence; the value of each vector representation of each word is partially based on the nearby terms surrounding the word, which provides the context. The decoder portion of the transformer architecture is similar to the encoder but can convert a vector into a sequence of text. The original Bidirectional Encoder Representations from Transformers (BERT) model was introduced by Devlin et al [44] at Google in 2018. The BERT model is trained on the natural language processing tasks of masked language modeling on English Wikipedia and the Brown Corpus texts and it predicts the missing word in a sentence and performs next-sentence prediction. Masked language modeling allows the BERT model to understand the relationships between words, whereas

next-sentence prediction allows the BERT model to understand the relationships between sentences.

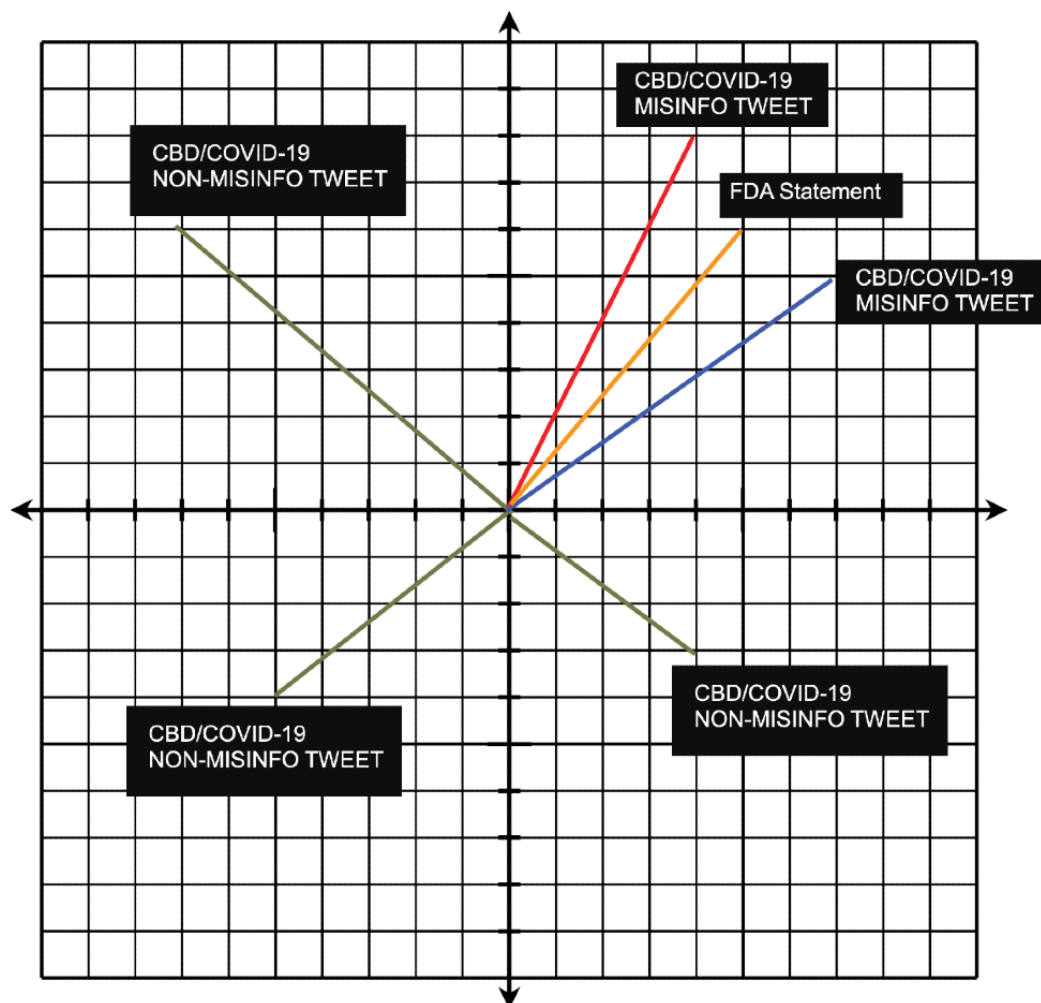
In this study, we used a transformer language model to encode each of the commercial CBD tweets that we collected. Specifically, we used the Sentence-T5 model because it is a state-of-the-art language model that has outperformed other models in semantic textual similarity [45]. We also computed the encoding for each statement containing misinformation taken from the FDA website into vectors of size 768. We then calculated the cosine similarity (Equation 1) between the encoding vectors of tweets (A) and the encoding vectors of the misinformation quotes taken from the FDA Warning Letters (B).

$$\text{cosine similarity} = S_c(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Figure 3 provides a theoretical illustration of how we isolated tweets that made false claims using the quotes taken from the FDA Warning Letters to find contextual similarities using cosine similarity. This example shows that when the tweets and quotes from Warning Letters are encoded into vectors, tweets containing CBD or COVID-19 misinformation will be closer in cosine distance to the FDA quotes than tweets without similar false claims about CBD and COVID-19. Using cosine similarity as the distance of similarity, we expected that the shortest cosine distances would contain more misinformation, that is, tweets that were contextually similar to the FDA's samples.

Conversely, we expected the tweets with the longest cosine distances to contain less misinformation. Using these points, we identified a threshold at which we could confidently identify sets of tweets that mostly contained misinformation because of their semantic similarity to an established example of misinformation. The optimal threshold should be the point at which the maximum number of tweets deemed as misinformation is captured while minimizing the number of false positives (ie, tweets that do not contain misinformation) being captured.

**Figure 3.** Representation of contextual similarity between cannabidiol (CBD) tweets and quotes from Food and Drug Administration (FDA) Warning Letters.



## Results

### Analysis of Misinformation

After annotating the tweets for misinformation, we observed that approximately 19% (938/4937) of the tweets contained misinformation related to both CBD and COVID-19. Table 2 shows some of the terms most frequently associated with CBD

or COVID-19 misinformation. Aside from CBD-related terms, terms such as “help,” “boost,” “health,” “virus,” “sanitizer,” and “immune system” were among the most frequently occurring terms. We observed an increase in CBD or COVID-19 conversations beginning in February 2020. Misinformation related to CBD and COVID-19 peaked in March 2020; although it appeared to taper down, it did not stop.

**Table 2.** Top n-grams in the cannabidiol or COVID-19 misinformation tweets.

Term	Frequency
immune	257
cbdoil	244
system	187
immune system	181
products	157
help	152
hemp	139
health	129
virus	127
new	127
cbd cbdoil	125
natural	119
wellness	114
boost	112
oil	99
hand	96
sanitizer	92
cbdproducts	88
cbd products	85
hand sanitizer	83
use	81
immunity	78

### Detecting Misinformation

Using 27 misleading quotes taken from the FDA Warning Letters and converting them into vector form and then converting each tweet into vector form, we calculated the cosine similarity. Using the cosine distance, we then counted the

number of tweets that were labeled as misinformation compared with the tweets that were considered nonmisinformation. We observed that the nearest tweets indeed contained misinformation, whereas the most distant tweets did not (Table 3).

**Table 3.** Measuring the cosine distance between sentence vectors of statement 0 and the tweets.

Tweet	Position	Misinformation		Cosine distance
		Yes	No	
“With the growing concern of the COVID-19 virus we understand the importance of boosting the immune system. CBD <sup>a</sup> is a natural way to do that. The Healing Leaf wants to help make CBD more available and lessen costs for those interested. Please come in or call and place your orders!” <MASKED-URL>	Most similar	✓		0.069577
“CBD may reduce cytokine storm and inflammation in COVID-19\n” <MASKED-URL>	Second most similar	✓		0.077852
“They snuck a shipping ban on vape products into the last covid relief bill. All vape products. CBD, nicotine, delta 8, doesn’t matter.”	Second most distant		✓	0.296933
“Final point: When COVID hit, we LOWERED our prices and never put them back up. Funny how reddit and Review haters never use their real names and 9/10 when you find out their real name - they run for the hills or perhaps just back to Mam’s house and their keyboard...”	Most distant		✓	0.297551

<sup>a</sup>CBD: cannabidiol.

We calculated the cosine distance between the sentence vector of each of the 27 statements that we extracted from the FDA Warning Letters, along with each of the tweets, so that we could determine a cosine distance threshold in which we collected the most tweets that contained CBD or COVID-19 misinformation while minimizing false positives (non-CBD or COVID-19 misinformation). Figures 4 and 5 show a consistent observation that as the cosine distance increased, the percentage (recall) and frequency of the tweets, respectively, containing misinformation increased. Figure 6 indicates that as the cosine distance increased, the precision of the tweets containing misinformation decreased. However, if the cosine distance was too small, few to no misinformation tweets were captured, and not all FDA

statements performed equally in capturing misinformation tweets.

We also examined the top 5 FDA statements that captured large amounts of misinformation with high precision (Textbox 3). Figures 7-9 are equivalent to Figures 4-6 but include only the statements displayed in Textbox 3. From these figures, we can see that at a cosine distance between 0.10 and 0.13, we were able to capture between 21.9% (206/938) and 28.9% (272/938) of the misinformation tweets with a precision of above 80%. Specifically, statement 8 was able to capture 34.4% (323/938) of tweets at a cosine distance of 0.105, and 84.8% (274/938) of these tweets were labeled as misinformation. The 274 captured tweets represented 29.2% of the 938 misinformation in our data set.

Figure 4. Cosine distance versus proportion of misinformation tweets captured.

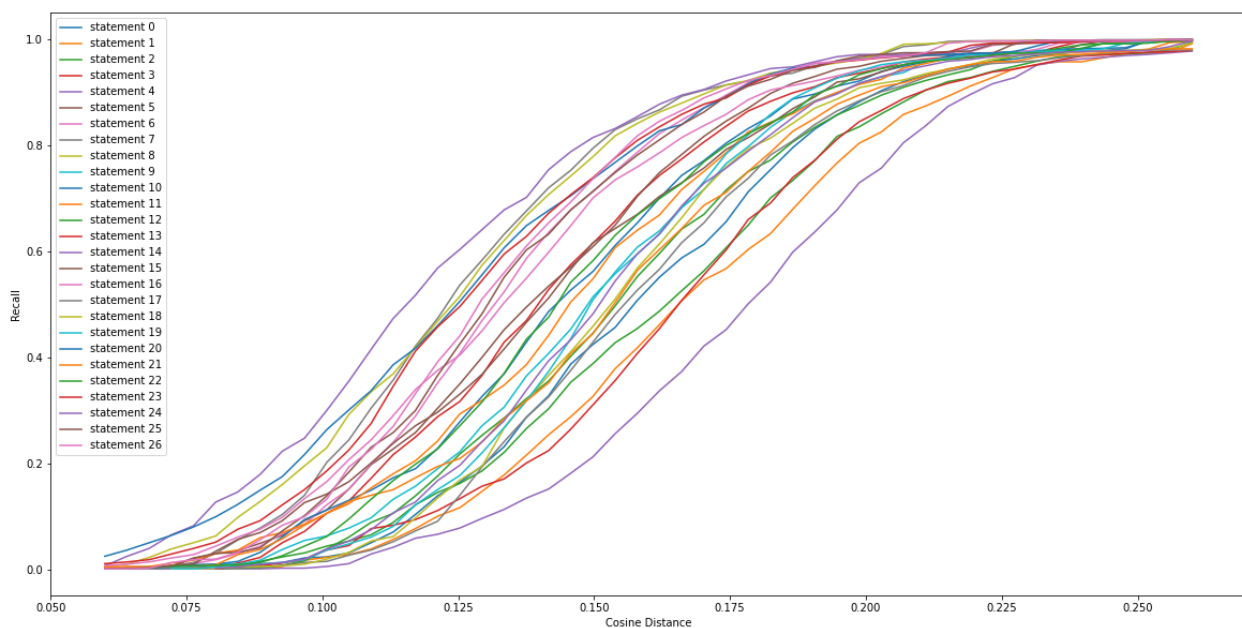
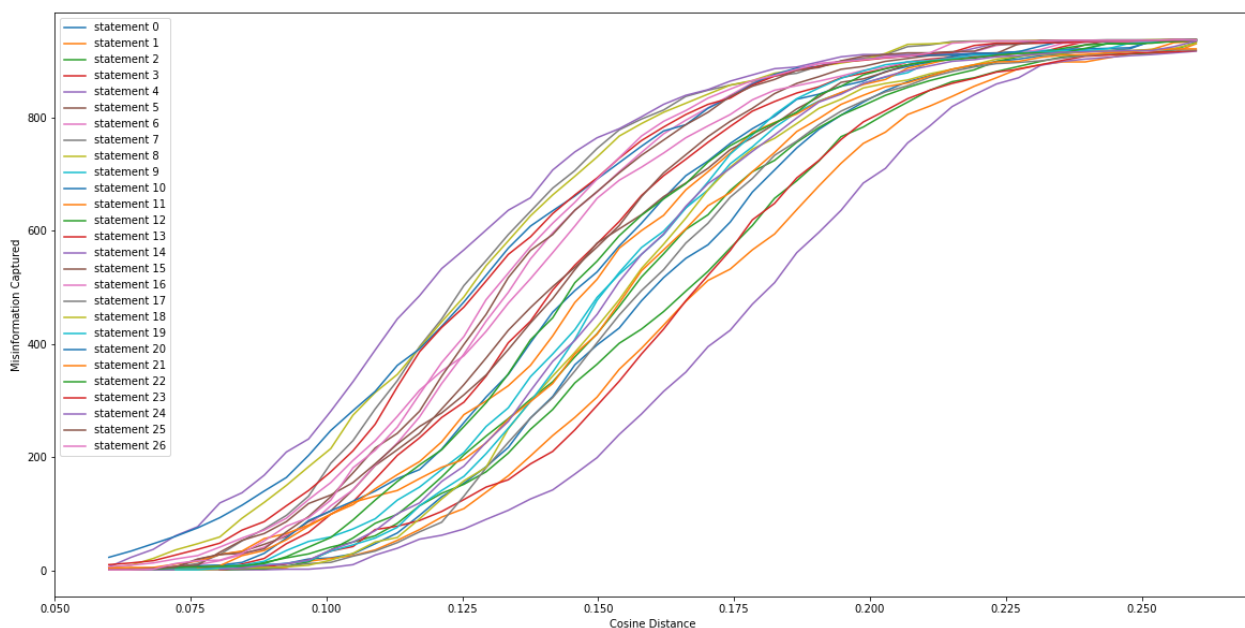
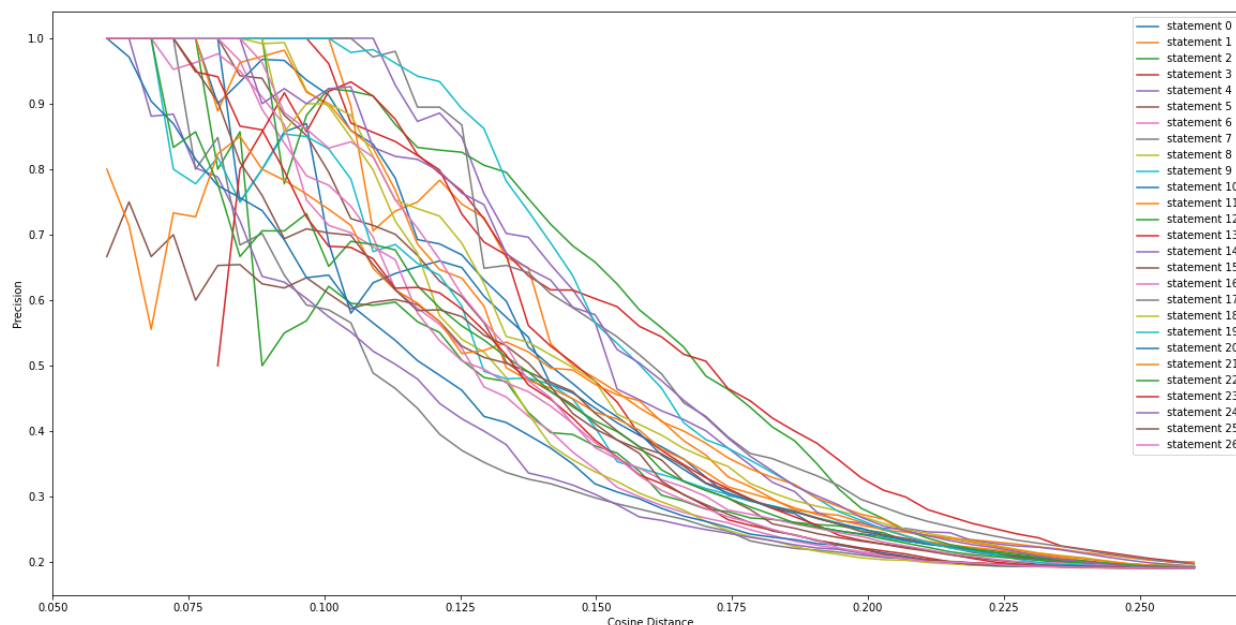


Figure 5. Cosine distance versus number of misinformation tweets captured.

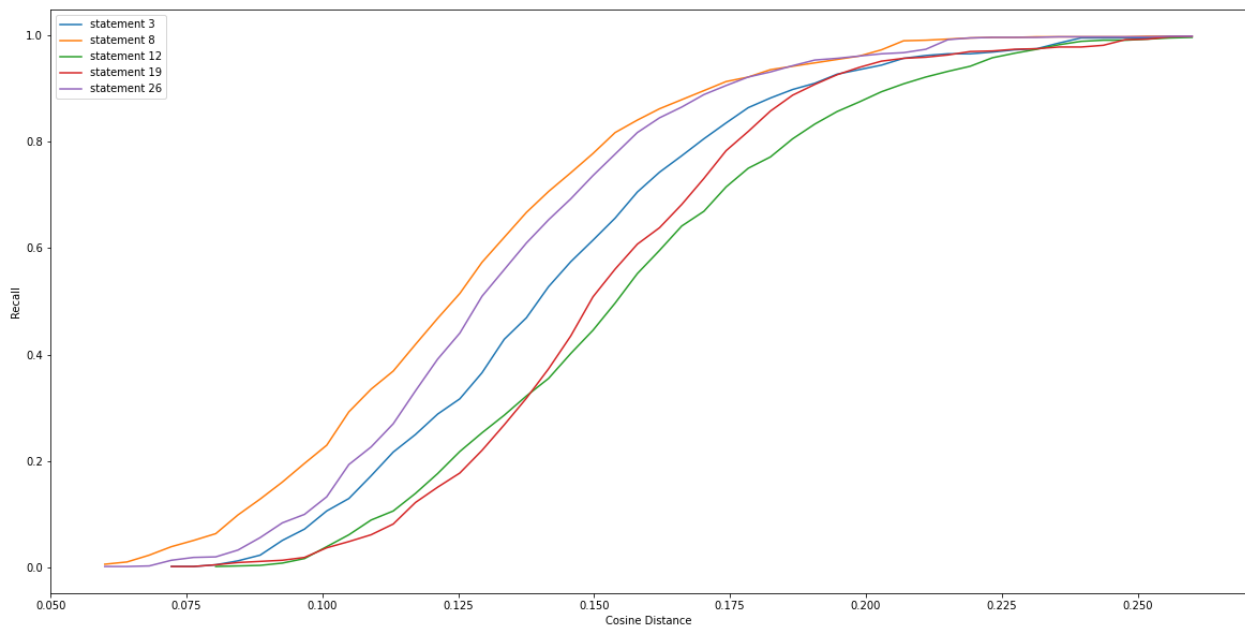




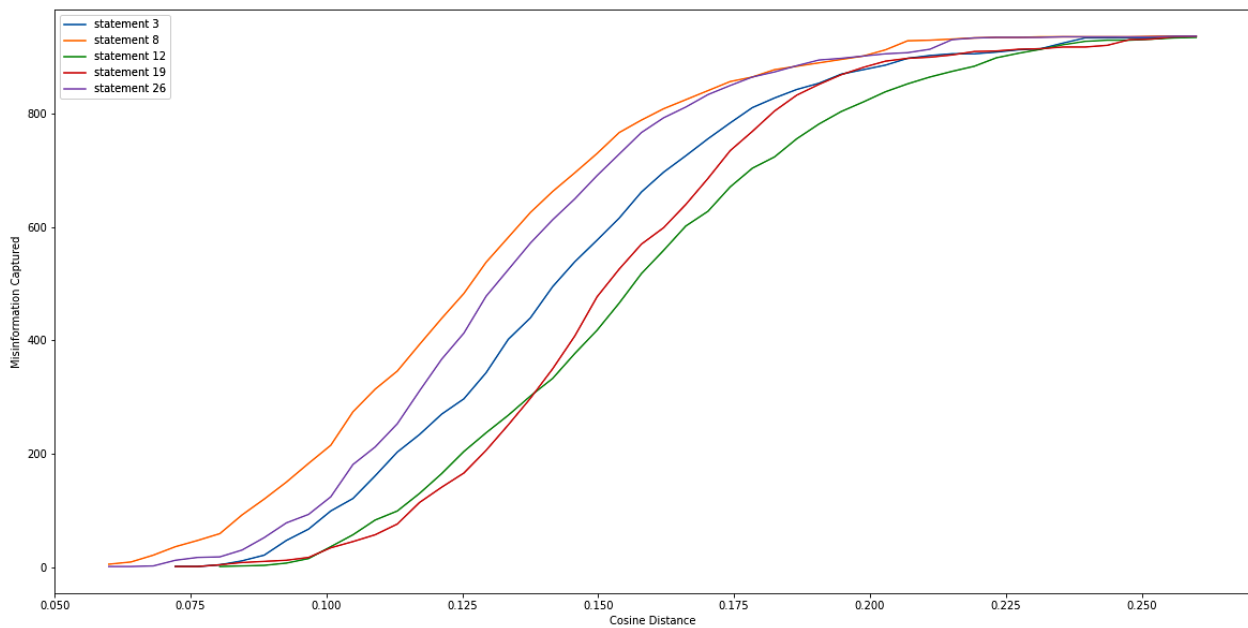
**Figure 6.** Cosine distance versus misinformation precision of tweets captured.**Textbox 3.** Food and Drug Administration (FDA) statements that captured the largest amount of misinformation with the highest amount of precision.

- Statement 3
  - “Firstly, the research performed to date has shown that CBD can reduce a number of pro-inflammatory cytokines (numerous different types of substances, such as interferon, interleukin, and growth factors, which are secreted by certain cells of the immune system and have an effect on other cells) including IL-6, the one reduced by other drugs being studied for COVID-19. CBD was also shown to reduce interleukin (IL)-2, IL-1 $\alpha$  and  $\beta$ , interferon gamma, inducible protein-10, monocyte chemoattractant protein-1, macrophage inflammatory protein-1 $\alpha$ , and tumor necrosis factor- $\alpha$  – all of which are associated with the pathology of severe cases of COVID-19. In addition to reducing these pro-inflammatory cytokines, CBD has also been shown to increase the production of interferons, a type of signaling protein that activates immune cells and prevents viruses from replicating”
- Statement 8
  - “DML CBD: Immune Boost Pack... ALERT: There is no cure or treatment for COVID19. With this in mind, many doctors claim the best defense is to boost the body’s immune system. DML CBD aims to help our customers in an attempt to boost the immune system... WHY TO BUY THE BOOST PACK: Studies suggest that CBD can help fight off inflammation, boost the immune system, and help battle against certain harmful bacteria. Some research suggests it can help suppress the cytokine storm inside the body that can cause great illness and sometimes death... NOTE: The cytokine storm is often triggered in patients with COVID19. Please note there is no proven cure or treatment for COVID19... There has never been a more important time than to boost your immune system. To help our customers get a full CBD experience that aims to boost your immune system, we offer the ‘DML CBD Immune Boost’ package...”
- Statement 12
  - “What is COVID-19? Coronavirus is referred to as a novel cause for viral pneumonia because it’s a virus we haven’t seen before and have developed no immunity to... What Happens If You Get Infected and What Can Help? ...Regardless of the shape you’re in at this moment, there may be ways you can prepare and protect your body from developing a more severe response to infection. Explore the solutions included in NoronaPak below! [graphic with the following text] ‘Selenium, Cannabidiol (CBD), Vitamin-C, Zinc, Vitamin-D, N-Acetylcysteine’... Supplementation with selenium results in changes in the gene expression that is required for protein biosynthesis in lymphocytes, the infection-fighting cells that are crucial to the immune system being able to identify infection and mount an immune response... Selenium is not only important in boosting the immunity of the individual but also to slow the development of more virulent strains of some viral pathogens... CBD may suppress the productions of cytokines in the setting of infection”
- Statement 12
  - “In the wake of the current epidemic, it is now more important than ever to keep your immune system as healthy as can be... Here are 5 key ways to strengthen your immune system during the outbreak...Take supplements such as CBD”
- Statement 26
  - “Crush Corona... While scientists around the world are working 24/7 to develop a COVID-19 vaccine, it will take many more months of testing before it’s approved and available. However, there’s something you can do right now to strengthen your immune system. Take CBD... CBD can help keep your immune system at the stop of its game... We want everyone to take CBD and take advantage of its potential to help prepare your body to fight a coronavirus infection. So, we’re making all of our products more affordable”

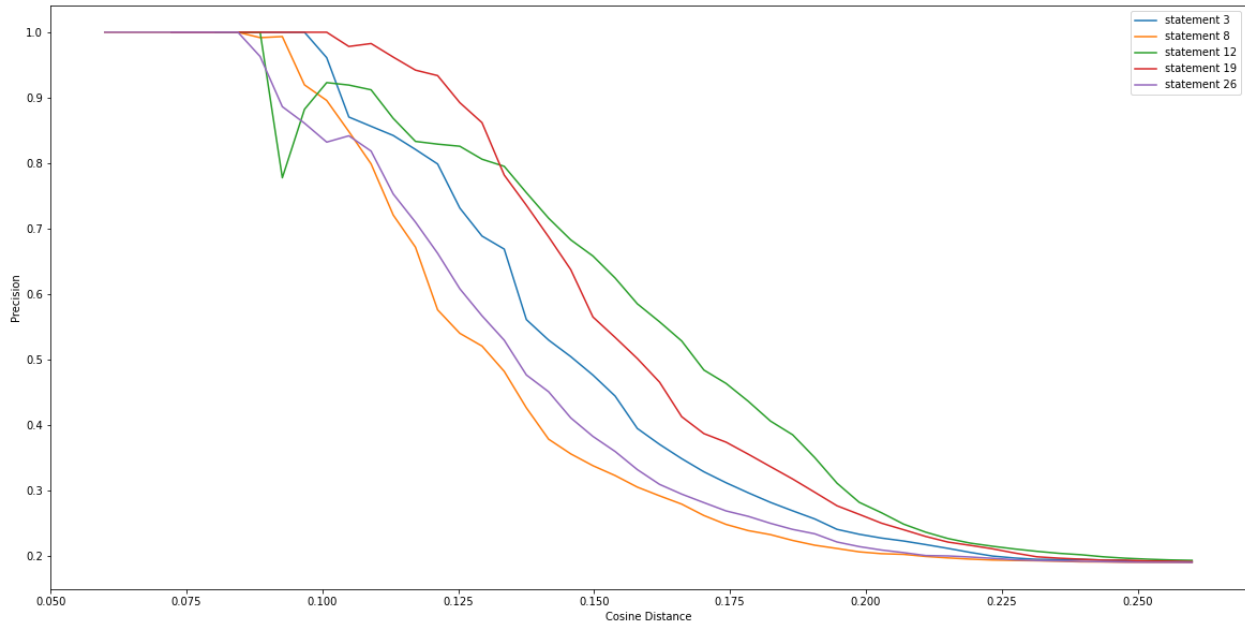
**Figure 7.** Cosine distance versus percentage of misinformation tweets captured by top 5 performing Food and Drug Administration (FDA) statements, which captured large amount of misinformation with high precision.



**Figure 8.** Cosine distance versus number of misinformation tweets captured by top 5 performing Food and Drug Administration (FDA) statements, which captured large amount of misinformation with high precision.



**Figure 9.** Cosine distance versus misinformation precision of tweets captured.



**Application to Other Contexts**

To further demonstrate the flexibility of our methods, we applied it to a corpus of tweets collected in 2019 using only the terms “CBD” and “cannabidiol,” and misleading quotes from FDA Warning Letters regarding autism and Alzheimer disease (Figure

10). As shown in Table 4, we observed that the tweets most similar to the misinformation samples suggested that CBD could alleviate the symptoms of autism and Alzheimer disease. The most distant tweets did not make false claims about CBD’s ability to treat those conditions.

**Figure 10.** Warning Letter taken from Food and Drug Administration (FDA) website regarding cannabidiol (CBD) as a treat for teething, autism, attention deficit hyperactivity disorder (ADHD), and Alzheimer disease.

**Table 4.** Most similar and most distant (in cosine similarity) tweets to misinformation quotes regarding Alzheimer disease.

FDA <sup>a</sup> misinformation quote	Most similar tweet	Most distant tweet
“CBD <sup>b</sup> oil may have neuroprotective properties and may protect against neurological conditions, such as Parkinson’s and Alzheimer’s disease”	@<MASKED-USER> @<MASKED-USER> @<MASKED-USER> @<MASKED-USER> @<MASKED-USER> “On the up side, some evidence suggests THC <sup>c</sup> and CBD may be neuroprotective, so there’s a rationale for some MMJ for you. Alzheimer’s prophylaxis”	@<MASKED-USER> “do not believe he was the co-owner and while the store is definitely shitty, the sign did not say CBD cures autism”
“Possible uses for CBD include helping with skin problems such as acne, autism, ADHD <sup>d</sup> , and even cancer. It’s often used in conjunction with traditional treatments to provide extra help. Children can use high amounts of CBD safely and without any risk.”	“CBD is recommended as a treatment for conditions such as seizures, depression and anxiety, and symptoms such as sleeplessness, inflammation, acne, and pain. It has also proven to be effective in treating autistic children. \nSource: <MASKED-URL> <MASKED-URL>”	“4) Sydney is demonstrating that she understands what effectual customer demands are, which is going straight to the source to demand change. She did not see a phone number she could call on the sign that promoted CBD as a cure-all for autism, so she went in to ask for one”

<sup>a</sup>FDA: Food and Drug Administration.

<sup>b</sup>CBD: cannabidiol.

<sup>c</sup>THC: tetrahydrocannabinol.

<sup>d</sup>ADHD: attention deficit hyperactivity disorder.

## Discussion

### Principal Findings

COVID-19–related misinformation can have fatal consequences [46]. Those consuming this misinformation may have misconceptions about how the virus is transmitted, disease symptoms, or health effects; they may communicate misinformation to others who subsequently spread it and put themselves and others at risk. Although there have been some preliminary studies on the benefits of cannabis in treating the symptoms of COVID-19, findings are uncertain, and at the time of this writing, cannabis is not an approved treatment by the FDA [47]. Therefore, claiming that CBD products can unequivocally treat or prevent COVID-19 is a federal violation in the United States, where the FDA has made numerous attempts to warn web-based retailers making false claims about their products [1]. Our study demonstrates how FDA Warning Letters can be used with transformer language models to identify tweets containing misinformation that are semantically similar to Warning Letters. Our approach reduces the time, labor, and potential monetary costs of other text classification methods. To our knowledge, this is the first study to use FDA Warning Letters as a foundation for identifying misinformation, particularly as it relates to CBD and COVID-19, in web-based social networks.

Transformer language models, such as the one used herein, are powerful tools that have been used for a number of purposes, such as translation [48,49] and text classification [50,51], and have even been used to generate descriptions of images in text [52]. Because of their ability to understand and summarize natural language, several studies have used these models to identify web-based misinformation, which often includes nuanced language and requires techniques that recognize not only semantic but also contextual similarities. Kumar et al [22] used Twitter to build a multilabel tweet classifier system in their study using a RoBERTa-large transformer language model to identify COVID-19–related misinformation. Their model could identify 4 classes: irrelevant, conspiracy, true information, and

false information, and it achieved an  $F_1$ -score of 76%. Although this study focused on any class of misinformation that was identified in the FDA Warning Letters, future studies could further classify the letters by type and evaluate the accuracy of our approach.

Although this work was built on only Twitter platform, it is possible that it could also be applied to other social network platforms. Serrano et al [25] used transformer-based language models to identify YouTube videos containing COVID-19–related misinformation via comments posted on the videos. They built a text classifier to identify conspiracy-related content and concluded that YouTube videos containing misinformation were accompanied by user comments with a high percentage of conspiracy-related content [25]. Future studies can assess the performance of our method on other platforms.

In a recent approach to concept drift (changes in data and meaning over time) in Twitter data streams, Bechini et al [53] trained a semantic-based classifier using the BERT language model to examine the change in opinions about vaccines within a corpus of Italian tweets; this model outperformed other strategies, such as retraining the ensemble approaches. This suggests that transformer-based models, such as the one described herein, for identifying commercial tweets can be resilient over time. In addition, given more extensive (eg, “firehose” access) and future access to Twitter data, our misinformation tweet classifier could identify newer and current tweets that were not included in our data set, as well as identify tweets making similar violations in near real time. Furthermore, as previously noted, this technology can be applied to other forms of misinformation that threaten public health and safety.

Our approach to identifying tweets that make false claims about CBD and COVID-19 used quotes extracted from FDA Warning Letters to identify tweets that are semantically and contextually similar based on the cosine distance of sentence vectors. Compared with the approaches that require a large amount of data annotation, this substantially reduces the time required to identify the tweets making false health-related claims and flag

them with a high amount of confidence. This is attributed to 2 factors: we used minimal manually annotated data for validation purposes and we used a simple calculation of cosine similarity between tweets and quotes.

Our study not only illustrates the scope of misleading information about CBD and COVID-19 in particular but also demonstrates an efficient and affordable approach to identifying other instances of this widespread problem—an approach that can be used by government entities, social networks, and message board administrators concerned with minimizing false advertising and misinformation and the potential threats they pose.

### Limitations

This study had several limitations. First, it was built only on the Twitter social networking platform. In addition, although we did not explicitly acknowledge “bots” on Twitter, this was implicitly addressed during the annotation process where the tweets that appeared to be machine-generated were not considered cases of commercial CBD or misinformation. In addition, this model was trained on a collection of CBD tweets from before the COVID-19 pandemic, that is, before words like “COVID,” “COVID-19,” “corona,” and “coronavirus” would have been associated with “CBD” and “cannabidiol.” In this case, we applied a CBD commercial tweet classifier that was

trained on tweets that were primarily authored in 2019 to a collection of tweets between 2020 and 2021. Although we did observe satisfactory results in testing this model’s extraction of commercial CBD tweets that also mentioned COVID-19, we acknowledge that concept drift is always a potential factor in classifying streaming and should always be considered.

### Conclusions

There is a clear and pronounced advertising presence on Twitter of loosely regulated substances touted to treat COVID-19, although this type of self-treatment lacks evidentiary support. Twitter is a medium known for rapid spread of medical misinformation, perhaps especially concerning substances like CBD [54-57]. The COVID-19 pandemic has been yet another opportunity for CBD marketers and sellers to mislead the public about CBD’s role in treating or preventing the disease. Illegitimate web-based CBD sellers pose a public threat by spreading misinformation, selling unregulated products, and generally sidestepping regulations. Our approach to addressing this issue identified a semantic relationship between tweets containing false claims about CBD in treating COVID-19 and FDA Warning Letters. Using transformer language models and quotes from FDA Warning Letters to other CBD advertisers, this framework can be easily adapted to find misinformation about other conditions and substances, thereby potentially serving a crucial purpose in benefiting public health.

### Acknowledgments

The authors wish to thank Mr Steve Claas at the University of Kentucky College of Public Health Office of Scientific Writing for reviewing this manuscript.

### Authors' Contributions

JT, MK, and RV-S conceived and designed the study and cowrote and revised the final manuscript. JT and AGB collected and analyzed the data.

### Conflicts of Interest

None declared.

### References

1. Fraudulent Coronavirus Disease 2019 (COVID-19) products. U.S. Food & Drug Administration. 2022 Oct 26. URL: <https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-covid-19-products> [accessed 2023-01-02]
2. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JY, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: cross-sectional online study. *J Med Internet Res* 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: [10.2196/22205](https://doi.org/10.2196/22205)] [Medline: [33048825](https://pubmed.ncbi.nlm.nih.gov/33048825/)]
3. Kearney MD, Chiang SC, Massey PM. The Twitter origins and evolution of the COVID-19 “plandemic” conspiracy theory. *HKS Misinfo Review*. 2020 Oct 9. URL: <https://misinforeview.hks.harvard.edu/article/the-twitter-origins-and-evolution-of-the-covid-19-plandemic-conspiracy-theory/> [accessed 2023-01-02]
4. Ayers JW, Chu B, Zhu Z, Leas EC, Smith DM, Dredze M, et al. Spread of misinformation about face masks and COVID-19 by automated software on facebook. *JAMA Intern Med* 2021 Sep 01;181(9):1251-1253 [FREE Full text] [doi: [10.1001/jamainternmed.2021.2498](https://doi.org/10.1001/jamainternmed.2021.2498)] [Medline: [34096988](https://pubmed.ncbi.nlm.nih.gov/34096988/)]
5. Skafle I, Nordahl-Hansen A, Quintana DS, Wynn R, Gabarron E. Misinformation about COVID-19 vaccines on social media: rapid review. *J Med Internet Res* 2022 Aug 04;24(8):e37367 [FREE Full text] [doi: [10.2196/37367](https://doi.org/10.2196/37367)] [Medline: [35816685](https://pubmed.ncbi.nlm.nih.gov/35816685/)]
6. Malinowska B, Baranowska-Kuczko M, Kicman A, Schlicker E. Opportunities, challenges and pitfalls of using cannabidiol as an adjuvant drug in COVID-19. *Int J Mol Sci* 2021 Feb 17;22(4):1986 [FREE Full text] [doi: [10.3390/ijms22041986](https://doi.org/10.3390/ijms22041986)] [Medline: [33671463](https://pubmed.ncbi.nlm.nih.gov/33671463/)]

7. Brown JD. Cannabidiol as prophylaxis for SARS-CoV-2 and COVID-19? Unfounded claims versus potential risks of medications during the pandemic. *Res Social Adm Pharm* 2021 Jan;17(1):2053 [FREE Full text] [doi: [10.1016/j.sapharm.2020.03.020](https://doi.org/10.1016/j.sapharm.2020.03.020)] [Medline: [32273254](https://pubmed.ncbi.nlm.nih.gov/32273254/)]
8. Khalsa JH, Bunt G, Maggirwar SB, Kottitil S. COVID-19 and Cannabidiol (CBD). *J Addict Med* 2021;15(5):355-356 [FREE Full text] [doi: [10.1097/ADM.0000000000000771](https://doi.org/10.1097/ADM.0000000000000771)] [Medline: [33323690](https://pubmed.ncbi.nlm.nih.gov/33323690/)]
9. Iseger TA, Bossong MG. A systematic review of the antipsychotic properties of cannabidiol in humans. *Schizophr Res* 2015 Mar;162(1-3):153-161. [doi: [10.1016/j.schres.2015.01.033](https://doi.org/10.1016/j.schres.2015.01.033)] [Medline: [25667194](https://pubmed.ncbi.nlm.nih.gov/25667194/)]
10. Smolev ET, Rolf L, Zhu E, Buday SK, Brody M, Brogan DM, et al. "Pill pushers and CBD oil"-a thematic analysis of social media interactions about pain after traumatic brachial plexus injury. *J Hand Surg Glob Online* 2021 Jan;3(1):36-40 [FREE Full text] [doi: [10.1016/j.jhsg.2020.10.005](https://doi.org/10.1016/j.jhsg.2020.10.005)] [Medline: [33537664](https://pubmed.ncbi.nlm.nih.gov/33537664/)]
11. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. *J Med Internet Res* 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: [10.2196/17187](https://doi.org/10.2196/17187)] [Medline: [33470931](https://pubmed.ncbi.nlm.nih.gov/33470931/)]
12. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science* 2018 Mar 09;359(6380):1146-1151. [doi: [10.1126/science.aap9559](https://doi.org/10.1126/science.aap9559)] [Medline: [29590045](https://pubmed.ncbi.nlm.nih.gov/29590045/)]
13. Fake news in the time of C-19. Tortoise. 2020 Mar 23. URL: <https://www.tortoisemedia.com/2020/03/23/the-infodemic-fake-news-coronavirus/> [accessed 2023-01-02]
14. Trew B. Coronavirus: hundreds dead in Iran from drinking methanol amid fake reports it cures disease. Independent. 2020 Mar 27. URL: <https://www.independent.co.uk/news/world/middle-east/iran-coronavirus-methanol-drink-cure-deaths-fake-a9429956.html> [accessed 2023-01-02]
15. Shabad R. Twitter removes tweet highlighted by Trump falsely claiming COVID-19 'cure'. NBC News. 2020 Jul 28. URL: <https://www.nbcnews.com/politics/donald-trump/twitter-removes-tweet-highlighted-trump-falsely-claiming-covid-cure-n1235075> [accessed 2023-01-02]
16. Ozawa S, Billings J, Sun Y, Yu S, Penley B. COVID-19 treatments sold online without prescription requirements in the United States: cross-sectional study evaluating availability, safety and marketing of medications. *J Med Internet Res* 2022 Feb 16;24(2):e27704 [FREE Full text] [doi: [10.2196/27704](https://doi.org/10.2196/27704)] [Medline: [34662286](https://pubmed.ncbi.nlm.nih.gov/34662286/)]
17. Mackey TK, Li J, Purushothaman V, Nali M, Shah N, Bardier C, et al. Big data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: infoveillance study on twitter and Instagram. *JMIR Public Health Surveill* 2020 Aug 25;6(3):e20794 [FREE Full text] [doi: [10.2196/20794](https://doi.org/10.2196/20794)] [Medline: [32750006](https://pubmed.ncbi.nlm.nih.gov/32750006/)]
18. Yazdany J, Kim AH. Use of hydroxychloroquine and chloroquine during the COVID-19 pandemic: what every clinician should know. *Ann Intern Med* 2020 Jun 02;172(11):754-755 [FREE Full text] [doi: [10.7326/M20-1334](https://doi.org/10.7326/M20-1334)] [Medline: [32232419](https://pubmed.ncbi.nlm.nih.gov/32232419/)]
19. Della Porta A, Bornstein K, Coye A, Montrieff T, Long B, Parris MA. Acute chloroquine and hydroxychloroquine toxicity: a review for emergency clinicians. *Am J Emerg Med* 2020 Oct;38(10):2209-2217 [FREE Full text] [doi: [10.1016/j.ajem.2020.07.030](https://doi.org/10.1016/j.ajem.2020.07.030)] [Medline: [33071096](https://pubmed.ncbi.nlm.nih.gov/33071096/)]
20. FDA cautions against use of hydroxychloroquine or chloroquine for COVID-19 outside of the hospital setting or a clinical trial due to risk of heart rhythm problems. U.S. Food & Drug Administration. 2020 Jan 7. URL: <https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or> [accessed 2023-01-02]
21. Al-Rakhami MS, Al-Amri AM. Lies kill, facts save: detecting COVID-19 misinformation in Twitter. *IEEE Access* 2020;8:155961-155970 [FREE Full text] [doi: [10.1109/ACCESS.2020.3019600](https://doi.org/10.1109/ACCESS.2020.3019600)] [Medline: [34192115](https://pubmed.ncbi.nlm.nih.gov/34192115/)]
22. Kumar S, Pranesh RR, Carley KM. A fine-grained analysis of misinformation in COVID-19 tweets. Research Square. 2021 Jun 7. URL: <http://paperpile.com/b/4ySh57/T7tx> [accessed 2023-01-02]
23. Hayawi K, Shahriar S, Serhani MA, Taleb I, Mathew SS. ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public Health* 2022 Feb;203:23-30 [FREE Full text] [doi: [10.1016/j.puhe.2021.11.022](https://doi.org/10.1016/j.puhe.2021.11.022)] [Medline: [35016072](https://pubmed.ncbi.nlm.nih.gov/35016072/)]
24. Schroeder DT, Schaal F, Filkukova P, Pogorelov K, Langguth J. WICO Graph: a labeled dataset of twitter subgraphs based on conspiracy theory and 5G-corona misinformation tweets. In: Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021). Setúbal, Portugal: SCITEPRESS – Science and Technology Publications; 2021.
25. Serrano J, Papakyriakopoulos O, Hegelich S. NLP-based feature extraction for the detection of COVID-19 misinformation videos on youtube. In: Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020. 2020 Presented at: 1st Workshop on NLP for COVID-19 at ACL 2020; Jul, 2020; Virtual URL: <https://aclanthology.org/2020.nlpCOVID19-acl.17>
26. Choudrie J, Banerjee S, Kotecha K, Walambe R, Karende H, Ameta J. Machine learning techniques and older adults processing of online information and misinformation: a covid 19 study. *Comput Human Behav* 2021 Jun;119:106716 [FREE Full text] [doi: [10.1016/j.chb.2021.106716](https://doi.org/10.1016/j.chb.2021.106716)] [Medline: [34866770](https://pubmed.ncbi.nlm.nih.gov/34866770/)]
27. Janmohamed K, Soale A, Forastiere L, Tang W, Sha Y, Demant J, et al. Intersection of the web-based vaping narrative with COVID-19: topic modeling study. *J Med Internet Res* 2020 Oct 30;22(10):e21743 [FREE Full text] [doi: [10.2196/21743](https://doi.org/10.2196/21743)] [Medline: [33001829](https://pubmed.ncbi.nlm.nih.gov/33001829/)]
28. Mackey TK, Purushothaman V, Haupt M, Nali MC, Li J. Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter. *Lancet Digit Health* 2021 Feb;3(2):e72-e75 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30318-6](https://doi.org/10.1016/S2589-7500(20)30318-6)] [Medline: [33509386](https://pubmed.ncbi.nlm.nih.gov/33509386/)]

29. Tian L, Zhang X, Wang Y, Liu H. Early detection of rumours on twitter via stance transfer learning. In: *Advances in Information Retrieval*. Cham: Springer; Apr 8, 2020.
30. About warning and close-out letters. U.S. Food & Drug Administration. URL: <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/about-warning-and-close-out-letters> [accessed 2023-01-02]
31. Regulatory procedures manual. U.S. Food & Drug Administration. URL: <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/compliance-manuals/regulatory-procedures-manual> [accessed 2023-01-02]
32. Warning letters. U.S. Food & Drug Administration. URL: <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/compliance-actions-and-activities/warning-letters> [accessed 2023-01-02]
33. Ormerod M, Martínez Del Rincón J, Devereux B. Predicting semantic similarity between clinical sentence pairs using transformer models: evaluation and representational analysis. *JMIR Med Inform* 2021 May 26;9(5):e23099 [FREE Full text] [doi: [10.2196/23099](https://doi.org/10.2196/23099)] [Medline: [34037527](https://pubmed.ncbi.nlm.nih.gov/34037527/)]
34. A social networking service scraper in Python. GitHub. URL: <https://github.com/JustAnotherArchivist/snsrape> [accessed 2023-01-02]
35. Turner J, Kantardzic M, Vickers-Smith R. Infodemiological examination of personal and commercial tweets about cannabidiol: term and sentiment analysis. *J Med Internet Res* 2021 Dec 20;23(12):e27307 [FREE Full text] [doi: [10.2196/27307](https://doi.org/10.2196/27307)] [Medline: [34932014](https://pubmed.ncbi.nlm.nih.gov/34932014/)]
36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. *arXiv* 2017 [FREE Full text]
37. Warning letter Avazo-healthcare, llc. U.S. Food & Drug Administration. URL: <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/avazo-healthcare-llc-610282-11302020> [accessed 2023-01-02]
38. Warning letter for our vets llc dba patriot supreme. U.S. Food & Drug Administration. URL: <https://tinyurl.com/ms68eh8v> [accessed 2023-01-02]
39. Warning letter Nova Botanix LTD DBA CanaBD. U.S. Food & Drug Administration. URL: <https://tinyurl.com/58tae34n> [accessed 2023-01-02]
40. Warning letter project 1600 inc. U.S. Food & Drug Administration. URL: <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/project-1600-inc-608349-06182020> [accessed 2023-01-02]
41. U.S. Food & Drug Administration. URL: <https://tinyurl.com/ymhrbkr9> [accessed 2023-01-02]
42. Warning letter CBD online store. U.S. Food & Drug Administration. URL: <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/cbd-online-store-606587-04072020> [accessed 2023-01-02]
43. Warning letter Sanit Technologies LLC dba Durisan. U.S. Food & Drug Administration. URL: <https://tinyurl.com/2p8v8ncx> [accessed 2023-01-02]
44. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv* 2019 [FREE Full text]
45. Ni J, Ábrego G, Constant N, Ma J, Hall K, Cer D, et al. Sentence-T5: scalable sentence encoders from pre-trained text-to-text models. *arXiv* 2021 [FREE Full text]
46. Love JS, Blumenberg A, Horowitz Z. The parallel pandemic: medical misinformation and COVID-19 : primum non nocere. *J Gen Intern Med* 2020 Aug;35(8):2435-2436 [FREE Full text] [doi: [10.1007/s11606-020-05897-w](https://doi.org/10.1007/s11606-020-05897-w)] [Medline: [32410129](https://pubmed.ncbi.nlm.nih.gov/32410129/)]
47. Byraredy SN, Mohan M. SARS-CoV2 induced respiratory distress: can cannabinoids be added to anti-viral therapies to reduce lung inflammation? *Brain Behav Immun* 2020 Jul;87:120-121 [FREE Full text] [doi: [10.1016/j.bbi.2020.04.079](https://doi.org/10.1016/j.bbi.2020.04.079)] [Medline: [32360437](https://pubmed.ncbi.nlm.nih.gov/32360437/)]
48. Maruyama T, Yamamoto K. Extremely low resource text simplification with pre-trained transformer language model. In: *Proceedings of the 2019 International Conference on Asian Language Processing (IALP)*. 2019 Presented at: 2019 International Conference on Asian Language Processing (IALP); Nov 15-17, 2019; Shanghai, China URL: <http://paperpile.com/b/4ySh57/cYQM> [doi: [10.1109/IALP48816.2019.9037650](https://doi.org/10.1109/IALP48816.2019.9037650)]
49. Voita E, Sennrich R, Titov I. The bottom-up evolution of representations in the transformer: a study with machine translation and language modeling objectives. *arXiv* 2019 [FREE Full text]
50. Zhang J, Chang WZ, Yu HF, Dhillon IS. Fast multi-resolution transformer fine-tuning for extreme multi-label text classification. *arXiv* 2021 [FREE Full text]
51. Lu H, Ehwerhemuepha L, Rakovski C. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. *BMC Med Res Methodol* 2022 Jul 02;22(1):181 [FREE Full text] [doi: [10.1186/s12874-022-01665-y](https://doi.org/10.1186/s12874-022-01665-y)] [Medline: [35780100](https://pubmed.ncbi.nlm.nih.gov/35780100/)]
52. Zhou Y, Zhang Y, Hu Z, Wang M. Semi-autoregressive transformer for image captioning. In: *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 2021 Presented at: 2021 IEEE/CVF

- International Conference on Computer Vision Workshops (ICCVW)q; Oct 11-17, 2021; Montreal, BC, Canada URL: <http://paperpile.com/b/4ySh57/JW4t> [doi: [10.1109/iccvw54120.2021.00350](https://doi.org/10.1109/iccvw54120.2021.00350)]
53. Bechini A, Bondielli A, Ducange P, Marcelloni F, Renda A. Addressing event-driven concept drift in twitter stream: a stance detection application. *IEEE Access* 2021;9:77758-77770 [FREE Full text] [doi: [10.1109/ACCESS.2021.3083578](https://doi.org/10.1109/ACCESS.2021.3083578)]
54. Pang RD, Dormanesh A, Hoang Y, Chu M, Allem J. Twitter posts about cannabis use during pregnancy and postpartum: a content analysis. *Subst Use Misuse* 2021;56(7):1074-1077. [doi: [10.1080/10826084.2021.1906277](https://doi.org/10.1080/10826084.2021.1906277)] [Medline: [33821757](https://pubmed.ncbi.nlm.nih.gov/33821757/)]
55. Walsh-Buhi ER. Social media and cancer misinformation: additional platforms to explore. *Am J Public Health* 2020 Oct;110(S3):S292-S293. [doi: [10.2105/AJPH.2020.305949](https://doi.org/10.2105/AJPH.2020.305949)] [Medline: [33001721](https://pubmed.ncbi.nlm.nih.gov/33001721/)]
56. Cresswell L, Espin-Noboa L, Murphy MS, Ramlawi S, Walker MC, Karsai M, et al. The volume and tone of Twitter posts about cannabis use during pregnancy: protocol for a scoping review. *JMIR Res Protoc* 2022 Mar 29;11(3):e34421 [FREE Full text] [doi: [10.2196/34421](https://doi.org/10.2196/34421)] [Medline: [35348465](https://pubmed.ncbi.nlm.nih.gov/35348465/)]
57. Bellon-Harn ML, Ni J, Manchaiah V. Twitter usage about autism spectrum disorder. *Autism* 2020 Oct;24(7):1805-1816. [doi: [10.1177/1362361320923173](https://doi.org/10.1177/1362361320923173)] [Medline: [32508126](https://pubmed.ncbi.nlm.nih.gov/32508126/)]

## Abbreviations

**BERT:** Bidirectional Encoder Representations from Transformers

**CBD:** cannabidiol

**FDA:** Food and Drug Administration

*Edited by T Mackey; submitted 30.03.22; peer-reviewed by A Pal, SF Tsao; comments to author 16.05.22; revised version received 07.09.22; accepted 30.11.22; published 23.01.23*

*Please cite as:*

*Turner J, Kantardzic M, Vickers-Smith R, Brown AG*

*Detecting Tweets Containing Cannabidiol-Related COVID-19 Misinformation Using Transformer Language Models and Warning Letters From Food and Drug Administration: Content Analysis and Identification*

*JMIR Infodemiology* 2023;3:e38390

URL: <https://infodemiology.jmir.org/2023/1/e38390>

doi: [10.2196/38390](https://doi.org/10.2196/38390)

PMID: [36844029](https://pubmed.ncbi.nlm.nih.gov/36844029/)

©Jason Turner, Mehmed Kantardzic, Rachel Vickers-Smith, Andrew G Brown. Originally published in *JMIR Infodemiology* (<https://infodemiology.jmir.org>), 23.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Infodemiology*, is properly cited. The complete bibliographic information, a link to the original publication on <https://infodemiology.jmir.org/>, as well as this copyright and license information must be included.