People Still Care About Facts: Twitter Users Engage More with Factual Discourse than Misinformation

Luiz Giovanini[1[0000−0002−2617−0847]⋆], Shlok Gilda[1[0000−0002−9355−4381]∗], Mirela Silva[1[0000−0001−5021−0311]⋆⋆], Fabrício Ceschin[2⋆⋆], Prakash Shrestha[1], Christopher Brant[1], Juliana Fernandes[1[0000−0002−8391−8460]], Catia S. Silva[1], André Grégio[2], and Daniela Oliveira[1]

[1] University of Florida, Gainesville, FL, USA, 32611
{lfrancogiovanini@, shlokgilda@, msilva1@, prakash.shrestha@, g8rboy15@, juliana@jou., catiaspsilva@ece., daniela@ece.}ufl.edu,
[2] Federal University of Paraná, Curitiba, Paraná, Brazil, 81530-000
{fjoceschin, gregio}@inf.ufpr.br

Abstract. Misinformation entails disseminating falsehoods that lead to society's slow fracturing via decreased trust in democratic processes, institutions, and science. The public has grown aware of the role of social media as a superspreader of untrustworthy information, where even pandemics have not been immune. In this paper, we focus on COVID-19 misinformation and examine a subset of 2.1M tweets to understand misinformation as a function of engagement, tweet content (COVID-19- vs. non-COVID-19-related), and veracity (misleading or factual). Using correlation analysis, we show the most relevant feature subsets among over 126 features that most heavily correlate with misinformation or facts. We found that (i) factual tweets, regardless of whether COVID-related, were more engaging than misinformation tweets; and (ii) features that most heavily correlated with engagement varied depending on the veracity and content of the tweet.

Keywords: Engagement · Misinformation · Social Media.

1 Introduction

Disinformation refers to false or deceptive content distributed via any communication medium (e.g., word-of-mouth, print, Internet, radio, broadcast) by an adversary who aims to hurt a target (usually a country, political party, or community) via the spread of propaganda and promotion of societal division, thus casting doubt in democratic processes, government institutions, and on science. Over the past few years, our society has grown wearily aware of the highly polarized schism that has developed beyond the context of mere political discourse.

⋆ These authors contributed equally.
⋆⋆ These authors contributed equally.

The perceived extremities of our thoughts and opinions are now intimately meshing with falsehoods and outright lies, calling into question the integrity of our government agencies', political representatives', and our own individual handling of public health crises, such as the COVID-19 pandemic [39].

*Mis*information, however, is closely related to disinformation and differs only in the lack of purposeful intent to harm, often coupled with the raw ignorance of the individual spreading such misleading facts. COVID-19-related misinformation primarily comes from domestic sources; we have seen politicians, pundits, and personalities pushing misleading narratives [6] that may prevent society from controlling the spread of the coronavirus, potentially increasing the number of deaths. With the advent of the COVID vaccines, misinformation has unequivocally been used to discredit its effectiveness, preventing efficient immunization and fueling further hyperpartisanship.

Engagement is a crucial dimension in disseminating falsehoods. Avram et al. [5] showed that higher social engagement results in less fact-checking and verification, especially for less credible content. This paper investigates the relationship between misinformation and user engagement in COVID-19-related tweets. We use the term *misinformation* to refer to tweets spreading deceptive content, even though some tweets may have been created with malice. Using a curated dataset of 2.1M tweets labeled as `fact` or `misinformation` for COVID-related and general topics, we aim to answer the following research questions:

① **RQ1:** Are COVID-19 misinformation tweets more engaging than COVID-19 factual tweets?
② **RQ2:** Are general topic misinformation tweets more engaging than general topics factual tweets?
③ **RQ3:** Which features are most correlated with engagement in COVID-19 vs. general topics misinformation tweets?
④ **RQ4:** Which features are most correlated with engagement in COVID-19 vs. general topics factual tweets?

We measured engagement in COVID-19-related tweets by combining the number of likes and retweets. After preprocessing the tweets, we analyzed our dataset with statistical and correlational methods. Our study found that: (i) factual tweets were more engaging than misinformation tweets, regardless of their topic; (ii) features correlated with engagement varied depending on the tweet's veracity and topic; yet (iii) syntactical features of informal speech and punctuation strongly correlated with general and COVID-related factual tweets, as well as COVID-related misinformation while (iv) user metadata strongly correlated with general topic misinformation but not COVID-19 misinformation; and (v) semantic features, such as sentiment and writing with clout, strongly correlated with factual COVID-related tweets but not misinformation. These findings suggest that addressing misinformation should be targeted toward specific issues rather than using a one-size-fits-all approach.

To our knowledge, prior work [37,31,8,34,24,43,18] has yet to study users' engagement related to factual and misinformation tweets relative to COVID- and general-related topics. This paper thus makes the following contributions:

1. We analyze Twitter discourse on COVID-19 and non-COVID-19 topics to discover whether misinformation tweets are more engaging than factual tweets.
2. We identify discriminating characteristics of a tweet and its author that can distinguish factual and misinformation tweets based on tweet engagement.
3. To support the broader research community, we offer guidance in acquiring the same datasets we employed, although we are not able to directly supply the dataset due to certain restrictions. Our dataset, derived from nine different sources, covers around 2.1M tweets on COVID-19 and various other topics. It encompasses a rich variety of features and labels, obtained through diverse analyses such as those focused on misinformation/factual content, sociolinguistic factors, moral aspects, and sentiment. Researchers eager to work with these datasets or replicate our study are encouraged to contact the authors[3].

Our paper is organized as follows. Section 2 reviews prior works on misinformation and public health and considers the added value of our work. Section 3 discusses our dataset, its curation process, and the preprocessing and feature extraction steps taken. Section 4 then analyses our cleaned datasets' results via statistical tests. Section 5 discusses the takeaways and limitations of our analyses and the future directions for this line of work. Section 6 concludes the paper.

2 Related Work

Intending to understand the nuances that correlate engagement to COVID-19 and other topics of misinformation in the Twittersphere, a few unique approaches have produced intriguing results. This section provides an overview of literature relevant to our work.

Various researchers have explored the presence, prevalence, and sentiment of misinformation on social media of COVID-19 discourse [34,18,37,8,43,24,1], user's susceptibility and psychological perceptions on this public health crisis [30,38], the predictors of fake news [4,19], and the role of bots [24,43] on spreading COVID-19 misinformation. For instance, Sharma et al. [34] examined Twitter data to identify misinformation tweets leveraging state-of-the-art fact-checking tools (e.g., Media Bias/Fact Check, NewsGuard, and Zimdars) along with topics, sentiments, and emerging trends in the COVID-19 Twitter discourse. Singh et al. [37] found that misinformation and myths on COVID-19 are discussed at a lower volume than other pandemic-specific themes on Twitter. They also concluded that information flow on Twitter shows a spatiotemporal relationship with infection rates. Jiang et al. [20] examined the usage of hashtags in 2.3M tweets in the United States and observed that the American public frames the pandemic as a core political issue. Cinelli et al. [8] went beyond Twitter

[3] In order to comply with Twitter's Terms of Service (https://developer.twitter.com/en/developer-terms/agreement-and-policy), we omitted the tweet's raw text, as well as any features that could potentially reveal the users' identity.

and analyzed data from four other social media platforms: Instagram, YouTube, Reddit, and Gab, finding different volumes of misinformation on each platform.

Huang et al. [18] analyzed ∼67.4M tweets and observed that news media and government officials' tweets are highly engaging and that most discussion on misinformation originates from the United States. Unlike this work which explored the kind of users involved and the location of dissemination of highly engaging tweets, this present paper aims to identify the set of a tweet and user characteristics that can predict factual/misinformation tweets and engagement with factual/misinformation tweets.

Although studies on COVID-19 misinformation exist, few have focused on measuring users' engagement and discriminating features, as proposed in this paper. Al-Rakhami and Al-Amri [1] collected 409K COVID-related tweets and used entropy- and correlation-based ranking to distinguish between misinformation and factual information, but they did not examine engagement features. We curated a feature list with 126 features, including textual content, to understand which features contribute most to engagement. Our methodology differs from Al-Rakhami and Al-Amri's, who assumed that Twitter users with large followings are less likely to spread misinformation. However, recent studies [9,11] show that verified users and anti-vaxxers are responsible for a significant portion of misinformation; indeed, we found a positive correlation between followers and engagement with general topic misinformation.

Some studies have analyzed engagement metrics in the context of misinformation on social media in general (e.g., [36,41]). Vosoughi et al.[41] found that fake or false news tend to have higher engagement than verified ones on Twitter, contrasting our results. However, methodological differences between our works could explain this discrepancy. Our engagement analysis combined retweets and favorites, whereas Vosoughi et al.[41] measured diffusion relative to retweet count. Our dataset also contained a larger number of tweets from nine unique datasets, including non-COVID-related false and factual information. Additionally, we analyzed regular users' tweets and replies that did not contain URLs, while the authors specifically looked at fake news with verified true/false URLs. Lastly, we could not collect several tweets of our curated datasets using the Twitter API due to limitations (see Sec. 5.1). This could indicate that, in the three years since Vosoughi et al.'s [41] work, Twitter may have improved its ability to cull high-engagement misinformation tweets.

3 COVID Misinformation and Factual Datasets: Preprocessing and Feature Engineering

We curated data from multiple sources to compose four Twitter datasets used in our analysis for this paper: (1) COVID-19 misleading claims, (2) COVID-19 factual claims, (3) misleading claims on general topics, and (4) factual claims on general topics. We specifically combined different sources of data in each dataset to avoid biasing the results and to improve the generalizability of our findings. For example, our datasets of COVID-19 claims include discourse related to the

spread of the virus, vaccine, etc. The two latter datasets were created to understand how user engagement with COVID-19 claims (misleading and factual) differs from engagement with other claims (e.g., politics, violence, terrorism). This section details our process for building the four datasets mentioned above and the steps taken for data preprocessing and feature extraction.

3.1 Dataset Selection

Several Twitter datasets can be found in the literature, with some designed explicitly for misinformation analysis. These datasets include ground truth labels of *true/factual* and *fake/misleading* for tweets, replies, and/or news articles included in the tweets via URLs. Ground truth labels are typically assigned manually through human annotators; however, automatic annotation strategies are sometimes employed to reach more labeled data. Below, we discuss publicly available Twitter datasets for misinformation analysis on different narratives (including COVID-19) and how we leveraged them to compose the datasets used in our analysis.

COVID-19 Tweets We found five Twitter datasets potentially relevant for analyzing COVID-19 misinformation, which we combined to compose our datasets of COVID-19 *misleading* and *factual* claims.

Dataset 1. Shashi et al. [33] released a dataset[4] containing 1,736 tweets mentioning Coronavirus-related news articles that have been fact-checked by over 92 professional fact-checking organizations and mentioned on Snopes and/or Poynter between January and July 2020. The tweets were classified into four categories based on the veracity of the claims: *false* ($N = 1,345$), *partially false* ($N = 315$), *true* ($N = 41$), and *other* ($N = 35$). We included only the tweets from the first two categories in our dataset of COVID-19 *misleading* claims, while the *true* tweets were included in our dataset of COVID-19 *factual* claims.

Dataset 2. Schroeder et al. [32] created a dataset[5] consisting of tweets linking COVID-19 with 5G conspiracy theories. They collected COVID-related tweets posted between January and May 2020 and filtered for those that mentioned 5G. A random sample of 3,000 tweets was labeled manually as either *5G-corona conspiracy*, *other conspiracy*, or *non-conspiracy*, after which the authors automatically labeled the rest of the tweets based on the subgraphs extracted from the three groups. The resulting dataset contained ∼19K tweets promoting COVID-19 5G conspiracies, ∼38.7K tweets promoting other COVID-related conspiracies, and ∼157K tweets that did not promote any conspiracy. We included tweets from the first two groups in our COVID-19 *misleading* claims dataset and excluded those that did not promote conspiracies, as they contained both—factual and misleading claims.

[4] https://github.com/Gautamshahi/Misinformation_COVID-19
[5] https://datasets.simula.no/wico-graph/

Dataset 3. The Covid-19 Healthcare Misinformation Dataset (CoAID)[6] released by Cui and Lee [12] includes news articles and social media posts related to COVID-19 alongside ground truth labels of *fake claim* and *factual claim* manually assigned by human coders. We leveraged 484 *fake claim* tweets (e.g., "only older adults and young adults are at risk") and 8,092 *factual claim* tweets (e.g., "5G mobile networks do not spread COVID-19") tweeted by the WHO official account.

Dataset 4. Paka et al. [27] published the COVID-19 Twitter fake news (CTF) dataset[7], consisting of a mixture of labeled and unlabeled tweets related to COVID-19. We focused only on the labeled part, comprising 45,261 tweets, of which 18,555 are labeled as *genuine* and 26,706 as *fake*. However, the dataset was not entirely available, and the authors released a sample of 2,000 *fake* and 2,000 *genuine* tweets, which we included in our datasets of COVID-19 *misleading*, and *factual* claims, respectively.

Dataset 5. Muric et al. [26] released a dataset[8] of tweets related to anti-vaccine narratives, including falsehoods and conspiracies surrounding the COVID-19 vaccine. The dataset contains over 1.8 million tweets tweeted between October 2020 and April 2021, containing keywords indicating opposition to the COVID-19 vaccine. Additionally, the authors collected more than 135 million tweets from 70K accounts actively spreading anti-vaccine narratives, which may restrict the diversity of the data. To avoid this, we considered only the first part of their dataset in our study, which contains tweets posted by various users. We included such tweets in our COVID-19 *misleading* claims dataset.

General Topics Tweets We combined four other sources of data to compose two diverse datasets of *misleading* and *factual* claims about general topics (e.g., politics, terrorist conflicts, entertainment, etc.).

Dataset 6. Mitra and Gilbert [25] released CREDBANK, a large-scale crowdsourced dataset of approximately 60M tweets covering 96 days starting from October 2014. All tweets were related to 1,049 real-world news events; 30 annotators from Amazon Mechanical Turk analyzed each tweet for credibility. We selected 18 events rated *certainly accurate* by all 30 annotators for a total of 1,943,827 tweets.

Dataset 7. The Russian Troll Tweets Kaggle dataset[9] contains 200K tweets from malicious accounts connected to Russia's Internet Research Agency (IRA) posted between July 2014 and September 2017. A team reconstructed this dataset at NBC News[10] after Twitter deleted data from almost 3K accounts believed to

[6] https://github.com/cuilimeng/CoAID
[7] https://github.com/williamscott701/Cross-SEAN
[8] https://github.com/gmuric/avax-tweets-dataset
[9] https://www.kaggle.com/vikasg/russian-troll-tweets?select=tweets.csv
[10] https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731

be connected with the IRA in response to an investigation of the House Intelligence Committee into how Russia may have influenced the 2016 U.S. election.

Dataset 8. Vo and Lee [40] released a dataset[11] of tweets that were fact-checked based on news articles from two popular fact-checking websites (Snopes and Politifact). The authors originally collected 247,436 fact-checked tweets posted between May 2016 through 2018. After discarding certain tweets (non-English, removed by Twitter, etc.), their final dataset consisted of 73,203 fact-checked tweets, where 59,208 were labeled as *fake* and 13,995 as *true*, which we included in our datasets of *misleading* and *factual* claims, respectively.

Dataset 9. Jiang et al. [21] released a dataset[12] of 2,327 tweets from Twitter, labeled across a spectrum of fact-check ratings including *true*, *mostly true*, *half true*, *mostly false*, *false*, and *pants on fire*. We focused on purely misleading and factual claims and thus included only *true* ($N = 231$) in our *factual* claims dataset, and both *false* ($N = 1130$) and *pants on fire* ($N = 134$) tweets in our *misleading* claims dataset.

3.2 Data Collection & Stratified Random Sampling

First, we discarded repeated tweet IDs from the four composed datasets. We then used the Twitter API to collect these tweets along with metadata related to the tweets themselves (e.g., language, lists of hashtags, symbols, user mentions, and URLs included), the users/authors of the tweets (e.g., name, profile description, account date of creation, number of followers, number of friends), and the tweet engagement (e.g., number of retweets and number of likes). However, we were able to retrieve only a portion of tweets per each dataset. Many tweets were no longer available/accessible by the time of the data collection (especially those containing misleading claims), most likely because they had been deleted by either Twitter or the user. Moreover, we discarded non-English language tweets and tweets containing no text or very short texts. Upon collecting the entire dataset, we dropped 416,283 entries with null values for the combined engagement metric—this likely was due to errors during poor parsing of the json strings after collecting the entire datasets; nonetheless, this step left us with 2,116,397 total tweets (summarized in Table 1).

This data imbalance is not ideal for statistical analyses as it introduces biases, but it is, unfortunately, part of the misinformation phenomenon. COVID-related tweets are often misleading due to the rapidly evolving scientific research, leading to a rumor-prone environment [3]. To reduce the imbalance, we used stratified random sampling to obtain sample populations representing each dataset's engagement distribution. In other words, instead of randomly selecting data from each of the four datasets, we sampled subgroups, i.e., *strata*, of $n \approx 4,556$ from each dataset according to the distribution of combined engagement. This n was chosen as it is 50% of the smallest population size across our datasets (i.e.,

[11] https://github.com/nguyenvo09/LearningFromFactCheckers
[12] https://shanjiang.me/resources/#misinformation

Table 1: Descriptive statistics of our final four datasets based on the combined engagement metric.

	Factual		**Misinformation**	
	COVID-Related	*General Topics*	*COVID-Related*	*General Topics*
N	9,111	1,243,913	828,501	32,243
n_{strata}	4,814	4,448	4,533	4,147
μ	368.5	9,791.6	2,214.3	3,014.7
σ	7,157.9	73,305.6	10,051.9	28,727.4
Mean Rank	2407.5	2244.5	2267.0	2074.0

$N = 9,111$ for COVID-related factual tweets) and allowed us to maintain variability across all class sizes. We repeated this process 10 times, obtaining 10 stratified random samples of $17,982$ tweets each. Figure 1 compares the original datasets with one of the stratified random samples, demonstrating that we stayed true to the original distribution of engagement.

3.3 Data Preprocessing

Before feature extraction, the full text of the collected tweets was preprocessed by removing numbers (e.g., "1 million" or "12,345" become " million" and ","), emojis, hashtags (e.g., "#COVID"), mentions (e.g., "@WHO"), and URLs. Other typical NLP preprocessing steps, such as tokenization, removal of stop words, and lemmatization, were not performed, as both LIWC and sentiment analysis packages can work with raw text.

3.4 Feature Extraction

From the cleaned dataset, we extracted a total of 126 features per tweet, including features derived from the metadata (i.e., tweet- and user-related descriptors), addressing sociolinguistic (e.g., cognitive and structural components, such as formal and logical language) and moral frames (e.g., fairness or reciprocity), as well as sentiment characteristics of the tweet texts.

Tweet Metadata, User Metadata, and Engagement We extracted the following features from the collected Twitter metadata:

- Six **tweet-related** features: # of likes, # of retweets, # of hashtags, # links/URLs, # of combined engagement (i.e., # retweets + # likes), and # of emojis in the tweet.
- Twelve **user-related** features: # of followers, # of friends, # of lists, # of favorited tweets, verified (binary), presence of profile image (binary), use of default profile image (binary) or default profile (binary), whether geolocation is enabled (binary), whether the user has an extended profile (binary) or background tile (binary), and # of tweets made by the user.

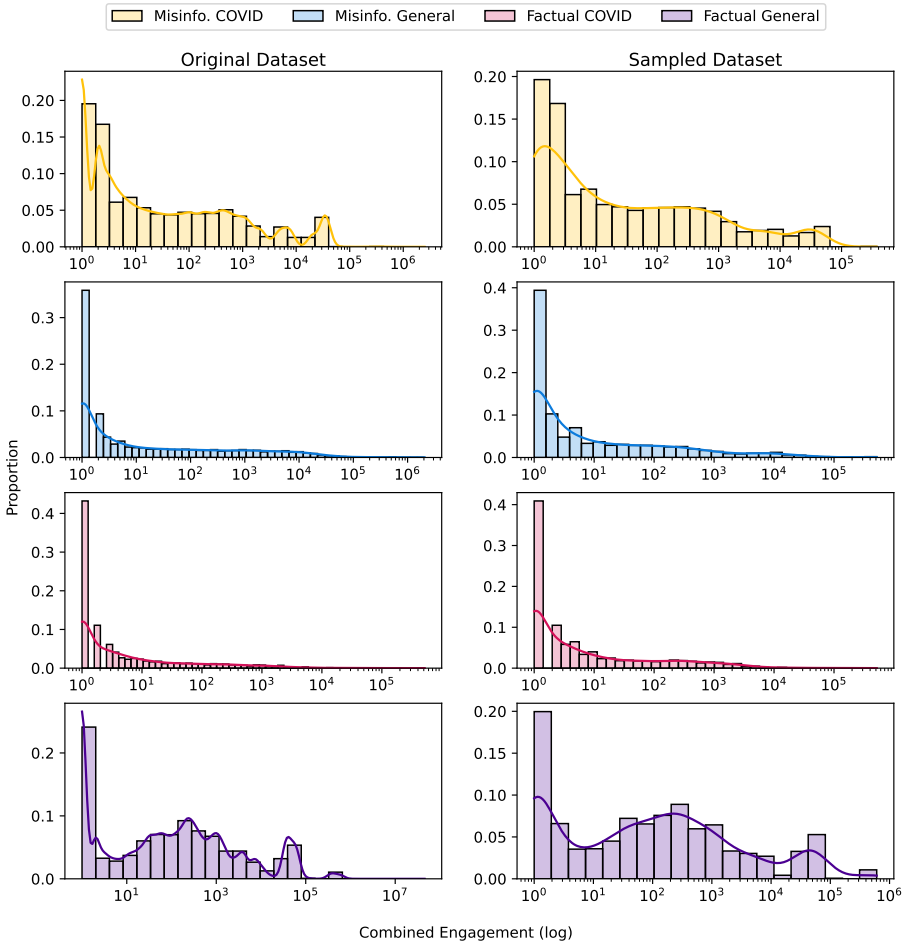

Fig. 1: Comparison of the distribution of combined engagement (log-normalized) versus the proportion of its occurrence for each dataset.

We combined likes and retweets to form an engagement metric, but it was left-skewed, so we log-normalized it using Aldous and Jansen's method [2]. The method suggests a 4-level scale to measure engagement on Twitter, where retweets are the highest level of engagement (level-4) and likes are level-2. Commenting (level-3) is more public than liking but less than retweeting (since retweeting is a deliberate effort to amplify the reach of the content through different networks), and viewing (level-1) is the most private. Our dataset lacked engagement metrics for levels 1 and 3, so we analyzed the likes and retweets combined as a single metric.

To capture sentiment and emotions in tweets, we implemented an emoji and emoticon counter, but we later decided to disregard emoticons due to a high occurrence of false positives. Many combinations of regular punctuations were incorrectly identified as emoticons, leading to misclassification. Therefore, we only counted for emojis.

Sociolinguistic Analysis We performed a sociolinguistic analysis on the collected tweets using the Linguistic Inquiry and Word Count (LIWC) software (version 2015) [28]. This tool estimates the rate at which certain emotions, moods, and cognition (e.g., analytical thinking) are present in a text based on word counts (e.g., the words "nervous," "afraid," and "tense" counted as expressing anxiety). More specifically, we extracted 93 features related to emotional, cognitive, and structural components from the collected tweets, including:

- Four **language metrics**: total number of words, average number of words per sentence, number of words containing more than six letters, and number of words found in the LIWC dictionary.
- Eighty-five **dimensions**, including function words (e.g., pronouns, articles, prepositions), grammar characteristics (e.g., adjectives, comparatives, numbers), affect words (e.g., positive and negative emotions), social words (e.g., family, friends, male/female referents), cognitive process (e.g., insight, certainty), core needs (e.g., power, risk/prevention focus), time orientation (e.g., past/present/future focus), personal concerns (e.g., home, money, death), informal speech (e.g., swear words, netspeak), and punctuation (e.g., periods, commas, question marks). These features reflect the percentage of total words per dimension (e.g., "positive emotions" equal to 7.5 means that 7.5% of all words in the tweet were positive emotion words).
- Four **summary variables** expressed in a scale ranging from 0 (very low) to 100 (very high): (i) analytical thinking; (ii) clout; (iii) authenticity; (iv) emotional tone.

Moral Frames Analysis We measured moral frames using the moral foundations dictionary [15] dictionary in LIWC. Based on moral foundations theory [17], the authors aggregated 295 words for each of five moral intuitions encompassing 11 total features, which encompass psychological preparations for reacting to issues about **harm/care**, **fairness/reciprocity**, **ingroup/loyalty**, **authority/respect**, and **purity/sanctity**.

Sentiment Analysis For sentiment analysis, we used VADER [14], a rule-based NLP library available with NLTK [22]. Among the outputs generated by VADER, we used the *compound score*, a uni-dimensional normalized, weighted composite score. A compound score ≥ 0.05 denotes a positive sentiment, between -0.05 and 0.05 denotes a neutral sentiment, and ≤ -0.05 denotes a negative sentiment. We extracted three binary sentiment features for each collected tweet: positive, negative, and neutral.

3.5 Correlation Analysis

To investigate the correlation of engagement with COVID- and non-COVID-related misinformation and factual tweets (RQs 3 and 4), we used Pearson's correlation coefficient, r, to measure feature importance. However, as r only captures linear relationships, we employed another method to identify non-linear correlations. We used the Alternating Conditional Expectations (ACE) algorithm to find each feature's fixed point of Maximal Correlation (MC). The ACE algorithm transforms variables to maximize r for the dependent and independent variables, making it robust against noisy data and capable of detecting non-linear correlations more accurately than r [13]. Note that MC ranges from 0 to 1, indicating the polarity of the correlation. We used this method to supplement our analysis and provide a more comprehensive understanding of the relationships between engagement and tweets on COVID- and non-COVID-related misinformation and factual information.

Additionally, the Pearson correlation coefficient is biased such that the simple mean of r of all 10 samples would underestimate the true r. Therefore, performing a Fisher z-transformation correction of the rs allows us to reduce bias and more accurately estimate the population correlation [10]. In other words, we report the average Pearson's correlation coefficient, r_z, i.e., the inverse z-transform of the averaged z-values over all the 10 samples. Additionally, we rely on the Fisher method by the sum of logs to combine the p-values obtained for each sample into a single metric.

4 Results

This section details the statistical and correlation analyses performed on our curated dataset to answer each of our four research questions and their results. All statistical tests were performed based on a 1% significance level ($\alpha = .01$). Tables 2 and 3 summarize our results.

4.1 RQ1: Are COVID-19 misinformation tweets more engaging than COVID-19 factual tweets?

We investigated the difference in engagement between factual and misinformation tweets related to COVID-19. Firstly, we checked whether the combined engagement metric for factual and misinformation tweets followed a normal

distribution using the Shapiro-Wilk ($p < .001$) and D'Agostino's K-squared ($p < .001$) tests. The results showed that the distribution was non-normal and heavily skewed towards zero for most tweets. Additionally, we found that the distribution was not homogeneous between the two groups ($W = 378.89$, $p < .001$), so we used non-parametric tests to analyze the log-normalized combined engagement metric for each group. The Two-Sample Kolmogorov-Smirnov test results indicated that the engagement distribution of COVID-19 factual tweets was significantly different from that of COVID-19 misinformation tweets ($KS = 0.21, p < .001$). These results were consistent across all stratified random samples, indicating that the strata adequately reflected the distribution of combined engagement.

Table 2: Summary results for statistical tests conducted on engagement metrics and bot/user account labels.

Data	Measure	Measurement Statistics	
Combined Engagement (raw)	Shapiro-Wilk	Factual COVID-Related	$W = 0.7875$***
		Misinformation General Topics	$W = 0.8946$***
		Factual General Topics	$W = 0.9374$***
		Misinformation General Topics	$W = 0.7969$***
Combined Engagement (log-norm)	Levene	Factual vs. Misinformation COVID-Related	$W = 378.89$***
		Factual vs. Misinformation General Topics	$W = 359.59$***
	Two-Sample Kolmogorov-Smirnov	Factual vs. Misinformation COVID-Related	$K_2 = 0.2133$***
		Factual vs. Misinformation General Topics	$K_2 = 0.3459$***
	Mann-Whitney U	Factual vs. Misinformation COVID-Related	$U = 7,662,279$***, $r = 0.35$
		Factual vs. Misinformation General Topics	$U = 5,725,193$***, $r = 0.31$

*** Significant at $p < .001$

A comparison of the mean distribution of factual and misinformation tweets was desirable, given the notable differences in the overall populations ($\mu_{COVID,factual} = 368.5$ and $\mu_{COVID,misinfo} = 2,214.3$. However, due to the non-normality and skew of these variables, we opted to conduct the Mann-Whitney U-test and compare the mean ranks of the two samples. For each strata, factual COVID-19 tweets ($n = 4,814$) had a larger average mean rank ($2,407.5$) than misinformation tweets ($n = 4,533, \mu_{rank} = 2,267.0$). Therefore, the combined engagement of the factual tweets was statistically and significantly higher than the misinformation tweets $U = 7,662,279$, $p < .001$), indicating that factual COVID-19 tweets tend to be more engaging than COVID-19 misinformation tweets. Given that $U_{max} = n_{strata,1} \times n_{strata,2} = 21,821,862$, we can convert the U-statistic to an effect size, $r = U/U_{max} = 0.35$. In simpler words, there is a medium probability that a combined engagement value from the factual tweets will be greater than misinformation tweets.

COVID-19 factual tweets were statistically and significantly more engaging than misinformation tweets about COVID-19.

4.2 RQ2: Are general topic misinformation tweets more engaging than general topic factual tweets?

We repeated the analyses conducted for **RQ1**, finding that the combined engagement metrics also do not follow normal distribution based on the Shapiro-Wilk ($p < .001$) and D'Agostino's K-squared ($p < .001$) tests, and that the distribution of the data was not homogeneous for the two groups ($W = 359.59, p < .001$). The Two-Sample Kolmogorov-Smirnov test also showed that the distribution between factual and misinformation general topic tweets was significantly different ($KS = 0.35, p < .001$).

The Mann-Whitney U-test revealed that the average mean rank for combined engagement was higher for factual general topic tweets ($n = 4,448, \mu_{rank} = 2,244.5$) compared to misinformation tweets ($n = 4,147, \mu_{rank} = 2,074$). As a result, we concluded that factual general topic tweets have significantly higher combined engagement than misinformation tweets ($U = 5,745,193.0$, $p < .001$). The U-statistic was converted to an effect size of $r = 0.31$, suggesting a medium probability that combined engagement from factual general topic tweets will be higher than that of misinformation tweets.

> Factual tweets were statistically and significantly more engaging that misinformation tweets about general topics.

4.3 RQ3: Which features are most correlated with engagement in COVID-19 vs. general topics misinformation tweets?

Our correlation analysis found that only a few of the extracted features were strongly correlated ($r_{MC,z} \geq 0.5$) with the log-normalized combined engagement metric. For COVID-related misinformation combined engagement, we observed a strong correlation with LIWC-based grammar features (i.e., use of informal speech, punctuation, impersonal pronouns) and word count, with correlation coefficients ranging from $[0.50, 0.75]$. On the other hand, for general topic misinformation, only three features showed a strong correlation, all related to user metadata: the number of followers ($r_{MC,z} = 0.73$), the number of public lists of which that a user is a member ($r_{MC,z} = 0.66$), and whether the user is verified ($r_{MC,z} = 0.53$).

> The top features related to engagement for COVID-19 and general topics misinformation were, respectively, the tweet's grammar (e.g., use of informal speech) and user metadata (e.g., verified user).

4.4 RQ4: Which features are most correlated with engagement in COVID-19 vs. general topics factual tweets?

Compared to the other groups, factual COVID-related tweets showed several strong correlations. The highest correlation ($r_{MC,z} = 0.91$) was using third-person singular words, a feature not strongly correlated with any other group,

Table 3: Summary of correlation analysis between the log normalized combined engagement metric and all features. Only r_z values indicating a moderate correlation (> 0.5) and with a combined MC p-value $< .01$ are shown.

Feature Type	Feature	r_z	(MC) r_z
Factual: COVID-Related			
LIWC	Affective Processes	0.53	0.71
	All Punctuation	-0.05	0.58
	Assent (Informal Language)	0.65	0.74
	Clout	0.36	0.56
	Colon (Punctuation)	0.34	0.54
	Dictionary Words	0.13	0.56
	Past Focus	0.49	0.66
	Informal Speech	0.62	0.72
	Insight (Cognitive Processes)	0.32	0.68
	Male Referents (Social Words)	0.77	0.88
	Netspeak (Informal Language)	0.66	0.77
	Positive Emotion (Affect Words)	0.52	0.78
	Person Pronouns (Linguistic Dimensions)	0.31	0.56
	Question Marks (All Punctuation)	-0.31	0.53
	Reward (Drives)	0.33	0.67
	Sad (Affect Words)	0.48	0.65
	3rd Person Singular (Function Words)	0.81	0.91
	Words $>$ 6 Letters	-0.26	0.59
	Social Words	0.41	0.63
	Time (Relativity)	0.21	0.51
Sentiment	VADER Compound	0.19	0.66
Factual: General Topics			
LIWC	Assent (Informal Speech)	0.36	0.68
	Colons (All Punctuation)	0.20	0.52
	Informal Speech	0.29	0.62
	Netspeak (Informal Speech)	0.32	0.63
	Prepositions (Function Words)	0.02	0.54
Misinformation: COVID-Related			
LIWC	Assent (Informal Speech)	0.26	0.75
	Colons (All Punctuation)	0.34	0.75
	Informal Speech	0.19	0.69
	Impersonal Pronouns	0.06	0.64
	Netspeak (Informal Speech)	0.26	0.73
	Quotation Marks (All Punctuation)	0.10	0.50
	Word Count	-0.10	0.51
Misinformation: General Topics			
User Metadata	Followers Count	0.28	0.73
	Listed Count	0.30	0.66
	User Verified	0.53	0.53

while the second-highest correlation ($r_{MC,z} = 0.88$) was related to male referents. Additionally, factual COVID tweets were strongly correlated with effective processes ($r_{MC,z} = 0.71$) and emotion, as measured by LIWC ($r_{MC,z,positive} = 0.78$ and $r_{MC,z,sad} = 0.65$) and VADER ($r_{MC,z} = 0.66$). Only one LIWC summary variable, Clout ($r_{MC,z} = 0.71$), indicated confidence and leadership in writing and appeared among any of the groups. In contrast, factual general topics tweets only strongly correlated with LIWC's grammar features, such as informal speech, punctuation, and prepositions, similar to the strongly correlated features for COVID-related misinformation.

> The top features related to engagement for COVID-19 factual tweets pertained to grammar (e.g., use of netspeak), emotion (both positive and negative), and the writer's confidence, whereas general topic tweets pertained solely to grammar (e.g., use of colons or prepositions).

5 Discussion

In this paper, we set out to answer four research questions relating to COVID-19 and general topics tweets as a function of the combined engagement metric. This section summarizes the takeaways and limitations of our work and suggests possible future research directions.

First, it is essential to note that distinguishing between factual and misinformation tweets is challenging as research has shown that automatic detection of misinformation is a nuanced and open research problem in the machine learning field [44] and social media platforms are inherently rooted in big data that is unstructured and noisy [35]. Such problems exacerbate the difficulty of detecting misinformation. The digital revolution and the integration of social media into our daily lives have been leveraged as tools for the faster propagation of disinformation campaigns. Research has shown that humans are poor at detecting deception [16], and our ability to detect digital fake news is "bleak" [42]. Understanding how machines can detect highly engaging dis/misinformation will provide a first line of defense against deception in the online sphere. Government agencies and organizations can use this knowledge to convey critical public health information to the general populace. For example, with respect to the Italian Ministry of Health, Lovari [23] found that keeping the public constantly informed via dissemination of information in understandable forms (e.g., data and visuals) helps reduce the spread of misinformation.

Therefore, the primary purpose of this work was to point researchers toward potentially impactful metadata that could give inklings towards purposeful or unintentional false information. Importantly, we found that misinformation tweets about general topics strongly correlated with the users' metadata; these features all contained a positive polarity in terms of r_z, potentially indicating that influential users were responsible for generating engagement with general topics misinformation.

Assuming that real Twitter accounts are more likely to be verified and have several followers, we can infer that misinformation tweets by seemingly real and influential users can offer a perceived sense of credibility. However, this can be even more deceiving to the average user in the context of misinformation [44].

As such, the semantic content of the tweet itself (based on LIWC analysis) appears not to be relevant to engagement (except for factual COVID tweets). Instead, the *syntax* was highly correlated with engaging tweets for factual COVID tweets and factual and misinformation general topics. Interestingly, we found that tweet sentiment was not relevant to predict engagement, except in the context of truthful COVID-related tweets.

In stark contrast, we found that engagement with COVID-related factual tweets differed from engagement with other types of tweets. Engagement with factual tweets was highly correlated with sentiment and cognitive processing-related keywords, indicating that tweets appealing to pathos were more engaging. In contrast, fewer complex words (i.e., > 6 letters) and question marks were associated with high engagement (strong MC correlation—$r_{MC,z} = 0.59$ and 0.53, respectively), suggesting that clear and straightforward language drives engagement with misinformation. This highlights the importance of understanding and addressing different types of misinformation on a per-issue basis rather than lumping them together.

We also found that factual tweets were statistically more engaging than misinformation tweets, regardless of the tweet's context (general topics or COVID-19). To our knowledge, our study is the first to analyze engagement in COVID tweets relative to veracity and other topics. Surprisingly, we did not find that the # of ULRs in the tweet was a strongly correlated feature. We suspected that URLs could increase the veracity of the information presented in the tweet, thus helping distinguish factual information from misinformation and reinforcing false claims in misinformation tweets, increasing their engagement.

5.1 Limitations & Future Works

In light of the contributions made by our research, it is incumbent upon us to acknowledge the concomitant limitations of our study and delineate potential paths for future exploration. This section discusses these limitations and outlines promising trajectories for further research.

Dataset imbalance and representativeness Our dataset was imbalanced, with factual general topics and COVID-related misinformation dominating over factual COVID-related and general misinformation tweets. We generated 10 stratified random samples to address this issue, but factual COVID tweets lacked variety and exhibited stronger correlations than the other groups. This limits the generalizability of our findings and could lead to overfitting in machine learning models. Future studies could generate larger synthetic datasets or adopt down-sampling strategies to overcome this.

Our meta-analysis of nine datasets included 2.1M tweets, but we could not collect some tweets removed by Twitter, potentially favoring high-engagement

factual tweets. Future studies could conduct a time-series analysis of tweets to understand the relationship between engagement and truthfulness and identify factors contributing to tweet removal.

Although we demonstrated the impact of different meta features on engagement in tweets, future studies could take a more nuanced approach by comparing the impact of veracity and tweet context across different topics, such as COVID and measles vaccine hesitancy, or specific events and controversies associated with misinformation, such as the 2020 U.S. General Election.

Feature engineering, feature selection, and classification models Our research provides a foundation for future studies in machine learning, but there are still many other features to explore beyond the ones we analyzed. For example, studies have found that emojis can help determine Twitter sentiment, and automated feature extractors like Word Embedding, TF-IDF, Word2Vec, BERT, and GloVe could be used to predict misinformation and tweet engagement. Additionally, investigating how tweets are written may be a more straightforward approach than fact-checking every claim.

While previous works have studied the prevalence of COVID-19 misinformation on Twitter and characterized the role of bots in spreading misinformation, more research should examine how automatic adversaries spread misinformation. Investigating demographic attributes and their impact on engagement with falsehoods may also prove fruitful.

Two similar features were deemed negatively correlated with engagement for misinformation and factual COVID-related tweets: the tweet's length, as measured by word count, and the use of words > 6 words. Historically, we have seen the use of short texts, lots of images, a touch of sex, and a tendency towards sensationalism used as a recipe for propaganda success, leveraged by the KGB, Stasi, and CIA [29]. The presence of an image and the amount of text (and, therefore, information that a user must process) in a tweet might be leveraged by both disinformation campaigns and reputable sources alike to help users quickly digest information. Additionally, this suggests that users are likelier to engage with an image over words, especially considering that sociolinguistic and sentiment features were not of utmost importance in predicting engagement. While we did not measure for the presence of an image, few studies (e.g., [7]) have conducted exploratory research on visual misinformation videos, and we advise future work to consider this dimension in their work.

Another limitation of our study was our reliance on pairwise correlation analysis. Future work could benefit from utilizing multivariate analyses such as principal component analysis (PCA) to identify the most relevant features tailored to specific models. Additionally, examining the correlation between groups of features could help in the feature selection process.

6 Conclusion

This paper curated a dataset of 2.1M COVID-19- and non-COVID-related misinformation and factual tweets to investigate misinformation as a function of veracity, content, and engagement. Via the use of statistical and correlation analyses, we offer the following conclusions: (i) misinformation tweets were less engaging than factual tweets; (ii) features for general and COVID-related tweets varied in correlation to engagement based on veracity; for example, user metadata features (e.g., followers count) were most strongly associated with engagement for general misinformation, which COVID-related misinformation correlated most with grammar-related features present in the tweet's text. We propose several directions and suggestions for future works on misinformation in the online sphere. In particular, our insights on what features can aid with predicting high engagement can be leveraged for defense approaches against misinformation, such as increasing the engagement of factual tweets, especially those coming from verified government accounts and reputable organizations (e.g., WHO, NIH), thus contributing to factual public health information reaching the masses.

Acknowledgements This work was supported by the National Science Foundation under Grant No. 2028734, by the University of Florida Seed Fund award P0175721, and by the Embry-Riddle Aeronautical University award 61632-01/PO# 262143. This material is based upon work supported by (while serving at) the National Science Foundation.

References

1. Al-Rakhami, M.S., Al-Amri, A.M.: Lies Kill, Facts Save: Detecting COVID-19 Misinformation in Twitter. IEEE Access **8**, 155961–155970 (Aug 2020)
2. Aldous, K.K., An, J., Jansen, B.J.: View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. Proceedings of the International AAAI Conference on Web and Social Media **13**(01), 47–57 (Jul 2019)
3. Allport, G.W., Postman, L.: The psychology of rumor. Journal of Clinical Psychology (1947)
4. Apuke, O.D., Omar, B.: Fake news and covid-19: Modelling the predictors of fake news sharing among social media users. Telematics and Informatics p. 101475 (2020)
5. Avram, M., Micallef, N., Patil, S., Menczer, F.: Exposure to social engagement metrics increases vulnerability to misinformation. arXiv preprint arXiv:2005.04682 (2020)
6. Bell, B., Gallagher, F.: Who Is Spreading COVID-19 Misinformation and Why. https://abcnews.go.com/US/spreading-covid-19-misinformation/story?id=70615995 (May 2020), accessed: 2020-11-21
7. Brennen, J.S., Simon, F.M., Nielsen, R.K.: Beyond (mis) representation: Visuals in covid-19 misinformation. The International Journal of Press/Politics (2020)

8. Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F., Scala, A.: The covid-19 social media infodemic. arXiv preprint arXiv:2003.05004 (2020)
9. Cohen, J.: Verified Twitter Users Shared An All-Time-High Amount of Fake News in 2020. https://www.pcmag.com/news/verified-twitter-users-shared-an-all-time-high-amount-of-fake-news-in-2020 (Feb 2021), accessed: 2021-9-4
10. Corey, D.M., Dunlap, W.P., Burke, M.J.: Averaging Correlations: Expected Values and Bias in Combined Pearson rs and Fisher's z Transformations. Journal of General Psychology **125**(3), 245–261 (Jul 1998)
11. for Countering Digital Hate, C.: The disinformation dozen: Why platforms must act on twelve leading online anti-vaxxers. Counterhate. com (2021)
12. Cui, L., Lee, D.: Coaid: Covid-19 healthcare misinformation dataset (2020)
13. Deebani, W., Kachouie, N.N.: Ensemble Correlation Coefficient. In: International Symposium on Artificial Intelligence and Mathematics (2018)
14. Gilbert, C., Hutto, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media (ICWSM-14). vol. 81 (2014)
15. Graham, J., Haidt, J., Nosek, B.A.: Liberals and conservatives rely on different sets of moral foundations. Journal of Personality and Social Psychology **96**(5), 1029–1046 (May 2009)
16. Granhag, P.A., Andersson, L.O., Strömwall, L.A., Hartwig, M.: Imprisoned knowledge: Criminals' beliefs about deception. Legal and Criminological Psychology **9**(1), 103–119 (Feb 2004)
17. Haidt, J., Graham, J.: When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize. Soc. Justice Res. **20**(1), 98–116 (Mar 2007)
18. Huang, B., Carley, K.M.: Disinformation and misinformation on twitter during the novel coronavirus outbreak. arXiv preprint arXiv:2006.04278 (2020)
19. Islam, A.N., Laato, S., Talukder, S., Sutinen, E.: Misinformation sharing and social media fatigue during covid-19: An affordance and cognitive load perspective. Technological Forecasting and Social Change **159** (2020)
20. Jiang, J., Chen, E., Yan, S., Lerman, K., Ferrara, E.: Political polarization drives online conversations about covid-19 in the united states. Human Behavior and Emerging Technologies **2**(3), 200–211 (2020)
21. Jiang, S., Wilson, C.: Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. Proceedings of the ACM on Human-Computer Interaction **2**(CSCW), 1–23 (2018)
22. Loper, E., Bird, S.: Nltk: The natural language toolkit. arXiv preprint cs/0205028 (2002)
23. Lovari, A.: Spreading (dis) trust: Covid-19 misinformation and government intervention in italy. Media and Communication **8**(2), 458–461 (2020)
24. Memon, S.A., Carley, K.M.: Characterizing covid-19 misinformation communities using a novel twitter dataset. arXiv preprint arXiv:2008.00791 (2020)
25. Mitra, T., Gilbert, E.: Credbank: A large-scale social media corpus with associated credibility annotations. In: Ninth International AAAI Conference on Web and Social Media (2015)
26. Muric, G., Wu, Y., Ferrara, E.: Covid-19 vaccine hesitancy on social media: building a public twitter data set of antivaccine content, vaccine misinformation, and conspiracies. JMIR public health and surveillance **7**(11), e30642 (2021)

27. Paka, W.S., Bansal, R., Kaushik, A., Sengupta, S., Chakraborty, T.: Cross-sean: A cross-stitch semi-supervised neural attention model for covid-19 fake news detection. Applied Soft Computing **107** (2021)
28. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015. Tech. rep. (2015)
29. Rid, T.: Active measures: The secret history of disinformation and political warfare. Farrar, Straus and Giroux (2020)
30. Roozenbeek, J., Schneider, C.R., Dryhurst, S., Kerr, J., Freeman, A.L., Recchia, G., van der Bles, A.M., van der Linden, S.: Susceptibility to misinformation about covid-19 around the world. Royal Society Open Science **7**(10) (2020)
31. Schild, L., Ling, C., Blackburn, J., Stringhini, G., Zhang, Y., Zannettou, S.: "go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of covid-19. arXiv preprint arXiv:2004.04046 (2020)
32. Schroeder, D.T., Pogorelov, K., Schaal, F., Filkukova, P., Langguth, J.: Wico graph: a labeled dataset of twitter subgraphs based on conspiracy theory and 5g-corona misinformation tweets. In: ICAART 2021 : 13th International Conference on Agents and Artificial Intelligence. OSF Preprints (2021)
33. Shahi, G.K., Dirkson, A., Majchrzak, T.A.: An exploratory study of covid-19 misinformation on twitter. Online Social Networks and Media **22** (2021)
34. Sharma, K., Seo, S., Meng, C., Rambhatla, S., Liu, Y.: Covid-19 on social media: Analyzing misinformation in twitter conversations. arXiv preprint arXiv:2003.12309 (2020)
35. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake News Detection on Social Media: A Data Mining Perspective. SIGKDD Explor. Newsl. **19**(1), 22–36 (Sep 2017)
36. Silva, M., Giovanini, L., Fernandes, J., Oliveira, D., Silva, C.S.: What makes disinformation ads engaging? a case study of facebook ads from the russian active measures campaign. Journal of Interactive Advertising pp. 1–20 (2023)
37. Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E., Wang, Y.: A first look at covid-19 information and misinformation sharing on twitter. arXiv preprint arXiv:2003.13907 (2020)
38. Swami, V., Barron, D.: Analytic thinking, rejection of coronavirus (covid-19) conspiracy theories, and compliance with mandated social-distancing: Direct and indirect relationships in a nationally representative sample of adults in the united kingdom. OSF Preprints (2020)
39. Tagliabue, F., Galassi, L., Mariani, P.: The "pandemic" of disinformation in covid-19. SN Compr Clin Med **2**, 1287–1289 (Aug 2020)
40. Vo, N., Lee, K.: Learning from fact-checkers: Analysis and generation of fact-checking language. In: The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (2019)
41. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
42. Wineburg, S., McGrew, S., Breakstone, J., Ortega, T.: Evaluating information: The cornerstone of civic online reasoning. Stanford Digital Repository. Retrieved January **8**, 2018 (2016)
43. Yang, K.C., Torres-Lugo, C., Menczer, F.: Prevalence of low-credibility information on twitter during the covid-19 outbreak. arXiv preprint arXiv:2004.14484 (2020)
44. Zhou, X., Zafarani, R.: A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Computing Surveys **53**(5), 1–40 (Sep 2020)