

The Discussions of Monkeypox Misinformation on Social Media

Or Elroy ¹  and Abraham Yosipof ^{2,3,*} ¹ Department of Computer Science, University of Oregon, Eugene, OR 97403, USA; ore@uoregon.edu² Faculty of Information Systems and Computer Science, College of Law & Business, Ramat-Gan, P.O. Box 852, Bnei Brak 5110801, Israel³ International Institute for Applied Systems Analysis, 2361 Laxenburg, Austria

* Correspondence: yosipof@iiasa.ac.at or aviyo@clb.ac.il

Abstract

The global outbreak of the monkeypox virus was declared a health emergency by the World Health Organization (WHO). During such emergencies, misinformation about health suggestions can spread rapidly, leading to serious consequences. This study investigates the relationships between tweet readability, user engagement, and susceptibility to misinformation. Our conceptual model posits that tweet readability influences user engagement, which in turn affects the spread of misinformation. Specifically, we hypothesize that tweets with higher readability and grammatical correctness garner more user engagement and that misinformation tweets tend to be less readable than accurate information tweets. To test these hypotheses, we collected over 1.4 million tweets related to monkeypox discussions on X (formerly Twitter) and trained a semi-supervised learning classifier to categorize them as misinformation or not-misinformation. We analyzed the readability and grammar levels of these tweets using established metrics. Our findings indicate that readability and grammatical correctness significantly boost user engagement with accurate information, thereby enhancing its dissemination. Conversely, misinformation tweets are generally less readable, which reduces their spread. This study contributes to the advancement of knowledge by elucidating the role of readability in combating misinformation. Practically, it suggests that improving the readability and grammatical correctness of accurate information can enhance user engagement and consequently mitigate the spread of misinformation during health emergencies. These insights offer valuable strategies for public health communication and social media platforms to more effectively address misinformation.



Academic Editor: Rüdiger Pryss

Received: 21 July 2025

Revised: 19 August 2025

Accepted: 22 August 2025

Published: 25 August 2025

Citation: Elroy, O.; Yosipof, A. The Discussions of Monkeypox Misinformation on Social Media. *Data* **2025**, *10*, 137. <https://doi.org/10.3390/data10090137>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: misinformation; readability; NLP; monkeypox

1. Introduction

The worldwide outbreak of the monkeypox virus was declared a global health emergency by the World Health Organization (WHO) between May 2022 and May 2023 [1].

Social media platforms are a primary source of information during times of emergency, and as such are also used by authorities to communicate reliable news, updates, and medical instructions. Social media is also a ground for the growth and distribution of misinformation in times of emergency [2]. Misinformation can take various forms, such as conspiracy theories, rumors, fake news, and others. The ability of misinformation to spread globally and easily through social media can adversely influence how people perceive and mitigate various risks or behave in emergencies. The incorrect perception or insufficient mitigation of risks during an emergency can have severe consequences and may even lead

to an unnecessary loss of life [3–6]. Fighting the spread of misinformation and promoting reliable information is, therefore, a very important effort.

X (formerly Twitter) is one of the most popular social media platforms. Users on X (formerly Twitter) mainly share their thoughts, creative ideas, and reactions to other users through short texts [7]. Prior studies found that high virality of tweets is associated with low lexical tweet density during a crisis [8], and that posts on social media with easy-to-read text tend to achieve a higher performance in terms of engagement and awareness of users [9]. One of the core factors of an increase or decrease in the spread of misinformation and not-misinformation on social media is user engagement. Positive user engagement amplifies the spread of content [10,11].

Unlike traditional media, social media typically provides quantified metrics of how many users have engaged with each piece of content [12]. When cues are shown, a larger number of others have engaged with a post, and users are more likely to share and like that post, and the presence of social cues increases the sharing of true over false news, relative to a control without social cues [12].

The readability of text is an important factor in how users perceive the information it provides. Participants trusted the information more in texts that were more difficult to read, regardless of whether the information presented was correct or incorrect [13].

However, the various aspects of readability have not been discussed and compared for tweets spreading misinformation and tweets not spreading misinformation. An understanding and comparison of linguistic understandability, i.e., lexicon-related and grammar-related indicators, are still missing. Following prior research, the level of influence of tweet readability on the user engagement with misinformation and not-misinformation tweets remains unknown.

This work uses different natural language processing methods to classify a large dataset of over 1.4 M tweets related to misinformation regarding the monkeypox outbreak. The classified dataset is analyzed using numerous readability metrics to assess the usefulness and importance of readability in fighting misinformation. The findings enable policy recommendations regarding how, when, and where to focus the fight against misinformation in order to achieve efficient results.

This study is conceptually grounded in diffusion theory, which explains how new ideas spread through a population. We investigate whether using principles from the diffusion of innovation theory, which explains how an innovation (in this case, information and misinformation related to monkeypox) diffuses in a social network [14].

Existing research on the diffusion theory mostly emphasizes network structures, user susceptibility, and social cues as the main drivers of information spread, while giving less attention to the textual properties of the messages themselves [15]. Our work extends diffusion theory by investigating readability and grammatical correctness as drivers of engagement, showing how linguistic form shapes the spread of both accurate and inaccurate health information online. Most computational models of news sharing assume that users forward content primarily based on prior exposure within their network [16]. However, recent social science research highlights that sharing decisions are also influenced by user attributes such as demographics and behaviors, message features such as title and content, and the broader social context, such as follower counts and tie strength [7].

The innovation of this study lies in integrating readability and grammatical correctness into the framework of misinformation diffusion, thereby providing both theoretical and practical insights into how message form influences engagement and spread.

We therefore present the following research questions to determine the level of influence the readability of tweets has on the user engagement they receive. Answering

these research questions would allow us to derive policy recommendations to increase user engagement with not-misinformation and therefore fight misinformation more efficiently.

1. *How do the temporal patterns of monkeypox misinformation and not-misinformation tweets evolve over time?* [RQ1]
2. *How do the readability and grammatical accuracy of monkeypox misinformation tweets compare to those of not-misinformation tweets?* [RQ2]
3. *What is the impact of tweet readability on user engagement with monkeypox misinformation and not-misinformation tweets?* [RQ3]

2. Literature Review

2.1. Misinformation

Misinformation associated with health suggestions that prevent someone from acquiring diseases is more likely to be shared during health emergencies, when people pay more attention to their health condition [6]. Misinformation is more likely to be shared with others to fill the information gap when updates from authorities or the media are missing or incomplete [4]. Health-related misinformation, as well as the general discussion on social media, was studied extensively [17]. Research in the field of misinformation has focused on the dynamics of the discussion on social media and analyzed different aspects of misinformation and conspiracy theories [18]. Previous studies investigated the differences between misinformation and not-misinformation tweets [11,19–23]. Other studies developed machine learning methods to identify misinformation tweets and mine different metadata of tweets [11,24]. The metadata and characteristics of tweets and their authors are useful to achieve a better understanding of behavioral patterns. Beskow and Carley [25] used the number of users that follow the author and the number of users the author follows as an indication of whether the author is a robot or not. O'Donovan, et al. [26] and Gupta, et al. [27] found that the metadata of tweets, such as URLs (Uniform Resource Locator), mentions, retweets, and tweet length, may serve as indicators for credibility. Elroy and Yosipof [11] found that certain features, such as the presence of a URL and the sentiment score of a tweet, can help identify conspirative tweets.

2.2. User Engagement

There are several definitions for user engagement. Jacques [28] defined user engagement as *“a user’s response to an interaction that gains, maintains, and encourages their attention, particularly when they are intrinsically motivated”*. O’Brien [29] defined user engagement as *“a category of user experience characterized by attributes of challenge, positive affect, endurance, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control”*. Atfield, et al. [30] defined it as *“the emotional, cognitive and behavioural connection that exists, at any point in time and possibly over time, between a user and a resource.”*

User engagement on X (formerly Twitter) is typically quantified by the number of retweets, likes, quotes, and replies. The reading ease of posts on social media was found to be correlated with the awareness and engagement of users [9]. Likes and retweets are more relevant engagement metrics than quotes and replies for measuring the extent to which scholarly tweets of scientific papers are engaged with by X (formerly Twitter) users [31]. The usefulness and importance of readability metrics in identifying misinformation are a subject of interest in existing research. Papakyriakopoulos and Goodman [32] investigated the impact of X (formerly Twitter) labels on misinformation spread and user engagement, as measured by likes, retweets, quotes, and replies. Segev [33] investigated how the expression of feelings and the use of personal pronouns on X (formerly Twitter) are associated with user engagement and found a significant difference between retweeting (sharing) and replying

to a tweet. Zhao and Buro [34] explored the effect of content types on user engagement and compared the number of likes and retweets between text-only and multimedia tweets.

2.3. Readability

The definition of readability is the ease of understanding or comprehension due to the style of writing, and is an attribute of clarity. The focus is on writing style and vocabulary rather than issues such as content, design, and organization, and consideration is not given to the correctness of the information delivered [35].

The Flesch Reading Ease (FRE) and Flesch–Kincaid Grade Level (FKGL) are commonly used tests to assess the readability level of text. The readability is generally higher (more difficult) for false information than true information, and the average number of syllables per word and the average number of characters per word work best with short documents [36]. The difficulty of reading text affects the trust given to the information it provides [13].

Readability metrics were also previously used as features for the classification of text as fake news [37,38]. Grammatical correctness is another metric that can be used to assess the readability of text. The Corpus of Linguistic Acceptability (CoLA) [39] is a well-recognized dataset of sentences from published linguistics literature that are labeled as grammatical or ungrammatical. CoLA is used as one of the tasks in the General Language Understanding Evaluation (GLUE) benchmark for natural language understanding [40]. BERT models fine-tuned on the CoLA dataset were previously used to determine whether news is grammatically correct, as part of a system to detect fake news [41].

2.4. Synthesis

Three strands of literature, namely misinformation, user engagement, and readability, converge to form the foundation of this study. Prior work on misinformation highlights how false health content spreads rapidly in emergencies, often filling gaps left by official communication [11,19]. Research on user engagement shows that the degree to which users interact with content is a central mechanism driving its visibility and diffusion on social media. Studies on readability demonstrate that textual clarity and grammatical correctness affect how information is processed and trusted. Taken together, these insights suggest that misinformation does not spread solely because of social or network factors but also because of characteristics embedded in the text itself. Readability and grammar influence whether users engage with a message, and engagement in turn determines the extent of its diffusion. While each of these factors has been studied separately, few works have integrated them to explain how misinformation and accurate information compete for visibility [21,42,43]. This study addresses that gap by examining the relationship between readability, user engagement, and misinformation in the context of the monkeypox outbreak.

3. Methodology

To answer the first research question, we collected 1,437,724 tweets related to the discussion of monkeypox on X (formerly Twitter) (see the Section 3.1) and trained a semi-supervised learning classifier with a dynamic confidence threshold to classify the dataset (see the Section 3.3).

To answer the second and third research questions, we further calculated three features that indicate the readability and grammar level of each tweet. We used a generalized linear model with negative binomial distribution (see the Section 3.5) to evaluate the relationship between tweet readability levels, grammar correctness, and user engagement.

3.1. Dataset

We collected a dataset of tweets using X's (formerly Twitter) academic research API, which enables access to the full archive of X (formerly Twitter). The dataset includes all tweets in English that contain the term "monkeypox" between 15 May 2022 and 24 August 2022 and excludes retweets. This period aligns with the early emergence and heightened discussion of monkeypox, capturing the initial rise in public attention and the evolving responses over the first few critical months, ensuring the dataset reflects the key stages of early discourse without extending beyond the point when interest and coverage began to stabilize. The search criteria are simple and very effective in filtering tweets related to the discussion on monkeypox, while enabling us to collect a large amount of data without much noise.

For performance optimization, the dataset was limited to tweets that were 350 characters or shorter after preprocessing. The preprocessed dataset consisted of 1,437,724 tweets related to discussion on monkeypox that were posted by 504,474 users.

3.2. Natural Language Processing

A fundamental task in analyzing differences between misinformation and not-misinformation tweets is to identify tweets that are misinformation and tweets that are not misinformation. The substantially large dataset required us to label and classify the tweets using a machine learning algorithm [11,19,44,45] to train a classifier that can classify the tweets, and a set of features was needed.

A classifier can be based on different features of the text, such as embedding, the presence of specific keywords, hashtags, or the metadata of tweets. The language model BERT—Bidirectional Encoder Representations from Transformers [46]—provides superior results for different NLP tasks, including word embedding [47,48]. Multiple variations of BERT with different strengths and weaknesses are available for a variety of tasks.

BERT embeddings are widely used in research that analyzes misinformation. For example, Micallef et al. used BERT embeddings to investigate and counter misinformation in tweets related to COVID-19 over a period of five months. Baruah, et al. [49] used the BERT embeddings of all tweets by authors to detect spreaders of misinformation.

RoBERTa was pretrained using different design decisions than used for BERT that improve the performance and state-of-the-art results on different datasets [50]. Evaluations of RoBERTa found that it provides better results than BERT [51–53]. Models that were pretrained on domain-specific data provide better embeddings [54].

Sentence-BERT is a modification of the pre-trained BERT network that uses Siamese and triplet network structures on top of the BERT model and is fine-tuned based on high-quality sentence interface data to learn more sentence-level information [55]. Sentence-BERT transforms word embeddings into sentence embeddings.

Elroy and Yosipof (2022) [11] transformed BERT word embeddings into sentence embeddings using Sentence-BERT to train a classifier and classify a dataset of over 300 K tweets related to the COVID-19 5G conspiracy theory. Elroy, Erokhin, Komendantova, and Yosipof [19] fine-tuned a RoBERTa model to classify and analyze a dataset of over 1.4 M tweets related to discussions on monkeypox misinformation. Karande, et al. [56] performed stance detection to analyze the credibility of information on social media at the early stage of publication and before its dissemination through social media.

3.3. Classification

To train a classifier and classify the complete dataset into misinformation, not-misinformation, or neutral, we hand-labeled 3386 randomly selected tweets, based on facts provided by the World Health Organization [1]. The labeling was carried out by two

experts in the field, and disagreements between the classifications were resolved by a third person. Table 1 presents the number and examples of tweets in each category.

Table 1. Number and examples of tweets in each category.

Category	Number of Tweets	Example Tweet
Misinformation	1128	“Monkeypox release is PSYCHOLOGICAL TERRORISM to keep humanity paralyzed with FEAR”
Not-misinformation	737	“WHO clarifies monkeypox not a sexually transmitted disease”
Neutral	1521	“New York City investigating possible monkeypox case”

A relatively small number of labeled samples, about 0.2% of the dataset, can produce decent results. However, a model that is trained on a very small subset of the dataset is likely to be biased. To address this issue, we used proxy-label semi-supervised learning (SSL) methodology and added dynamic confidence requirements for the pseudo-labels. The first advantage of this approach is that it balances an imbalanced dataset of pseudo-labels. Second, instead of treating pseudo-labels equally and down-sampling the dataset by randomly removing data, the methodology prioritizes pseudo-labels with higher prediction confidence and therefore increases the overall quality of the pseudo-labels.

To train a machine learning classifier, we calculated the word embeddings and transformed those into sentence embeddings using RoBERTa and Sentence-BERT, respectively, resulting in 768 features for each tweet. We trained an SSL model based on a k-nearest-neighbors (kNN) model with hyper-parameters that were optimized in Elroy and Yosipof [57]. We therefore ran kNN with the 5 nearest neighbors using the labeled dataset, where predictions with a probability of over >0.8 were assumed to be correct for all three labels and qualified as pseudo-labels.

To address imbalances in the pseudo-labeled dataset, we applied a dynamic confidence requirement for pseudo-labels. The probability threshold was initially set to 0.8 for all groups. The threshold for the larger groups was increased, therefore reducing the number of qualifying pseudo-labeled tweets in those groups, until a balance was achieved with the smallest group. In cases where the dynamic threshold reached 1 and a group was still bigger than the smallest group, the subset was downsampled by randomly removing a subset of the pseudo-labels. The outcome was a balanced dataset.

We pseudo-labeled 106,343 tweets as misinformation with a confidence level of >0.8 , 119,668 tweets as not-misinformation with a confidence level of >0.8 , and 134,735 tweets as neutral or irrelevant with a confidence level of 1. The larger groups were downsampled to 106,343 tweets in each group, and finally, the manually labeled tweets were appended. The training dataset for the classifier, therefore, consisted of 107,471 tweets labeled as misinformation, 107,080 tweets labeled as not-misinformation, and 107,864 tweets labeled as neutral or irrelevant. We evaluated the model using 5-fold cross-validation on a balanced dataset of 331,620 tweets to achieve an average F1 score of 0.938 ± 0.011 , an average precision of 0.940 ± 0.01 , and an average recall of 0.938 ± 0.011 .

3.4. Readability and Grammar Features

We enriched the dataset with three readability and grammar indices for each tweet, namely the Flesch Reading Ease (FRE) and Flesch–Kincaid Grade Level (FKGL) scores and the probability of the tweets being grammatically correct, using a model that was fine-tuned on the Corpus of Linguistic Acceptability (CoLA) dataset.

The *Flesch Reading Ease (FRE)* test provides a score between 0 and 100, representing the difficulty of reading the given text based on several constants and the number of words, sentences, and syllables. Higher scores indicate the text is easier to read, e.g., text scored

between 90 and 100 can be easily understood by an average 11-year-old. Lower scores indicate the text is harder to read, where a score below 30 is very difficult to read and is best understood by university graduates. Let W be the total number of words in the text, S be the total number of sentences, and Y be the total number of syllables; the FRE score is calculated according to Equation (1).

$$206.835 - 1.015\left(\frac{W}{S}\right) - 84.6\left(\frac{Y}{W}\right) \quad (1)$$

The *Flesch Kincaid Grade Level (FKGL)* test provides a grade-level score, such that text with an FKGL score of 6 means that the reader needs to have a sixth-grade reading level to understand it. The ease of interpretation of the score makes it more usable in various applications. Text intended for the public should generally aim for a grade level of around 8, equivalent to a schooling age of 13 to 14, which is the mean US adult reading level [58]. The FKGL score is calculated according to Equation (2). Let W be the total number of words in the text, S be the total number of sentences, and Y be the total number of syllables.

$$0.39\left(\frac{W}{S}\right) + 11.8\left(\frac{Y}{W}\right) - 15.59 \quad (2)$$

The *Grammar Correctness Index (GCI)* represents the grammatical correctness of the data. To calculate the GCI, we used a RoBERTa model that was fine-tuned on the CoLA dataset [59]. The model predicts whether text is grammatically correct or incorrect, as well as the probability of the prediction. The prediction probabilities of the model were normalized, such that the GCI is the grammatical correctness of the text on a scale of 0 to 1.

3.5. Estimation Approach

In order to evaluate the relationship between readability, grammar level, and user engagement, we evaluated three dependent variables, namely the number of retweets, likes, and replies of misinformation and not-misinformation tweets, separately. Considering the dependent variables were discrete, we used a count data model. We implemented a negative binomial regression model, rather than a Poisson model, as the latter assumes equality between the conditional mean and conditional variance [60], which does not characterize the distribution of the retweet, like, and reply variables. We used four control variables and three readability and grammar variables. The four control variables were the number of users the user is following, the number of followers, a binary variable for the presence of a URL in the tweet, and a binary variable for the presence of media in the tweet. The three readability and grammar features were the *FRE*, *FKGL*, and *GCI*.

4. Results

We assigned the unlabeled dataset to the classifier that was trained on the labeled and pseudo-labeled dataset (see the Section 3.3). The results show that 425,656 of the tweets were misinformation, 524,602 were not misinformation, and 487,466 were neutral or irrelevant to the discussion on misinformation related to the monkeypox virus. The results of the classifier can be seen in Table 2.

Table 2. The number of tweets and users in each category.

Category	Number of Tweets	# Users	Tweets/User
Misinformation	425,656	235,655	1.806
Not-misinformation	524,602	249,417	2.103
Neutral	487,466	168,473	2.893

The comparison between the daily tweet frequency in the misinformation and not-misinformation groups reveals three peaks in both groups. Figure 1 presents the daily number of tweets in each group.

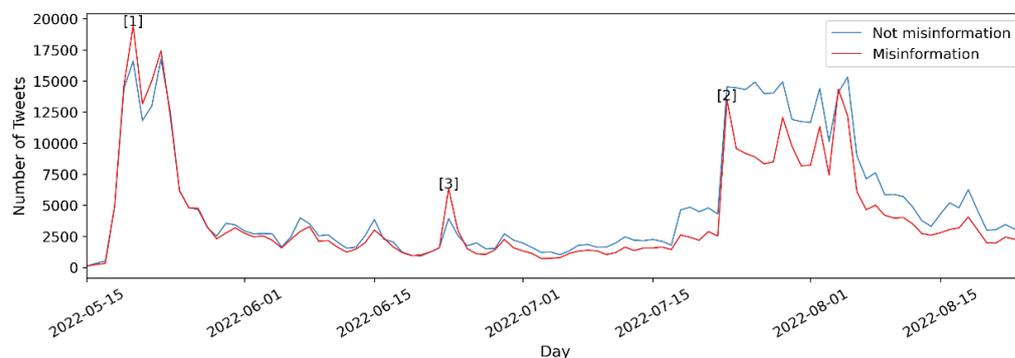


Figure 1. Daily frequency time series of misinformation and not-misinformation tweets. The red line represents the misinformation tweets, and the blue line represents the not-misinformation tweets.

The two major peaks (Figure 1, annotations 1 and 3, respectively) are likely related to the epidemiological evolution of the virus. The first peak was during the second half of May 2022, when, according to data from the WHO, the number of new cases reported globally started to grow exponentially [61]. The second peak, in July and August 2022, took place during a peak of new confirmed cases. During the first peak (Figure 1, annotation 1) in May 2022, the misinformation group dominated the conversation. However, after the end of this peak, the not-misinformation group showed, almost consistently, more volume until the end of the examined period. One exception can be seen during a local peak (Figure 1, annotation 3), where the misinformation group had seen a one-sided peak in interest. The shift in dominance became clearer at the beginning of the second peak (Figure 1, annotation 2), possibly marking the last cycle of misinformation [11].

We tested for cross-correlation between the daily tweet frequency of the misinformation and not-misinformation groups. The results show that the time series are correlated at time t with an $r = 0.947$.

We used Ordinary Least Squares (OLS) regression to investigate whether the daily tweet frequency of each group at time t could be predicted using the other group's frequency at times t , $t - 1$, $t - 7$, and $t - 14$. We applied the OLS time series model in Equation (3) to estimate the effect of the daily misinformation tweet frequency on the daily not-misinformation tweet frequency, and vice versa. Table 3 presents the results of the OLS regressions. The regression results show that both time series can be significantly predicted using the other time series at times t , $t - 1$, and $t - 7$. These findings mean that the tweet frequency of each group on a given day can be predicted using the other group's tweet frequency for the time periods of today, yesterday, and one week ago. The coefficient of determination in both models was found to be very high.

$$y_t = a + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 x_{t-7} + \beta_4 x_{t-14} \quad (3)$$

The regression results show that the daily tweet frequencies of misinformation and not-misinformation were significantly and mutually affected by each other at time t ($\beta = 1.01$, p -value < 0.001 , and $\beta = 0.93$, p -value < 0.001 , respectively). An increase in the daily tweet frequency of misinformation at time $t - 1$ would positively affect the daily tweet frequency of not-misinformation at time t ($\beta = 0.37$, p -value < 0.001). However, an increase in the not-misinformation at time $t - 1$ would negatively affect the daily tweet frequency of misinformation at time t ($\beta = -0.26$, p -value < 0.001). An increase in either daily tweet

frequency at time $t - 7$ would mean a decrease in the other daily tweet frequency at time t ($\beta = -0.07$, p -value < 0.01 for the prediction of misinformation, and $\beta = -0.07$, p -value < 0.01 for the prediction of not-misinformation). The behavior at time $t - 14$ of either daily tweet frequency would not significantly affect the other daily tweet frequency. In addition, the Granger causality test was performed between the misinformation and the not-misinformation tweet frequencies at different lags. Our analysis shows that misinformation tweets Granger-cause not-misinformation tweets at a lag of 7 days. These findings support the hypothesis that community responses are triggered promptly following the spread of misinformation, contributing to the overall discourse. Conversely, not-misinformation tweets also Granger-cause misinformation tweets at the same lag, suggesting an intricate feedback loop in the online discourse.

Table 3. OLS regression results for the daily frequency of misinformation and not-misinformation tweets.

	Model 1 Misinformation_t	Model 2 Not-Misinformation_t
Constant	93.22	172.54
Misinformation _t		1.01 ***
Misinformation _{t-1}		0.37 ***
Misinformation _{t-7}		-0.07 *
Misinformation _{t-14}		0
Not-misinformation _t	0.93 ***	
Not-misinformation _{t-1}	-0.26 ***	
Not-misinformation _{t-7}	-0.07 **	
Not-misinformation _{t-14}	-0.02	
N	88	88
R ²	0.945	0.95

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

However, these interactions may also be influenced by external factors, such as the progression of the monkeypox outbreak, media coverage, or announcements by public health authorities. These drivers could simultaneously affect both misinformation and not-misinformation tweet volumes, potentially confounding the observed relationships. We compared the daily average GCI, FRE, and FKGL scores in both the misinformation and not-misinformation groups, and tested the significance of the differences using a statistical paired t -test. Figure 2 presents the daily average GCI, FRE, and FKGL scores of tweets in each group. The first graph (Figure 2A) is the daily average Grammar Correctness Index (GCI) score of tweets spreading misinformation and tweets that are not misinformation. The second graph (Figure 2B) is the daily average Flesch Reading Ease (FRE) score of tweets spreading misinformation and tweets that are not misinformation. The third graph (Figure 2C) is the daily average Flesch–Kincaid Grade Level (FKGL) score of tweets spreading misinformation and tweets that are not misinformation.

The GCI and FRE scores over time show a clear difference between the misinformation and not-misinformation groups, with consistently higher scores observed for the not-misinformation group. However, for the FKGL score, the differences over time are not immediately visible.

The reading ease of tweets in the not-misinformation group was more stable over time, with a standard deviation of only 1.112 for the daily average FRE score, compared to a standard deviation of 2.344 for the misinformation group.

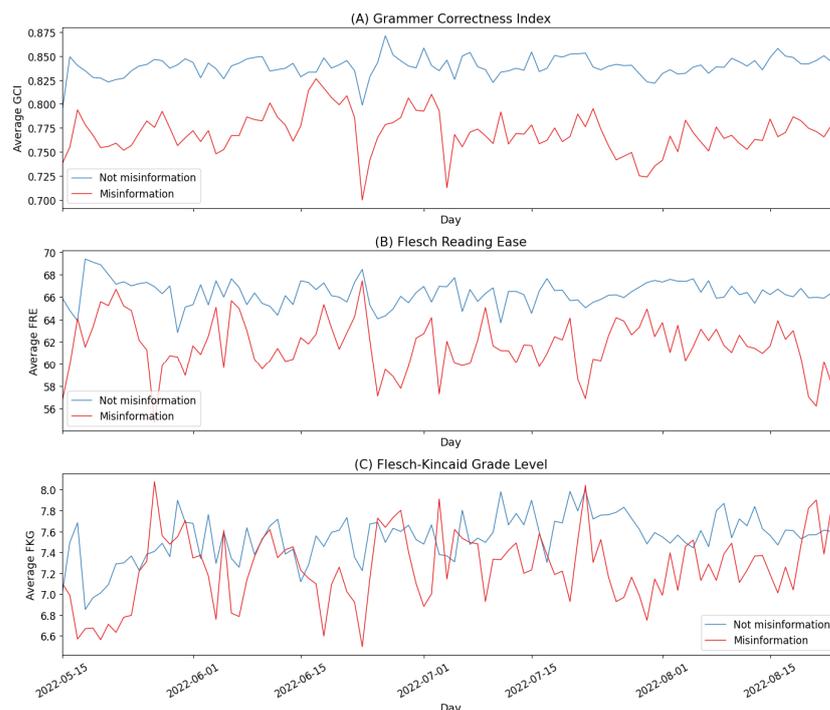


Figure 2. Daily average GCI, FRE, and FKGL scores of misinformation and not-misinformation tweets. The red line represents misinformation tweets, and the blue line represents not-misinformation tweets.

We found significant differences between the daily average FRE, FKGL, and GCI scores of the misinformation group and not-misinformation group. The results show that all daily average scores of tweets in the not-misinformation group were significantly higher than those in the misinformation group. Table 4 presents the paired *t*-test results between the daily average FRE, FKGL, and GCI scores for tweets in both groups.

Table 4. Paired *t*-test between the daily average FRE, FKGL, and GCI scores of tweets in the not-misinformation and misinformation groups.

	Not-Misinformation (Mean ± SD)	Misinformation (Mean ± SD)	<i>t</i> -Test	DF
FRE	66.338 ± 1.112	61.644 ± 2.344	21.616 ***	101
FKGL	7.543 ± 0.219	7.25 ± 0.339	8.883 ***	101
GCI	0.839 ± 0.011	0.77 ± 0.021	35.566 ***	101

*** *p* < 0.001.

The daily average FRE scores in the not-misinformation group (66.338 ± 1.112) were significantly higher (*t*-test = 21.616, *p*-value < 0.001) than those in the misinformation group (61.644 ± 2.344). The daily average GCI score was also significantly higher (*t*-test = 35.566, *p*-value < 0.001) in the not-misinformation group (0.839 ± 0.011) than in the misinformation group (0.77 ± 0.021). These findings mean that the not-misinformation tweets were easier to read and are well written. We also found significantly higher FKGL scores (*t*-test = 8.883, *p*-value < 0.001) for tweets in the not-misinformation (7.543 ± 0.219) group than for those in the misinformation group (7.25 ± 0.339). While higher FKGL scores mean more user engagement, they also mean the text requires a higher level of education to properly understand it. The average FKGL scores of tweets in the misinformation and not-misinformation groups show that misinformation tweets typically required lower grade levels to understand.

The level of user engagement drawn by tweets was measured by the number of retweets, replies, and likes they receive from other users. We assessed the significance of

readability metrics of each tweet for the prediction of the level of engagement it drew. The models were built separately for each group. A generalized linear model with negative binomial distribution was used as all of the target variables inherently fit the negative binomial distribution. In addition to the exploratory variables, namely the GCI, FKGL, and FKGL scores, we also used the number of users following the posting user, the number of users the posting user is following, and whether the tweet is accompanied by a URL or media as control variables. The number of users following the posting user and the number of users the posting user is following were normalized using the log of the value plus one. Table 5 presents the coefficients, standard deviation of the error, and *p*-values of the regression models.

Table 5. Regression models using readability metrics to predict the number of retweets, likes, replies, and tweets received.

Not-Misinformation			
Variable	Retweets	Replies	Likes
Constant	−5.814 *** (0.025)	−3.616 *** (0.028)	−3.24 *** (0.02)
Followers	0.773 *** (0.001)	0.465 *** (0.001)	0.621 *** (0.001)
Following	−0.104 *** (0.001)	−0.086 *** (0.002)	−0.093 *** (0.001)
Has URL	−0.953 *** (0.005)	−0.961 *** (0.005)	−1.451 *** (0.004)
Has Media	1.216 *** (0.007)	0.789 *** (0.007)	1.242 *** (0.006)
GCI	0.432 *** (0.01)	0.297 *** (0.011)	0.437 *** (0.008)
FRE	0.011 *** (0)	0.009 *** (0)	0.016 *** (0)
FKGL	0.085 *** (0.001)	0.04 *** (0.001)	0.097 *** (0.001)
Pseudo R ²	0.8924	0.5965	0.916
Constant	−5.814 *** (0.025)	−3.616 *** (0.028)	−3.24 *** (0.02)
Misinformation			
Variable	Retweets	Replies	Likes
Constant	−6.572 *** (0.025)	−5.406 *** (0.031)	−2.917 *** (0.019)
Followers	0.724 *** (0.001)	0.57 *** (0.001)	0.712 *** (0.001)
Following	−0.19 *** (0.002)	−0.069 *** (0.002)	−0.235 *** (0.001)
Has URL	−0.873 *** (0.006)	−1.002 *** (0.007)	−1.543 *** (0.004)
Has Media	1.794 *** (0.006)	1.014 *** (0.008)	2.103 *** (0.005)
GCI	0.01 (0.009)	0.372 *** (0.011)	−0.405 *** (0.007)
FRE	0.03 *** (0)	0.013 *** (0)	0.02 *** (0)
FKGL	0.183 *** (0.001)	0.07 *** (0.002)	0.121 *** (0.001)
Pseudo R ²	0.8271	0.6292	0.9177

*** *p* < 0.001.

As expected, the number of followers the posting user has positively and significantly affected the engagement the tweet received. However, following more users negatively and significantly affected the level of engagement, which generally aligns with the characteristics of recognized users that typically have many followers but do not follow many other users. The number of users a user can follow is also limited in certain ways to prevent abuse, which is often associated with automated robots.

The presence of a URL in a tweet usually aims to redirect the reader to other, external websites, which, if successful, disengages the user from the platform and, as can be expected, reduces engagement on the platform. Conversely, attaching media to a tweet draws more attention to the platform and therefore significantly increases the level of engagement on the platform. These results were consistent for all six models that were built.

The results show that all of the tested independent variables, namely the GCI, FRE, and FKGL scores, shared significant and positive coefficients in the not-misinformation group. That is, higher-grammatical-correctness, easier-to-read, and higher-grade-level scores led to more engagement in all tested forms from other users regarding not-misinformation tweets. For tweets in the misinformation group, the results show similar behavior regarding the FRE and FKGL scores. However, the results for the effect of the GCI score on the level of engagement are mixed, where the coefficient was positive for engagement in the form of replies, but negative for engagement in the form of likes. The results are insignificant for engagement in the form of retweets. To further explore the influence of poster identity on tweet engagement, we stratified the dataset into three groups based on the number of followers: less than 20, between 20 and 22,000, and more than 22,000. Separate regression models were run for each group to assess whether readability indices (FRE, FKGL) and grammatical correctness (GCI) predicted engagement metrics (likes, retweets, and replies).

The analysis revealed similar results: readability and grammatical correctness remained significant predictors of engagement across all follower groups.

5. Discussion

Countering misinformation and conspiracy theories may demand different forms of engagement. Countering misinformation likely requires more discussion and debate to explain and correct misconceptions, whereas countering conspiracy theories likely requires spreading more valid information to dismiss the conspiracy. In the context of the not-misinformation group, we found that regardless of the desired form of engagement, all three metrics were important and significantly and positively affected the level of engagement.

Users typically interact only with like-minded others, in communities dominated by a single view [62], and adapt to the dominant opinion within the respective media outlet. Misinformation poses a risk of dominating the discussion with false information and contributing to public confusion and panic during an outbreak. Tweets that are not misinformation play a crucial role in mitigating the impact of false information and providing accurate and reliable information to the public. According to the findings related to RQ1, fighting misinformation is not enough on its own, and a policy of acting against misinformation as soon as it is spotted on a platform, before it becomes the dominant opinion, is essential.

Social media in general and X (formerly Twitter) in particular are characterized by incorrect grammar, spelling mistakes, and standard and non-standard abbreviations [63,64]. The difference in grammatical correctness between the groups could be caused by many different reasons. While some causes may be attributed to the characteristics of the writers, other causes may be intentional, such as attempts to contain longer texts within the length limitation. Basic reading ease tests are not designed for tweets, which commonly make use of abbreviations, hashtags, and more, which can affect the test results. However, all

groups were reasonably assumed to be affected in the same way. We used these scores to compare between groups empirically. We found that there were statistically significant differences between the FRE, FKGL, and GCI scores of tweets in the misinformation and not-misinformation groups, where the scores were higher in the not-misinformation group.

Our findings can also be interpreted through the lens of the diffusion of innovation theory, which posits that the characteristics of a message, such as clarity, complexity, and perceived advantage, play a central role in determining its adoption and spread. Costantini and Fuse [65] emphasized that readability is crucial for the comprehension of and trust in vaccination information. Our results suggest that the diffusion of health-related content, whether accurate or misleading, is not determined solely by network effects or social influence but also by the linguistic features of the messages themselves. Our contribution extends this perspective by showing that readability and grammatical correctness actively shape user engagement, which in turn drove the diffusion process during the monkeypox outbreak.

Theoretical and Practical Implications

The reading ease scores for tweets in both groups showed that not-misinformation tweets were consistently easier to read, with a single exception at the beginning of the discussion. Higher scores carry different meanings in the different tests: a higher FRE score means the text is easier to read, while a higher FKGL score means the text is better understood by a higher grade level, and hence harder to understand. The results indicate that to maximize engagement from users, a text should be targeted towards a higher grade level and remain easy to read at the same time. For tweets in the not-misinformation group, grammatical correctness also plays a key role.

The number of followers of a user positively and significantly affects the engagement their tweets receive, and following more users negatively affects the level of engagement. One possible explanation for this is that not having a substantial number of followers indicates a user is not well recognized, and following too many users may indicate the user is abusive. This finding supports previous research by Jackson and Yariv [66], who found that a large enough initial group enables a message to spread to a significant portion of the population; otherwise, the behavior collapses, so no one in the population chooses to adopt the behavior. Translated into a policy recommendation, this suggests that users in charge of publishing official or authoritative information should aim to have more followers, and at the same time, follow as few users as possible.

Furthermore, the usage of URLs reduces user engagement, which may imply that the use of URLs should be avoided. However, when a user disengages from X (formerly Twitter) because they follow a link, they shift their focus to an external platform that was referred to in the tweet. Redirecting a user's focus towards a different website could help in educating certain users with valuable and confirmed information, but at the same time, may reduce the user engagement and therefore the exposure the tweet receives in general.

The results show that the GCI, FRE, and FKGL all positively and significantly increase user engagement for tweets in the not-misinformation group. The results are mixed for the misinformation group. This finding means that higher grammatical correctness, ease of reading, and higher grade level result in more user engagement for not-misinformation tweets. Generally, tweets that received high readability scores also received high user engagement. For example, in the not-misinformation group, the tweet *"monkeypox carries an isolation period of 3–4 weeks and that is obviously more sick days than most workers get per year. so this is going to be very very bad"* had CoLA, FRE, and FKGL scores of 0.986, 73.17, and 6.8, respectively, and received 493 replies, 25,286 retweets, and 152,159 likes. Likewise, the tweet *"Monkeypox is neither an STD nor a gay disease. DOH, in its statement yesterday, emphasized that ANYONE MAY GET MONKEYPOX. Warn. Do not stigmatize and harm."* had

CoLA, FRE, and FKGL scores of 0.98, 71.1, and 5.5, respectively, and received 273 replies, 11,042 retweets, and 37,470 likes. The same is true for tweets in the misinformation group, where, for instance, the tweet *“When smallpox and monkeypox stocks are going up 70% like they just did today, you can bet there is something going on behind the scenes. Looks like they are doing it to you again.”* had CoLA, FRE, and FKGL scores of 0.992, 79.6, and 6.4, respectively, and received 803 replies, 6709 retweets, and 21,914 likes. This finding is also true for tweets in the irrelevant group, such as *“BREAKING: USA becomes the first country in the world to surpass 5000 monkeypox cases after the CDC added 282 new cases. 30 days ago the USA had less than 400 cases.”*, which had CoLA, FRE, and FKGL scores of 0.963, 81.12, and 5.8, respectively, and received 432 replies, 10,750 retweets, and 33,063 likes.

Therefore, to maximize user engagement for tweets that promote correct information, it can be recommended to use easy-to-read text, with a higher grade level difficulty, and to keep grammatical correctness to a maximum. The types of text that shared these features in the context of this work were typically academic research and perhaps medical analyses. These findings are also in line with the results of Withall and Sagi [13].

6. Conclusions

We found that in order to fight misinformation effectively, measures have to be taken as early as possible, and preferably before misinformation dominates the discussion. Considering the structure of social media platforms, substantial user engagement is an important factor for success in this task.

Following the findings and discussion in this work, it can be recommended that not-misinformation tweets aspire for maximal user engagement. To increase user engagement, not-misinformation tweets are advised to be written in language that is easy to read (higher FRE score), requiring a higher grade level to understand (higher FKGL score), and that is more grammatically correct (higher GCI score). In the context of this work, certain types of text inherently share these features, such as academic research and other scientific papers, like medical analyses. It is therefore recommended to make use of such content more often.

We also found that using URLs in tweets reduces user engagement, possibly because the focus of the user is shifted outside of the platform. The tradeoff between using and not using URLs in not-misinformation tweets, and thus that between the wider dissemination of not-misinformation through higher user engagement and being deeply thorough in providing more information to the user, should be considered when making policy decisions. The use of media in not-misinformation, however, was found to significantly increase user engagement, likely because it draws the users' attention. The benefit of using media is that it does not disengage the user from X (formerly Twitter), and does not have an evident downside, and it is therefore recommended.

This work is limited to the discussion in English regarding the monkeypox virus outbreak that took place on X (formerly Twitter) over about 4 months. This work focuses on specific metrics for user engagement and readability. Furthermore, short or sarcastic tweets may be misclassified, as brevity and irony obscure their intended meaning, limiting the model's accuracy. Future works can expand the analysis to other languages, topics, and platforms, such as Facebook, TikTok, or others, as well as consider other metrics for user engagement, such as the number, variety, or characteristics of the users who retweeted, liked, or replied to the tweets investigated in this work. Future works could also investigate the difference in the impact of the wide and deep dissemination of not-misinformation, possibly at different times during the life cycle of misinformation. A valuable direction for future research would be to incorporate user-centered methods, such as surveys, interviews, or focus groups, to directly capture how audiences perceive and respond to these tweets.

Combining such audience-oriented approaches with textual analysis could yield a more comprehensive understanding of engagement.

Author Contributions: Conceptualization: O.E. and A.Y. Resources (dataset): O.E. and A.Y. Methodology: O.E. and A.Y. Investigation (analysis): O.E. and A.Y. Writing—original draft: A.Y. and O.E. Writing—review and editing: A.Y. and O.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 101021746, CORE (science and human factor for resilient society).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The tweet dataset was collected from X (formerly Twitter) using limited academic research API access. The tweet IDs are available at <https://github.com/elroyo/mpox> (accessed on 21 August 2025). Other data that supports the results is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

BERT	Bidirectional Encoder Representations from Transformers
CoLA	Corpus of Linguistic Acceptability
FKGL	Flesch–Kincaid Grade Level
FRE	Flesch Reading Ease
GCI	Grammar Correctness Index
GLUE	General Language Understanding Evaluation
kNN	k-Nearest-Neighbors
NLP	Natural Language Processing
OLS	Ordinary Least Squares
RoBERTa	Robustly Optimized BERT Approach
SSL	Semi-Supervised Learning
URL	Uniform Resource Locator
WHO	World Health Organization

References

1. World Health Organization. Questions and Answers: Monkeypox. Available online: <https://www.who.int/news-room/questions-and-answers/item/monkeypox> (accessed on 2 February 2023).
2. Lazer, D.M.J.; Baum, M.A.; Benkler, Y.; Berinsky, A.J.; Greenhill, K.M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. The science of fake news. *Science* **2018**, *359*, 1094–1096. [CrossRef] [PubMed]
3. Kwanda, F.A.; Lin, T.T. Fake news practices in Indonesian newsrooms during and after the Palu earthquake: A hierarchy-of-influences approach. *Inf. Commun. Soc.* **2020**, *23*, 849–866. [CrossRef]
4. Peary, B.D.; Shaw, R.; Takeuchi, Y. Utilization of social media in the east Japan earthquake and tsunami and its effectiveness. *J. Nat. Disaster Sci.* **2012**, *34*, 3–18. [CrossRef]
5. Peng, Z. Earthquakes and coronavirus: How to survive an infodemic. *Seismol. Res. Lett.* **2020**, *91*, 2441–2443. [CrossRef]
6. Zhou, C.; Xiu, H.; Wang, Y.; Yu, X. Characterizing the dissemination of misinformation on social media in health emergencies: An empirical study based on COVID-19. *Inf. Process. Manag.* **2021**, *58*, 102554. [CrossRef]
7. Elroy, O.; Komendantova, N.; Yosipof, A. Cyber-echoes of climate crisis: Unraveling anthropogenic climate change narratives on social media. *Curr. Res. Environ. Sustain.* **2024**, *7*, 100256. [CrossRef]
8. King, K.K.; Wang, B. Diffusion of real versus misinformation during a crisis event: A big data-driven approach. *Int. J. Inf. Manag.* **2023**, *71*, 102390. [CrossRef]

9. Gkikas, D.C.; Tzafilkou, K.; Theodoridis, P.K.; Garmpis, A.; Gkikas, M.C. How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in Facebook. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100067. [[CrossRef](#)]
10. Ibrahim, N.F.; Wang, X.; Bourne, H. Exploring the effect of user engagement in online brand communities: Evidence from Twitter. *Comput. Hum. Behav.* **2017**, *72*, 321–338. [[CrossRef](#)]
11. Elroy, O.; Yosipof, A. Analysis of COVID-19 5G Conspiracy Theory Tweets Using SentenceBERT Embedding. In *International Conference on Artificial Neural Networks*; Springer Nature: Cham, Switzerland, 2022; pp. 186–196.
12. Epstein, Z.; Lin, H.; Pennycook, G.; Rand, D. How many others have shared this? Experimentally investigating the effects of social cues on engagement, misinformation, and unpredictability on social media. *arXiv* **2022**, arXiv:2207.07562. [[CrossRef](#)]
13. Withall, A.; Sagi, E. The impact of readability on trust in information. In Proceedings of the Annual Meeting of the Cognitive Science Society, Virtual, 26–29 July 2021.
14. Rogers, E. *Diffusion of Innovations Simon and Schuster*; Science and Education Publishing: Newark, NJ, USA, 2010.
15. Joy, A.; Pathak, R.; Shrestha, A.; Spezzano, F.; Winiecki, D. Modeling the diffusion of fake and real news through the lens of the diffusion of innovations theory. *ACM Trans. Soc. Comput.* **2024**, *7*, 1–24. [[CrossRef](#)]
16. Elroy, O.; Woo, G.; Komendantova, N.; Yosipof, A. A dual-focus analysis of wikipedia traffic and linguistic patterns in public risk awareness Post-Charlie Hebdo. *Comput. Hum. Behav. Rep.* **2024**, 100580. [[CrossRef](#)]
17. Wang, Y.; McKee, M.; Torbica, A.; Stuckler, D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Soc. Sci. Med.* **2019**, *240*, 112552. [[CrossRef](#)] [[PubMed](#)]
18. Erokhin, D.; Yosipof, A.; Komendantova, N. COVID-19 Conspiracy Theories Discussion on Twitter. *Soc. Media + Soc.* **2022**, *8*, 20563051221126051. [[CrossRef](#)] [[PubMed](#)]
19. Elroy, O.; Erokhin, D.; Komendantova, N.; Yosipof, A. Mining the Discussion of Monkeypox Misinformation on Twitter Using RoBERTa. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, León, Spain, 14–17 June 2023; pp. 429–438.
20. Gerts, D.; Shelley, C.D.; Parikh, N.; Pitts, T.; Watson Ross, C.; Fairchild, G.; Vaquera Chavez, N.Y.; Daughton, A.R. “Thought I’d share first” and other conspiracy theory tweets from the COVID-19 infodemic: Exploratory study. *JMIR Public Health Surveill.* **2021**, *7*, e26527. [[CrossRef](#)] [[PubMed](#)]
21. Vicari, R.; Elroy, O.; Komendantova, N.; Yosipof, A. Persistence of misinformation and hate speech over the years: The Manchester Arena bombing. *Int. J. Disaster Risk Reduct.* **2024**, *110*, 104635. [[CrossRef](#)]
22. Kim, M.G.; Kim, M.; Kim, J.H.; Kim, K. Fine-tuning BERT models to classify misinformation on garlic and COVID-19 on Twitter. *Int. J. Environ. Res. Public Health* **2022**, *19*, 5126. [[CrossRef](#)]
23. Rajdev, M.; Lee, K. Fake and spam messages: Detecting misinformation during natural disasters on social media. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Singapore, 6–9 December 2015; Volume 1, pp. 17–20.
24. Batzdorfer, V.; Steinmetz, H.; Biella, M.; Alizadeh, M. Conspiracy theories on Twitter: Emerging motifs and temporal dynamics during the COVID-19 pandemic. *Int. J. Data Sci. Anal.* **2022**, *13*, 315–333. [[CrossRef](#)]
25. Beskow, D.M.; Carley, K.M. Bot-hunter: A tiered approach to detecting & characterizing automated activity on twitter. In Proceedings of the SBP-BRiMS: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Washington, DC, USA, July 10–13 2018; Volume 3, p. 3.
26. ODonovan, J.; Kang, B.; Meyer, G.; Höllerer, T.; Adalii, S. Credibility in context: An analysis of feature distributions in twitter. In Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, Amsterdam, The Netherlands, 3–5 September 2012; pp. 293–301.
27. Gupta, A.; Kumaraguru, P.; Castillo, C.; Meier, P. Tweetcred: Real-time credibility assessment of content on twitter. In Proceedings of the International Conference on Social Informatics, Barcelona, Spain, 10–13 November 2014; pp. 228–243.
28. Jacques, R.D. *The Nature of Engagement and Its Role in Hypermedia Evaluation and Design*; South Bank University: London, UK, 1996.
29. O’Brien, H. Theoretical perspectives on user engagement. In *Why Engagement Matters: Cross-Disciplinary Perspectives of User Engagement in Digital Media*; Springer Nature: Berlin/Heidelberg, Germany, 2016; pp. 1–26.
30. Attfeld, S.; Kazai, G.; Lalmas, M.; Piwowarski, B. Towards a science of user engagement (position paper). In Proceedings of the WSDM Workshop on User Modelling for Web Applications, Hong Kong, China, 9–12 February 2011.
31. Fang, Z.; Costas, R.; Wouters, P. User engagement with scholarly tweets of scientific papers: A large-scale and cross-disciplinary analysis. *Scientometrics* **2022**, *127*, 4523–4546. [[CrossRef](#)]
32. Papakyriakopoulos, O.; Goodman, E. The impact of Twitter labels on misinformation spread and user engagement: Lessons from Trump’s election tweets. In Proceedings of the ACM Web Conference 2022, Lyon, France, 25–29 April 2022; pp. 2541–2551.
33. Segev, E. Sharing Feelings and User Engagement on Twitter: It’s All About Me and You. *Soc. Media + Soc.* **2023**, *9*, 20563051231183430. [[CrossRef](#)]

34. Zhao, M.; Buro, K. Comparisons between text-only and multimedia tweets on user engagement. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020; pp. 3825–3831.
35. DuBay, W.H. *Smart Language: Readers, Readability, and the Grading of Text*; ERIC: Washington, DC, USA, 2007.
36. Tavakoli, M.; Alani, H.; Burel, G. On the Readability of Misinformation in Comparison to the Truth. *CEUR Workshop Proc.* **2023**, *3370*, 63–72.
37. Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; Mihalcea, R. Automatic detection of fake news. *arXiv* **2017**, arXiv:1708.07104. [[CrossRef](#)]
38. Shrestha, A.; Spezzano, F. Characterizing and predicting fake news spreaders in social networks. *Int. J. Data Sci. Anal.* **2021**, *13*, 385–398. [[CrossRef](#)]
39. Warstadt, A.; Singh, A.; Bowman, S.R. *Cola: The Corpus of Linguistic Acceptability (with Added Annotations)*; FDA: Silver Spring, MD, USA, 2019.
40. Wang, A.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; Bowman, S.R. GLUE: A multi-task benchmark and analysis platform for natural language understanding. *arXiv* **2018**, arXiv:1804.07461.
41. Sultana, R.; Nishino, T. Fake News Detection System: An implementation of BERT and Boosting Algorithm. In Proceedings of the 38th International Conference on Computers and Their Applications, Tokyo, Japan, 20–22 March 2023; Volume 91, pp. 124–137.
42. Kuo, H.-Y.; Chen, S.-Y. Predicting user engagement in health misinformation correction on social media platforms in taiwan: Content analysis and text mining study. *J. Med. Internet Res.* **2025**, *27*, e65631. [[CrossRef](#)] [[PubMed](#)]
43. Chen, R.; Chen, G.; Zhang, L.; Xie, R.; Chen, R. An analysis of the factors influencing engagement metrics within the dissemination of health science misinformation. *Front. Public Health* **2025**, *13*, 1571210. [[CrossRef](#)] [[PubMed](#)]
44. Dallo, I.; Elroy, O.; Fallou, L.; Komendantova, N.; Yosipof, A. Dynamics and characteristics of misinformation related to earthquake predictions on Twitter. *Sci. Rep.* **2023**, *13*, 13391. [[CrossRef](#)]
45. Daniulaityte, R.; Chen, L.; Lamy, F.R.; Carlson, R.G.; Thirunarayan, K.; Sheth, A. “When ‘bad’ is ‘good’”: Identifying personal communication and sentiment in drug-related tweets. *JMIR Public Health Surveill.* **2016**, *2*, e6327. [[CrossRef](#)]
46. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
47. González-Carvajal, S.; Garrido-Merchán, E.C. Comparing BERT against traditional machine learning text classification. *arXiv* **2020**, arXiv:2005.13012.
48. Piskorski, J.; Haneczok, J.; Jacquet, G. New Benchmark Corpus and Models for Fine-grained Event Classification: To BERT or not to BERT? In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 6663–6678.
49. Baruah, A.; Das, K.A.; Barbhuiya, F.A.; Dey, K. Automatic Detection of Fake News Spreaders Using BERT. *CLEF (Work. Notes)* **2020**. Available online: https://ceur-ws.org/Vol-2696/paper_237.pdf (accessed on 21 August 2025).
50. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
51. Adoma, A.F.; Henry, N.-M.; Chen, W. Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In Proceedings of the 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 18–20 December 2020; pp. 117–121.
52. Naseer, M.; Asvial, M.; Sari, R.F. An empirical comparison of bert, roberta, and electra for fact verification. In Proceedings of the 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Jeju Island, Republic of Korea, 20–23 April 2021; pp. 241–246.
53. Tarunesh, I.; Aditya, S.; Choudhury, M. Trusting roberta over bert: Insights from checklisting the natural language inference task. *arXiv* **2021**, arXiv:2107.07229.
54. Müller, M.; Salathé, M.; Kummervold, P.E. COVID-twitter-bert: A natural language processing model to analyse COVID-19 content on twitter. *arXiv* **2020**, arXiv:2005.07503. [[CrossRef](#)] [[PubMed](#)]
55. Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv* **2019**, arXiv:1908.10084. [[CrossRef](#)]
56. Karande, H.; Walambe, R.; Benjamin, V.; Kotecha, K.; Raghu, T. Stance detection with BERT embeddings for credibility analysis of information on social media. *PeerJ Comput. Sci.* **2021**, *7*, e467. [[CrossRef](#)]
57. Elroy, O.; Yosipof, A. Semi-Supervised Learning Classifier for Misinformation Related to Earthquakes Prediction on Social Media. In Proceedings of the International Conference on Artificial Neural Networks, Heraklion, Greece, 26–29 September 2023; pp. 256–267.
58. Eltorai, A.E.M.; Sharma, P.; Wang, J.; Daniels, A.H. Most American Academy of Orthopaedic Surgeons’ Online Patient Education Material Exceeds Average Patient Reading Level. *Clin. Orthop. Relat. Res.* **2015**, *473*, 1181–1186. [[CrossRef](#)]
59. Krishna, K.; Wieting, J.; Iyyer, M. Reformulating unsupervised style transfer as paraphrase generation. *arXiv* **2020**, arXiv:2010.05700. [[CrossRef](#)]
60. Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*; Cambridge University Press: Cambridge, UK, 2013; Volume 53.

61. World Health Organization. Monkeypox—United Kingdom of Great Britain and Northern Ireland. Available online: <https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON381> (accessed on 2 February 2023).
62. Williams, H.T.; McMurray, J.R.; Kurz, T.; Lambert, F.H. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Glob. Environ. Change* **2015**, *32*, 126–138. [[CrossRef](#)]
63. Agarwal, S.; Sureka, A. Using knn and svm based one-class classifier for detecting online radicalization on twitter. In Proceedings of the Distributed Computing and Internet Technology: 11th International Conference, ICDCIT 2015, Bhubaneswar, India, 5–8 February 2015; pp. 431–442.
64. Neelakandan, S.; Paulraj, D. A gradient boosted decision tree-based sentiment classification of twitter data. *Int. J. Wavelets Multiresolution Inf. Process.* **2020**, *18*, 2050027. [[CrossRef](#)]
65. Costantini, H.; Fuse, R. Health Information on COVID-19 Vaccination: Readability of online sources and newspapers in Singapore, Hong Kong, and the Philippines. *J. Media* **2022**, *3*, 228–237. [[CrossRef](#)]
66. Jackson, M.O.; Yariv, L. Diffusion on social networks. *Econ. Publique/Public Econ.* **2006**, *16*. Available online: <https://ideas.repec.org/p/clt/sswopa/1251.html> (accessed on 21 August 2025). [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.