Telling a Lie: Analyzing the Language of Information and Misinformation during Global Health Events

Ankit Aich and Natalie Parde
Natural Language Processing Laboratory
Department of Computer Science
University of Illinois at Chicago
{aaich2, parde}@uic.edu

Abstract

The COVID-19 pandemic and other global health events are unfortunately excellent environments for the creation and spread of misinformation, and the language associated with health misinformation may be typified by unique patterns and linguistic markers. Allowing health misinformation to spread unchecked can have devastating ripple effects; however, detecting and stopping its spread requires careful analysis of these linguistic characteristics at scale. We analyze prior investigations focusing on health misinformation, associated datasets, and detection of misinformation during health crises. We also introduce a novel dataset designed for analyzing such phenomena, comprised of 2.8 million news articles and social media posts spanning the early 1900s to the present. Our annotation guidelines result in strong agreement between independent annotators. We describe our methods for collecting this data and follow this with a thorough analysis of the themes and linguistic features that appear in information versus misinformation. Finally, we demonstrate a proof-of-concept misinformation detection task to establish dataset validity, achieving a strong performance benchmark (accuracy = 75%; F_1 = 0.7).

Keywords: Dataset, Misinformation, Disinformation

1. Introduction

Just like viruses themselves, misinformation from small clusters of individuals can quickly evolve to global, community spread. Information may propagate rapidly via word of mouth, social media, and news articles, creating confusion and dangerous situations. This phenomenon has been recognized as an *infodemic* by the World Health Organization.[1] Despite the urgency of developing automated methods capable of identifying and extinguishing misinformation before it spreads too widely, few resources exist to facilitate the development of such models, especially as they pertain to generalizable settings beyond the COVID-19 pandemic.

In this work, we introduce a new, large dataset to stimulate such research. We collect 2.8 million news articles and (when available) social media posts covering the following diseases: small pox, the 1918 Spanish flu, MERS, SARS, H1N1, Ebola, HIV (AIDS), and COVID-19. News articles from traditional print media range from the early 1900s to the present, and are sourced from long-standing print houses including *The New York Times*,[2] the *British Broadcasting Corporation*,[3] *Cable News Network (CNN)*,[4] *The Washington Post*,[5] and *Al Jazeera*,[6] as well as the *Centers for Disease Control and Prevention (CDC)*[7] and the *World Health Organization (WHO)*. Social media coverage of more recent diseases is sourced from *Twitter*.[8] Our primary contributions are as follows:

- We analyze prior work investigating health information (with a focus on the ongoing COVID-19 pandemic), existing misinformation datasets, and automated detection of misinformation, identifying gaps and important future directions (§2).
- We introduce a new, large (2.8 million instances) collection of news articles and social media posts covering (mis)information spanning a wide range of diseases, and a smaller subset (4752 instances) of gold standard annotations indicating article/post misinformation status (§4 and §5).
- We conduct an analysis of the collected data (§6) and demonstrate validity of this dataset for future modeling and classification approaches by establishing strong performance (accuracy = 75%) on a misinformation detection benchmarking task (§7).

We describe these contributions in further detail in the remainder of this paper. We release our dataset publicly to interested researchers to spur rapid growth in this important research area.[9]

2. Prior Work

2.1. Health (Mis)Information

Many health information datasets have emerged since the start of the COVID-19 pandemic, including

[1] https://www.who.int/news-room/spotlight/let-s-flatten-the-infodemic-curve
[2] https://www.nytimes.com
[3] https://www.bbc.com
[4] https://www.cnn.com
[5] https://www.washingtonpost.com
[6] https://www.aljazeera.com
[7] https://www.cdc.gov
[8] https://twitter.com
[9] https://github.com/ankitaich09/MISINFORMATION

those covering Portuguese tweets and Brazilian news (De Melo and Figueiredo, 2020), English, Spanish, and Portuguese tweets (Aguilar-Gallegos et al., 2020), and a variety of specific health and health-adjacent topics. These centered on research investigating psychological health outcomes (Haider and Al-Salman, 2020), healthcare expenditures (Zhou et al., 2020), student learning (Trung et al., 2020), patents and antiviral therapy (Machuca-Martinez et al., 2020), genome modeling (Barbosa and Fernandes, 2020), and COVID-19 mortality (Li et al., 2020).

Recently, several groups have sought to examine misinformation in the context of COVID-19 specifically. Shaar et al. (2021) released a dataset of Arabic, Bulgarian, and English tweets annotated along numerous dimensions of misinformation status as part of the *NLP4IF Workshop Shared Task On Fighting the COVID-19 Infodemic*, and Hossain et al. (2020) released a dataset of English tweets annotated for stance pertaining to a set of known COVID-19 misconceptions. Although these datasets offer valuable resources for learning to model misinformation related to COVID-19, they do not offer corresponding data associated with other global health events. Liu et al. (2020) assembled 30 million tweets covering multiple diseases over the last decade including COVID, Cholera, and H1N1, making their dataset perhaps closest to ours; however, they do not focus on misinformation or provide fact-checked news articles.

2.2. Misinformation Datasets

A larger variety of data exists for learning to model misinformation in general. Nørregaard et al. (2019) created a dataset for the study of misinformation in news articles, collecting data from 194 different news sources for almost a year and labeling each article for reliability, bias, transparency, and consumer trust. Their sources included both mainstream printhouses and various online sources, including those popular with conspiracy theorists.

Kinsora et al. (2017) created a dataset labeling misinformation in medical communities, collecting data by employing information retrieval techniques to extract information from health discussion forums. Memon and Carley (2020) sought to characterize misinformation communities online, examining data from Twitter communities and identifying both interesting sociological correlations and intriguing linguistic patterns, such as the increased use of narrative among informed versus misinformed people. A shortcoming of these existing misinformation datasets is that they cover relatively short timespans, precluding comparative analysis of misinformation across time. Our dataset, encompassing major health events from the 1900s through the present, takes a step toward filling this gap.

2.3. Misinformation Detection

Beyond dataset creation, some groups have also sought to automatically detect the presence of misinformation in varied data sources. Much of this work is reviewed in the extensive survey by Almaliki (2019), although they placed no special emphasis on health misinformation. Ahmed et al. (2018) analyzed misinformation-induced panic in tweets about Ebola and H1N1, finding various common themes across tweets; however, they did not explore other forms of news media. Chew and Eysenbach (2010) explored misinformation search terms associated with H1N1 from 2009-2010, utilizing specific terms and non-textual data to categorize tweets (including those linking to news articles). Goodall et al. (2011) examined print media during the first five months of the H1N1 virus in 2009, using sources obtained from not only newspapers, but also the CDC to find a parity between officially-sanctioned guidelines and everyday media messaging.

Vlachos and Riedel (2014) employed the use of journalist fact-checked news articles in their work, using clustering to assign labels along a five-point scale (*True, Mostly True, Half True, Mostly False, False*), indicating the veracity of a piece of information. Ciampaglia et al. (2015) used Wikipedia to build a knowledge graph, assigning connections between entities and facts obtained from related wiki pages. They employed this graph for fact checking, evaluating paths leading up to a new fact and assigning scores to them. Very recently, Serrano et al. (2020) detected misinformation in YouTube videos during the COVID-19 pandemic, leveraging information from video comments. To date, no existing approaches have tackled both news and social media articles from current and historical health events. We establish a proof of concept for doing so using our new dataset in §7.

3. Defining Misinformation

We define *misinformation* as the intentional or unintentional spread of inaccurate information, including through unchecked opinions. This differs from *disinformation*, where intent is a required factor.[10] Misinformation has been studied across specific domains including politics (Dunaway, 2021), healthcare (Peterson et al., 2020), and social sciences (Anspach and Carlson, 2020), and in general (Lee and Shin, 2021). We constrain our interests to accurate information and misinformation pertaining to major global health events from the early 1900s to the present.

4. Data Collection

We include both traditional print media and social media posts in our dataset. To source the print media, we downloaded digitized news articles that were available either freely or via paid subscription for our included diseases: *COVID-19*, *Middle East Respiratory Syndrome (MERS)*, *Severe Acute Respiratory Syndrome (SARS)*, *Ebola*, *H1N1 (Swine Flu)*, *Human Immunodeficiency Virus (HIV*, which causes *Acquired Immunodeficiency Syndrome AIDS)*, *Spanish Influenza*, and *Smallpox*. We

[10] https://www.dictionary.com/e/misinformation-vs-disinformation-get-informed-on-the-difference/

Source	Average Length	EASE OF READING
New York Times	8522.35	40.89
Washington Post	4933.33	26.15
CDC	10506.53	28.04
WHO	4594.68	20.49
CNN	5319.16	23.64
Al Jazeera	4411.06	7.94
BBC	4818.11	9.46

Table 1: Average length and EASE OF READING score per news source. Average length is provided as the average number of characters in an article from the specified source.

searched five news sources: *The New York Times*, *British Broadcasting Corporation* (BBC), *Cable News Network* (CNN), *The Washington Post*, and *Al Jazeera News*. All sources are in English, but offer global impact and broad coverage of international events. We also downloaded articles from the official websites of the *World Health Organization* (WHO) and the *Centers for Disease Control and Prevention* (CDC).

Collected articles had an average length of 6813.3 characters and 330 words. Table 1 shows the average length and EASE OF READING score for each of the news sources in our data. The EASE OF READING score is obtained from the Flesch reading ease (Flesch, 1979) index, using the Python textstat API.[11] It provides a quantitative and easily comparable measure of reading fluency, which is known to influence comprehension (Klauda and Guthrie, 2008) and engagement (Mills et al., 2013). The index ranges from 0-100, with scores of 100 indicating texts that are easiest to read. The articles sourced from the CDC have the highest average length, and the articles sourced from The New York Times have the highest average EASE OF READING score.

We scraped social media posts from Twitter using the following queries: *HIV*, *COVID*, *SARS CoV-2*, *covid-19*, *coronavirus*, *H1N1 OR H2N2 OR H3N3*, *Ebola*, *Asian Influenza*, *Spanish Flu*, *SARS*, and *MERS*. In total, our dataset has 2,800,500 datapoints, distributed with approximately one-half covering COVID-19, one-ninth covering each of Ebola, SARS, and MERS, and the remainder covering smallpox, the 1918 flu epidemic, H1N1, and AIDS.

We release our data publicly to foster further work in this area. For tweets, in keeping with Twitter's data sharing terms, we release the tweet IDs and a script to download the corresponding tweet text. For news articles, we release full text for articles in the public domain, or headlines and sources for those currently under copyright protection. Data collection was approved and exempted from further review by the Institutional Review Board at the University of Illinois at Chicago.

[11] `https://pypi.org/project/textstat`

Description	Class
Known misinformation, or information unverified during fact-checking	1
Information verified by < 4 sources	2
Known facts, or information supported by ≥ 4 sources	3

Table 2: Annotation guidelines (news). *Sources* refers to independent news sources retrieved when conducting a web search for the information claim. *Class 1* also includes articles containing insufficient text to assess validity.

Description	Class
Personal opinions, known misinformation, or information unverified during fact-checking	1
Information verified by < 5 sources	2
Known facts, or information supported by ≥ 5 sources	3

Table 3: Annotation guidelines (tweets).

5. Annotation

We randomly sampled subsets of both news and social media data for manual annotation of misinformation status, collecting annotations until sufficient coverage for a balanced benchmarking dataset of 1500 instances was reached. Given the naturally uneven class distributions in our full dataset, this resulted in a total of 4752 instances (652 news articles and 4100 social media posts) receiving manual annotation. We formalized annotation guidelines through discussion among members of the research team. Each instance was double-annotated from a pool of three trained annotators (computer science graduate and undergraduate students with strong or native English proficiency), with one annotator labeling both news and tweets and the other two each labeling one of those groups.

Instances were assigned to *Class 1*, *Class 2*, or *Class 3* as specified in Table 2 and Table 3 for news articles and tweets, respectively. An article or post's misinformation status was determined based on available knowledge at the time of publication. For instance, information provided by an article that was supported by numerous sources at the time the article was published could be labeled as *Class 3*, even if that information was disproved several months later. Inter-annotator agreement was measured using an averaged pairwise Cohen's kappa (McHugh, 2012), achieving strong agreement with $\kappa = 0.81$ (Viera and Garrett, 2005).

In Figures 1 and 2, we illustrate the relative distributions of tweets and news articles across annotation classes. We observe a large but expected difference in annotation

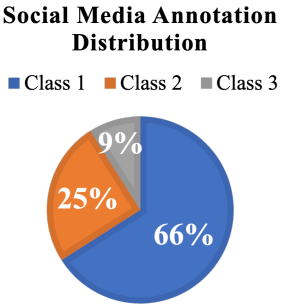

Figure 1: Distribution of manual annotations across *Classes 1*, *2*, and *3* in social media data.

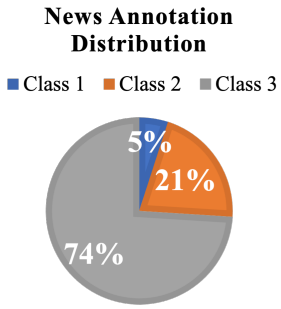

Figure 2: Distribution of manual annotations across *Classes 1*, *2*, and *3* in news data.

distribution between the two data sources. Specifically, a majority of social media posts were labeled as misinformation (*Class 1*), whereas a majority of news articles were labeled as accurate information (*Class 3*).

6. Analysis

Leveraging the manually annotated subsets of data, we examine the differences between *Class 3* and *Class 1* information, focusing on both linguistic (§6.1) and thematic (§6.2) characteristics.

6.1. Linguistic Analysis

We observed differences in both semantic and syntactic preference in *Class 3* versus *Class 1* news. The presence of personal pronouns in *Class 3* news is low to non-existent; the tokens *I* or *we* were not found in any articles, and the tokens *you*, *she*, *he*, and *they* were each present at a rate of less than 2%. Comparing this to an equivalent-sized sample of *Class 1* news yields a clear difference: the frequency of first-person pronouns (e.g., *I* or *we*) increases to 10%, although the frequencies of second- or third-person pronouns remain nearly the same.

We also observed marked differences in text complexity. To measure this and other psycholinguistic factors, we used Linguistic Analysis and Word Count (Tausczik and Pennebaker, 2010), a proprietary tool developed to measure various layers of psycholinguistic and affective attributes. We find that the average analytical complexity of *Class 3* news articles is 91.08%, whereas for *Class 1* news this falls to 38%. We also compute TRUST SCORES for each text using the NRC Word-Emotion Lexicon (Mohammad and Turney, 2010), finding a correlation with assigned class. *Class 3* news articles show an average trust score of ≥ 600, whereas *Class 1* news articles show an average score of ≤ 150.

Finally, we found that topic shifts or inconsistencies may be a valuable marker of *Class 1* health news. *Class 1* articles also often included politically polarized n-grams in addition to those associated with health topics. Across *Class 1* news articles, the highest-frequency unigrams were *COVID19*, *TRUMP*, and *BIDEN*, following stopword removal. For *Class 3* news, the highest-frequency unigrams were *COVID19*, *MEASLES*, and *EBOLA*.

Topic	Keywords
1	smallpox, drug, people, year, patient, treatment, risk, give, develop, human
2	health, test, home, hospital, day, ebola, contact, symptom, die, return
3	flu, pandemic, swine flu, coronavirus, state, people, government, influenza, reopen, way
4	measle, case, virus, vaccine, health, people, outbreak, report, child, accord

Table 4: Top 10 keywords, ordered by computed weight from highest to lowest, for each identified topical theme associated with *Class 3* tweets.

Topic	Keywords
1	state, flu, fast, people, read, spread, call, world, supply, fact
2	com, twitter, covid, pic, pandemic, status, plan, reopen, work, stop
3	com, covid, coronavirus, death, case, new, twitter, pic, day, virus
4	coronavirus, crisis, grow, outbreak, economy, global, fin, ensure, bird, subprime

Table 5: Top 10 keywords, ordered by computed weight from highest to lowest, for each identified topical theme associated with *Class 1* tweets.

6.2. Thematic Analysis

We use latent Dirichlet allocation (LDA) to separately model topics present in news articles and social media posts, analyzing overarching themes. We train the LDA model on groups of *Class 3* and *Class 1* data after preprocessing text and removing all stopwords, punctuation, and symbols. We show the top identified themes in tweets, represented by most strongly associated keywords, in Tables 4 and 5.

	Class 1	Class 3
Tweets	tested positive covid19	World Health Org
News	case tally cross	Disease Control Prevention

Table 6: Top trigrams associated with unverified (*Class 1*) and supported (*Class 3*) information for both tweets and news articles. Stopwords were removed prior to computing trigram frequencies.

Feature	Description
TRUST SCORE	The number of words associated with *trust* in the NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2010; Mohammad and Turney, 2013)
READABILITY INDEX	Automated readability index (Senter and Smith, 1967), computed as follows: $\text{ARI} = 4.71(\frac{c}{w}) + 0.5(\frac{w}{s}) - 21.43$
EASE OF READING	Flesch reading ease score (Flesch, 1979), computed as follows: $\text{EoR} = 206.835 - 1.015(\frac{w}{s}) - 84.6(\frac{l}{w})$
DIFFICULT WORDS	Raw number of *difficult words*, defined as words (a) containing greater than two syllables and (b) not included in the Python textstat *Easy Words List*
+ NRC SCORES	The number of *positive* words in the NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2010; Mohammad and Turney, 2013)
- NRC SCORES	The number of *negative* words in the NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2010; Mohammad and Turney, 2013)
LENGTH	The raw value, c

Table 7: Features included in the benchmark experiments. For equations, let c be the number of characters in a text, w be the number of words, s be the number of sentences, and l be the number of syllables.

We find that the keywords in *Class 3* news articles (*smallpox*, *health*, *measles*, and *flu*) adhere firmly to health concerns. The keywords for *Class 1* articles (*twitter*, *crisis*, and *state*) instead center on politics, media, and panic. The motif of personal versus general information as seen in §6.1 also reappears in our thematic analysis. We illustrate this in Table 6, showing the top trigrams associated with *Class 1* and *Class 3* data.

Model	Accuracy	F_1
Random Forest	0.6849	0.6330
KNN	0.6986	0.6164
Decision Tree	0.5890	0.5068
Logistic Regression	**0.7534**	**0.6996**
Ridge	0.6849	0.6438
SVM	0.3424	0.6575

Table 8: Results from model comparison, using a 95%/5% randomized train/test split of a balanced subset of the labeled dataset.

7. Proof of Concept

To further establish dataset validity, we define a proof-of-concept classification task. We extract identified features from our analysis for each instance in our labeled data subset and train a suite of classifiers to predict the (mis)information status of news articles. We define this as a binary classification problem (*Class 3* versus *Class 1*) and remove any inherent metadata (e.g., source of the article or post) that could potentially bias a classifier. We compute our features as shown in Table 7.

We experiment with a balanced subset of our annotated data (1500 instances total, evenly distributed across *Class 1* and *Class 3*), using a 95%/5% train/test split to maximize the modeling algorithms' available training data. We compare seven popular classifiers, all from Python's `sk-learn`[12] library and trained using default parameters and the features defined in Table 7: *Random Forest*, *K Nearest Neighbors*, *Decision Tree*, *Logistic Regression*, *Ridge*, *Support Vector Machine*, and *Multilayer Perceptron*. We report our results in Table 8. We find that the best-performing model is *Logistic Regression*, achieving an accuracy of 75% and an F_1 of 0.7 at distinguishing *Class 1* and *Class 3* data.

8. Conclusion and Future Directions

In this paper, we introduce a substantial new dataset (2.8 million datapoints) to facilitate the study of information and misinformation during global health events. Importantly, the dataset includes both news and social media coverage of a broad range of health events, distinguishing it from existing datasets focusing on COVID-19 misinformation such as those created by Shaar et al. (2021) and Hossain et al. (2020). We collect manual misinformation annotations for a subset of 4752 articles and social media posts, and conduct a thorough analysis of the linguistic and thematic differences between information and misinformation manifesting in this subset. Finally, we establish dataset validity and a performance benchmark by training a suite of classifiers on the annotated data using features uncovered in our analysis, achieving high accuracy (75%) and F_1 (0.7) scores at distinguishing between health information and

[12] `https://scikit-learn.org`

misinformation. It is our hope that our dataset will spur further exploration of this important research area. Examples of downstream applications that may benefit from this work include healthcare dialogue systems (Valizadeh and Parde, 2022) and other clinical systems in need of feature-based or fine-grained detection of health misinformation, such as those geared towards health providers or caregivers (Kaelin et al., 2021).

9. Acknowledgements

This work was supported in part by a startup grant from the University of Illinois at Chicago. We thank the anonymous reviewers for their helpful comments.

10. Bibliographical References

Aguilar-Gallegos, N., Romero-García, L. E., Martínez-González, E. G., García-Sánchez, E. I., and Aguilar-Ávila, J. (2020). Dataset on dynamics of coronavirus on twitter. *Data in Brief*, 30:105684.

Ahmed, W., Bath, P. A., Sbaffi, L., and Demartini, G. (2018). Moral panic through the lens of twitter: An analysis of infectious disease outbreaks. In *Proceedings of the 9th International Conference on Social Media and Society*, SMSociety '18, page 217–221, New York, NY, USA. Association for Computing Machinery.

Almaliki, M. (2019). Online misinformation spread: A systematic literature map. In *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*, ICISDM 2019, page 171–178, New York, NY, USA. Association for Computing Machinery.

Anspach, N. and Carlson, T. (2020). What to believe? social media commentary and belief in misinformation. *Political Behavior*, 42, 09.

Barbosa, R. D. and Fernandes, M. A. (2020). Chaos game representation dataset of sars-cov-2 genome. *Data in Brief*, 30:105618.

Chew, C. and Eysenbach, G. (2010). Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak. *PloS one*, 5:e14118, 11.

Ciampaglia, G., Shiralkar, P., Rocha, L., Bollen, J., Menczer, F., and Flammini, A. (2015). Computational fact checking from knowledge networks. *PloS one*, 10, 01.

De Melo, T. and Figueiredo, C. M. (2020). A first public dataset from brazilian twitter and news on covid-19 in portuguese. *Data in Brief*, 32:106179.

Dunaway, J., (2021). *Polarisation and misinformation*, pages 131–141. Routledge, 02.

Flesch, R. F. (1979). *How to write plain English: A book for lawyers and consumers*. Harpercollins.

Goodall, C., Sabo, J., Cline, R., and Egbert, N. (2011). Threat, efficacy, and uncertainty in the first 5 months of national print and electronic news coverage of the h1n1 virus. *Journal of health communication*, 17:338–55, 12.

Haider, A. S. and Al-Salman, S. (2020). Dataset of jordanian university students' psychological health impacted by using e-learning tools during covid-19. *Data in Brief*, 32:106104.

Hossain, T., Logan IV, R. L., Ugarte, A., Matsubara, Y., Young, S., and Singh, S. (2020). COVIDLies: Detecting COVID-19 misinformation on social media. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online, December. Association for Computational Linguistics.

Kaelin, V. C., Valizadeh, M., Salgado, Z., Parde, N., and Khetani, M. A. (2021). Artificial intelligence in rehabilitation targeting the participation of children and youth with disabilities: Scoping review. *J Med Internet Res*, 23(11):e25745, Nov.

Kinsora, A., Barron, K., Mei, Q., and Vydiswaran, V. G. V. (2017). Creating a labeled dataset for medical misinformation in health forums. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 456–461.

Klauda, S. L. and Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational psychology*, 100(2):310.

Lee, E.-J. and Shin, S., (2021). *Debunking misinformation*, pages 470–479. Routledge, 02.

Li, J., Wang, L., Guo, S., Xie, N., Yao, L., Cao, Y., Day, S. W., Howard, S. C., Graff, J. C., Gu, T., Ji, J., Gu, W., and Sun, D. (2020). The data set for patient information based algorithm to predict mortality cause by covid-19. *Data in Brief*, 30:105619.

Liu, J., Singhal, T., Blessing, L. T., Wood, K. L., and Lim, K. H. (2020). Epic30m: An epidemics corpus of over 30 million relevant tweets. *arXiv preprint arXiv:2006.08369*.

Machuca-Martinez, F., Amado, R. C., and Gutierrez, O. (2020). Coronaviruses: A patent dataset report for research and development (r&d) analysis. *Data in brief*, 30:105551.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.

Memon, S. A. and Carley, K. M. (2020). Characterizing covid-19 misinformation communities using a novel twitter dataset.

Mills, C., D'Mello, S., Lehman, B., Bosch, N., Strain, A., and Graesser, A. (2013). What makes learning fun? exploring the influence of choice and difficulty on mind wandering and engagement during learning. In H. Chad Lane, et al., editors, *Artificial Intelligence in Education*, pages 71–80, Berlin, Heidelberg. Springer Berlin Heidelberg.

Mohammad, S. and Turney, P. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34, Los Angeles, CA, June. Association for Computational Linguistics.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Nørregaard, J., Horne, B. D., and Adalı, S. (2019). Nela-gt-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 630–638.

Peterson, J., Swire-Thompson, B., and Johnson, S. (2020). What is the alternative? responding strategically to cancer misinformation. *Future Oncology*, 16, 06.

Senter, R. and Smith, E. A. (1967). Automated readability index. Technical report, Cincinnati University, Ohio.

Serrano, J. C. M., Papakyriakopoulos, O., and Hegelich, S. (2020). Nlp-based feature extraction for the detection of covid-19 misinformation videos on youtube. In *Proceedings of the ACL 2020 Workshop on Natural Language Processing for COVID-19 (NLP-COVID)*, 07.

Shaar, S., Alam, F., Da San Martino, G., Nikolov, A., Zaghouani, W., Nakov, P., and Feldman, A. (2021). Findings of the NLP4IF-2021 shared task on fighting the COVID-19 infodemic and censorship detection. In *Proceedings of the Fourth Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, NLP4IF@NAACL' 21, Online, June. Association for Computational Linguistics.

Tausczik, Y. and Pennebaker, J. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29:24–54, 03.

Trung, T., Hoang, A.-D., Nguyen, T. T., Dinh, V.-H., Nguyen, Y.-C., and Pham, H.-H. (2020). Dataset of vietnamese student's learning habits during covid-19. *Data in Brief*, 30:105682.

Valizadeh, M. and Parde, N. (2022). The AI doctor is in: A survey of task-oriented dialogue systems for healthcare applications. In *Proceedings of the 2022 Conference of the Association for Computational Linguistics*, Dublin, Ireland, May. Association for Computational Linguistics.

Viera, A. and Garrett, J. (2005). Understanding interobserver agreement: The kappa statistic. *Family medicine*, 37:360–3, 06.

Vlachos, A. and Riedel, S. (2014). Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, MD, USA, June. Association for Computational Linguistics.

Zhou, Y., Zhang, Z., Wang, B., Ren, G., Qi, H., and Wang, X. (2020). Construction time, cost and testing data of a prefabricated isolation medical unit for covid-19. *Data in Brief*, 32:106068.