ClimateMiSt: Climate Change Misinformation and Stance Detection Dataset

YeonJung Choi[0000−0001−5889−6565], Lanyu Shang[0000−0002−7480−6889], and Dong Wang[0000−0002−9599−8023]

School of Information Sciences, University of Illinois Urbana-Champaign, Champaign IL, USA
{yc55, lshang3, dwang24}@illinois.edu

Abstract. Climate change has been a worldwide concern for more than 50 years, and climate change misinformation has also become a critical issue as it questions the causes and effects of climate change, thereby disrupting climate action. Climate misinformation has been a major obstacle to mitigating climate change and its effects, aggravating the issue and polarizing the public. In this paper, we introduce ClimateMiSt, a new climate change misinformation and stance detection dataset consisting of social media data with manually verified labels. The data is collected from Twitter/X and our dataset contains 146,670 tweets. We implement state-of-the-art baseline models for both misinformation and stance detection on our dataset and discover that GPT-4 outperforms them in both tasks. To the best of our knowledge, ClimateMiSt is the first dataset focused on climate change that includes both veracity and stance annotations collected from a social media platform. Our novel dataset can be used for climate change misinformation and stance detection, and it can further contribute to research in this field.

Keywords: Climage Change · Climate Change Dataset · Misinformation Detection · Stance Detection · Online Social Media

1 Introduction

The emergence of social networking platforms (e.g., Twitter/X, Facebook/Meta) has significantly changed the way people interact and communicate online, leading to a whole new wave of applications and reshaping existing information ecosystems [23,29]. However, the popularity of social media platforms simultaneously prompted the growth of misinformation, resulting in detrimental societal effects such as undermining public trust and support and putting people's lives at risk [9,19,25,31]. Climate change is no exception as a target of misinformation propagation. Although climate scientists have reached a near-unanimous consensus on anthropogenic climate change [8], widespread public disbelief persists due to misinformation about this scientific conclusion [12,22]. Climate misinformation reduces public acceptance and understanding of climate change, contributes to polarization, and impedes support for mitigation policies [12,20]. In this paper, we propose a novel Climate Change Misinformation and Stance Detection Dataset (ClimateMiSt) that aims to address climate change misinformation issues by facilitating relevant studies.

Since climate change became a matter of "real concern" from a US President's Advisory Committee panel in 1965,[1] numerous efforts have been made worldwide to mitigate its effects, such as United Nations Framework Convention on Climate Change (UNFCCC) and Paris Agreement [5]. However, these efforts have not resolved the climate change issue. On July 27th, 2023, the UN Secretary-General warned that "the era of global boiling has arrived."[2] Recent extreme droughts and heatwaves (including in Europe and the US) and wildfires in California and Canada are all attributed to anthropogenic climate change [15].

Goldberg *et al.* analyze consecutive election cycles from 1990 to 2018 and find that oil and gas companies systematically provide financial support to anti-environmental politicians [16]. In addition, Farrell examines a network of 164 organizations (e.g., think tanks, foundations, etc.) and 4,556 individuals associated with these organizations who participated in the climate change counter-movement. He identifies that organizations with corporate funding are more likely to polarize the climate change issue [14]. He also suggests that science is being privatized due to the increasing influence of corporate wealth on scientific issues. These studies imply that the climate change issue is highly political and that climate change misinformation is created and disseminated by prominent politicians and organizations. That is, climate misinformation is repeated and amplified by people with power, influence, or recognition, from where it reaches a wider public. Considering its unique characteristics, detecting and correcting climate change misinformation is an urgent matter.

Moreover, the general public lacks the expertise and skills to evaluate the veracity of a claim, leading them to rely more on heuristics. Hence, people make judgments based on the character of those who speak about climate change, rather than on the claim itself [7]. This tendency leads people to be more vulnerable to climate misinformation. Considering that falsehood reaches people about six times faster than the truth [28], it is significantly important to accurately identify misleading information on climate change. In addition, studying the stance on climate change is critical as it directly affects policy making, public awareness, and participation. However, there are fewer constructions of climate change misinformation datasets and relevant research compared to other domains. For example, several recent efforts have been made to construct data repositories for climate change [11,4] However, these datasets do not specifically target climate change misinformation: they lack either veracity annotations (e.g., misinformation/non-misinformation) or stance annotations (e.g., favor/against).

To address these limitations, we present ClimateMiSt, a novel dataset specifically targeted for climate misinformation and stance detection tasks. Our goal is to construct a comprehensive dataset for climate misinformation that can be utilized in misinformation or stance detection within the climate change domain. In particular, we extensively collect data from one of the widely used social media platforms and provide veracity and stance annotations for our dataset. Additionally, we experiment with multiple benchmark models to evaluate their performance on our dataset. The evaluation results indicate that GPT-4 achieves the best performance for both misinformation and stance detection tasks. We expect our study to facilitate relevant research in the field, such

[1] https://www.bbc.com/news/science-environment-15874560

[2] https://news.un.org/en/story/2023/07/1139162

as advances in misinformation and stance detection models within the climate change domain. This, in turn, can help people distinguish misinformation from factual information and raise awareness of climate change issues. We summarize the key contributions of our paper as follows:

- ClimateMiSt is the first comprehensive dataset for both misinformation and stance detection in the context of climate change, consisting of social media data.
- We implement state-of-the-art baselines, including generative AI models like GPT and Llama, and evaluate their performance on both misinformation and stance detection tasks using our collected dataset.

2 Related Works

2.1 Misinformation Detection

Several studies of climate change focus on either dataset construction [11] or analyses on the misinformation itself [2,6]. Diggelmann *et al.* introduce the CLIMATE-FEVER dataset, designed to verify climate change-related claims based on the methodology of FEVER [11]. However, the dataset does not provide misinformation or stance annotations for English tweets. On the other hand, Al-Rawi *et al.* study public discourses around climate change and global warming by collecting 6.8 million tweets referencing "fake news" and find that discussions about climate change/global warming and fake news are highly polarized [2]. Coan *et al.* investigate the role of misinformation in the climate change debate, suggesting that conservative think tanks and climate contrarian blogs damage the credibility of climate science and scientists through conspiratorial messaging [6].

2.2 Studies on Stance Detection

With the proliferation of online content, many natural language processing tasks, including stance detection, have been extensively investigated. Reveilhac and Schneider present a rule-based stance detection model that is easily replicable across various targets [24]. Gómez-Suta *et al.* propose a two-phase classification system for stance detection that leverages topic modeling features and provides explanations for stance labels using these features [17]. Upadhyaya *et al.* propose a novel framework called Sentiment and Temporal Aided Stance Detection (STASY) to classify tweets' stance on climate change as either "denier" or "believer" [27]. While the other aforementioned studies lack their own datasets for stance detection, Upadhyaya *et al.* create the "CLiCS" dataset for the climate change domain [27]. However, CLiCS consists only of tweets with stance annotations and lacks veracity annotations.

3 Data

In this section, we provide a detailed overview of the data collection and annotation process for ClimateMiSt. In particular, we detail the social media data collection and annotation process in Section 3.1. We provide comparison results with other datasets in Section 3.2. An overview of the data collection process is illustrated in Figure 1.

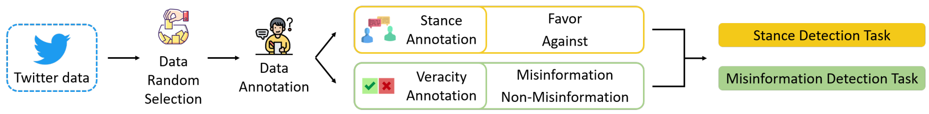

FIG. 1: Overview of Data Collection Process

3.1 Social Media Data

We choose Twitter/X as the social media platform to collect data as Twitter is one of the most popular social media platforms with around 450 million monthly active users.[3] Also, users can follow others and retweet their tweets, allowing information to be easily propagated. Twitter API v2[4] is used to crawl all English tweets that mention "climate change" (excluding retweets), from January 1st, 2022 to September 30th, 2022. For each tweet, we collect a set of attributes, including "author_id", "created_at", "id", "text", "attachments", "url", "verified[5]", etc. A total of 147,957 tweets are collected, with 1,287 being duplicates. In total, we crawl 146,670 unique tweets.

Veracity Annotation Due to labor constraints, 2,008 tweets are randomly selected from the 146,670 collected tweets and manually annotated as either misinformation (0) or non-misinformation (1). In our study, we define misinformation as false or inaccurate information about climate change issues, and non-misinformation as true or accurate information about them. We exclude any tweets that are 1) too political (extreme right or left inclination), 2) not related to climate change (a mere mention of "climate change"), or 3) lacking sufficient context.

For any tweets, we search for information that could support or rebut the tweet. We look for news articles (e.g., CNN, BBC), reports (e.g., UN, IPCC), or journal papers (e.g., Elsevier, Springer) from reputable sources for verification. If any information is found supporting the statement, the tweet is labeled as non-misinformation. If the information rebuts the statement, the tweet is labeled as misinformation. In total, we annotate 1,140 tweets as misinformation and 868 tweets as non-misinformation.

Stance Annotation For the same tweets used in the veracity annotation, we also manually annotate them as either "against (0)" or "favor (1)." Regardless of the veracity of each tweet, if the writer denies the existence of (anthropogenic) climate change, we label it as "against." If the writer accepts it, we label it as "favor."

In total, we annotate 980 tweets as "against" and 1,028 tweets as "favor." Among all "favor" tweets, about 82% are "non-misinformation," while approximately 18% are "misinformation." In contrast, among "against" tweets, 97.45% are "misinformation," and only 2.55% are "non-misinformation."

[3] https://www.demandsage.com/twitter-statistics/

[4] https://developer.twitter.com/en/docs/twitter-api

[5] The 'verified' account refers to the blue Verification badge on Twitter which indicates that an account of public interest is authentic.

We also calculate Cramer's V to measure the association between two categorical variables: veracity labels and stance labels [1]. The value of Cramer's V ranges from 0 (no association between the variables) to 1 (complete association between the variables). The Cramer's V between veracity labels and stance labels is 0.8017 (with a p-value of 1.274e-28), indicating a strong and significant association between the two variables.

In summary, we collect a total of 146,670 tweets and randomly select 2,008 of them for annotation. We apply two different labels: veracity and stance. Our dataset includes 1,140 tweets labeled as misinformation and 868 as non-misinformation. Additionally, there are 980 tweets labeled "against" and 1,028 labeled "favor" in terms of stance.

3.2 Comparison with Existing Datasets

We compare the properties of ClimateMiSt with those of other existing climate change misinformation datasets in Table 1. One important observation is that only our dataset includes political stance information. While every study in Table 1 provides climate change misinformation data, each has distinct differences from our dataset.

TABLE 1: Comparison of Properties of Different Climate Change Misinformation Datasets

Dataset	Data Source	Language	Political Stance	Total Amount (Annotated Total)	Time Range
CLIMATE-FEVER [11]	Social Media	English	✗	1,535 (-)	✗
Twitter-COMMs [4]	Social Media	English	✗	212,665 (-)	06/01/2016 - 09/31/2021
Al-Rawi *et al.* [2]	Social Media	English	✗	over 6.8 million (-)	11/27/2019 - 02/14/2020
Coan *et al.* [6]	Social Media	English	✗	249,413 (-)	1998 - 2020
ClimateMiSt (Ours)	Social Media	English	✓	146,670 (5,353)	01/01/2022 - 09/30/2022

CLIMATE-FEVER contains 7,675 annotated claim-evidence pairs, including 1,535 verifiable claims, with labels for each claim-evidence pair [11]. However, the evidence sentences are only retrieved from Wikipedia, which may raise credibility and validity issues. Moreover, the labels refer to the relationship between the claim and the evidence, not the claim's veracity. In addition, this dataset lacks stance annotations and does not include any experiments on misinformation or stance detection tasks. Similarly, Twitter-COMMs is a large-scale multi-modal dataset with a total of 884,331 tweets, including 212,665 tweets related to climate change [4]. However, this dataset does not include any veracity or stance annotations, nor does it involve misinformation or stance detection tasks. Al-Rawi *et al.* and Coan *et al.* also collect large amounts of data from different sources to study contrarian claims or fake news about climate change and analyze the characteristics of these discourses [2,6]. However, they do not include any annotations for misinformation or stance detection tasks. Compared to these datasets, ClimateMiSt is a comprehensive climate change dataset designed for both misinformation and stance detection tasks. It is manually labeled with veracity and stance annotations, and we have experimented with our dataset on both misinformation and stance detection tasks.

4 Experiment

In this section, we perform misinformation and stance detection on our dataset using several state-of-the-art baseline models. The evaluation results show that GPT-4 achieves the best performance in both tasks.

4.1 Baselines and Experiment Setup

The baselines consist of generative large language models (LLMs) as well as models specifically targeted for misinformation detection and text classification. We select GPT-4[6] and Llama 3[7] as our generative LLMs, employing few-shot learning by providing three to four sample tweets for each label in every task. We use supervised learning for all other models, which include: dEFEND (Explainable FakE News Detection) [26], SHINE (HIerarchical heterogeNEous graph representation learning method for STC) [30], TextING (Text classification method for INductive word representations via Graph neural networks) [32], BERT Text Classification (Bidirectional Encoder Representations from Transformers) [10], and RoBERTa Text Classification (Robustly Optimized BERT-Pretraining Approach) [21].

TABLE 2: Hyperparameters for Baseline Models

Model	(Graph/Text) Embedding Dimension	Batch Size	Epoch	Learning Rate	Dropout Rate	Max Length
dEFEND	100	20	50	-	-	-
SHINE	1000	-	-	1e-3	0.7	-
TextING	300	4096	200	0.005	0.5	-
BERT	768	16	50	1e-6	0.5	512
RoBERTa	768	16	50	1e-5	0.5	150

For supervised classifiers, the entire annotated dataset is divided into training, development, and test sets in an 8:1:1 ratio, resulting in 1,606 training samples, 201 development samples, and 201 test samples. We run the benchmark models on Ubuntu 20.04 using four NVIDIA A16 and four NVIDIA L40s. Additionally, we use PyTorch version 1.13.1 and TensorFlow version 1.12.0. Specific parameters for each benchmark model are detailed in Table 2.

4.2 Misinformation Detection Performance Analysis

The misinformation detection results on our dataset are shown on the left side of Table 3. We observe that the best-performing model is GPT-4, achieving a classification F1 score of 0.9590, while RoBERTa performs the best among all non-LLMs. GPT-4 outperforms RoBERTa by 2.22%, 2.15%, 1.97%, and 2.36% in terms of F1 score, accuracy, precision, and recall, respectively.

[6] https://openai.com/research/gpt-4
[7] https://llama.meta.com/llama3/

TABLE 3: Baseline Model Performances with Our Dataset on Misinformation Detection (left) and Stance Detection (right)

	Model	F1 score	Accuracy	Precision	Recall
Generative LLMs	**GPT-4**	**0.9590**	**0.9602**	**0.9627**	**0.9561**
	Llama 3	0.9267	0.9303	0.9462	0.9176
Supervised Classifiers	dEFEND	0.8010	0.8010	0.8010	0.8010
	SHINE	0.8448	0.8458	0.8454	0.8538
	TextING	0.8831	0.8657	0.8870	0.8793
	BERT	0.8977	0.9005	0.8992	0.8965
	RoBERTa	0.9382	0.9403	0.9441	0.9341

	Model	F1 score	Accuracy	Precision	Recall
Generative LLMs	**GPT-4**	**0.8953**	**0.8955**	**0.9002**	**0.8958**
	Llama 3	0.8123	0.8159	0.8443	0.8166
Supervised Classifiers	dEFEND	0.7214	0.7214	0.7214	0.7214
	SHINE	0.8396	0.8408	0.8433	0.8387
	TextING	0.7831	0.7960	0.7957	0.7708
	BERT	0.7904	0.7910	0.7940	0.7908
	RoBERTa	0.8752	0.8756	0.8803	0.8753

4.3 Stance Detection Performance Analysis

The stance detection results are displayed on the right side of Table 3. Similar to the misinformation detection results, the best-performing model for stance detection is GPT-4, with a classification F1 score of 0.8953. Other text classification models, such as RoBERTa and SHINE, also show decent performance in terms of both accuracy and F1 score. One noticeable point is that the overall performance of all baseline models has decreased in the stance detection task. This is possibly due to the characteristics of stance annotation, which is strongly related to the writers' opinions, whereas veracity annotation is contingent on facts.

5 Discussion and Future Work

Experiments on our dataset indicate that generative LLMs achieve the best performance compared to other supervised classifiers, even with few-shot learning. This can be attributed to the billions to trillions of parameters and the massive training data used in training generative LLMs. Hence, generative LLMs could be used for future automatic data annotation, thereby enhancing misinformation detection tasks.

Our study can facilitate relevant climate change misinformation tasks, including 1) constructing datasets for both misinformation and stance detection on climate change, and 2) implementing and evaluating benchmark baselines, including generative AI models such as GPT-4 and Llama 3. This will help people distinguish misinformation from factual information about climate change issues, thereby gradually alleviating the climate change problem.

Moreover, our research can be further investigated in a few ways. Our dataset currently contains only textual data (single modality), but modern data often comprises multiple modalities (e.g., text, image, video, or audio). Incorporating multimodal features into our dataset will further contribute to extending relevant studies and enhancing the performance of both misinformation and stance detection. For example, leveraging other modalities such as image or audio, which exist alongside text, can improve detection performance by providing additional information (e.g., visual representation of images [13], MFCC from audio [18]) for classifying veracity and stance. Moreover, we could further improve model performance by developing human-AI hybrid solutions that incorporate human intelligence (e.g., crowdsourcing) to identify and verify posts likely to be misclassified by LLMs [3]. Additionally, zero-shot learning on generative LLMs could be explored to investigate its impact on misinformation and stance detection performance.

6 Conclusion

Climate change misinformation has undermined mitigation efforts and contributed to public polarization. It is crucial to address this misinformation immediately, but climate change misinformation datasets have not been extensively studied so far. Through this study, we provide a novel climate change dataset specifically designed for both misinformation detection and stance detection tasks, consisting of social media data. We experiment with several state-of-the-art baseline models for misinformation detection and stance detection, and provide benchmark performance results. Both the misinformation detection and stance detection results show that generative LLMs, specifically GPT-4, excel in both tasks. We hope that our study opens the door to further in-depth research on climate change misinformation detection.

Acknowledgement

This research is supported in part by the National Science Foundation under Grant No. IIS-2202481, CHE-2105032, IIS-2130263, CNS-2131622, CNS-2140999. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

1. Akoglu, H.: User's guide to correlation coefficients. Turkish journal of emergency medicine **18**(3), 91–93 (2018)
2. Al-Rawi, A., O'Keefe, D., Kane, O., Bizimana, A.J.: Twitter's fake news discourses around climate change and global warming. Frontiers in Communication **6** (11 2021). https://doi.org/10.3389/fcomm.2021.729818
3. Arous, I., Dolamic, L., Yang, J., Bhardwaj, A., Cuccu, G., Cudré-Mauroux, P.: Marta: Leveraging human rationales for explainable text classification. In: AAAI Conference on Artificial Intelligence. pp. 5868–5876 (2021)
4. Biamby, G., Luo, G., Darrell, T., Rohrbach, A.: Twitter-comms: Detecting climate, covid, and military multimodal misinformation. CoRR **abs/2112.08594** (2021), `https://arxiv.org/abs/2112.08594`
5. Blau, J.: The Paris Agreement: climate change, solidarity, and human rights. Springer (2017)
6. Coan, T., Boussalis, C., Cook, J., Nanko, M.O.: Computer-assisted detection and classification of misinformation about climate change (Mar 2021). https://doi.org/10.31235/osf.io/crxfm, `osf.io/preprints/socarxiv/crxfm`

7. Cook, J., Ellerton, P., Kinkead, D.: Deconstructing climate misinformation to identify reasoning errors. Environmental Research Letters **13**, 024018 (02 2018). https://doi.org/10.1088/1748-9326/aaa49f
8. Cook, J., Oreskes, N., Doran, P., Anderegg, W., Verheggen, B., Maibach, E., Carlton, J., Lewandowsky, S., Skuce, A., Green, S., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., Rice, K.: Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. Environmental Research Letters **11**(4), 48002 (Apr 2016). https://doi.org/10.1088/1748-9326/11/4/048002
9. Cui, L., Seo, H., Tabar, M., Ma, F., Wang, S., Lee, D.: Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. p. 492–502. KDD '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3394486.3403092, https://doi.org/10.1145/3394486.3403092
10. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR **abs/1810.04805** (2018), http://arxiv.org/abs/1810.04805
11. Diggelmann, T., Boyd-Graber, J.L., Bulian, J., Ciaramita, M., Leippold, M.: CLIMATE-FEVER: A dataset for verification of real-world climate claims. CoRR **abs/2012.00614** (2020), https://arxiv.org/abs/2012.00614
12. Ding, D., Maibach, E., Zhao, X., Roser-Renouf, C., Leiserowitz, A.: Support for climate policy and societal action are linked to perceptions of scientific agreement. Nature Climate Change **1**, 462–466 (11 2011). https://doi.org/10.1038/NCLIMATE1295
13. Fang, Y., Wang, W., Xie, B., Sun, Q., Wu, L., Wang, X., Huang, T., Wang, X., Cao, Y.: Eva: Exploring the limits of masked visual representation learning at scale. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19358–19369 (2023)
14. Farrell, J.: Corporate funding and ideological polarization about climate change. Proceedings of the National Academy of Sciences **113**(1), 92–97 (2016). https://doi.org/10.1073/pnas.1509433112, https://www.pnas.org/doi/abs/10.1073/pnas.1509433112
15. Fountain, H.: Hotter summer days mean more sierra nevada wildfires, study finds. The New York Times (2021), https://www.nytimes.com/2021/11/17/climate/climate-change-wildfire-risk.html?smid=url-share
16. Goldberg, M.H., Marlon, J.R., Wang, X., van der Linden, S., Leiserowitz, A.: Oil and gas companies invest in legislators that vote against the environment. Proceedings of the National Academy of Sciences **117**(10), 5111–5112 (2020). https://doi.org/10.1073/pnas.1922175117, https://www.pnas.org/doi/abs/10.1073/pnas.1922175117
17. Gómez-Suta, M., Echeverry-Correa, J., Soto-Mejía, J.A.: Stance detection in tweets: A topic modeling approach supporting explainability. Expert Syst. Appl. **214**(C) (mar 2023). https://doi.org/10.1016/j.eswa.2022.119046, https://doi.org/10.1016/j.eswa.2022.119046
18. Hamza, A., Javed, A.R.R., Iqbal, F., Kryvinska, N., Almadhor, A.S., Jalil, Z., Borghol, R.: Deepfake audio detection via mfcc features using machine learning. IEEE Access **10**, 134018–134028 (2022)
19. Islam, M.S., Sarkar, T., Khan, S.H., Kamal, A.H.M., Hasan, S.M., Kabir, A., Yeasmin, D., Islam, M.A., Chowdhury, K.I.A., Anwar, K.S., Chughtai, A.A., Seale, H.: Covid-19-related infodemic and its impact on public health: A global social media analysis. The American journal of tropical medicine and hygiene **103**(4), 1621 (2020)

20. van der Linden, S., Leiserowitz, A., Rosenthal, S., Maibach, E.: Inoculating the public against misinformation about climate change. Global Challenges **1**(2), 1600008 (2017). https://doi.org/https://doi.org/10.1002/gch2.201600008, https://onlinelibrary.wiley.com/doi/abs/10.1002/gch2.201600008
21. 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. CoRR **abs/1907.11692** (2019), http://arxiv.org/abs/1907.11692
22. Oreskes, N., Conway, E.: Defeating the merchants of doubt. Nature **465**, 686–7 (06 2010). https://doi.org/10.1038/465686a
23. Reis, J.C.S., Correia, A., Murai, F., Veloso, A., Benevenuto, F.: Explainable machine learning for fake news detection. In: Proceedings of the 10th ACM Conference on Web Science. p. 17–26. WebSci '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3292522.3326027, https://doi.org/10.1145/3292522.3326027
24. Reveilhac, M., Schneider, G.: Replicable semi-supervised approaches to state-of-the-art stance detection of tweets. Information Processing & Management **60**(2), 103199 (2023). https://doi.org/10.1016/j.ipm.2022.103199
25. Shang, L., Chen, B., Vora, A., Zhang, Y., Cai, X., Wang, D.: Socialdrought: A social and news media driven dataset and analytical platform towards understanding societal impact of drought. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 18, pp. 2051–2062 (2024)
26. Shu, K., Cui, L., Wang, S., Lee, D., Liu, H.: defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 395–405 (2019)
27. Upadhyaya, A., Fisichella, M., Nejdl, W.: Towards sentiment and temporal aided stance detection of climate change tweets. Information Processing & Management **60**(4), 103325 (2023)
28. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018). https://doi.org/10.1126/science.aap9559, https://www.science.org/doi/abs/10.1126/science.aap9559
29. Wang, D., Kaplan, L., Abdelzaher, T.F.: Maximum likelihood analysis of conflicting observations in social sensing. ACM Transactions on Sensor Networks (ToSN) **10**(2), 1–27 (2014)
30. Wang, Y., Wang, S., Yao, Q., Dou, D.: Hierarchical heterogeneous graph representation learning for short text classification. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. pp. 3091–3101. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic (Nov 2021). https://doi.org/10.18653/v1/2021.emnlp-main.247, https://aclanthology.org/2021.emnlp-main.247
31. Zhang, D., Wang, D., Vance, N., Zhang, Y., Mike, S.: On scalable and robust truth discovery in big data social media sensing applications. IEEE transactions on big data **5**(2), 195–208 (2018)
32. Zhang, Y., Yu, X., Cui, Z., Wu, S., Wen, Z., Wang, L.: Every document owns its structure: Inductive text classification via graph neural networks. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (2020)