Current Research in Environmental Sustainability 7 (2024) 100256

Contents lists available at ScienceDirect

Current Research in Environmental Sustainability

journal homepage: www.sciencedirect.com/journal/current-research-in-environmental-sustainability

Cyber-echoes of climate crisis: Unraveling anthropogenic climate change narratives on social media

Or Elroy [a,b], Nadejda Komendantova [b], Abraham Yosipof [b,c,*]

[a] Department of Computer Science, University of Oregon, Eugene, OR, USA
[b] International Institute for Applied Systems Analysis, Laxenburg, Austria
[c] Faculty of Information Systems and Computer Science, College of Law & Business, Ramat-Gan, Israel

ARTICLE INFO

Keywords:
Climate change
Clustering
Natural language processing
Social media
Policy
Misinformation

ABSTRACT

Social media platforms have a key role in spreading narratives about climate change, and therefore it is crucial to understand the discussion about climate change in social media. The discussion on anthropogenic climate change in general, and social media specifically, has multiple different narratives. Understanding the discourses can assist efforts of mitigation, adaptation, and policy measures development. In this work, we collected 333,635 tweets in English about anthropogenic climate change. We used Natural Language Processing (NLP) and machine learning methods to embed the semantic meaning of the tweets into vectors, cluster the tweets, and analyze the results. We clustered the tweets into four clusters that correspond to four narratives in the discussion. Analyzing the behavioral dynamics of each cluster revealed that the clusters focus on the discussion of whether climate change is caused by humans or not, scientific arguments, policy, and conspiracy. The research results can serve as input for media policy and awareness-raising measures on climate change mitigation and adaptation policies, and facilitating future communications related to climate change.

1. Introduction

Social media platforms are among the primary sources of information people use and trust when searching for information about climate change. Social media platforms are easy to use and nowadays their outreach is almost universal. They also offer room for participation when everyone can become not only a receiver but also a spreader of information. Online social media platforms are a significant source of information as well as a forum for public debate, and their potential to shape individual attitudes and behaviors is widely recognized.

Climate change and climate-related news are among some of the most significant discourses online (Das and Chakraborty, 2022; Falkenberg et al., 2022) and extreme weather events generate vivid attention to climate change (Sisco et al., 2017). Despite the variety of information available on social media, users typically interact only with like-minded others, in communities dominated by a single view (Williams et al., 2015) and adapt to the dominant opinion within the respective media outlet. Therefore, user comment sections serve as echo chambers rather than as corrective mechanisms (Walter et al., 2018). Participation in climate conversations with friends and family was found to be shaping beliefs and feelings about global warming and leads people to learn influential facts, stronger perceptions of scientific agreement, and enter people into a pro-climate social feedback loop (Goldberg et al., 2019). Messages between like-minded users typically carry positive sentiment but mixed-attitude communities in which skeptics and activists frequently interact carry negative sentiment (Williams et al., 2015). Climate change denial is more visible in user comment sections in countries where the climate change debate reflects the scientific consensus on climate change and user comments create niches of denial (Walter et al., 2018).

Social media platforms in general, and X (formerly Twitter) specifically, provide an ideal atmosphere and fertile ground for the creation and dissemination of misinformation and disinformation quickly (Hilary and Dumebi, 2021; Lazer et al., 2018). Misinformation is false or inaccurate information according to the best factual evidence that is available at a given point in time, regardless of an intention to mislead or deceive (Komendantova et al., 2021). Misinformation and conspiracy theories offer people explanations and a sense of control over a situation of uncertainty (Batzdorfer et al., 2022). The spread of misinformation through social media is also getting almost a universal character. Misinformation, which is not a new phenomenon and has existed for centuries, is now being spread by social media within seconds, and its

* Corresponding author at: International Institute for Applied Systems Analysis, Laxenburg, Austria.
E-mail addresses: yosipof@iiasa.ac.at, aviyo@clb.ac.il (A. Yosipof).

https://doi.org/10.1016/j.crsust.2024.100256
Received 3 February 2024; Received in revised form 4 June 2024; Accepted 4 June 2024

spread can get a viral character while creating misperceptions, preconditions, prejudices, erosion of trust in public institutions, and the lack of willingness to participate in climate change mitigation or adaptation measures. A conspiracy theory is misinformation according to which a group of people are secretly trying to cause harm or achieve something when other explanations are more probable. A conspiracy theory typically opposes the consensus among qualified professionals. Misinformation is spread by way of rumors, conspiracy theories, fake news, and fear-mongering, and can significantly affect what people perceive as a risk, and how they believe they should react to it (Erokhin et al., 2022). The ability of misinformation to spread globally and easily through social media raises the risk of worsening the harms of different types of emergency situations (Kwanda and Lin, 2020; Peary et al., 2012; Peng, 2020; Zhou et al., 2021). False news may reach up to 100 times more people than true news (Essebo, 2022). The belief in conspiracy theories is motivated by a humane need to rationalize events. The lack of authoritative sources with reliable information in case of an emergency event, combined with circumstantial evidence, fosters uncertainty and thus misinformation and conspiracy narratives (Aschwanden et al., 2018; Ortiz-Martínez et al., 2020). Believing in misinformation and conspiracy theories led to the unnecessary loss of life during the COVID-19 pandemic. Hence, fighting the spread of misinformation and conspiracy theories and countering them with reliable information is critical during climate change emergency events. "Prebunking" is a psychological inoculation strategy whereby the public is forewarned about the possibility of being misled by misinformation, and then later exposed to small doses of misinformation along with strong countering statements (Fraser et al., 2021). Analysis of the climate change discussion on social media can help understand where policy measures are needed to prebunk and debunk misinformation more efficiently (Dallo et al., 2023).

Climate change action is a contested policy issue, meaning that various opinions and views exist about its impacts and if and how it should be addressed. These views are formed by various sources of information and influence the actions of people regarding climate change mitigation and adaptation. The discussions related to climate change online in general and on social networks in specific are not well understood (Williams et al., 2015). A broad range of "climate contrarian" views was observed during the United Nations Conference of the Parties on Climate Change (COP), emphasizing the theme of political hypocrisy as a topic of cross-ideological appeal, and contrarian views and accusations of hypocrisy have become key themes in the X (formerly Twitter) climate discussion (Falkenberg et al., 2022). Some segments of the population are still doubtful about the human impact on climate change, despite the broad consensus among scientists and journalists on the subject (Walter et al., 2018). The sentiment of the discussion about climate change on X (formerly Twitter) was found to be overall negative, especially when users react to political or extreme weather events (Dahal et al., 2019). These claims that deny the existence of climate change, the relation between the human factor to climate change, or otherwise insist it is somehow related to a variety of other causes, some of which are intentional and accompanied by a secret agenda, are therefore misinformation.

It is important to understand discussions about climate change in social media and their development because such understanding can help identifying major discourses about climate change which include perceptions, views, and beliefs. Such understanding can enable better sharing of successful practices with other stakeholders, which is important for achieving sustainable development goals (Matsui et al., 2022). Findings about the relative influence of people's appraisals of risk and the potential to respond in ways that reduce risk on behavior have been mixed, and the lack of consistent findings suggests a potential need to refine frameworks for the context of climate change (Fischer et al., 2022). The importance and urgency of addressing various climate change challenges have been emphasized in prior research (Jerneck et al., 2011), and some of the most recent challenges involve misinformation on social media.

The understanding of climate change discourses can help understand various positions regarding climate change mitigation and adaptation, and further develop policy measures that will address these positions. This can also contribute to fair and inclusive climate change policy when various existing views are being addressed and policy measures target a variety of existing perceptions while leaving no one behind. We, therefore, investigate the climate change narratives in the discussion on X (formerly Twitter). The analysis of various narratives would further contribute to developing policy recommendations to increase climate change awareness and the efficiency of implementing climate change mitigation and adaptation measures.

In this work, we answer two primary research questions. First, what are the main narratives in the discussion about climate change on social media. Second, how do major climate change events and reports influence public awareness and involvement in the discussion. Following previous research, we hypothesize that major climate change events, such as extreme weather events and significant reports, will trigger increased attention to climate change discussions on social media across the various narratives.

2. Methodology

The methodology includes the following sections: (1) Data: we collected the tweets related to anthropogenic climate change on X (formerly Twitter); (2) Embedding: we embedded the tweets using RoBERTa, and the word embeddings were transformed into sentence embeddings; (3) Clustering: we clustered the sentence embeddings using machine learning methods and analyzed the results.

2.1. Data

To answer the research questions, we collected tweets in English between January 1, 2022, and May 30, 2023, using X (formerly Twitter)'s academic research API, which enables access to the full archive of X (formerly Twitter). The data collected excludes retweets. The period was carefully chosen to include both regular times and special climate change events, including extreme or emergency events, or important publications and conferences related to climate change.

We designed a search query by using keywords related to man-made and climate change, hence targeting the discussion of the human cause concerning climate change. The keywords were selected according to the sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) (Lee et al., 2023). The following search query was used: *[[climate AND [change OR changes OR crisis OR emergency]] OR "global warming"] AND [human OR anthropogenic OR "man made" OR "man-made"]*. The search query is not limited to a certain stance but instead aims for data that is relevant to every perspective to find and analyze different narratives in the discussion.

To optimize the performance of the model, we preprocessed the collected dataset to discard tweets that are longer than 350 characters, and to remove mentions of users and links, from the remaining tweets. The resulting dataset consists of 333,635 tweets.

2.2. Embedding

We applied the word embedding method for the semantic meaning of each tweet using RoBERTa (Robustly Optimized BERT Pretraining Approach). Word embedding is an NLP methodology in which the semantic meaning of each word in the text is embedded using a vector of a certain length, resulting in a list of vectors for each sentence. Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) provides superior results for different NLP tasks, including word embedding (González-Carvajal and Garrido-Merchán, 2020; Piskorski et al., 2020). Sentence embeddings combine the semantic meanings of multiple word embeddings into a single vector that represents the semantic meaning of the whole sentence (Reimers and Gurevych, 2019).

Different variations of models based on BERT exist, each of which comes with its strengths and weaknesses. RoBERTa was proven to provide improved performance and state-of-the-art results (Adoma et al., 2020; Naseer et al., 2021; Tarunesh et al., 2021). BERT-based models are often used to embed the semantic meaning of tweets (Elroy and Yosipof, 2022; Uthirapathy and Sandanam, 2023), and RoBERTa in particular (Dallo et al., 2023; Elroy et al., 2023; Elroy and Yosipof, 2023). We computed the word embedding of each tweet using RoBERTa-base and converted the word embeddings to sentence embeddings using Sentence-BERT, resulting in a vector of 768 features per tweet.

2.3. Clustering

Clustering is an unsupervised machine learning methodology that assigns objects within the dataset to clusters according to similarity measures, therefore finding a structure in a collection of unlabeled data (Madhulatha, 2012). A variety of clustering algorithms is used in current research, and k-Means is one of the first and most popular. The k-Means algorithm is based on the idea that each sample is assigned to one of the k clusters according to its similarity to the cluster's centroid (Fung, 2001). One of the major drawbacks of k-Means is that it is very sensitive to the provided initial clusters (Fung, 2001). k-Means++ is an initialization algorithm that obtains a proper set of initial centers that is provably close to the optimum solution (Arthur and Vassilvitskii, 2007; Bahmani et al., 2012). After presenting a general model of clustering techniques, the properties of the clusters are to be examined.

The k-Means algorithm can be used for a wide variety of data types and is also quite efficient. However, k-Means is not suitable for data with natural clusters that are globular, clusters of different sizes and densities, or data that contains outliers. This shortcoming can be overcome by leveraging k-Means' ability to find pure sub-clusters if a large enough number of clusters is specified. That is, by finding a large number of smaller clusters such that each of them represents a part of a natural cluster, and putting them together in a post-processing step. Other algorithms like Density-Based Spatial Clustering of Applications with Noise (DBSCAN) are often useful for density-based clustering and for discovering clusters of arbitrary shapes. However, in this case, calculating the dissimilarity matrix between tweets shows that the dataset is uniformly distributed. Therefore, the advantages of DBSCAN become weaknesses in this dataset.

The clustering was performed based on the 768-features long vectors of the embeddings provided by RoBERTa. The process of clustering the dataset was carried out in three steps, namely using a large number of smaller clusters to overcome the aforementioned limitation of k-Means, manually reviewing the clusters, and finally merging the clusters using hierarchical clustering. First, to overcome the limitation of k-Means, we clustered the data using a large number of clusters, e.g., 20, 40, 60. Measuring the quality of the clusters has shown that increasing the number of clusters did not lead to significant changes in the silhouette and sum of squared distances. We, therefore, used 20 clusters. Second, following the clustering process, we manually reviewed and characterized the 20 clusters using n-grams and TF-IDF with different lengths of phrases which exhibited very similar results. Additionally, random samples of each cluster were examined by two annotators. Finally, we used hierarchical clustering using Ward's method which minimizes the variance of the clusters being merged, to cluster the centroids of the 20 clusters. Fig. 1 presents the dendrogram of the hierarchical clustering of the 20 clusters provided by k-Means. After reviewing the dendrogram and following manual inspection of the 20 clusters, we found that four main clusters were formed and therefore merged the clusters accordingly.

In addition, to find the optimal number of clusters, we evaluated the results of k-Means using two to nine clusters and found that four clusters provide optimal results according to the elbow method based on the distortion score, i.e., the sum of square distances from each point to its assigned center (Satopaa et al., 2011).

3. Analysis and results

The tweets were assigned to four clusters as previously explained in the methodology. The four clusters correspond to four narratives that we identified throughout the analysis presented in this section as Anthropogenic, Scientific, Policy, and Conspiracy. Random samples of tweets from each cluster were examined at the end of the analysis by two annotators to decide whether they fit under the narrative associated with the cluster they are in. We calculated Cohen's kappa coefficient, representing the level of agreement between the annotators, to be 0.77.

To visualize the clusters of tweets, we used t-SNE to reduce the dimensionality of the sentence embeddings from 768 to 2-D. Fig. 2 visualizes the tweets using t-SNE for dimensionality reduction, colored by their cluster association. This visualization highlights the clear separation between the clusters.

Table 1 presents the number of tweets, unique authors, and average tweets per author in each cluster. Each tweet is uniquely associated with a single cluster, whereas authors could have posted multiple tweets, each associated with a different cluster. Therefore, while tweets may only appear in one cluster, an author can appear in more than one cluster but no more than once in each cluster.

3.1. Clusters

We investigated the frequency of phrases in each cluster.

The focus of the discussion in the *Anthropogenic* cluster (Fig. 1, Blue) is on whether anthropogenic climate change is real or not. We checked for the most frequently used terms in the cluster using n-grams and TF-IDF, and found some of the prominent phrases in this cluster to be

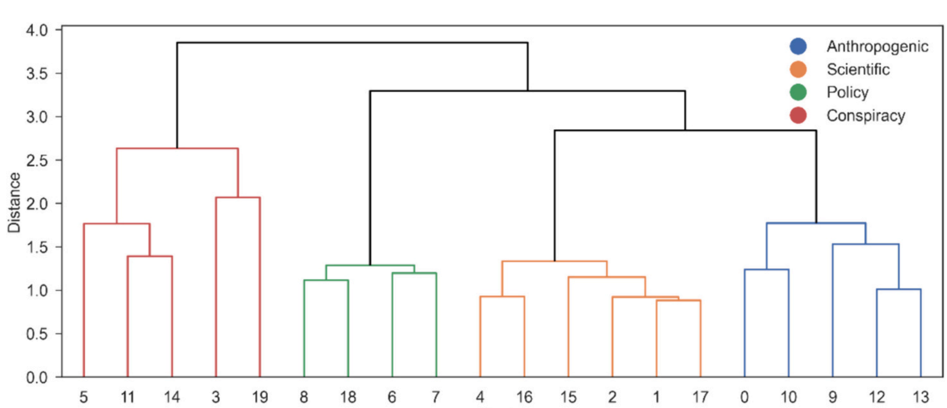

Fig. 1. Dendrogram for the hierarchical clustering of the 20 clusters provided by k-Means.

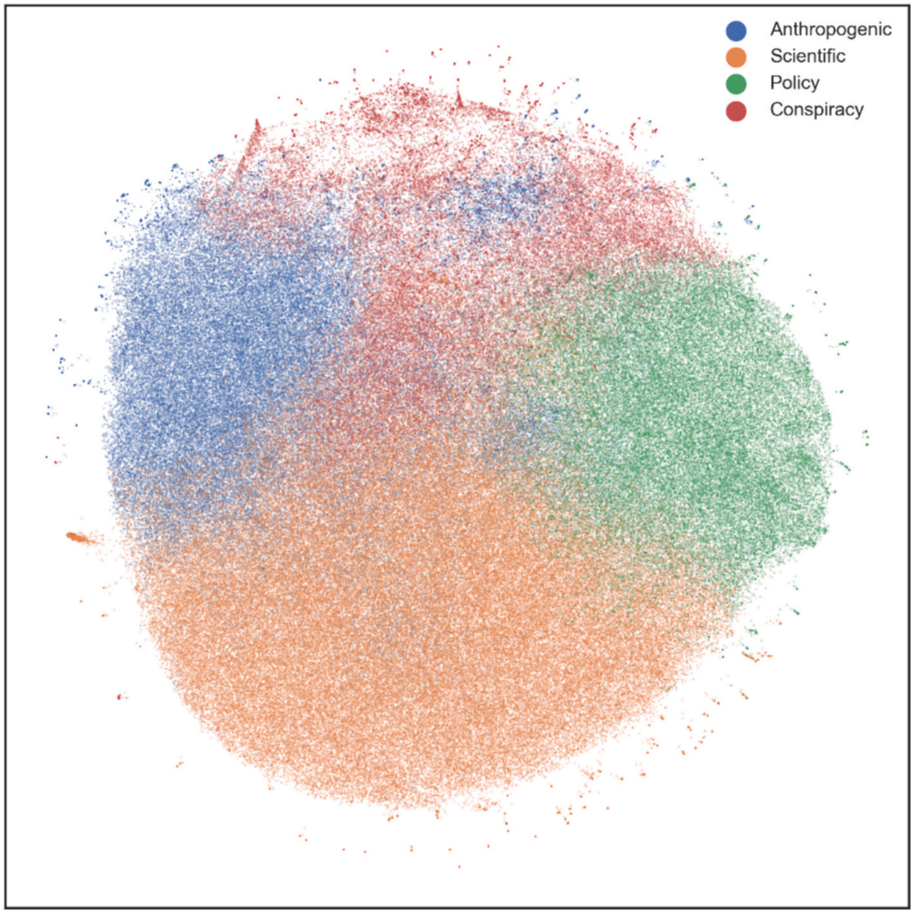

Fig. 2. Visualization of the tweets in the dataset by clusters. Blue dots represent tweets in the Anthropogenic cluster, orange dots represent tweets in the Scientific cluster, green dots represent tweets in the Policy cluster, and red dots represent tweets in the Conspiracy cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Descriptive statistics of the clusters, including the number of tweets, unique authors, and average tweets per author.

Cluster	# of tweets	# of unique authors	Average tweets/author
Anthropogenic	83,820	50,468	1.661
Scientific	148,433	84,959	1.747
Policy	57,832	33,310	1.736
Conspiracy	43,550	25,732	1.692

anthropogenic hoax, anthropogenic real, scientific consensus, scam believers, green agenda. The narratives in this cluster vary between users claiming climate change does not exist, such as *"The weather crisis is scientifically generated. There's no Climate Change! It's man made!"*, to users claiming climate change is real, but without relation to the human factor, such as *"Climate change is real and natural, man made climate change is a hoax."*, and others claiming climate change as well as its anthropogenic nature is real. Despite the provocative nature of this cluster, it has the lowest number of tweets per author with an average of 1.661.

The *Scientific* cluster (Fig. 1, Orange) focused on tweets that rely on scientists and scientific findings to discuss the narrative of anthropogenic climate change. Prominent phrases according to n-grams and TF-IDF are scientific consensus, verifiable facts, fossil fuels, annual emission, CO_2 cause, interview scientist, science say. Tweets in this cluster argue against conspiracy theories, such as *"All of the science says you're wrong. The following page lists the nearly 200 worldwide scientific organizations that hold the position that climate change has been caused by human action."* and *"The science says otherwise but go ahead and believe whatever you want to believe. The fact that *human* influenced change in climate is happening will always be a fact."* This cluster exhibited the largest number of tweets, unique authors, and highest tweets per user on average.

The *Policy* cluster (Fig. 1, Green) mainly discusses policy matters, such as the importance of climate change mitigation to reduce greenhouse gas emissions, e.g., *"There are multiple, feasible and effective options to reduce greenhouse gas emissions and adapt to human-caused climate change, and they are available now, said scientists in the latest Intergovernmental Panel on Climate Change (IPCC) report released today."* The Policy cluster is neither the smallest nor the largest one in terms of tweets, number of users, or average tweets per user. Notable frequent phrases used in this cluster according to n-grams and TF-IDF are greenhouse gas, greenhouse effect, carbon dioxide, weather events, extreme weather, and heat wave.

The *Conspiracy* cluster (Fig. 1, Red) expressed higher interest in terms related to conspiracy theories, according to Coan et al. (2021). Tweets in this cluster are generally dismissive of the anthropogenic factor in climate change, and some blame officials for artificially creating the concept, such as in *"the main goal of this climate change hoax is to introduce carbon taxation and passports and thus huge amount of restrictions including human live span limits."* and *"How valuable is human life? Why do money and profit rule our world? Why aren't they looking for the real cause of climate change?"* This cluster is the smallest in terms of the number of tweets and the number of unique authors.

3.2. Time series

To identify the public interest and awareness regarding climate change, we analyzed the time series of the clusters. Fig. 3 shows the daily tweet frequency of each cluster.

On February 28, 2022 (Fig. 3, annotation 1), the IPCC released its Sixth Assessment Report (AR6). AR6 states that human-induced climate change is causing dangerous and widespread disruption in nature and affecting the lives of billions of people worldwide, despite efforts to reduce the risks. Scientists in the IPCC mentioned that people and ecosystems that are least able to cope are being hit the hardest (Lee et al., 2023). AR6 provides various scenarios of the future of life on earth, characterized by ecosystem collapse, species extinction, and climate hazards such as heatwaves and floods (Atwoli et al., 2022). All clusters, except for the Scientific cluster, share a common peak in tweet frequency on this date. The peak is most notable in the Policy cluster, which discusses various goals of climate change mitigation, such as the goal of reducing greenhouse gas emissions to reach the target of $<1.5\ ^{\circ}$C global warming to avoid catastrophic effects across all regions of the globe (Atwoli et al., 2022).

Another peak that constitutes the highest one in the Anthropogenic and Scientific clusters can be observed on July 19–20, 2022 (Fig. 3, annotation 2), following a record-breaking heat wave across the globe (Holley and Lee, 2022). The heat wave seriously affected humans and livestock alike (Cooke and Rivero, 2023). Extreme weather events generate attention to climate change on X (formerly Twitter) (Sisco et al., 2017). The severe heat wave reignited discussion about climate change, the causes thereof, and different ways of dealing with it. While this peak is observed in all clusters, and despite its significance in the Anthropogenic and Scientific clusters, it is negligible in the Policy and Conspiracy clusters.

On September 22, 2022 (Fig. 3, annotation 3), a California law legalized composting human remains in an attempt to tackle climate change through the final disposition of human bodies in a way that will not contribute emissions into the atmosphere (Daily Mail, 2022). The passing of the law triggered the second-highest peak in the Conspiracy cluster, with tweets questioning the necessity and reasoning of the law, and some mocking it, such as *"Grandmas, get out now before you become a new batch of tomatoes in Malibu! CA Will Allow Human Composting After Death to Combat Climate Change"*.

Protests are known to take place ahead of COP meetings, and 2022 was no different. Multiple protests took place during October 2022, and several reached the headlines in mass media. Some of the more severe incidents took place on October 27, 2022 (Fig. 3, annotation 4), such as an activist attempting to glue his head to the Johannes Vermeer's Girl With a Pearl Earring (ART News, 2022), and two others throwing cans of tomato soup on Van Gogh's famous Sunflowers painting (CNN, 2022).

November 7–8, 2022 (Fig. 3, annotation 5) marked the beginning of the 2022 Climate Change Conference of the United Nations Framework Convention on Climate Change (COP27). COP27 is an important conference that health journal editors perceived as capable of delivering justice to vulnerable countries, and publicly called for action before the conference (Atwoli et al., 2022). This finding is in line with previous research in which the attention that COP26, the same conference of the previous year, received from mainstream media has led to an increase of contrarian views in the climate change conversation on social media (Falkenberg et al., 2022).

The peak observed on March 20–21, 2023 (Fig. 3, annotation 6), is related to the Climate Change 2023: Synthesis Report by the United Nations Environment Programme, which is a main scientific input to COP28 and the review of progress towards the Paris Agreement goals (United Nations Environment Programme, 2023). The report outlined that the 1.5 degrees Celsius limit is still achievable and that critical action is required. The release of the report has led to the highest peak in interest in the Conspiracy cluster. The publication of the report is also followed, as could be expected, by increased interest in the Policy cluster.

The peak of April 22–23, 2023 (Fig. 3, annotation 7), is associated

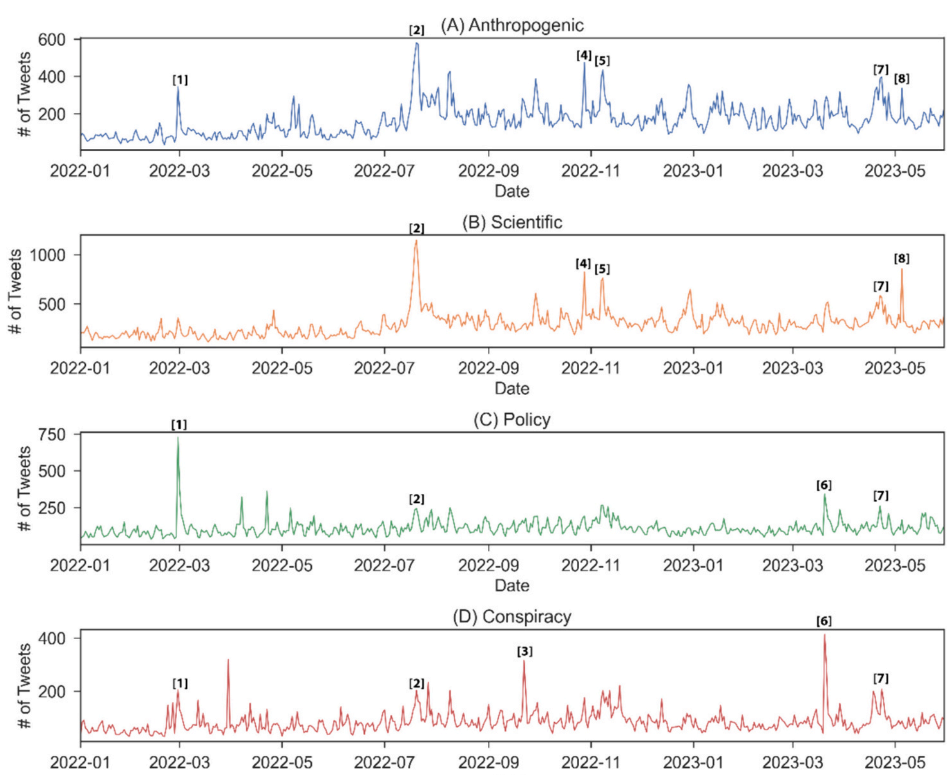

Fig. 3. Daily tweet frequency in each cluster. (A) the daily tweet frequency in the Anthropogenic cluster. (B) the daily tweet frequency in the Scientific cluster. (C) the daily tweet frequency in the Policy cluster. (D) The daily tweet frequency in the Conspiracy cluster.

with Earth Day, an annual event to demonstrate support for environmental protection. Earth Day of 2023 was featured in mass media due to multiple demonstrations by climate change activists (Reuters, 2023).

On the days before May 5, 2023 (Fig. 3, annotation 8), a massive heatwave struck parts of southwestern Europe and North Africa and brought extremely high temperatures. The heatwave was accompanied by record-breaking temperatures in some countries and a peak in the Anthropogenic and Scientific clusters.

3.3. Usage of web links

Analyzing the metadata of tweets inside each cluster provides valuable insights into the behavioral dynamics within the clusters. The use of URLs in tweets may have different meanings, depending on the context, such as an attempt to convey legitimacy or otherwise reinforce the content of the tweet with external validation (Elroy et al., 2023; Elroy and Yosipof, 2022). Fig. 4A shows the number of tweets that use a URL and the number of tweets that do not use a URL in each cluster. We found that tweets in the Anthropogenic and Scientific clusters use fewer URLs to back their stances. This finding is reasonable in light of the narratives of these clusters, where in the Anthropogenic cluster, the discussion is mostly casual, and in the Scientific cluster, the discussion relies on scientifically accredited claims. In contrast, URLs are used more frequently in the Policy and Conspiracy clusters. The finding that tweets in the Conspiracy cluster frequently use URLs is in line with previous research that found that tweets promoting conspiracies are more likely to use external resources to back their claims (Elroy and Yosipof, 2022).

Fig. 4B shows the most commonly referenced websites in each cluster. Comparing the most referenced websites in each cluster allows us to analyze behavioral dynamics in each cluster. We found that the Policy cluster mostly references websites that discuss policy matters, and most notably the IPCC AR6 report. As can be seen in Fig. 4B, these websites include the IPCC's website, as well as other websites that discuss the report and other policy matters, such as certain news agencies like The Guardian, The New York Times, The Washington Post, and others, as well as the Nature Journal's website.

The Anthropogenic and Conspiracy clusters reference websites that serve user content, such as YouTube, Rumble, and WordPress. In addition, tweets in the Conspiracy cluster cite websites that promote the narrative that climate change is not caused by humans, such as

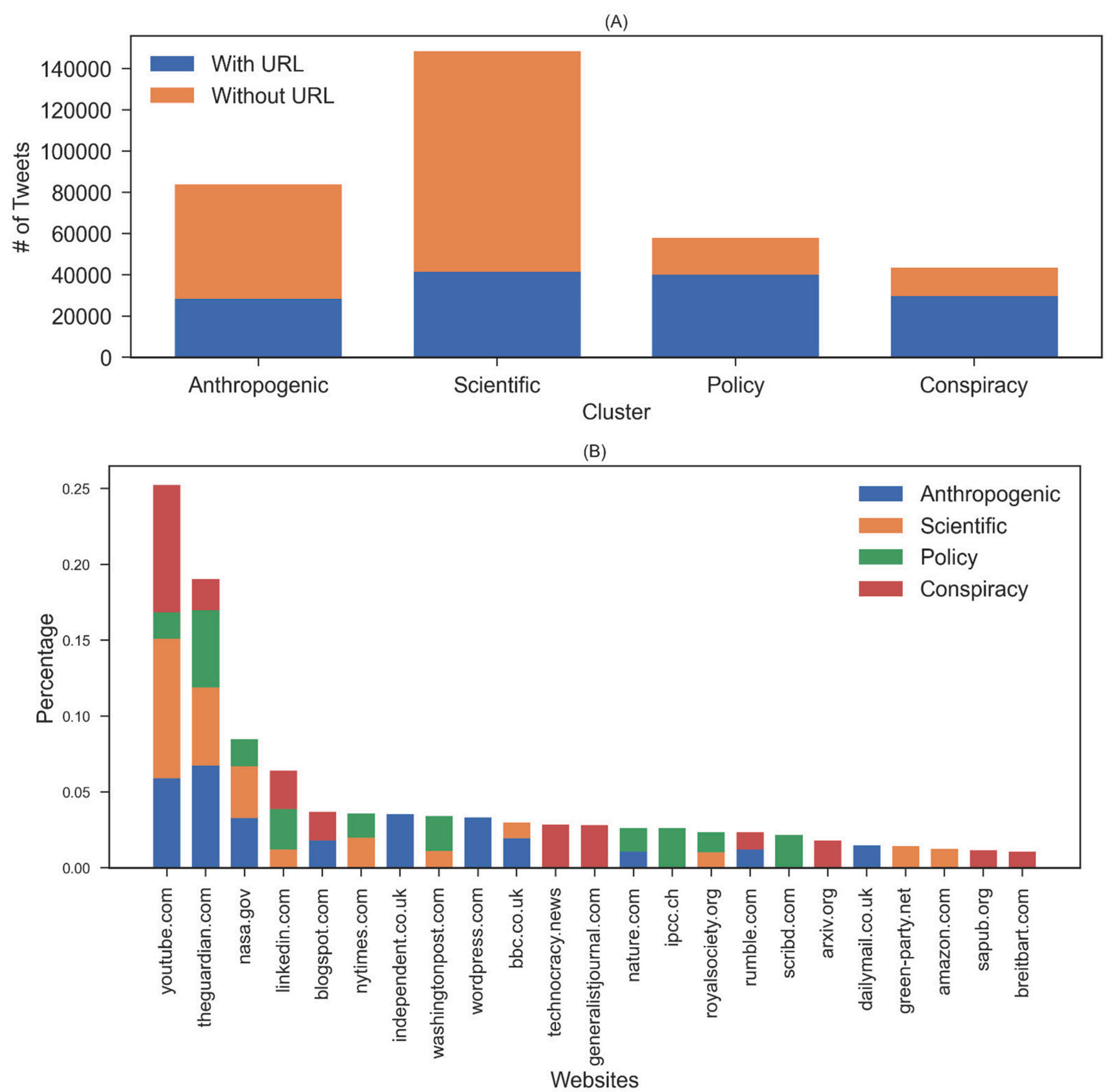

Fig. 4. (A) The number of tweets in each cluster that reference a URL is represented in blue, and the number of tweets in each cluster that do not reference a URL is represented in orange. (B) The most commonly referenced websites in each cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Technocracy News, for reinforcement.

YouTube is a prominent source of online content that is frequently referred to in all four clusters, and for that reason and others, content posted on YouTube cannot be easily distinguished as credible or not. An in-depth analysis of the tweets found that >500 tweets in the conspiracy cluster cited Tim Ball's books to support their claims. Tim Ball is a well-known climate change skeptic according to the Climate Disinformation Database (DeSmog, 2020). In addition, >100 tweets cited Patrick Moore, another climate change skeptic according to the Climate Disinformation Database (DeSmog, 2020), that promotes the notion that climate change is fake science. The similar behavioral dynamics in the Anthropogenic and Conspiracy clusters can be attributed to the similar nature of the discourses in these clusters. While users typically only interact with like-minded users in communities dominated by a single view, these clusters are of the mixed-attitude communities in which skeptics and activists frequently interact between them (Williams et al., 2015).

4. Discussion

The results enable us to identify four different narratives in the anthropogenic climate change discussion on X (formerly Twitter), namely Anthropogenic, Scientific, Policy, and Conspiracy.

The Policy cluster focuses on specialized terms or policy measures such as climate change, rights, impacts, or health. It also has a more global perspective while talking about the planet and the world. The discourse within the different clusters is driven by a substantially different number of unique users, however, users are generally as active as in the other clusters. The Anthropogenic and Conspiracy clusters mainly focus on the causes of climate change and question the existence of climate change in general, the human factor in climate change, and other scientifically based facts. The Policy cluster received increased attention when the IPCC AR6 report was published on February 28, 2022. We found that other than the official IPCC website, tweets in this cluster mostly reference websites of major news agencies, that likely reported about the report or summarized parts of the report. This finding is in line with the findings of previous research that found that people rely on media representations to help interpret and understand the complex issues surrounding climate science, governance, and decision-making (O'Neill et al., 2015). Furthermore, different media outlets frame their articles differently according to how they wish to be perceived by their target audience (O'Neill et al., 2015). Our finding is also in line with the previous research from O'Neill et al. (2015), who found that media outlets favor particular frames, like The Guardian which positioned itself as a liberal voice with a high level of coverage, much of which is framed as a political or ideological struggle over solutions or strategy to address climate change.

This work supports the findings in previous research that extreme weather events generate attention to climate change on X (formerly Twitter) (Sisco et al., 2017). In addition to extreme weather, we found that other major events also attract significant attention and increase the daily frequency of tweets during and after these events, such as the release of IPCC annual reports on climate change.

The fact that 97% of climate scientists agree that climate change is happening and is largely caused by humans (Hornsey and Lewandowsky, 2022), is commonly used by users to support their stance. Experiments show that the 97% heuristics are more likely to accept the existence of human-caused climate change and, in turn, to support policy interventions than those who are not (Cook et al., 2017; Lewandowsky et al., 2013). It is therefore sensible that users who participated in the Scientific cluster use this fact as a tool to persuade, e.g., *"CO_2 contributes to climate change, as evidence by 10,000s of research papers from 10,000 scientists. …"*.

The Anthropogenic cluster presents a substantial narrative in the discussion, which focuses on whether climate change is man-made or not, or the level of contribution humans have in the current situation or could have to a possible solution. The important question in this regard is 'Why would people want to reject anthropogenic climate change?', and any answer to that question helps our understanding of conspiracy theories and the discussion regarding anthropogenic climate change in particular (Hornsey and Lewandowsky, 2022). One explanation is that although the discussion regarding climate change is rich in scientific research, it may still be uneasy to accept that climate change is an inconvenient truth in that the solution implies painful sacrifices (Hornsey and Fielding, 2017). Rejecting scientific results is typically easier than taking actions that are possibly painful (Campbell and Kay, 2014), e.g., by clinging to baseless conspiracy theories or shifting standards of proof as a function of how convenient the evidence is (Hornsey et al., 2018).

Additionally, the results show that tweets in the Conspiracy cluster tend to refer to websites that provide content uploaded by individual users, mostly without any verification of its source or facts, such as by scientists recognized to be climate change deniers. This finding is in line with previous works that found evidence of coordinated and well-funded spread of misinformation about climate change (Hornsey et al., 2018). This coordination and funding is also reflected in the fact that most environmentally skeptical books are published or financed by conservative think tanks (Dunlap and Jacques, 2013; Jacques et al., 2008), and that climate-denying think tanks and advocacy organizations receive multi-million funding annually (Brulle, 2014). The Conspiracy cluster is the smallest of the four clusters in terms of the number of tweets and the number of unique authors. It is still significant enough to justify further development of media policy measures such as awareness measures, education, critical thinking, and artificial intelligence tools to check information online and raise attention to conspiracy narratives.

The research on the different narratives in the discussion about climate change on social media emphasizes the challenges related to climate change adaption and mitigation measures, and awareness of the population to climate change as part of the process.

Practical suggestions focus on the communication strategies of official organizations, such as governments. Such strategies should address the need to raise awareness, adequately communicate, and improve the readiness of the public for climate change adaptation and increase participation in mitigation measures. These strategies include communication through social media, traditional media, and official channels such as governmental websites.

Another practical suggestion is to help educate the public through the dissemination of correct information. The discussion on climate change is comprised of many subjects, and many participants are involved in the discussion, especially on social media. Following our finding that users tend to be more involved in the discussion when significant reports such as the IPCC report are issued, the issuance of IPCC reports may provide an optimal time window for dissemination of such information.

Conspiracy theories are continuously present on social media, and climate change conspiracy theories are no exception. For this reason, measures are continuously needed to counteract them, such as providing adequate, reliable, and correct information. As such, communication strategies should address misinformation and conspiracy theories on various social media platforms by monitoring multiple platforms. Monitoring social media for the narratives and the change of narratives over time can provide a baseline for decision support tools that are essential to every policy process.

5. Conclusion

In this work, we collected 333,635 tweets in English about anthropogenic climate change. We used Natural Language Processing (NLP) and machine learning methods to embed the semantic meaning of the tweets into vectors, cluster the tweets, and analyze the results.

We discovered and analyzed four different narratives in the discussion of anthropogenic climate change on social media, namely

Anthropogenic, Scientific, Policy, and Conspiracy. Our results show that the existence of climate change, the human factor leading to the current situation, and the solution are contested issues that attract controversial discourses.

We found evidence to support the hypothesis that major events, such as extreme weather events and significant climate change reports and conferences, generate attention to the discussion about climate change on social media and increase the daily frequency of tweets in different clusters. We also found that users on both sides are likely to use scientific facts and resources to support their stances.

Our results can serve as input for media policy and awareness-raising measures on climate change mitigation and adaptation policies. The findings about attention to climate change after extreme weather events, as well as on controversies in discourses about the existence of climate change and its causes, can serve as input for media policy and awareness-raising measures on climate change mitigation and adaptation policies, and facilitate future communications related to climate change. These results can also provide input to communication policy and how climate change, its causes, and consequences could be communicated to various social groups considering such issues as trust in various sources of information, including science, and existing level of awareness about climate change, particular climate change mitigation, and adaptation measures. These results can also help improve the communication and awareness-raising measures to address various discourses. Furthermore, certain policy measures are needed to address misinformation and conspiracy theories while providing tools for pre-bunking and debunking incorrect information about the existence of climate change, its causes, and its consequences.

Future studies could seek to expand the research to other, different languages and investigate whether the narratives found in this work can also be observed in languages or regions that speak languages other than English. Future works can further investigate specific clusters identified in this work, such as using supervised classification methodology to analyze climate change denial. Future works may use other unsupervised learning methods, such as topic modeling. The data used in this work is limited to the discussion related to anthropogenic climate change on X (formerly Twitter) over a specific period. Discussion on other social media platforms, such as Reddit, Facebook, or TikTok, can also be investigated, as well as over different or longer periods. Applying the research knowledge provided in this work in future communications of climate change, such as future IPCC reports, could facilitate the communication of the problems, the consequences, and possible solutions to the public.

CRediT authorship contribution statement

Or Elroy: Conceptualization, Resources, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Nadejda Komendantova:** Writing – original draft, Writing – review & editing. **Abraham Yosipof:** Conceptualization, Resources, Methodology, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing 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.

Data availability

The tweets dataset was collected from X (formerly Twitter) using limited academic research API access. The dataset can be retrieved from X (formerly Twitter) according to the query explained in the methodology. Other data that supports the results is available from the corresponding author upon reasonable request.

Acknowledgements

This research has 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).

References

Adoma, A.F., Henry, N.-M., Chen, W., 2020. Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp. 117–121.

ART News, 2022. Climate Activist Attempts to Glue Head to 'Girl with a Pearl Earring'. Retrieved December 17, 2023 from. https://www.artnews.com/art-news/news/climate-activist-attempts-to-glue-head-to-girl-with-a-pearl-earring-1234644651.

Arthur, D., Vassilvitskii, S., 2007. K-means++ the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035.

Aschwanden, A., Demir, C., Hinselmann, R., Kasser, S., Rohrer, A., 2018. Zika and travel: public health implications and communications for blood donors, sperm donors and pregnant women. Travel Med. Infect. Dis. 21, 77.

Atwoli, L., Erhabor, G.E., Gbakima, A.A., Haileamlak, A., Ntumba, J.M.K., Kigera, J., Laybourn-Langton, L., Mash, R., Muhia, J., Mulaudzi, F.M., Ofori-Adjei, D., Okonofua, F., Rashidian, A., El-Adawy, M., Sidibé, S., Snouber, A., Tumwine, J., Yassien, M.S., Yonga, P., Zakhama, L., Zielinski, C., 2022. COP27 Climate change conference: Urgent action needed for Africa and the world. Lancet Oncol 23 (12), 1486–1488. https://doi.org/10.1016/S1470-2045(22)00645-3.

Bahmani, B., Moseley, B., Vattani, A., Kumar, R., Vassilvitskii, S., 2012. Scalable k-means ++. arXiv preprint arXiv:1203.6402.

Batzdorfer, V., Steinmetz, H., Biella, M., Alizadeh, M., 2022. Conspiracy theories on twitter: emerging motifs and temporal dynamics during the COVID-19 pandemic. Int. J. Data Sci. Anal. 13 (4), 315–333.

Brulle, R.J., 2014. Institutionalizing delay: foundation funding and the creation of US climate change counter-movement organizations. Clim. Chang. 122, 681–694.

Campbell, T.H., Kay, A.C., 2014. Solution aversion: on the relation between ideology and motivated disbelief. J. Pers. Soc. Psychol. 107 (5), 809.

CNN, 2022. Fossil Fuel Protesters Charged After Tomato Soup Thrown on Van Gogh's 'Sunflowers' in London Gallery. Retrieved December 17, 2023 from. https://edition.cnn.com/style/article/oil-protest-van-gogh-sunflower-soup-intl-scli-gbr/index.html.

Coan, T.G., Boussalis, C., Cook, J., Nanko, M.O., 2021. Computer-assisted classification of contrarian claims about climate change. Sci. Rep. 11 (1), 22320.

Cook, J., Lewandowsky, S., Ecker, U.K., 2017. Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. PLoS One 12 (5), e0175799.

Cooke, A.S., Rivero, M.J., 2023. Livestock heat stress risk in response to the extreme heat event (heatwave) of July 2022 in the UK. bioRxiv. https://doi.org/10.1101/2023.05.18.541284, 2023.2005.2018.541284.

Dahal, B., Kumar, S.A., Li, Z., 2019. Topic modeling and sentiment analysis of global climate change tweets. Soc. Netw. Anal. Min. 9, 1–20.

Daily Mail, 2022. Decomposing HUMAN Remains can Legally be used as Compost from 2027 Thanks to New California Law Aimed at Tackling Climate Change. Retrieved December 17, 2023 from. https://www.dailymail.co.uk/sciencetech/article-11236485/HUMAN-remains-used-compost-2027-new-California-law-tackle-climate-change.html.

Dallo, I., Elroy, O., Fallou, L., Komendantova, N., Yosipof, A., 2023. Dynamics and characteristics of misinformation related to earthquake predictions on twitter. Sci. Rep. 13 (1), 13391. https://doi.org/10.1038/s41598-023-40399-9.

Das, S., Chakraborty, S., 2022. Perception of United Nations climate change conference in social networks. In: 2022 IEEE 19th India Council International Conference (INDICON), pp. 1–6.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

DeSmog, 2020. Climate Disinformation Database. Retrieved May 29, 2024 from https://www.desmog.com/climate-disinformation-database/.

Dunlap, R.E., Jacques, P.J., 2013. Climate change denial books and conservative think tanks: exploring the connection. Am. Behav. Sci. 57 (6), 699–731.

Elroy, O., Yosipof, A., 2022. Analysis of COVID-19 5G conspiracy theory tweets using sentence BERT embedding. In: International Conference on Artificial Neural Networks, pp. 186–196.

Elroy, O., Yosipof, A., 2023. Semi-supervised learning classifier for misinformation related to earthquakes prediction on social media. In: International Conference on Artificial Neural Networks, pp. 256–267.

Elroy, O., Erokhin, D., Komendantova, N., Yosipof, A., 2023. Mining the discussion of Monkeypox misinformation on twitter using RoBERTa. In: IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 429–438.

Erokhin, D., Yosipof, A., Komendantova, N., 2022. COVID-19 conspiracy theories discussion on twitter. Soc. Media Soc. 8 (4) https://doi.org/10.1177/20563051221126051, 20563051221126051.

Essebo, M., 2022. Storying COVID-19: fear, digitalisation, and the transformational potential of storytelling. Sustain. Sci. 17 (2), 555–564.

Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, N., Larosa, F., Sas, M., Mekacher, A., Pearce, W., Zollo, F., Quattrociocchi, W., 2022. Growing polarization around climate change on social media. Nat. Clim. Chang. 12 (12), 1114–1121.

Fischer, A.P., Russo, M., Powers, G., 2022. Behavioral adaptation to climate change: new insights on psychosocial frameworks from the context of managed forests. Sustain. Sci. 17 (3), 1023–1036.
Fraser, T., Morikawa, L., Aldrich, D.P., 2021. Rumor has it: the role of social ties and misinformation in evacuation to nearby shelters after disaster. Clim. Risk Manag. 33, 100320.
Fung, G., 2001. A Comprehensive Overview of Basic Clustering Algorithms.
Goldberg, M.H., van der Linden, S., Maibach, E., Leiserowitz, A., 2019. Discussing global warming leads to greater acceptance of climate science. Proc. Natl. Acad. Sci. 116 (30), 14804–14805.
González-Carvajal, S., Garrido-Merchán, E.C., 2020. Comparing BERT against Traditional Machine Learning Text Classification. arXiv preprint arXiv:2005.13012.
Hilary, I.O., Dumebi, O.-O., 2021. Social media as a tool for misinformation and disinformation management. Linguist. Cult. Rev. 5 (S1), 496–505.
Holley, D., Lee, S.H., 2022. Forecasting extreme heat in the UK during July 2022. Weather 77 (9), 320–321.
Hornsey, M.J., Fielding, K.S., 2017. Attitude roots and Jiu Jitsu persuasion: understanding and overcoming the motivated rejection of science. Am. Psychol. 72 (5), 459.
Hornsey, M.J., Lewandowsky, S., 2022. A toolkit for understanding and addressing climate scepticism. Nat. Hum. Behav. 6 (11), 1454–1464.
Hornsey, M.J., Harris, E.A., Fielding, K.S., 2018. Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations. Nat. Clim. Chang. 8 (7), 614–620.
Jacques, P.J., Dunlap, R.E., Freeman, M., 2008. The organisation of denial: conservative think tanks and environmental scepticism. Environ. Polit. 17 (3), 349–385.
Jerneck, A., Olsson, L., Ness, B., Anderberg, S., Baier, M., Clark, E., Hickler, T., Hornborg, A., Kronsell, A., Lövbrand, E., 2011. Structuring sustainability science. Sustain. Sci. 6, 69–82.
Komendantova, N., Ekenberg, L., Svahn, M., Larsson, A., Shah, S.I.H., Glinos, M., Koulolias, V., Danielson, M., 2021. A value-driven approach to addressing misinformation in social media. Humanit. Soc. Sci. Commun. 8 (1), 1–12.
Kwanda, F.A., Lin, T.T., 2020. Fake news practices in Indonesian newsrooms during and after the Palu earthquake: a hierarchy-of-influences approach. Inf. Commun. Soc. 23 (6), 849–866.
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., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., Zittrain, J.L., 2018. The science of fake news. Science 359 (6380), 1094–1096. https://doi.org/10.1126/science.aao2998.
Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W., Connors, S., Denton, F., Diongue Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Zommers, Z., 2023. Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
Lewandowsky, S., Gignac, G.E., Vaughan, S., 2013. The pivotal role of perceived scientific consensus in acceptance of science. Nat. Clim. Chang. 3 (4), 399–404.
Madhulatha, T.S., 2012. An Overview on Clustering Methods. arXiv preprint arXiv:1205.1117.
Matsui, T., Suzuki, K., Ando, K., Kitai, Y., Haga, C., Masuhara, N., Kawakubo, S., 2022. A natural language processing model for supporting sustainable development goals: translating semantics, visualizing nexus, and connecting stakeholders. Sustain. Sci. 17 (3), 969–985.
Naseer, M., Asvial, M., Sari, R.F., 2021. An empirical comparison of bert, roberta, and electra for fact verification. In: 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 241–246.
O'Neill, S., Williams, H.T., Kurz, T., Wiersma, B., Boykoff, M., 2015. Dominant frames in legacy and social media coverage of the IPCC fifth assessment report. Nat. Clim. Chang. 5 (4), 380–385.
Ortiz-Martínez, Y., Garcia-Robledo, J.E., Vásquez-Castañeda, D.L., Bonilla-Aldana, D.K., Rodriguez-Morales, A.J., 2020. Can Google® trends predict COVID-19 incidence and help preparedness? The situation in Colombia. Travel Med. Infect. Dis. 37, 101703.
Peary, B.D., Shaw, R., Takeuchi, Y., 2012. Utilization of social media in the East Japan earthquake and tsunami and its effectiveness. J. Nat. Dis. Sci. 34 (1), 3–18.
Peng, Z., 2020. Earthquakes and coronavirus: how to survive an infodemic. Seismol. Res. Lett. 91 (5), 2441–2443.
Piskorski, J., Haneczok, J., Jacquet, G., 2020. 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, pp. 6663–6678.
Reimers, N., Gurevych, I., 2019. Sentence-Bert: Sentence Embeddings using Siamese Bert-Networks. arXiv preprint arXiv:1908.10084.
Reuters, 2023. Climate Change Activists Protest at Berlin Formula E Race. Retrieved December 17, 2023 from. https://www.reuters.com/sports/motor-sports/climate-change-activists-protest-berlin-formula-e-race-2023-04-23/.
Satopaa, V., Albrecht, J., Irwin, D., Raghavan, B., 2011. Finding a" kneedle" in a haystack: detecting knee points in system behavior. In: 2011 31st International Conference on Distributed Computing Systems Workshops, pp. 166–171.
Sisco, M.R., Bosetti, V., Weber, E.U., 2017. When do extreme weather events generate attention to climate change? Clim. Chang. 143, 227–241.
Tarunesh, I., Aditya, S., Choudhury, M., 2021. Trusting Roberta Over Bert: Insights from Checklisting the Natural Language Inference Task. arXiv preprint arXiv:2107.07229.
United Nations Environment Programme, 2023. Climate Change 2023: Synthesis Report. Retrieved December 17, 2023 from. https://www.unep.org/resources/report/climate-change-2023-synthesis-report.
Uthirapathy, S.E., Sandanam, D., 2023. Topic modelling and opinion analysis on climate change twitter data using LDA and BERT model. Procedia Comput. Sci. 218, 908–917.
Walter, S., Brüggemann, M., Engesser, S., 2018. Echo chambers of denial: explaining user comments on climate change. Environ. Commun. 12 (2), 204–217.
Williams, H.T., McMurray, J.R., Kurz, T., Lambert, F.H., 2015. Network analysis reveals open forums and echo chambers in social media discussions of climate change. Glob. Environ. Chang. 32, 126–138.
Zhou, C., Xiu, H., Wang, Y., Yu, X., 2021. Characterizing the dissemination of misinformation on social media in health emergencies: an empirical study based on COVID-19. Inf. Process. Manag. 58 (4), 102554.