A Comparative Analysis of Climate Change Discourse on Reddit and Twitter

Robert Dow,[1] **Yunyi Xu,** [2] **Ruoyun Zhang** [2]

[1] Department of Cognitive Science, McGill University
[2] Department of Computer Science, McGill University
{robert.dow},{yunyi.xu},{ruoyun.zhang}@mail.mcgill.ca

Abstract

Climate change is a crucial issue in our society. With the rise of public awareness of climate change, this matter is widely discussed throughout social media platforms. In this paper, we analyze climate change discourse on Twitter and Reddit as it occurs in the subreddit r/AskReddit. These two sources were chosen because of their popularity and distinct functionalities. This paper presents a comparative analysis of 2,500 combined tweets and Reddit posts from throughout 2021, along with 4,686 corresponding Reddit comments. The posts were collected by keyword search with the following words: *climate change, climate crisis, global warming, and greenhouse gas*. We categorized the collected data into 5 topics by manual annotation and then performed engagement, sentiment, and TF-IDF analyses. The results suggest that Twitter users show high civic engagement with climate change, as their discourse reflects sentiments and vocabularies that align with popular political trends. This is in part due to Twitter's role as an information hub, hosting scientific articles and news. In contrast, Reddit users engaged with climate change issues more broadly while personalizing responsibilities related to climate change and showing higher conversational engagement. This study prompts questions for future research that relate factors such as the presence of news and meaningful discussion on social media platforms to the politicization of ideas, responsibilities, and understanding of climate change.

1 Introduction

Anthropogenic climate change is a major issue of our time, with serious environmental, social, and economic implications (Carleton and Hsiang 2016). As governments and citizens around the world work towards finding and implementing new strategies to combat climate change, it is becoming increasingly important that we understand public opinion on the issue. By better understanding public reception of climate change information as it occurs in places like the news media, we can learn how to more effectively inform and engage the public (Gunster 2017). One important way we can gain such an understanding is by looking towards social media.

Social media has become increasingly influential in receiving and discussing news. Studies show that two out of three American adults use social media as a news source (Shearer and Gottfried 2017). It has also been found that social media contributes to an increased level of participation in political issues (Halpern, Valenzuela, and Katz 2017), and fast dissemination of educational content (González-Padilla and Tortolero-Blanco 2020). Similar situations happened during the 2019 Climate Strike. The initiator, Greta Thunberg, gained the public's attention by posting photos of her strikes on Instagram and Twitter (Wikipedia 2022). Moreover, it has been shown that internet searches relating to climate change have noticeably increased since 2019 (Sisco et al. 2021). For instance, over the period of 3 days in March 2019, 13,542 tweets with the hashtag #SchoolStrike4Climate were posted (Boulianne, Lalancette, and Ilkiw 2020).

We conducted this study to analyze how climate change is discussed on social media. In particular, we performed a comparative analysis to find out the similarities and differences in the ways of engaging with climate change online. We used data from Twitter and Reddit, two popular yet distinct social media platforms. Studies have shown that instead of a socially active platform, Twitter has become a pool of constantly updating information (Anger and Kittl 2011). Reddit, on the other hand, is an online forum which can provide a unique insight into how groups of individuals yield collective thinking (Medvedev, Lambiotte, and Delvenne 2019).

This study involved the collection and analysis of 1,000 Twitter posts, 1,500 Reddit posts from subreddit r/AskReddit, and 4,868 corresponding Reddit comments occurring throughout 2021. The posts were collected using a keyword search with the following words: *global warming, climate change, climate crisis,* and *greenhouse gas*. By manual annotation, each post was categorized into one of the following five topics: *Technical, Responsibilities, Forecasting, Denial and Belief,* and *Other* with topic categorization of Reddit comments being labeled the same as its corresponding post. We performed TF-IDF, engagement, and sentiment analyses across all datasets. We observed that the messaging on Twitter and Reddit are quite distinct. Posts on Twitter were found to be highly engaged with socio-political issues, whereas Reddit users engaged with climate change more broadly as it relates to the individual and humanity as a whole. Furthermore, posts on Reddit were found to have

higher levels of engagement and more negative sentiment than Twitter. In short, this research explores the dimensions of discussions around climate change, how this discourse differs across platforms, and the general attitude of citizens towards climate change.

2 Related Work

Studies investigating the public's attitudes towards climate change span multiple dimensions. A 2018 survey looking at belief of climate change suggested that 70% of Americans believed that climate change is happening, however only half of them viewed it as a personal risk (Leiserowitz et al. 2019). Another report showed that the majority of the Americans considered that climate change has been worsening over the past few years (Marlon 2014). Benegal (2018) studied Americans' opinions on climate change over 12 years and concluded that the Republican partisanship and the conservative ideology are associated strongly with climate change denial. Age also played a part, where the younger generations tend to pay more attention to climate change (Benegal 2018).

Many studies also focused on the discussions of climate change on social media. Olteanu et al. (2015) compared climate news on social media and in the mainstream media and found out that there were different triggers of news coverage between the media. Kirilenko and Stepchenkova (2014) analyzed the discussion of climate change on Twitter in multiple dimensions, including the geographic patterns, the central topics of discussion, and the major news events that trigger discussion. In a 2013 study, it was found that climate anomalies, especially hot temperatures, have a strong connection with the tweeting intensity of climate discourses (Molodtsova, Kirilenko, and Stepchenkova 2013). Social media messaging also has correlations to real-world events such as protests (Boulianne and Theocharis 2018). Boulianne, Lalancette, and Ilkiw (2020) analyzed the functions of tweets to the global dynamics of the climate strike that happened in March 2019.

Different from previous studies, our research involves a comparative analysis of climate change discourse across different social media platforms. Consequentially, our study highlights the differential roles of platforms in attracting, amplifying, and otherwise hosting climate change discourse.

3 Data Collection

3.1 Defining Climate Change Discourse Online

Discourse about climate change takes many forms. It often involves engagement with social, political, or environmental matters such as climate justice, environmental policy, or deforestation. More broadly, climate change discourse involves the synthesis or dissemination of thoughts, opinions, and even news as it relates to the Earth's change in climate and the corresponding ways that this change affects people.

This study aims to identify climate change discourse online by using a keyword search to classify Twitter and Reddit posts. This classification task was done using the following keywords: *climate change, climate crisis, global warming, and greenhouse gas*. By limiting our definition of climate change discourse to include only posts containing these keywords, we fail to capture all relevant discourse. However, this heuristic greatly reduces the likelihood of misclassification. Additionally, by excluding more specific phrases such as "deforestation" and "ocean acidification", we instead get a representation of how climate change is more generally contended with as a topic.

The climate change discourse data collected from Twitter and Reddit contain tweets as well as Reddit posts and comments from the subreddit "r/AskReddit". These two sources serve as distinct settings for conversations around climate change. r/AskReddit is a forum on Reddit in which people can ask questions about anything and be met with answers and comments from users across the platform. It is the third most popular subreddit, with over 35 million subscribers. It also hosts more posts containing the aforementioned keywords than any other subreddit. In contrast, the content and form of posts on Twitter are not constrained by the rules of a subreddit and instead vary broadly in their construction. Some notable and prevalent styles of tweets are the reproduction and dissemination of news, engagement in political discourse, and stream of consciousness thoughts and opinions. While posts from r/AskReddit represent what kinds of questions people have about climate change, Twitter posts offer a broader picture of climate change news and discourse. Together, these posts are used to characterize climate change discourse online.

3.2 Data Acquisition

The datasets used in this study were constructed using sampling techniques that aim to represent climate change discourse throughout 2021 on both Twitter and Reddit.

Reddit Datasets We collected two Reddit datasets for this study. Each was gathered using the Python Pushshift API Wrapper (PSAW) which allows the retrieval of archived Reddit posts from the Pushshift project (see Baumgartner et al. (2020)).

The first dataset (henceforth referred to as the Reddit Questions dataset) contains climate change related questions posted in the r/AskReddit subreddit during 2021 along with information about the number of comments associated with each post. The dataset was gathered by first collecting all Reddit posts from 2021 that were made in the subreddit and also contained any of the keywords mentioned in section 3.1. Of these 2,820 posts, we randomly sampled 1,500. The proportions at which each keyword appears in the dataset is outlined in Table 1.

The second Reddit dataset (henceforth referred to as the Reddit Comments dataset) was formed by collecting comments associated with the 1,500 posts. A maximum of 10 comments were collected for each post. Comments were collected in the order at which they were created. Following this procedure, there were a total of 4,686 comments remaining.

Twitter Dataset The Twitter dataset used in this study was constructed with the use of the Tweepy API (Roesslein 2020). Samples of the 200 most recent posts containing any of the keywords mentioned in section 3.1 were collected at random times and days within the first and last half of

each month. This procedure was intended to reduce biases towards a particular time of the day or year in creating a representative sample from 2021. In total, 4,800 posts were collected. The final dataset consisted of a randomly sampled subset of 1,000 posts along with metadata such as the number of Likes, Replies, and Retweets associated with each post. Statistics on the frequency at which each keyword appeared in the dataset are listed in Table 1.

	Climate change	Climate crisis	Global warming	Greenhouse gas
Proportion on Twitter	63.9%	8.4%	11.3%	.31%
Proportion on Reddit	71.1%	6.3%	23.3%	.01%

Table 1: Proportions at which keywords appeared in the Twitter and Reddit Questions datasets

4 Methods and Data Annotation

This section focuses on the methods of processing and analyzing the Twitter and Reddit datasets. First, we present the topics that are used for data annotation as well as the motivation behind the topic-selection process (4.1). Next, we provide detailed information on the procedure of data annotation (4.2). Finally, we discuss the implementation procedure of the TF-IDF algorithm (4.3) and sentiment analysis (4.4).

4.1 Topic Selection

Topics used to annotate the three datasets used in this study were chosen manually using an open coding technique. First, subsets of 200 posts were randomly generated from both the Twitter and Reddit Questions datasets. Next, each author of this paper generated 3-8 topics that mutually described posts in the two subsets. Subsequently, topics were discussed and considered based on their ability to explain prevalent and insightful modes of conversation. After considering topics that distinguished things such as science vs news and conjecture vs fact, we found it was most informative to select topics that differentiated the content of the text rather than its style. In doing so, we left pervasive sentiments and themes to materialize in our analysis.

Upon deliberation, the following 5 topics were chosen to be used across datasets: Technical, Responsibilities, Forecasting, Denial and Belief, and Other. The topics are defined as follows:

- Technical: posts related to the sharing of or inquiry about technical information such as news, scientific research, or generally about how climate change works.
- Responsibilities: posts that assert responsibility towards an individual or group for causing or counteracting climate change. Posts in this category include the discussion of blame, plans of action, and content relating to what one ought to do in the face of accelerating climate change.
- Forecasting: posts that raise questions or assertions about the future as it relates to climate change, including how it will affect human life.
- Denial and Belief: posts that mention climate change denialism or more broadly contend with the idea of climate change as something to be believed in. This includes both posts that assert denialist opinions as well as discussion or inquiry about climate change denialism.
- Other: posts that fit in no other category. These posts vary across a range of topics, and a small number of them mention climate change related keywords but do not meaningfully contend with the topic.

4.2 Data Annotation

Data annotation was carried out manually by the three authors of this paper for both the Twitter and Reddit Questions datasets. The Reddit Comments dataset was annotated automatically by inheriting the topic label of the post it is a comment of.

The annotation process involved categorizing posts into one of five topics which were used across datasets. Topics were prioritized in the following order: 1 - Denial and Belief, 2 - Responsibilities, 3 - Forecasting, 4 - Technical, and 5 - Other. This means that if a post mentioned more than one of these topics, its topic label would default to that with the lowest number. The intuition behind this system was that lower numbered topics were generally the most informative in characterizing a post when co-existing with another topic. For example, the Reddit post "Climate change deniers of Reddit, where do you think all of the CO2 and stinky stuff coming out of the hundreds of millions of cars driven every day actually goes?" falls into 1 - Denial and Belief, rather than in 4 - Technical. While the post is technical in nature, we find that the sentiment of climate change denial is more pervasive in defining its overall significance.

The annotation procedure was carried out in the same way for both datasets. Each dataset was divided up equally such that each post was annotated by two authors, and any ties were broken by the third. This triple annotation was done "blind" meaning that annotators were not able to view each other's label assessments while carrying out annotation. However, the first 100 posts of each dataset were an exception to this "blind" strategy which gave the annotators an opportunity to verify our mutual understanding of the topic definitions within the context of the datasets.

4.3 TF-IDF

Before performing the TF-IDF (term frequency-inverse document frequency) analysis, the three datasets were cleaned. This process involved lower-casing the content of the posts, removing punctuation, links, numbers, and any non-alphanumeric characters. Then, posts were further cleaned by filtering out all stop words using the stop list provided by Landauer and Dumais (1997). This process helped optimize the identification of meaningful words across posts.

Subsequently, the TF-IDF analysis was performed on each of the three datasets. The implementation of this algorithm used topics as documents but was otherwise standard. TF-IDF scores words more highly if they occur more

frequently in a topic, but decreases their score if they occur in more topics. By following this procedure, this analysis produced the top most relevant and informative words that characterized each topic.

4.4 Engagement Analysis

We measured engagement in this study as the frequency at which users commented on posts in the case of the Reddit Questions dataset and the rate at which users replied to, liked, or retweeted posts for the Twitter dataset. These data were processed by using statistical measurements such as means and medians. Data was also further analyzed by finding the min, max, and interquartile range and presenting the results in a box and whisker plot. This analysis sought to characterize engagement on each platform as well as across topics within the datasets.

4.5 Sentiment Analysis

We performed an automated sentiment annotation for the Twitter dataset and the Reddit Comments dataset with the use of a fine-tuned version of the RoBERTa-large model (Liu et al. 2019). This language model was fine-tuned on a series of texts including reviews and tweets by Hartmann et al. (2022) to optimize its performance on the binary sentiment classification task across different text types. Since the model is informed by punctuation and capitalization, each post and comment was input to the model as is. This method was used to classify the sentiments of each entry of the two datasets.

5 Results

5.1 Annotations

The distribution of posts into topics for each dataset is displayed in Table 2. All datasets were found to have the highest proportion of posts labeled as Technical or Responsibilities and the lowest proportion of posts labeled as Denial and Belief or Forecasting. There were only 2.2% of posts (22 in total) labeled as Forecasting in the Twitter dataset. Thus, the subsequent analyses of the forecasting Twitter posts have limited interpretability.

	Twitter	Reddit Questions	Reddit Comments
Responsibilities	29.1%	25.0%	25.5%
Technical	30.9%	26.1%	27.9%
Other	21.2%	24.1%	22%
Denial and Belief	16.6%	13.5%	12.7%
Forecasting	2.2%	11.2%	11.7%
Total (count)	1000	1500	4686

Table 2: Proportions of posts per topic

5.2 TF-IDF

The results of the TF-IDF analysis are displayed in Table 3 in decreasing order by their TF-IDF scores. Out of twelve highest-scoring words for each category, five words were selected for each topic. Words representing the results of the TF-IDF analysis were chosen by prioritizing nouns and adjectives as they were more informative and provided better understanding on the relationship between words and topics.

5.3 Engagement

Engagement was measured as it relates to the direct interaction of users with posts from both the Twitter and Reddit Questions dataset. Overall, our study found that there was higher engagement on Reddit than on Twitter.

Reddit Reddit engagement was measured in terms of the number of comments left on each post in the Reddit Questions dataset. The results are displayed in Figure 1. The X's in the plot represent the mean values, the color bars represent the interquartile ranges, and the thin lines in the color bars represent the median values. Note that outliers were removed from the Box and Whisker plot because they were not statistically important and they affected the readability of the plot.

Figure 1 illustrates that the levels of discussion around the topic Responsibilities are notably higher than the other topics with higher median, mean, interquartile range (IQR), and upper fence. Although the mean, IQR, and upper fence are also higher in Denial and Belief than in other topics, we found that the median of Denial and Belief is not preeminent (for reference, the median of Reddit comments is 5 with all topics combined).

Also, we noticed from Figure 1 that the values of the first quartile (Q1) are equally low across all topics which indicates that 25% of all posts recieve little or no engagement.

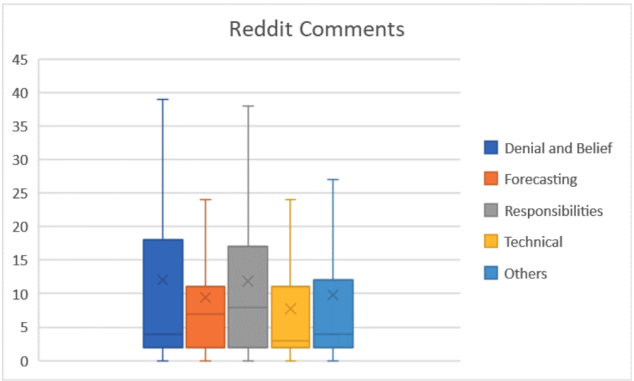

Figure 1: Box and Whisker plot representing the number of comments on Reddit posts by topic.

Twitter Twitter engagement was measured in respect to the number of Likes, Replies, and Retweets left on each post in the Twitter dataset. Overall, we found that over half of all tweets recieved no engagement across these metrics. Table

	Forecasting	Denial and Belief	Responsibilities	Technical	Other
Reddits Questions	century generation places wages plans	legitimate arguments real antivaccinators view	help individuals governments fight actions	water fires nature energy reduce	hunger wrong inequality future society
Reddit Comments	diets nutrition plantbased environment mood	operational mandate remedy depopulation governor	technical cared insinuation protecting liar	erosion devices banned downpours outbreaks	profitability outraged joes fallout footprints
Twitter	availability forests military intensity nationwide	fake hoax real anti concerned	infrastructure subsidize action urgent plan	alarm ocean methane country food	hate unvaccinated debate economics innovation

Table 3: Five out of top twelve TF-IDF words for each topic in the Reddit Questions, Reddit Comments, and in Twitter datasets.

4 reveals that Likes were the most popular form of engagement. Further analysis was done on engagement across topics, but results were not found to be statistically significant.

	Likes	Retweets	Replies
Average Count	5.09	1.39	0.41

Table 4: Average number of Likes, Replies, and Retweets per Twitter post.

5.4 Sentiment

Figures 2 and 3 display the sentiment classification results of the Twitter and Reddit Comments datasets respectively. The results of the automated sentiment classification were validated by manually annotating 100 randomly sampled posts from each dataset. 100 posts were found to have 93% agreement in the Twitter dataset, while 150 posts were found to have 87% agreement.

The results of the Twitter analysis show the average sentiment of climate change related tweets across topics was 53.1%, meaning more than half of tweets had a positive sentiment. Technical tweets were found to have the highest overall sentiment with a rate of 61.5% positive, while tweets labeled as Denial and Belief were found to have the lowest positive sentiment with only 29.8% positive.

While the average sentiment in the Twitter dataset varied between topics by 31.7%, there was only a variation of 4% in the Reddit Comments dataset. On average, 40.1% of comments were negative. Inversely to the Twitter dataset, the most positive topic was Denial and Belief with 43.3% positive posts while the most negative was Technical with only 39.3% postitive comments.

6 Discussion

From the results of the engagement analysis, it is noticeable that the measurable engagement levels in Reddit are

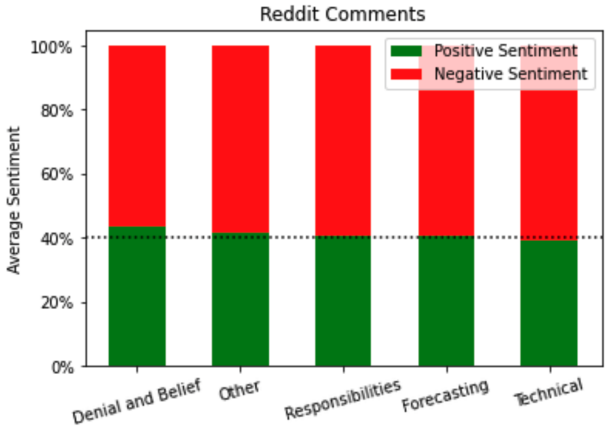

Figure 2: Average sentiment of Reddit comments by topic. Dotted line represents average sentiment over all topics.

higher than that in Twitter. We attribute this finding to a variation between the nature of these two social media platforms. Despite being a top social media platform, studies have suggested that there has been an enormous increase in bots on Twitter (Gilani et al. 2016), and that the platform is not "social" enough due to the low level of engagement in dialogues, thus making the platform an information pool (Anger and Kittl 2011). Reddit, on the other hand, is a forum oriented around collective discussion. In particular, we collected our data from r/AskReddit, a subreddit where each post is prompted with a question which encourages engagement. Overall, in comparison to Twitter we find that high levels of conversational engagement uniquely characterize r/AskReddit. This is particularly notable considering that the subreddit hosts 35 million subscribers and was found to be the most popular subreddit to post about climate change on the platform.

As we sought to further characterize these platforms, we looked towards the results of the TF-IDF analysis. In the

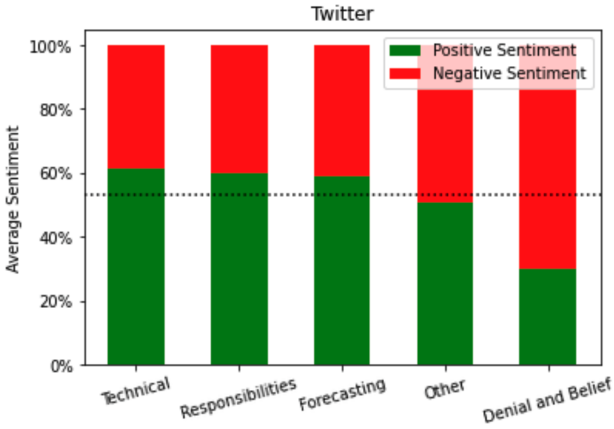

Figure 3: Average sentiment of Tweets by topic. Dotted line represents average sentiment over all topics.

case of Twitter, these results point to the interpretation of climate change related posts on the platform as largely concerned with civic engagement. We see that the responsibilities section is characterized by words such as "action", "infrastructure", and "subsidize" which indicate that these posts may engage with the responsibility of climate change as being more related to a broader social, economic, and political system rather than as an individual responsibility. These posts score above average in terms of sentiment in comparison to tweets in other topics, indicating that many of these posts may be geared towards a solutions-oriented approach to climate change which addresses what can and is being done rather than the ongoing conflict that these situations create.

In line with our observation that posts in the Twitter dataset engage with climate change at an infrastructural level, we also find that the posts relating to denial and belief tend to reflect the contentious socio-political views on this subject. Terms such as "hoax" and "fake" which are often used in polarizing political debate around climate change (Van Der Linden, Panagopoulos, and Roozenbeek 2020) characterize the largely negative discourse on this topic. This may be a result of the high prevalence of fake news circulation on Twitter (Vosoughi, Roy, and Aral 2018).

In contrast to the Twitter dataset, the Reddit datasets contained content that engaged with climate change more broadly as it relates to both the individual and humanity as a whole. The results of the TF-IDF analysis of the Reddit Comments dataset characterize Responsibilities as involving both "individuals" and "governments". Note here that "governments" relates broadly to all governments, not to a specific government. These comments also mention words such as "actions" and "help" which indicate conversations about agency. Notably, a recurring question on this topic asked, "What can an individual can do to combat climate change?" This topic comprised over a quarter of climate change related questions and had the highest overall engagement in terms of median number of comments. This result may be a reflection of the helplessness that many people feel in the face of rapidly accelerating climate change and their desire to do more.

Given that the Reddit Comments dataset was negative across all topics, in many ways our results brought to light the major concerns people generally have about climate change. Technical questions on Reddit tackled inquiries about natural resources and extreme weather events, using words like "water", "fires", "nature", and "energy". These questions were met with answers that note the harsh consequences of climate change, citing consequences like "erosion", "outbreaks", and "downpours" as well as the need for human measures such as bans on fossil fuels and meat products as noted by the keyword "banned". This was the most negative topic overall for the Reddit Comments dataset which contrasted the Twitter dataset where it was mostly positive. By our interpretation, this may reflect the contrast of how news on climate change avoids stoking fear while the general public feels gravely about the technical implications of climate change.

7 Conclusion

This study investigated the behaviors of climate change related discourse on both Twitter and Reddit. We collected data using selected keywords and manually annotated them into five topics. We then performed TF-IDF, engagement, and sentiment analyses. We observed that the behaviours between the Twitter and Reddit datasets were distinct. From TF-IDF, we noticed that users on Twitter were more prone to civic engagement, while discourse on Reddit tended to be more conversational, bringing a varied scope from personal perspectives to discourse about humanity as a whole. Our engagement analysis showed that Reddit noticeably had more conversational engagement while Twitter was more of an information pool. In terms of sentiment, we found that on average, comments from the subreddit r/AskReddit had a more consistently negative tone compared to Twitter where tone was overall more positive. This aligns with our finding that uniquely high valence technical posts on Twitter tend to be centered around news which seems to be framed in a largely positive way.

Due to the scope of this study, our findings are limited in a number of ways. Firstly, our analysis of Reddit was restricted to posts and comments collected from the subreddit r/AskReddit. This means that our findings are limited in terms of their ability to explain discourse more broadly across the platform. Additionally, we acknowledge that our method of identifying climate change discourse via keyword search limits the representativeness of our dataset by excluding users who engage with this topic without explicit mention of our listed keywords. Lastly, by using a conservative sample size for our datasets, our results are limited in their ability to fully represent trends of engagement, sentiment, and language usage across these platforms in 2021.

This study uniquely contributes to research on the social perception of climate change as it analyzes two different social media contexts in which climate change discourse is happening. The comparative nature of this exploratory study opens up questions about what influences climate change discourse on different social media platforms. It prompts

questions like: How does the relative pervasiveness of news across different platforms influence ideas and discussions? Does the politicization of climate change online shift perceptions of responsibility away from the individual and towards larger social and political systems? We hope that our research offers some insight into these questions and opens the door for deeper analyses into them.

References

Anger, I.; and Kittl, C. 2011. Measuring influence on Twitter. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, 1–4.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, 830–839.

Benegal, S. D. 2018. The impact of unemployment and economic risk perceptions on attitudes towards anthropogenic climate change. *Journal of Environmental Studies and Sciences*, 8.

Boulianne, S.; Lalancette, M.; and Ilkiw, D. 2020. "School Strike 4 Climate": Social Media and the International Youth Protest on Climate Change. *Media and Communication*, 8.

Boulianne, S.; and Theocharis, Y. 2018. Young People, Digital Media, and Engagement: A Meta-Analysis of Research. *Social Science Computer Review*.

Carleton, T. A.; and Hsiang, S. M. 2016. Social and economic impacts of climate. *Science*, 353(6304): aad9837.

Gilani, Z.; Almeida, M.; Farahbakhsh, R.; Wang, L.; and Crowcroft, J. 2016. Stweeler: A Framework for Twitter Bot Analysis. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 37–38.

González-Padilla, D. A.; and Tortolero-Blanco, L. 2020. Social media influence in the COVID-19 Pandemic. *International Brazilian Journal of Urology*, 46.

Gunster, S. 2017. Engaging climate communication. *Journalism and climate crisis: Public engagement, media alternatives*, 49.

Halpern, D.; Valenzuela, S.; and Katz, J. E. 2017. We Face, I Tweet: How Different Social Media Influence Political Participation through Collective and Internal Efficacy. *Journal of Computer-Mediated Communication*, 22.

Hartmann, J.; Heitmann, M.; Siebert, C.; and Schamp, C. 2022. More than a Feeling: Accuracy and Application of Sentiment Analysis.

Kirilenko, A. P.; and Stepchenkova, S. O. 2014. Public microblogging on climate change: One year of Twitter worldwide. *Global Environmental Change*, 26.

Landauer, T. K.; and Dumais, S. T. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2): 211.

Leiserowitz, A.; Maibach, E.; Rosenthal, S.; Kotcher, J.; Ballew, M.; Goldberg, M.; and Gustafson, A. 2019. Climate Change in the American Mind: December 2018.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Marlon, J. 2014. Extreme Weather and Climate Change in the American Mind in November 2013.

Medvedev, A. N.; Lambiotte, R.; and Delvenne, J.-C. 2019. The Anatomy of Reddit: An Overview of Academic Research. *Dynamics On and Of Complex Networks*, 3.

Molodtsova, T.; Kirilenko, A.; and Stepchenkova, S. 2013. Utilizing the social media data to validate 'climate change' indices. In *American Geophysical Union, Fall Meeting 2013*.

Olteanu, A.; Castillo, C.; Diakopoulos, N.; and Aberer, K. 2015. Comparing Events Coverage in Online News and Social Media:The Case of Climate Change. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, 288–297.

Roesslein, J. 2020. Twitter for python. *URL: https://github.com/tweepy/tweepy*.

Shearer, E.; and Gottfried, J. 2017. News Use Across Social Media Platforms 2017. *Pew Research Center's Journalism Project*.

Sisco, M. R.; Pianta, S.; Weber, E. U.; and Bosetti, V. 2021. Global climate marches sharply raise attention to climate change: Analysis of climate search behavior in 46 countries. *Journal of Environmental Psychology*, 75.

Van Der Linden, S.; Panagopoulos, C.; and Roozenbeek, J. 2020. You are fake news: political bias in perceptions of fake news. *Media, Culture & Society*, 42(3): 460–470.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science*, 359(6380): 1146–1151.

Wikipedia. 2022. Greta Thunberg. *Wikipedia, the free encyclopedia*.