

Climate Change: A Large-Scale Analysis of 11.5 Billion Reddit Posts

Corresponding Author: Professor Dariusz Jemielniak

Attachments originally supplied by the reviewers can be found at the end of this file, in order by reviewer number and revision.

Version 0:

Decision Letter:

**** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors ****

Dear Professor Jemielniak,

Your manuscript titled "Climate Change: A Large-Scale Analysis of 11.5 Billion Reddit Posts" has now been seen by 2 reviewers, and we include their comments at the end of this message. They find your work of interest, but some important points are raised. We are interested in the possibility of publishing your study in Communications Earth & Environment, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response that takes into account the points raised. Please highlight all changes in the manuscript text file.

In particular, please ensure that the revised manuscript meets the following editorial thresholds:

**** Expand your discussion section, analyzing the implications of the observed trends and correlating your findings and results with relevant previously published literature on climate change/global warming discussions in social media. Modify your conclusions accordingly.**

**** Provide a clear explanation of your methodology, including details about the analyzing tools used, and the specificities of your data set.**

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees' comments (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and the completed checklist:

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to receive your revised paper within six weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has

been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

Please do not hesitate to contact us if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Carolina Ortiz Guerrero
Associate Editor
Communications Earth & Environment

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

Editorial Policy: [Policy requirements](https://www.nature.com/documents/nr-editorial-policy-checklist.pdf) (Download the link to your computer as a PDF.)

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist: [Communications Earth & Environment formatting checklist](https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf)

and also in our style and formatting guide [Communications Earth & Environment formatting guide](https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

*** DATA: Communications Earth & Environment endorses the principles of the Enabling FAIR data project (<http://www.copdess.org/enabling-fair-data-project/>). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>.

In particular, the Data availability statement should include:

- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at <http://www.nature.com/sdata/policies/repositories>.

If a community resource is unavailable, data can be submitted to generalist repositories such as [figshare](https://figshare.com/) or [Dryad Digital Repository](http://datadryad.org/). Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at <http://www.nature.com/authors/policies/availability.html>.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

This article conducts an analysis of discussions surrounding climate change and global warming from 2005 to 2021 through the examination of Reddit posts. The findings suggest that, overall, conversations pertaining to climate change tend to manifest a more negative sentiment compared to those about global warming. Furthermore, it is observed that around the year 2012, the term "climate change" began to be employed more frequently than "global warming." This study provides a

detailed perspective on how public discourse on these critical environmental issues has evolved, reflecting shifts in public concerns and trends in the perception of the climate crisis.

1. Further detail on the Reddit platform should be discussed. It is not as globally recognized as platforms mentioned in the text (such as Twitter/ X). Who are its users? Could it be considered representative? Additionally, the differences with Twitter, which are only briefly mentioned, warrant further exploration.
2. The origin of the analyzed conversations should be clarified. Are they predominantly from the USA? This detail is not mentioned anywhere in the text.
3. Are the comments analyzed exclusively in English, or do they encompass multiple languages?
4. The discourse is noted to shift in 2002 (Bush administration as it is indicated in the text), yet the analyzed data originates from 2005. Implications should be addressed.
5. The terms 'Reddit submissions' and 'Reddit comments' are utilized without explanation. For readers unfamiliar with Reddit, these concepts may be challenging to grasp.
6. The tools 'VADER' and 'TextBlob' are mentioned without description. Readers unfamiliar with these analysis tools are left without understanding of their application or significance. I know that VADER, for example, ranges from -1 to 1. But this kind of aspects do not appear in the text to give you more feedback about the content of the paper (just an example).
7. It is uncertain whether the detailed description provided in the "selecting climate related posts" section is necessary with such a level of description.
8. In the results, Figures 1 and 2 are presented without immediate commentary, leaving them devoid of context.
9. The concept of 'TextBlob subjectivity' is mentioned, yet a more detailed explanation of the analysis conducted is warranted.
10. For Figure 2, an inquiry into the factors behind the spikes in % submissions around the years 2010, 2017, and 2019 should be made.
11. The term 'subreddits' is used in a context that may imply understanding, yet no definition or explanation is provided within the text.
12. Figures 3 and 4 are presented with minimal detailed commentary on the observed phenomena, primarily being described rather than analyzed in depth.
13. The observations made in the discussions section appear to extend beyond merely analyzing the terminological shift from "global warming" to "climate change." However, these findings are not emphasized within the article itself.
14. I miss a statistical/econometric analysis beyond the a simple description

My objective has been to highlight areas for improvement in providing clarity and context for readers, particularly those unfamiliar with the platforms and analytical tools mentioned.

Reviewer #2 (Remarks to the Author):

Dear authors and editors,

thank you very much for giving me the chance to review your paper in which you checked in how far discussions on climate change on the internet platform Reddit changed over time. I am certain that your paper is of high relevance to the scientific community and offers novel insights on an important topic. I have not worked with Reddit posts and large language data sets myself, hence I feel unable to comment on your methodological proceedings.

The comments I have on your paper revolve around your discussion section and conclusion predominantly. I have made specific comments within the manuscript which I will upload to this platform for greater clarity. Hence, my comments here are more general.

I feel that your discussion could use a more sustained engagement with previous research on the topic of social media discussions of climate change/ global warming. To me it was somewhat surprising to see a discussion section without references to other people's work. I believe there is much potential to enhance your valuable contribution if you could help the reader to better understand in how far your findings align/ differ from other research. I understand that the majority of work has been conducted on debates on X (former Twitter) or Facebook and not Reddit but still it would be useful to alert the reader to potential differences/ similarities.

You might want to consider some of these papers as a starting point for further research and reading on how to embed your

findings:

This paper could be of interest when considering specific platform cultures:

W. Pearce, et al.

The social media life of climate change: platforms, publics, and future imaginaries

WIREs Climate Change, 10 (2) (2018), pp. 1-13

Fownes et al., 2018

J. Fownes, C. Yu, B. Drew

Twitter and climate change

Sociology Compass, 12 (6) (2018), Article e12587

S. Hautea, P. Parks, B. Takahashi, J. Zeng

Showing they care (or don't): Affective publics and ambivalent climate activism on TikTok

Social Media + Society (2021)

M. Oz, P. Zheng, G.M. Chen

Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes

New Media & Society, 20 (9) (2018), pp. 3400-3419

This could potentially help shed light on how particular wordings change:

Esteves Gonçalves da Costa, B., & Cukierman, H. L. (2019). How anthropogenic climate change prevailed: A case study of controversies around global warming on Portuguese Wikipedia. *New Media & Society*, 21(10), 2261-2282.

<https://doi.org/10.1177/1461444819838227>

Maybe consider looking at more theoretical accounts of how/ when debates online decline:

P. Porten-Che  , C. Eilders

Spiral of silence online: How online communication affects opinion climate perception and opinion expression regarding the climate change debate

Studies in Communication Sciences, 15 (1) (2015), pp. 143-150

Another short remark pertains to the portrayal of "the public" being polarized on climate change. This might not be as extreme as the authors are stating in countries other than the US...

On public perceptions and political divides in the context of cc outside of the US see for example:

A.M. McCright, et al.

Political ideology and views about climate change in the European Union

Environmental Politics, 25 (2) (2015), pp. 338-358

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Communications Earth & Environment is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the Platform Support Helpdesk.

Author Rebuttal letter: The author's response to these comments can be found at the end of this file.

Version 1:

Decision Letter:

Dear Professor Jemielniak

Thank you for submitting your revised manuscript titled "Climate Change: A Large-Scale Analysis of 11.5 Billion Reddit Posts".

After careful consideration and discussion with my colleagues, we are sorry that we do not feel that the reviewers' comments have been sufficiently addressed to justify sending this revision back to the reviewers. Specifically, we are unable to conclude that the revised manuscript meets our first editorial threshold as outlined before ("Expand your discussion section, analyzing the implications of the observed trends and correlating your findings and results with relevant previously published literature on climate change/global warming discussions in social media. Modify your conclusions accordingly.")

When we request to strengthen the conclusions of a manuscript, the editors carefully assess whether appropriate material has indeed been incorporated into the revised manuscript before deciding whether to send it back to the referees.

We realize that the points raised during review are not trivial to address. Nevertheless, we feel that your manuscript has not been sufficiently strengthened, and that it would not be appropriate to send your manuscript back to the referees at this time.

We take this course of action to avoid unproductive rounds of review that can lower the chances of a manuscript obtaining a fair and objective evaluation.

Should future work allow you to include the required information in a revised manuscript, we would be happy to reconsider our decision. If appropriate, we can contact the reviewers for further comments at a later stage. In the meantime, however, we are regretfully returning your manuscript so that you can decide how best to proceed.

We are sorry that we cannot be more positive on this occasion.

Sincerely,

Carolina Ortiz Guerrero, Ph.D.
Associate Editor
Communications Earth & Environment

Version 2:

Decision Letter:

**** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors ****

Dear Professor Jemielniak,

Your manuscript appeal titled "Climate Change: A Large-Scale Analysis of 11.5 Billion Reddit Posts" has now been seen by our original reviewer #1, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*******Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *******

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commsenv/open-access>

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Carolina Ortiz Guerrero, Ph.D.
Associate Editor
Communications Earth & Environment

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

All the comments have been addressed.

** Visit Nature Research's author and referees' website at <http://www.nature.com/authors> for information about policies, services and author benefits**

Author Rebuttal letter: The author's response to these comments can be found at the end of this file.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Climate Change: A Large-Scale Analysis of 11.5 Billion Reddit Posts

Gabriele Fariello¹, Dariusz Jemielniak^{2,3}

¹Harvard University, USA, 1557 Massachusetts Avenue, 02138 Cambridge, MA,
fariello@g.harvard.edu

²Kozminski University, Poland, Jagiellonska 59, 03301 Warszawa,
darekj@kozminski.edu.pl

³Harvard University, USA, Berkman-Klein Center for Internet and Society, 1557
Massachusetts Avenue, 02138 Cambridge, MA, djemielniak@cyber.harvard.edu

Funding: Polish National Science Centre, Grant/Award Number: 2019/35/B/HS6/01056

Climate Change: A Large-Scale Analysis of 11.5 Billion Reddit Posts

Abstract

Purpose: Our purpose was to check if people discuss climate change more or less proportionally to other topics. We collected and analyzed 16 years and 11.5 billion Reddit posts, using sentiment analysis.



Findings: We show that “climate change” replaced “global warming” in 2013 in frequency. We also show that climate is becoming a less frequent topic of conversations. Moreover, we show that “global warming” is associated with more positive sentiment than “climate change.” Additionally, we discover that discussions about “global warming” use less subjective language than “climate change.”

Practical implications: The observation that more denialist-minded people also speak in more authoritative language, and do not perceive the climate changes as dangerous, is not a reason to worry, especially since their discourse is in decline. However, the decline in percentage of discussions related to climate is a dire warning.

Original value of the paper: This is the largest study of Reddit discussions in history in relation to popular environmental topics. It shows surprising language and interest changes, which is highly important for policymakers. Finally, as we release the database along with the paper, we offer a highly valuable database for all social scientists interested in internet research of climate topics.

Introduction

Global environmental change has been discussed in the media for many decades, with varied issue attention (Hase et al., 2021) and different framing from country to country (Vu, Liu and Tran, 2019). Climate journalism has emerged and evolved as a field (Schäfer and Painter, 2021).

“Global warming” and “climate change” are terms that are sometimes used interchangeably (Shehata and Hopmann, 2012; Broadbent *et al.*, 2016). Yet, they do have different semantic contexts and connotations: “Global warming” refers to a long-term increase in global temperatures, while “climate change” denotes changes in climate patterns, in particular involving the increased levels of carbon dioxide in the atmosphere (Lineman *et al.*, 2015). Both terms are highly politicized, though, and for a good reason.

“Climate change” has been used in academic papers since the fifties (Plass, 1956), but it has been more widely popularized in the public discourse by Frank Luntz, a Republican communication strategist, who in 2002 advised the Bush administration to use it instead of “global warming,” as a less frightening term (Burkeman, 2003), and to make the perceived lack of scientific certainty a vital issue in the debate about the environment.

It made a political sense even though for liberals, “global warming” and “climate change” are associated equally with rising temperatures and, overall, the heat impact on the planet, conservatives see “climate change” as more disconnected from it (Schuldt and Roh, 2014). It is also worth noting that the US media, in particular, are prone to relativization of the topic and frequently evoke scientific uncertainty and controversy regarding human-induced climate change (Shehata and Hopmann, 2012). Whether the Republican plan was successful or not, over the years, a noticeable shift in vocabulary from “global warming” to “climate change” in the popular press and academic papers alike occurred (Zehr, 2000; Sonnett, 2022).

The social perception of these terms on the Internet has been studied, e.g. by Google Trends (Jun, Yoo and Choi, 2018; Sisco and Weber, 2022). There have been also many studies of the perception of climate change and environmental challenges on the social media, with a heavy bias towards Twitter (Lineman *et al.*, 2015; Pearce *et al.*, 2019; Ganczewski and Jemielniak, 2022), possibly because of the ease of data access.

However, it has not been a subject of a systematic review of large-scale online discussions. It is essential, though, as online discussion fora provide a platform for contrasting arguments and reasoning, in contrast to Twitter, which relies on short, linguistically-restricted statements, and in contrast to just Google Trends, which only reflects people’s search interests. Online discussions are also a good proxy for general public interest and perceptions, and may help to understand the emergence of public spheres in climate-related topics (Neff and Jemielniak, 2022).

Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends (Fariello, Jemielniak and Sulkowski, 2021; Gaudette *et al.*, 2021). In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.

Methods

Data collection

We obtained a large dataset ($n=11.486 \times 10^9$) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files). We uncompressed each monthly file, and removed duplicates. We then used a series of regular expressions to identify all posts related to climate change while ignoring all instances of political, economic, financial, academic, work, social, job, or employment climate, as well as terms such as “climate control” used to refer to topics not related to climate change while still retaining all posts that contained explicit references to climate change, global warming, and related climate specific topics such as references to the Paris Accord, the Kyoto Protocol, Greta Thunberg, and others. The complete list and order of expressions used is included in the supplemental materials.

This resulted in 15.341×10^6 posts likely to be related to climate change (0.134%). These posts were then converted from JSON format to a more data science friendly CSV format retaining all the relevant information on which we performed language detection using both the polyglot (Al-Rfou, 2015) and langdetect Python libraries, keeping only posts that were determined to be in English by both ($n=1.501 \times 10^6$, 97.9% of matched posts). We subsequently concatenated and lemmatized the title and body of all the remaining posts using standard NLTK library methods (nltk.stem.WordNetLemmatizer) and removed all English stop words (Bird, Klein and Loper, 2009). Each post was then analyzed using the VaderSentiment (Hutto and Gilbert, 2014), TextBlob sentiment, TextBlob naïve Bayes, and textstat python libraries among several other interesting metrics such as readability and language complexity scores provided in the supplemental data.

We then analyzed the unigram, bigrams, and trigrams and subsequently analyzing the occurrences of the “climate” unigram when not followed by “change” and the “climate change” bigram as well as the “global” unigram when not followed by “warming” and the “global warming” bigram. We accomplished this by replacing all permutations of occurrences of “climate change” with “climate-change” and “global warming” with “global-warming” after lemmatization and performing a unigram analysis. For each of the resultant unigrams, we analyzed sentiment, polarity, and readability subdivided by the top 40 subreddit, and looked at the relative trends for the unigrams across each year.

Selecting Climate Related Posts

On the original JSONL formatted data downloaded from PushShift.io, we extract all posts whose title text and body text, when concatenated together, matched any one of the following regular expressions:

- `\bclimat(?:e|olog|arian|ard|is[mt]|caust|sci)\w*\bclimist\w*`
- `\b(?:global\w*|environ\w*|wea?ther\w*|oceans?|seas?|atmos\w*|historic\w*)[\W_]*(?:warm\w*|cool\w*|temper\w*|heat\w*|hot[estr]*\b|cold[estr]*\b)`

- 121 • \b(?:oceans?|seas?)[\W_]*(?:level\w*|temper\w*|acid\w*|ris[ei]\w*|heat[ingswave]*|ph\b)
- 122 • \b(?:level\w*|temper\w*|acid\w*|ris[ei]\w*|heat[ingswave]*|ph\b)\w*[\W_]*(?:oceans?|seas?)
- 123 • \b(?:green[estr]?|altern\w*|renewab\w*|wind|solar|hydro(?:!xide|ponic|phobi|fobi)\w*|geo[\W_]*ther\w*|nuclear\w*|nucular\w*|clean[estr]*)[\W_]*energ\w*
- 124 • \b(?:green[estr]?|wind|solar|hydro\w*|geother\w*|clean[estr]*)[\W_]*(?:tech\w*|solut\w*)
- 125 • \bclean[estr][\W_]*(?:tech\w*|energ\w*)
- 126 • \bgreen[\W_]*house\w*[\W_]*(?:gas\w*|effect\w*)
- 127 • \bcarbon[\W_]*(?:neutr|emi[st]|diox|foot)\w*
- 128 • \bco2\b
- 129 • \b(?:emit\w*|produc\w*|contrib\w*)[\W_]*carbon
- 130 • \b(?:polar|arc?tic(?:al|le)\w*|ant[\W_]*arc?tic\w*)[\W_]*(?:ice\w*|shelf\w*|temper\w*|warm\w*|cool\w*)
- 131 • \bthu[mn][mn]?b[eu]rg\b
- 132 • \bipcc\b
- 133 • \b(?:montreal|kyoto)[\W_]*proto\w*\b
- 134 • \bice[\W_]*(?:caps?\b|shelf\w*|sheets?)\b
- 135 • \bfossil[\W_]*fuels?\b
- 136 • \b[dr]e[\-]*forest\w*
- 137 • \brenewables\b
- 138 • \bwind[\W_]*turbin[es]*\b
- 139 • \bparis[\W_]*(?:accord|agreem)\w*

145 And all of the following regular expressions:

- 146 • (?<!politic\w*)\bclimat
- 147 • (?<!econom\w*)\bclimat
- 148 • (?<!financ\w*)\bclimat
- 149 • (?<!academ\w*)\bclimat
- 150 • (?<!work\w*)\bclimat
- 151 • (?<!social\w*)\bclimat
- 152 • (?<!job\w*)\bclimat
- 153 • (?<!employ[me]\w*)\bclimat
- 154 • \bclimat\w*\W*(?!control)

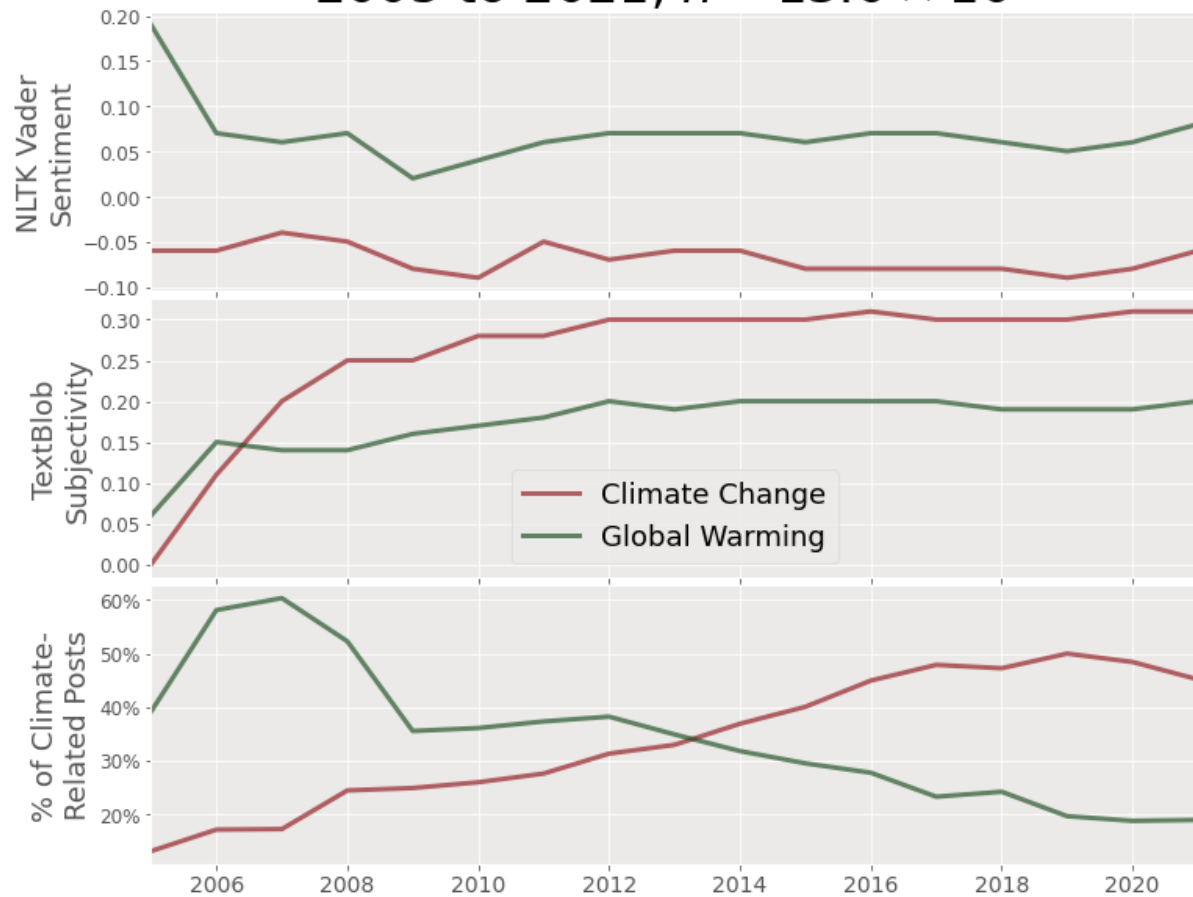
155 The former regular expressions include any posts with text tightly associated with climate
 156 change related topics, as well as bigrams such as “political climate” or “climate control” which
 157 refer to topics that are not specific to climate change while the latter regular expressions
 158 ensure that any items that did match “climate” did so because climate appeared at least once
 159 not preceded by words such as “political,” “financial,” “job,” or “academic” and not followed by
 160 “control.”

161

162

Results

Climate Related NGram Properties 2005 to 2021, $n = 13.0 \times 10^6$



163

164

Figure 1: The sentiment, subjectivity, and percentage of posts related to climate

165

Figure 1 shows the sentiment, subjectivity, and percentage of posts related to climate.

166

Percentage of All Posts Matching Climate-Change Related Terms from 2005 to 2021, $n = 11.5 \times 10^9$

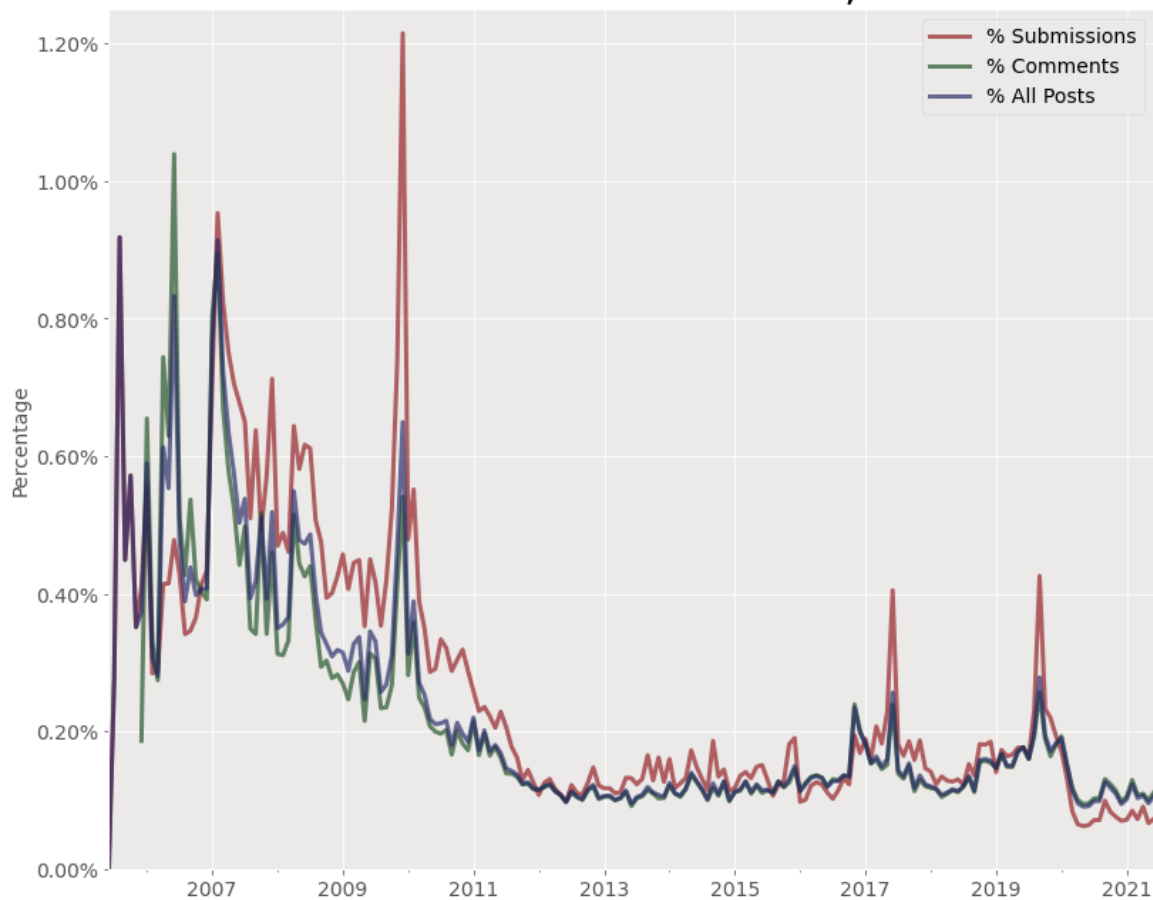


Figure 2: Percentage of all posts related to climate change

Figure 2 shows the relative popularity of climate-change-related topics in time.

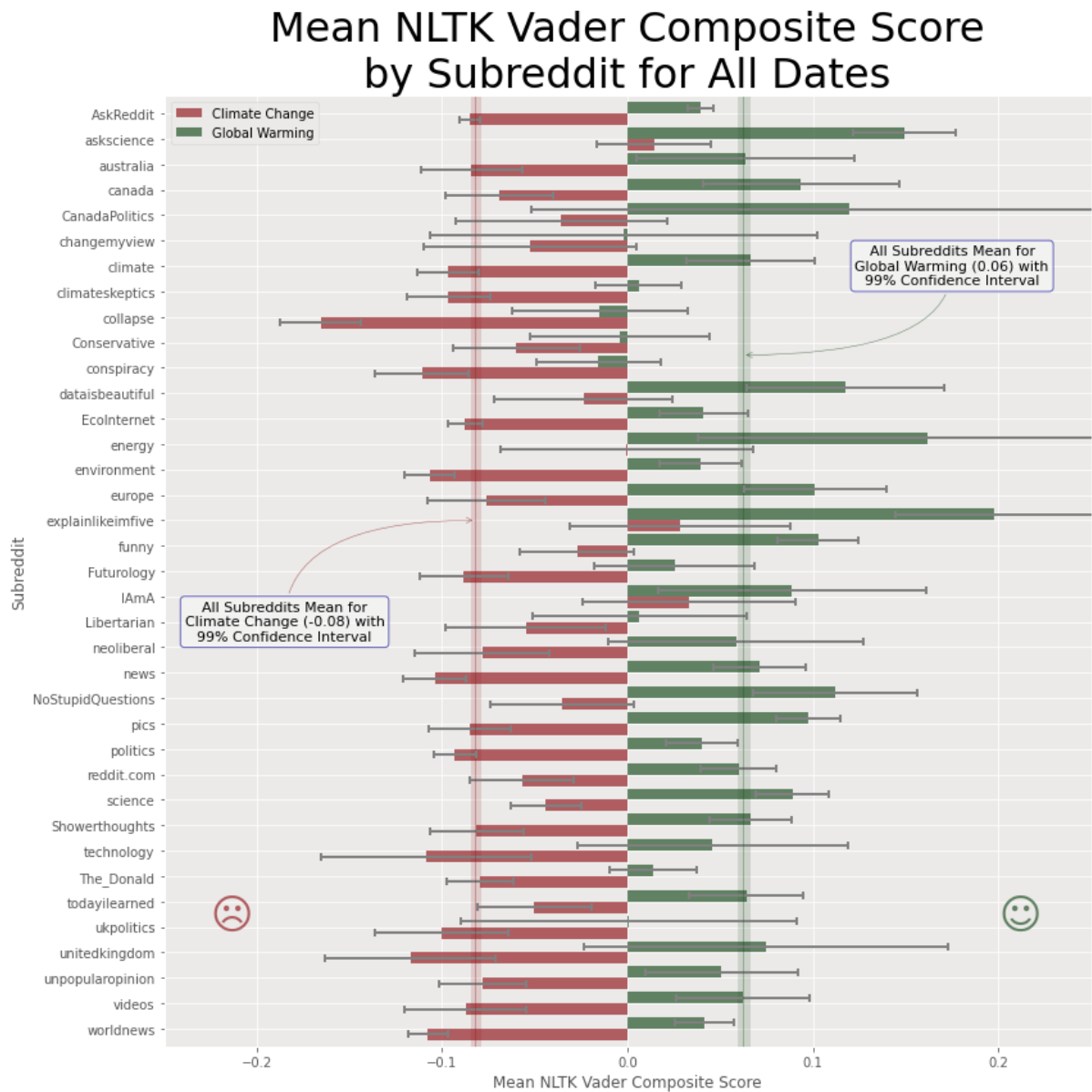


Figure 3: Mean NLTK Vader Composite Scores by Subreddit

Figure 3 depicts the mean NLTK Vader Sentiment Composite Score for climate-related posts containing the terms “climate change” or “global warming” across all subreddits, which are plotted as vertical lines with the shaded area representing the 99% confidence interval while the bars represent the mean for each of these for a given subreddit with the error bars showing the 99% confidence interval for that given subreddit and term. The subreddits (37 total) were selected by choosing the 50 highest volume subreddits with more than 100 climate-related posts.

Mean TextStat Difficult Words Score by Subreddit for All Dates

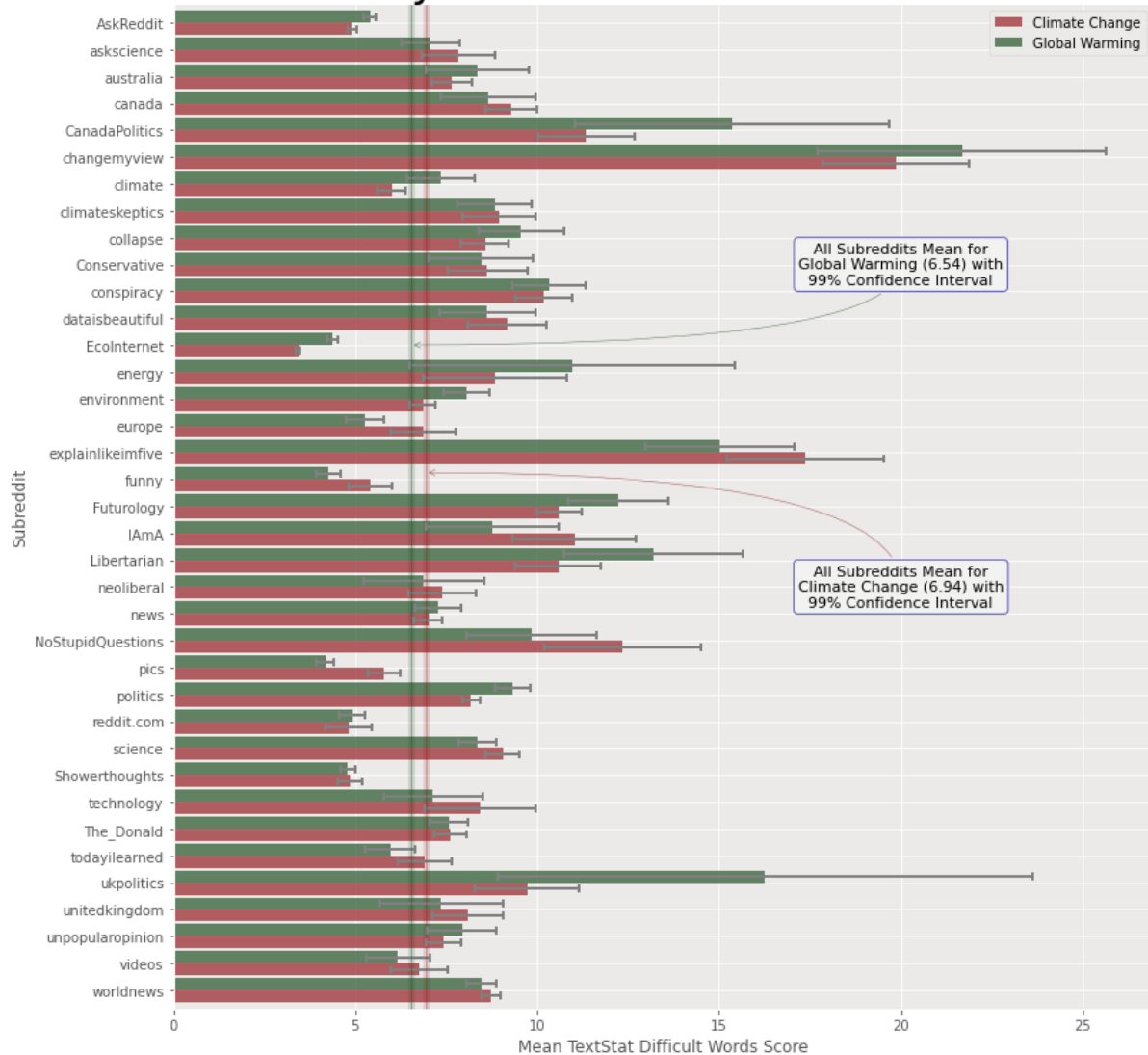


Figure 4: Mean TextStat Difficult Words Scores by Subreddit

In Figure 4, the mean TextState Difficult Words score for all posts containing the terms “climate change” and “global warming” across all subreddits are plotted as vertical lines with the shaded area representing the 99% confidence interval while the bars represent the mean for each of these for a given subreddit with the error bars showing the 99% confidence interval for that given subreddit and term. The subreddits (37 total) were selected by choosing the 50 highest volume subreddits with more than 100 climate-related posts.

In the supplemental materials, we provide the raw analytical data, which includes calculated values calculated for a variety of text analyses, allowing for considerable additional data exploration.

Google Search Trend Scores from 2004 to 2022 for Climate Change and Global Warming 6-Month Centered Rolling Average

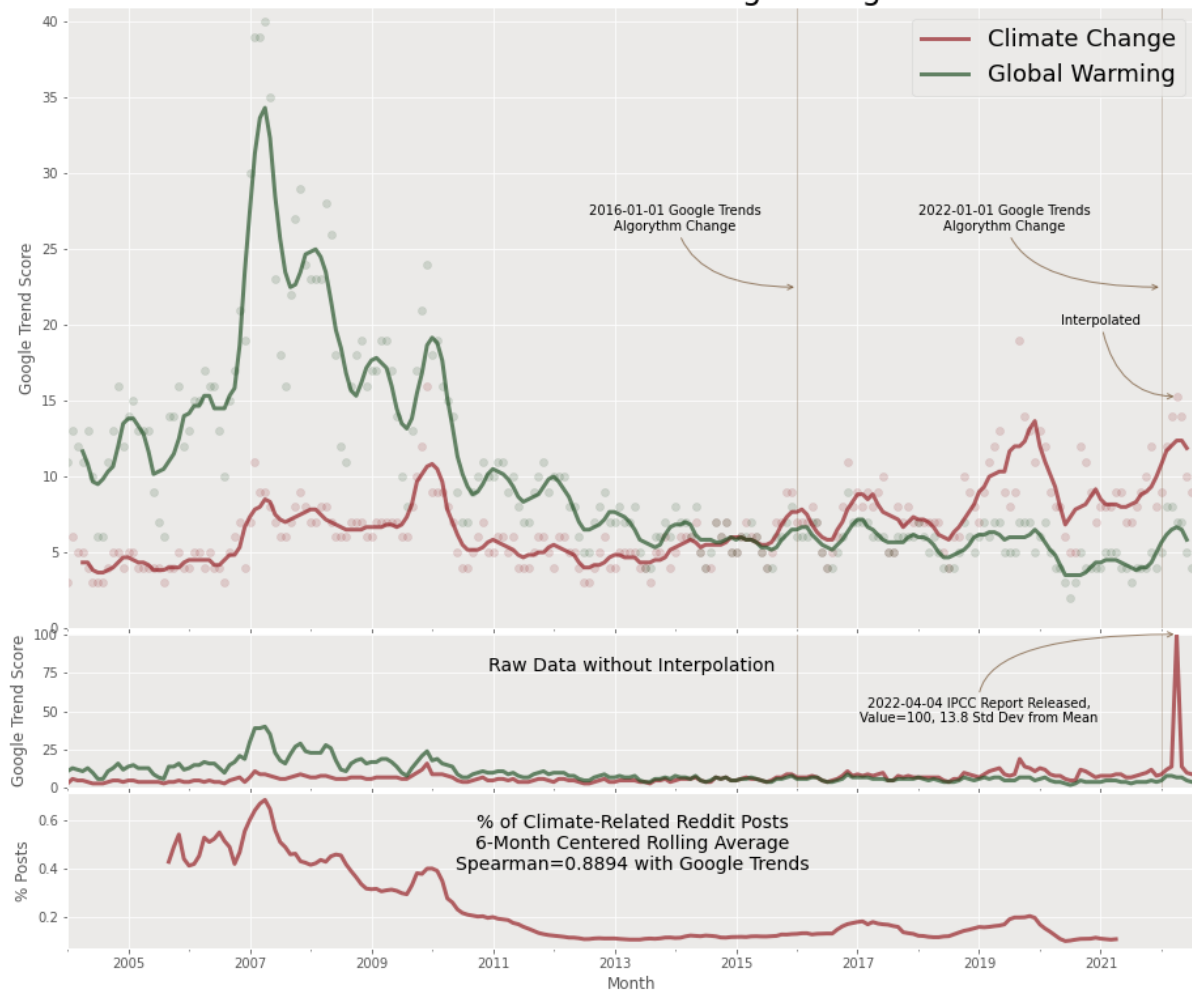


Figure 5: Google Search Trends

Figure 5 shows the Google search term trends for “climate change” and “global warming” (without quotes) for the 18 years from January 2004 to August 2022. Vertical lines represent the dates on which Google implemented algorithmic changes to how it tracks searches. In April 2022, the month during which the IPCC released its most recent and most alarming report, there was a large spike in searches reported; however, we are unable to ascertain if the two are related.

Discussion

Our first conclusion is that the shift from “global warming” to “climate change” indeed happened, but a decade later than the Republican initiative. Curiously, it was relatively abrupt, and around 2013, “climate change” overcame “global warming” in general discussions we analyzed (see Figure 1). It is also notable that Google search trends (Figure 5) show a similar switch in terminology occurring only two years later.

206 Our analysis found that, contrary to previous assumptions, “global warming” is associated with
207 more positive sentiment polarity than “climate change.” Across all subreddits, the sentiment of
208 posts using the term “global warming” was positive, while the sentiment of posts using the
209 term “climate change” was negative. 📊 This result could indicate that those writing about “global
210 warming” are less concerned about the climactic changes than those naming it “climate
211 change,” which could be correlated with the slow uptake of new terminology. Interestingly, the
212 subreddit “askscience,” which focuses on a scientific approach to the world, had the most
213 positive sentiment overall for posts using either “climate change” or “global warming.”

214 The subjectivity of posts about “climate change” was much higher, meaning that people writing
215 about it were expressing their views in a less authoritative tone, while people writing about
216 “global warming” were phrasing their posts as if they were describing objective facts. This
217 result could indicate that those writing about “climate change” are more conscious about
218 separating opinions from facts and is, generally, not surprising. Still, this finding indicates that
219 those who write about “global warming” may see their views as non-negotiable truths.

220 It is worth mentioning that during Donald Trump’s presidency, there was a small drop in both
221 sentiment polarity and subjectivity - people were writing about “climate change” and “global
222 warming” more negatively, and in language more typical for a description 🗣️ rather than
223 viewpoints. We do not have a good explanation for this phenomenon.

224 Moreover, it is also quite surprising that the discussions about “climate” in general have not
225 increased significantly since 2009, and they remain at a stable percentage of overall posts.
226 Only the last four years note an increase from about 20% to nearly 30%. They appear to be
227 nearly perfectly correlated with “global warming” conversations - the more people talk about
228 one, the less they do about the other (see Figure 2).

229 When the percentage of Reddit posts with climate-related topics is considered, we can even
230 observe an evident decline. While 2005-2010 even up to 1.4% of topics were related to
231 climate, after that period the numbers steadily decreased. It may be misleading, though: since
232 we are measuring the proportions of topics, the overall growth of popularity of Reddit is likely
233 largely contributing to the observed phenomenon. In other words, the absolute numbers of
234 climate-related discussions have not decreased (in fact, they increased steadily), but the
235 growth of discussions on other topics was more rapid. Nevertheless, we believe that this result
236 is a mild warning of the fact that climate is becoming a less and less popular topic
237 proportionally. 🗨️

238 On a side note, we observe that, considering the word difficulty, “ecointernet” subredding had
239 the least difficult words in sentences related to “climate change”, “global warming,” and
240 “climate” in general. “Changemyview” subreddit, on the other hand, had the most difficult
241 sentences - it may be that this subreddit, dedicated to confrontational discussions, relied more
242 largely on sources and academic language. One surprising finding from our study is that
243 “Explain to me like I’m 5” thread uses more complex words than other threads. In the
244 supplement, we provide additional figures, illustrating this phenomenon.

245 As seen in Figure 3, the terms “climate change” appear to be associated very consistently with
246 Vader sentiment scores that are less favorable than those of “global warming” irrespective of
247 the subreddit, even when the subreddits vary greatly on the validity of climate-related claims
248 (e.g., the Conservative, The_Donald, and conspiracy versus climate, askscience, and

science), suggesting that the relatively more positive sentiment association with “global warming” may persist across political views and scientific literacy. Unsurprisingly, several subreddits with a stronger focus on educating people tended to use more positive tones (e.g., askscience, explainlikeimfive, IAmA).

We note that posts with the terms “climate change” have a small but very statistically significant increase in the use of difficult words when scores are aggregated across all subreddits, but the trend does not hold well within and across subreddits (see Figure 4). The subreddit dedicated to attempting to change the original poster’s (OP) views had the highest frequency of difficult words with respect to climate-related posts. It is interesting to note the considerable irony that the subreddit ostensibly meant to explain challenging topics using simple terminology, explainlikeimfive, had, on average, the second-highest frequency of use of difficult words.

The most interesting findings are that the Google search trends (Figure 5) demonstrate a similar change in frequency of the terms from “climate change” in the searches before 2015 to “global warming” to those after. Although this occurred at a later date than that which we observe in the Reddit discussions circa 2013, it appears to be a consistent feature across two disparate Internet activities. Furthermore, the 6-month rolling averages for the sum of the two Google trend lines, which roughly approximates the overall climate-related searches over time, is strongly correlated with the percentage of Reddit climate-related discussions with a Spearman correlation coefficient of 0.89 further suggesting that the trends observed in Reddit discussions are reflective of broader trends across at least some Internet-based media.

Conclusions

Global environmental change is highly politicized and ideologized in the public discourse (Stecula and Merkley, 2019; Chinn, Hart and Soroka, 2020). The mainstream media have been covering it systematically for years. Even though disproportionate attention has been given to contrarians rather than experts (Petersen, Vincent and Westerling, 2019), and even though extreme weather phenomena has been affecting the media attention more than scientific accounts (Pianta and Sisco, 2020), the overall media attention has been growing (Barkemeyer *et al.*, 2017; Layna and Altamirano, 2022). Moreover, even the more conservative outlets have started acknowledging the burning problem (Schmid-Petri *et al.*, 2017).

In our article we have shown that the proportionate interest of the general public, as reflected in large-scale online discussions, has not been following the same path. Possibly because of exhausting the topic and being tired of continuous high alert state, other competing crises (Stoddart *et al.*, 2021), as well as simply other interesting topics, the discourse about climate change has taken less and less space proportionately.

Moreover, we show that discussions about “global warming” appear to use less sophisticated language and are more positive than those about “climate change.” This result suggests that perceptions of global environmental change manifest in language, which is a phenomenon that call for further exploration.

We also show that online discussions seem to precede Google searches in terms of language shift from “global warming” to “climate change”, which suggests that monitoring social media discussions may be a more immediate way of detecting changes in public attention and imaginary.

All of these results indicate that the window for mobilizing the general public's attention around global environmental change may be closing, and that the main debate around the topic among non-experts may be affected by the informal discussions on social media more than elsewhere.



Statements

Declaration of interest statement: 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 access statement: Our whole dataset is released simultaneously to the publication.

Ethics statement: The paper bases on aggregated, public, non-identifiable, non-personal data.

References

- Al-Rfou, R. (2015) *Polyglot: A Massive Multilingual Natural Language Processing Pipeline*. search.proquest.com. Available at: <https://search.proquest.com/openview/a6748f01943c73c42dd77d55625097aa/1?pq-origsite=gscholar&cbl=18750>.
- Barkemeyer, R. *et al.* (2017) 'Media coverage of climate change: An international comparison', *Environment and Planning C: Politics and Space*, 35(6), pp. 1029–1054.
- Bird, S., Klein, E. and Loper, E. (2009) *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. 'O'Reilly Media, Inc.'
- Broadbent, J. *et al.* (2016) 'Conflicting Climate Change Frames in a Global Field of Media Discourse', *Socius*, 2, p. 2378023116670660.
- Burkeman, O. (2003) 'Memo exposes Bush's new green strategy', *The guardian*. Available at: <https://amp.theguardian.com/environment/2003/mar/04/usnews.climatechange>.
- Chinn, S., Hart, P.S. and Soroka, S. (2020) 'Politicization and Polarization in Climate Change News Content, 1985-2017', *Science communication*, 42(1), pp. 112–129.
- Fariello, G., Jemielniak, D. and Sulkowski, A. (2021) 'Does Godwin's law (rule of Nazi analogies) apply in observable reality? An empirical study of selected words in 199 million Reddit posts', *New media & society*, p. 146144482110620.

325 Ganczewski, G. and Jemielniak, D. (2022) 'Twitter is garbage: A Thick Big Data exploration
 326 of #zerowaste hashtag on Twitter in relation to packaging and food packaging materials',
 327 *Packaging Technology & Science* [Preprint]. Available at: <https://doi.org/10.1002/pts.2685>.

328 Gaudette, T. *et al.* (2021) 'Upvoting extremism: Collective identity formation and the extreme
 329 right on Reddit', *New Media & Society*, 23(12), pp. 3491–3508.

330 Hutto, C. and Gilbert, E. (2014) 'VADER: A Parsimonious Rule-Based Model for Sentiment
 331 Analysis of Social Media Text', *Proceedings of the International AAAI Conference on Web
 332 and Social Media*, 8(1), pp. 216–225.

333 Jun, S.-P., Yoo, H.S. and Choi, S. (2018) 'Ten years of research change using Google
 334 Trends: From the perspective of big data utilizations and applications', *Technological
 335 forecasting and social change*, 130, pp. 69–87.

336 Layna, J. and Altamirano, L. (2022) 'Promises that Don't Work? COP26 and the Problems of
 337 Climate Change', *Engaging Science, Technology, and Society*, 8(3), pp. 118–129.

338 Lineman, M. *et al.* (2015) 'Talking about Climate Change and Global Warming', *PloS one*,
 339 10(9), p. e0138996.

340 Neff, T. and Jemielniak, D. (2022) 'How do transnational public spheres emerge? Comparing
 341 news and social media networks during the Madrid climate talks', *New Media & Society*, p.
 342 14614448221081426.

343 Pearce, W. *et al.* (2019) 'The social media life of climate change: Platforms, publics, and
 344 future imaginaries', *Wiley interdisciplinary reviews. Climate change*, 10(2), p. e569.

345 Petersen, A.M., Vincent, E.M. and Westerling, A.L. (2019) 'Discrepancy in scientific authority
 346 and media visibility of climate change scientists and contrarians', *Nature communications*,
 347 10(1), p. 3502.

348 Pianta, S. and Sisco, M.R. (2020) 'A hot topic in hot times: how media coverage of climate
 349 change is affected by temperature abnormalities', *Environmental research letters: ERL [Web
 350 site]*, 15(11), p. 114038.

351 Plass, G.N. (1956) 'The carbon dioxide theory of climatic change', *Tell'Us*, 8(2), pp. 140–
 352 154.

353 Schäfer, M.S. and Painter, J. (2021) 'Climate journalism in a changing media ecosystem:
 354 Assessing the production of climate change- related news around the world', *Wiley
 355 interdisciplinary reviews. Climate change*, 12(1). Available at:
 356 <https://doi.org/10.1002/wcc.675>.

357 Schmid-Petri, H. *et al.* (2017) 'A changing climate of skepticism: The factors shaping climate
 358 change coverage in the US press', *Public understanding of science*, 26(4), pp. 498–513.

359 Schuldt, J.P. and Roh, S. (2014) 'Media Frames and Cognitive Accessibility: What Do
 360 "Global Warming" and "Climate Change" Evoke in Partisan Minds?', *Environmental
 361 Communication*, 8(4), pp. 529–548.

362 Shehata, A. and Hopmann, D.N. (2012) 'Framing climate change', *Journalism Studies*,
 363 13(2), pp. 175–192.

364 Sisco, M.R. and Weber, E.U. (2022) 'Local temperature anomalies increase climate policy
 365 interest and support: Analysis of internet searches and US congressional vote shares',
 366 *Global environmental change: human and policy dimensions*, 76, p. 102572.

- 367 Sonnett, J. (2022) 'Climate change risks and global warming dangers: a field analysis of
368 online US news media', *Environmental Sociology*, 8(1), pp. 41–51.
- 369 Stecula, D.A. and Merkley, E. (2019) 'Framing Climate Change: Economics, Ideology, and
370 Uncertainty in American News Media Content From 1988 to 2014', *Frontiers in*
371 *Communication*. Available at: <https://doi.org/10.3389/fcomm.2019.00006>.
- 372 Stoddart, M.C.J. *et al.* (2021) 'Competing Crises? Media Coverage and Framing of Climate
373 Change During the COVID-19 Pandemic', *Environmental Communication*, pp. 1–17.
- 374 Vu, H.T., Liu, Y. and Tran, D.V. (2019) 'Nationalizing a global phenomenon: A study of how
375 the press in 45 countries and territories portrays climate change', *Global environmental*
376 *change: human and policy dimensions*, 58, p. 101942.
- 377 Zehr, S.C. (2000) 'Public representations of scientific uncertainty about global climate
378 change', *Public understanding of science*, 9(2), p. 85.
- 379

380 Appendices: Supplemental Data and MetaData

381 The data and the associated data used to produce the plots and findings is made freely
382 available in a [Reddit-Climate-Data-2005-2021](#) shared folder with the following contents:

383 [Reddit-Climate-Data-2005-2021/](#)

- 384 • [Reddit General Climate Related Plots 2005-2021.ipynb](#) (2.2 MB)
385 The Jupyter Notebook compatible Google Colab Notebook used to create all the plots, and
386 others, in this publication.
- 387 • [Headers Information.xlsx](#) (20KB)
388 A Microsoft Excel file containing information on the various headers (columns) found in the data
389 files.
- 390 • [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.subreddit.csv](#) (77MB)
391 A (large) summary of all analyzed data grouped by subreddit.
- 392 • [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.subreddit_year.csv](#) (184MB)
393 A (large) summary of all analyzed data grouped by subreddit and year.
- 394 • [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.year.csv](#) (31KB)
395 A summary of all analyzed data grouped by year.
- 396 • [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.day_of_week.csv](#) (14KB)
397 A summary of all analyzed data grouped by the day of the week (0 = Monday, 6=Sunday).
- 398 • [RedditBoth-All-Climate-English-matching.csv](#) (10KB)
399 A summary of the number of submissions, comments, and both that matched climate related
400 searches.
- 401 • [Unsorted-Top-25-NGrams-Analysis-Results-Merged-All-Subreddits.csv](#) (23MB)
402 A summary of the (at most) top 25 NGrams analyzed unigrams, bi-grams, and 4-grams. Recall
403 that “climate change” and “global warming” were masked as the unigrams “climate-change” and
404 “global-warming” respectively for this analysis thereby enabling a comparison of the features of
405 “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by
406 “warming” vs. “global” not followed by “warming”.
- 407 • [Unsorted-Top-25000-NGrams-Analysis-Results-Merged-All-Subreddits.csv](#) (51MB)
408 A summary of the (at most) top 25,000 NGrams analyzed unigrams, bi-grams, and 4-grams.
409 Recall that “climate change” and “global warming” were masked as the unigrams “climate-
410 change” and “global-warming” respectively for this analysis thereby enabling a comparison of the

- features of “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by “warming” vs. “global” not followed by “warming”.
- [All-Subreddits.search-term-.unsorted-ngrams.all-time.csv.yearly.merged-top-250.xlsx](#) (6MB)
A summary in Microsoft Excel format with heat maps of the (at most) top 250 NGrams analyzed unigrams, bi-grams, and 4-grams. Recall that “climate change” and “global warming” were masked as the unigrams “climate-change” and “global-warming” respectively for this analysis thereby enabling a comparison of the features of “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by “warming” vs. “global” not followed by “warming”.
 - [All-Subreddits.search-term-.unsorted-ngrams.all-time.csv.yearly-used-columns.xlsx](#) (65KB)
A summary in Microsoft Excel format with heat maps containing only items used for some subset of analyses containing a subset of the (at most) top 250 NGrams analyzed unigrams, bi-grams, and 4-grams. Recall that “climate change” and “global warming” were masked as the unigrams “climate-change” and “global-warming” respectively for this analysis thereby enabling a comparison of the features of “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by “warming” vs. “global” not followed by “warming”.
 - [Google-Trends-Global-Warming-and-Climate-Change-2004-2022.csv](#) (4KB)
The Google Trends data obtained at <https://trends.google.com/trends/explore?date=all&q=%2Fm%2F0d063v,Global%20Warming> on September 4, 2022
 - [All-Climate-Related-Posts-Raw-Data-2005-2021.csv.zst](#) (5.54GB)
The complete raw data and text analysis data for all posts containing climate related matches.

Many thanks for the comments and suggestions for improvement. We are including a detailed reply to each of them below (reviewers' comments in magenta, our replies in blue, and changes in black), as well as a revision-tracked copy of the manuscript below.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

This article conducts an analysis of discussions surrounding climate change and global warming from 2005 to 2021 through the examination of Reddit posts. The findings suggest that, overall, conversations pertaining to climate change tend to manifest a more negative sentiment compared to those about global warming. Furthermore, it is observed that around the year 2012, the term "climate change" began to be employed more frequently than "global warming." This study provides a detailed perspective on how public discourse on these critical environmental issues has evolved, reflecting shifts in public concerns and trends in the perception of the climate crisis.

1. Further detail on the Reddit platform should be discussed. It is not as globally recognized as platforms mentioned in the text (such as Twitter/ X). Who are its users? Could it be considered representative? Additionally, the differences with Twitter, which are only briefly mentioned, warrant further exploration.

Thank you for your observation. We have changed

Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends ([Fariello, Jemielniak and Sulkowski, 2021](#); [Gaudette et al., 2021](#)). In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.

To:

Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends ([Fariello, Jemielniak and Sulkowski, 2021](#); [Gaudette et al., 2021](#)). As of 2023, Reddit, with its unique structure centered around interest-based communities known as "subreddits," hosts over 850 million monthly active users (MAUs) (Curry, 2024) and 2.1 billion unique visitors (Tiano Bianchi, 2024), engaging in in-depth discussions and sharing a wide array of content. Notably, it is the 5th most visited social media platform globally (Dixon, 2024), with approximately 48% of visits (Tiago Bianchi, Feb 29, 2024) originating from the United States. Reddit represents approximately one third as much social media traffic as Twitter/X which serves as a hub for real-time updates and public discourse (Dixon, 2024). Although Facebook remains the dominant force with just over 3 billion monthly active users, catering to a broad demographic by facilitating personal connections and diverse content consumption of which approximately 24% are based in India and 12% are based in the United States (Dean, 2024), Reddit's model of specialized, user-moderated subreddits is particularly resonant among specific interest groups, steadily growing in its impact on public opinion and media trends. In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.

2. The origin of the analyzed conversations should be clarified. Are they predominantly from the USA? This detail is not mentioned anywhere in the text.

We hope that the text above clarifies that the plurality of posts on Reddit are from the United States. We only consider posts in English; however, there is no way for us to determine with any precision the origin of posts since Reddit does not provide geolocation or other information.

3. Are the comments analyzed exclusively in English, or do they encompass multiple languages?

Yes. This is explicitly addressed in the methods section which we have revised slightly.

4. The discourse is noted to shift in 2002 (Bush administration as it is indicated in the text), yet the analyzed data originates from 2005. Implications should be addressed.

Although we do not address how this intentional shift in nomenclature influenced the discourse prior to 2005 since the dataset begins in late 2005, this would be an interesting area for further research.

5. The terms 'Reddit submissions' and 'Reddit comments' are utilized without explanation. For readers unfamiliar with Reddit, these concepts may be challenging to grasp.

Thank you for catching the lack of clarity. We have changed:

We obtained a large dataset ($n=11.486 \times 10^9$) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files).

To:

We obtained a large dataset ($n=11.486 \times 10^9$) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files). Submissions are the top-most post that initiates a discussion within a group known as a subreddit while comments are replies to either a submission or another comment.

6. The tools 'VADER' and 'TextBlob' are mentioned without description. Readers unfamiliar with these analysis tools are left without understanding of their application or significance. I know that VADER, for example, ranges from -1 to 1. But this kind of aspects do not appear in the text to give you more feedback about the content of the paper (just an example).

Thank you again for this observation. We have added the following text to section 2.1:

VADER (Valence Aware Dictionary and sEntiment Reasoner): A lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in

social media. VADER uses a combination of a sentiment lexicon and a set of heuristic rules to estimate sentiment scores for textual data (Hutto & Gilbert, 2014). VADER's rule-based approach, which takes into account both the semantic meaning of words and the context in which they are used (including punctuation, capitalization, and modifiers), makes it highly effective for the nuanced understanding of sentiment in social media posts. This approach provides a robust framework for analyzing how sentiment regarding climate change and global warming is expressed in online discussions, contributing significantly to the study's findings on public engagement and perception related to these topics.

- **Sentiment Score:** VADER analyzes the text to calculate a compound score that indicates the overall sentiment. This compound score is a normalized, weighted composite score that ranges from -1 (extremely negative) to 1 (extremely positive). This scoring is beneficial in understanding the general sentiment of each post regarding climate change or global warming.

TextBlob: Provides a simple API for common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. In this study, we utilized TextBlob's sentiment analysis features to assess the polarity and subjectivity of the posts.

- **Polarity:** This refers to the emotional orientation of the words used in the text. Polarity is a float that lies between -1.0 and 1.0, where -1.0 indicates a highly negative sentiment, 0 indicates neutrality, and 1.0 indicates a highly positive sentiment. It essentially measures the degree to which the language used in a text leans towards a positive or negative emotion. This metric is crucial for understanding the general sentiment expressed in discussions about climate change and global warming on social media platforms.
- **Subjectivity:** This measures the amount of personal opinion and factual information contained in the text. The value ranges from 0 to 1, where 0 is very objective and 1 is highly subjective. Subjective text expresses personal feelings, views, or beliefs, whereas objective text provides factual information. Analyzing subjectivity helps in distinguishing between posts that are based on personal opinions and those based on objective reports, which is vital for evaluating how discussions on climate change are framed by the public.

In our analysis, these metrics enabled a deeper understanding of the emotional and subjective layers within the climate change discourse on Reddit, providing insights into how people emotionally engage with the topic and the extent to which discussions are opinion-driven or fact-based (Loria, 2020).

Loria, S. (n.d.). *TextBlob: Simplified Text Processing* — *TextBlob 0.18.0.post0 documentation*. Retrieved April 21, 2024, from <https://textblob.readthedocs.io/en/dev/>

7. It is uncertain whether the detailed description provided in the "selecting climate related posts" section is necessary with such a level of description.

Your point is well taken. We have revised Section 2.2 to be:

To ensure our analysis focused solely on content relevant to climate change, we developed a set of criteria to filter posts from the PushShift.io data repository. This involved using regular expressions to select posts that specifically mentioned climate change-related terms and exclude irrelevant contexts.

Methodology Summary: Our approach was twofold. First, we included posts that contained explicit references to climate-related topics, such as "global warming," "carbon emissions," and specific environmental policies (e.g., the Paris Accord). For example, one of our expressions

`\\bglobal\\w*[W_]*(?:warm\\w*|cool\\w*|temper\\w*|heat\\w*|hot[estr]*\\b|cold[estr]*\\b)` captures discussions around global temperature changes.

Second, we excluded posts where the term "climate" appeared in non-environmental contexts, such as "political climate" or "climate control," using negative lookbehind assertions in our regular expressions. For instance, the expression `(?<!politic\\w*)\\bclimat` ensures the word "climate" is not preceded by "political" or related modifiers.

Detailed List and Rationale: For readers interested in the complete list of regular expressions and their specific purposes, we have included these details in the Supplementary Materials. This includes a comprehensive breakdown of all filters used to accurately curate our dataset, ensuring that only relevant discussions were analyzed.

We hope that this simplification improves understandability and, given that we provide the explicit regular expressions in the supplemental information, we hope that there is ample information for others to reproduce this work.

8. In the results, Figures 1 and 2 are presented without immediate commentary, leaving them devoid of context.

We have expanded the information provided for Figures 1 and 2 as follows:

Change:

Figure 1 shows the sentiment, subjectivity, and percentage of posts related to climate.

To:

Figure 1 shows that the frequency of usage of the term "global warming" decreased over time being preferentially replaced with the term "climate change" (bottom) even though the sentiment associated with "climate change" remained more negative while the subjectivity for both terms increased over time (top) with "climate change" being associated with more subjective text than "global warming" since 2007 (middle).

Change:

Figure 2 shows the relative popularity of climate-change-related topics in time.

To:

Figure 2 shows the relative frequency of climate-change-related topics in time as a percentage of discussions on Reddit. The discussion frequency has been trending down over the time period analyzed.

9. The concept of 'TextBlob subjectivity' is mentioned, yet a more detailed explanation of the analysis conducted is warranted.

We hope that the expanded explanation that addresses your point number 6 which included additional sections explaining the TextBlob Polarity and Subjectivity values addresses this point as well.

10. For Figure 2, an inquiry into the factors behind the spikes in % submissions around the years 2010, 2017, and 2019 should be made.

We have added the following:

Although we do not explore further the possible reasons for spikes in the number of conversations around 2010, 2018, and 2020, these present interesting opportunities for further investigation. For example, it may be that the spike around 2020 was driven by global news media covering the reduction in carbon emissions observed during the start of the COVID-19 pandemic.

11. The term 'subreddits' is used in a context that may imply understanding, yet no definition or explanation is provided within the text.

We hope that the added explanation in the expanded text addressing your point number 1 addresses this as well.

12. Figures 3 and 4 are presented with minimal detailed commentary on the observed phenomena, primarily being described rather than analyzed in depth.

That is true. While the observed phenomena are interesting and worth reporting, we do not see an immediate explanation or interpretation we could offer. However, we believe that they provide an important context for the study. It is also important and relevant to show the differences between different subreddits, which is important for our discussion, as well as to allow further, more detailed studies to follow. We have added a sentence highlighting how this contrasts with the expectations of political consultants from 2002 ((Burkeman, 2003). We have additionally added a sentence in the paragraph addressing Figure 4 suggesting more research may be warranted.

13. The observations made in the discussions section appear to extend beyond merely analyzing the terminological shift from "global warming" to "climate change." However, these findings are not emphasized within the article itself.

We made our best effort to interpret the results of our analysis of climate-related discussions. We have tried to keep a reasonable balance between interpretive observations and just reporting data.

14. I miss a statistical/econometric analysis beyond the a simple description

There are additional statistical and econometric analyses that could be done, and we would encourage others to do so; however, we hope that the strength of this paper is in the size of the dataset: we're not relying on a sample of the dataset, but rather analyze it all, which is particularly useful for new and emerging phenomena. The paper reports the key immediately relevant observation of terminological shift, and the word difficulty across subreddits, based on a study on an unprecedented scale. We agree that additional statistical and econometric analyses will be useful, however, we also believe that the current ones, we've provided, are sufficiently interesting for the larger audience.

My objective has been to highlight areas for improvement in providing clarity and context for readers, particularly those unfamiliar with the platforms and analytical tools mentioned.

We greatly appreciate the time and care you have taken to review our paper.

Reviewer #2 (Remarks to the Author):

Dear authors and editors,

thank you very much for giving me the chance to review your paper in which you checked in how far discussions on climate change on the internet platform Reddit changed over time. I am certain that your paper is of high relevance to the scientific community and offers novel insights on an important topic. I have not worked with Reddit posts and large language data sets myself, hence I feel unable to comment on your methodological proceedings.

Thank you kindly for your comments.

The comments I have on your paper revolve around your discussion section and conclusion predominantly. I have made specific comments within the manuscript which I will upload to this platform for greater clarity. Hence, my comments here are more general.

Thank you for taking the time to comment directly on the manuscript. We will address each of your comments below.

I feel that your discussion could use a more sustained engagement with previous research on the topic of social media discussions of climate change/ global warming. To me it was somewhat surprising to see a discussion section without references to other people's work. I believe there is much potential to enhance your valuable contribution if you could help the reader to better understand in how far your findings align/ differ from other research. I understand that the majority of work has been conducted on debates on X (former Twitter) or Facebook and not Reddit but still it would be useful to alert the reader to potential differences/ similarities.

Thank you for pointing us to additional research which may be of interest and relevance to those interested in our research. We have added additional text and citations to Pearce et al. 2019 in addition to Pearce et al. 2018, Fownes et al. 2018, Hautea et al. 2021, Oz et al. 2018, and Porten-Chee and Eilders 2015.

Since we are focused on English only posts which are predominantly from North America and we are unable to ascertain how many non North American originating posts there are, we opted not further address non North American contexts.

Comments from the PDF:

1. "Maybe you could rephrase this a bit, as your findings are much more detailed than a simple more/less answer that you claim to be the purpose of your study?"

Thank you for the suggestion. We have changed:

Purpose: Our purpose was to check if people discuss climate change more or less proportionally to other topics. We collected and analyzed 16 years and 11.5 billion Reddit posts using sentiment analysis.

To:

Purpose: The aim of this study was to examine the evolution of discourse surrounding 'climate change' and 'global warming' over 16 years within 11.5 billion Reddit posts. By employing sentiment analysis and other computational linguistics techniques, we sought to identify shifts in discussion frequency, sentiment polarity, and linguistic complexity in relation to these environmental topics. Our analysis not only reveals a change of terminology used in online environmental discourse and offers insights into public perception and engagement with these critical issues over time.

2. "This seems a bit generic; could you specify?" (Lines 49-50)

Agreed. We have deleted the comment. (Line 56)

3. "Check language?"

Thank you for catching this. We have removed the offending item.

4. comment: "?" regarding a sentence fragment (Lines 155-157).

We updated the methods section to be more concise and hopefully easier to understand.

5. "Interesting! Can you unpack and discuss this finding a bit more? Why 2013? Administrative change in leading countries? New IPCC insights? COP meetings?"

We tried to find an "easy" way to download the IPCC reports from 2005 on, but unfortunately, for many of them we could not find the complete report as one PDF or document to download, but only a site where one has to download each subsection by going to many different pages. An initial analysis of the IPCC reports sections we did download suggests that "global warming" is either never or almost never used so far. Thus, we believe that this is not immediately informative for our study – although may be useful in future research.

6. "I think the discussion could do with more in-depth debates on your findings and previous research. I am missing more engagement with other studies on social media debates and the topic of climate change. Such a contextualisation would help the reader to better evaluate and understand the significance of your findings."

It would be wonderful if we could analyze other online platforms; however, at this time we do not have access to a complete or substantial portion of other forums such as X (Twitter) or Facebook to perform a compelling analysis to answer this question. It is worth noting that, unlike previously, with current EULA as well as anti-scraping regulations and tools it is also impossible for all practical purposes.

We have expanded the discussion section and additionally added:

In addition to highlighting novel findings associated with the terms evaluated with respect to sentiment, polarity, readability, and preferred nomenclature, our findings provide additional context to previous findings concerning climate related topics in online social media ([Pearce et al. 2019](#); [Fownes et al. 2018](#); [Hautea et al. 2021](#); [Oz et al. 2018](#)). We would like to encourage those with access to relevant Facebook, Twitter/X, or other social media to compare those data with our findings.

7. "Please give more details/ examples on what you mean with 'positive' and 'negative' here. It is not immediately clear to the reader."

We have added a detailed explanation of the TextBlob and VADER sentiment and polarity scores and categories.

8. "Or is the discussion shifting to other online platforms?"

As discussed above, we feel this would be an excellent line of investigation which may be hampered by the changes to the EULAs of providers such as Facebook and Twitter/X. We have added the following to the conclusion:

Further investigation into how discussions about climate change are shifting across various online platforms could provide a more comprehensive understanding of the evolution of public discourse on this critical issue.

9. "How important is the US presidency here? Is it possible to georeference the Reddit posts to get a better understanding of where people are located who are taking part in these discussions?"

Although interesting questions, one could attempt to determine the effect of the US presidency on discourse; however, it is not immediately apparent how we might best approach this question. We therefore leave it for future research. With respect to using geolocation information, but we cannot do the latter since Reddit does not provide that information.

Lack of peaks does not suggest a significant effect; however, may be interesting for further research.

10. "Can you define 'difficult words'?"

We have added an expanded explanation of "difficult words":

TextStat's Difficult Words analysis is a tool used to quantify the complexity of vocabulary in a text. It evaluates the readability of content by counting words not found on a predefined list of commonly understood words, typically based on the Dale-Chall word list ([Dale and Chall 1948](#)). This metric is particularly useful in assessing the accessibility of text to audiences with varying levels of literacy. In our study, we employed TextStat Difficult Words analysis to measure the linguistic complexity of discussions within Reddit posts concerning 'climate

change' and 'global warming.' By applying this analysis, we aimed to determine the sophistication of language that could potentially influence the comprehensibility and reach of the discourse on these environmental issues. The insights gained help to understand how the complexity of language in online discussions may affect public engagement and awareness.

11. "This is certainly true for the US and some other countries; however, not so very true for others (Sweden, Denmark...). Maybe re-phrase and make the statement somewhat less Anglo-centric?"

Point well taken. We have modified:

Global environmental change is highly politicized and ideologized in the public discourse (Stecula and Merkley, 2019; Chinn, Hart and Soroka, 2020).

To:

Within North America, the global environmental change is highly politicized and ideologized in the public discourse (Stecula and Merkley, 2019; Chinn, Hart and Soroka, 2020).

Also added:

It is important to note that climate related topics may not have the same level of politicization in all regions of the world ([McCright et al. 2016](#))

12. "I am not sure if I am convinced by this final conclusion. You have acknowledged before that the downward trend in Reddit discussions on cc is likely due to the increase in discussions on other topics. Hence, it might be a bit hasty to declare a 'closing window' here?"

Thank you for pointing this out. We have removed the claim.

Abstract

In the study of 16 years and 11.5 billion Reddit posts, we show that “climate change” replaced “global warming” in 2013 in frequency. We also show that, more recently, the climate has become a less frequent topic of online conversations and searches. Moreover, and contrary to previous findings, we show that “global warming” is increasingly associated with more positive sentiment than “climate change.” Additionally, we discover that discussions about “global warming” use less subjective language than “climate change.” Our findings show that global environmental change is a complicated topic, and while it is being discussed in the media more often, the engagement of the general public is difficult to maintain. This result brings a dire warning to policymakers.

1. Introduction

Global environmental change has been discussed in the media for many decades, with varied issue attention (Hase et al., 2021) and different framing from country to country (Vu, Liu and Tran, 2019). Climate journalism has emerged and evolved as a field (Schäfer and Painter, 2021).

“Global warming” and “climate change” are terms that are sometimes used interchangeably (Shehata and Hopmann, 2012; Broadbent *et al.*, 2016). Yet, they do have different semantic contexts and connotations: “Global warming” refers to a long-term increase in global temperatures, while “climate change” denotes changes in climate patterns, in particular involving the increased levels of carbon dioxide in the atmosphere (Lineman *et al.*, 2015). ~~Both terms are highly politicized, though, and for a good reason.~~

“Climate change” has been used in academic papers since the fifties (Plass, 1956), but it has been more widely popularized in the public discourse by Frank Luntz, a Republican communication strategist, who in 2002 advised the Bush administration to use it instead of “global warming,” as a less frightening term (Burkeman, 2003), and to make the perceived lack of scientific certainty a vital issue in the debate about the environment. Although we do not address how this intentional shift in nomenclature influenced the discourse prior to 2005 since the dataset begins in late 2005, this would be an interesting area for further research.

~~It made a political sense:~~ even though for liberals, “global warming” and “climate change” are associated equally with rising temperatures and, overall, the heat impact on the planet, conservatives see “climate change” as more disconnected from it (Schuldt and Roh, 2014). It is also worth noting that the US media, in particular, are prone to relativization of the topic and frequently evoke scientific uncertainty and controversy regarding human-induced climate change (Shehata and Hopmann, 2012). Whether the Republican plan was successful or not, over the years, a noticeable shift in vocabulary from “global warming” to “climate change” in the popular press and academic papers alike occurred (Zehr, 2000; Sonnett, 2022).

The social perception of these terms on the Internet has been studied, e.g. by Google Trends (Jun, Yoo and Choi, 2018; Sisco and Weber, 2022). There have been also many studies of the perception of climate change and environmental challenges on the social media, with a

Z komentarzem [A1]: Addressing Reviewer #1

heavy bias towards Twitter (Lineman *et al.*, 2015; Pearce *et al.*, 2019; Ganczewski and Jemielniak, 2022), possibly because of the ease of data access.

However, it has not been a subject of a systematic review of large-scale online discussions. It is essential, though, as online discussion fora provide a platform for contrasting arguments and reasoning, in contrast to Twitter, which relies on short, linguistically-restricted statements, and in contrast to just Google Trends, which only reflects people's search interests. Online discussions are also a good proxy for general public interest and perceptions, and may help to understand the emergence of public spheres in climate-related topics (Neff and Jemielniak, 2022).

Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends (Fariello, Jemielniak and Sulkowski, 2021; Gaudette et al., 2021). As of 2023, Reddit, with its unique structure centered around interest-based communities known as "subreddits," hosts over 850 million monthly active users (MAUs) (Curry, 2024) and 2.1 billion unique visitors (Tiano Bianchi, 2024), engaging in in-depth discussions and sharing a wide array of content. Notably, it is the 5th most visited social media platform globally (Dixon, 2024), with approximately 48% of visits (Tiago Bianchi, Feb 29, 2024) originating from the United States. Reddit represents approximately one third as much social media traffic as Twitter/X which serves as a hub for real-time updates and public discourse (Dixon, 2024). Although Facebook remains the dominant force with just over 3 billion monthly active users, catering to a broad demographic by facilitating personal connections and diverse content consumption of which approximately 24% are based in India and 12% are based in the United States (Dean, 2024). Reddit's model of specialized, user-moderated subreddits is particularly resonant among specific interest groups, steadily growing in its impact on public opinion and media trends.

In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.

~~Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends (Fariello, Jemielniak and Sulkowski, 2021; Gaudette et al., 2021). In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.~~

Z komentarzem [A2]: Addressing Reviewer 1 asking for further details about Reddit and its importance.

2. Material and methods

2.1 Data collection

We obtained a large dataset (n=11.486 x 10⁹) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files). Submissions are the top-most post that initiates a discussion within a group known as a subreddit while comments are replies to either a submission or another comment.~~We obtained a large dataset (n=11.486 x 10⁹) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions~~

Z komentarzem [A3]: Addressing Reviewer 1

(RS files) and Reddit Comments (RC files). We uncompressed each monthly file, and removed duplicates. We then used a series of regular expressions to identify all posts related to climate change while ignoring all instances of political, economic, financial, academic, work, social, job, or employment climate, as well as terms such as “climate control” used to refer to topics not related to climate change while still retaining all posts that contained explicit references to climate change, global warming, and related climate specific topics such as references to the Paris Accord, the Kyoto Protocol, Greta Thunberg, and others. The complete list and order of expressions used is included in the supplemental materials.

This resulted in 15.341×10^6 posts likely to be related to climate change (0.134%). These posts were then converted from JSON format to a more data science friendly CSV format retaining all the relevant information on which we performed language detection using both the polyglot (Al-Rfou, 2015) and langdetect Python libraries, keeping only posts that were determined to be in English by both ($n=1.501 \times 10^6$, 97.9% of matched posts). We subsequently concatenated and lemmatized the title and body of all the remaining posts using standard NLTK library methods (`nltk.stem.WordNetLemmatizer`) and removed all English stop words (Bird, Klein and Loper, 2009). Each post was then analyzed using the VaderSentiment (Hutto and Gilbert, 2014), TextBlob sentiment, TextBlob naïve Bayes, and textstat python libraries among several other interesting metrics such as readability and language complexity scores provided in the supplemental data.

VADER (Valence Aware Dictionary and sEntiment Reasoner): A lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of a sentiment lexicon and a set of heuristic rules to estimate sentiment scores for textual data (Hutto & Gilbert, 2014).

VADER's rule-based approach, which takes into account both the semantic meaning of words and the context in which they are used (including punctuation, capitalization, and modifiers), makes it highly effective for the nuanced understanding of sentiment in social media posts. This approach provides a robust framework for analyzing how sentiment regarding climate change and global warming is expressed in online discussions, contributing significantly to the study's findings on public engagement and perception related to these topics.

- Sentiment Score: VADER analyzes the text to calculate a compound score that indicates the overall sentiment. This compound score is a normalized, weighted composite score that ranges from -1 (extremely negative) to 1 (extremely positive). This scoring is beneficial in understanding the general sentiment of each post regarding climate change or global warming.Top of Form

TextBlob: Provides a simple API for common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. In this study, we utilized TextBlob's sentiment analysis features to assess the polarity and subjectivity of the posts.

- Polarity: This refers to the emotional orientation of the words used in the text. Polarity is a float that lies between -1.0 and 1.0, where -1.0 indicates a highly negative sentiment, 0 indicates neutrality, and 1.0 indicates a highly positive sentiment. It essentially measures the degree to which the language used in a text leans towards a positive or negative emotion. This metric is crucial for understanding the general

sentiment expressed in discussions about climate change and global warming on social media platforms.

- **Subjectivity:** This measures the amount of personal opinion and factual information contained in the text. The value ranges from 0 to 1, where 0 is very objective and 1 is highly subjective. Subjective text expresses personal feelings, views, or beliefs, whereas objective text provides factual information. Analyzing subjectivity helps in distinguishing between posts that are based on personal opinions and those based on objective reports, which is vital for evaluating how discussions on climate change are framed by the public.

In our analysis, these metrics enabled a deeper understanding of the emotional and subjective layers within the climate change discourse on Reddit, providing insights into how people emotionally engage with the topic and the extent to which discussions are opinion-driven or fact-based (Loria, 2020).

TextStat's Difficult Words analysis is a tool used to quantify the complexity of vocabulary in a text. It evaluates the readability of content by counting words not found on a predefined list of commonly understood words, typically based on the Dale-Chall word list (Dale and Chall 1948). This metric is particularly useful in assessing the accessibility of text to audiences with varying levels of literacy. In our study, we employed TextStat Difficult Words analysis to measure the linguistic complexity of discussions within Reddit posts concerning 'climate change' and 'global warming.' By applying this analysis, we aimed to determine the sophistication of language that could potentially influence the comprehensibility and reach of the discourse on these environmental issues. The insights gained help to understand how the complexity of language in online discussions may affect public engagement and awareness.

Z komentarzem [A4]: Addressing reviewer 1

We then analyzed the unigram, bigrams, and trigrams and subsequently analyzing the occurrences of the "climate" unigram when not followed by "change" and the "climate change" bigram as well as the "global" unigram when not followed by "warming" and the "global warming" bigram. We accomplished this by replacing all permutations of occurrences of "climate change" with "climate-change" and "global warming" with "global-warming" after lemmatization and performing a unigram analysis. For each of the resultant unigrams, we analyzed sentiment, polarity, and readability subdivided by the top 40 subreddit, and looked at the relative trends for the unigrams across each year.

2.2 Selecting Climate Related Posts

To ensure our analysis focused solely on content relevant to climate change, we developed a set of criteria to filter posts from the PushShift.io data repository. This involved using regular expressions to select posts that specifically mentioned climate change-related terms and exclude irrelevant contexts.

Methodology Summary: Our approach was twofold. First, we included posts that contained explicit references to climate-related topics, such as "global warming," "carbon emissions,"

and specific environmental policies (e.g., the Paris Accord). For example, one of our expressions

`\\bglobal\\w*[\\W_]*(?:warm\\w*|cool\\w*|temper\\w*|heat\\w*|hot[estr]*\\b|cold[estr]*\\b)`

captures discussions around global temperature changes.

Second, we excluded posts where the term "climate" appeared in non-environmental contexts, such as "political climate" or "climate control," using negative lookbehind assertions in our regular expressions. For instance, the expression `(?<!politic\\w*)\\bclimat` ensures the word "climate" is not preceded by "political" or related modifiers.

For readers interested in the complete list of regular expressions and their specific purposes, we have included these details in the Supplementary Materials. This includes a comprehensive breakdown of all filters used to accurately curate our dataset, ensuring that only relevant discussions were analyzed.

Z komentarzem [A5]: Addressing Reviewer 1 and 2

On the original JSONL formatted data downloaded from PushShift.io, we extract all posts whose title text and body text, when concatenated together, matched any one of the following regular expressions:

- `\\bclimat(?:olog|arian|ard|is[mt]|caust|sci)\\w*\\bclimist\\w*`
- `\\b(?:global\\w*|environ\\w*|wea?ther\\w*|oceans?|seas?|atmos\\w*|historic\\w*)[\\W_]*(?:warm\\w*|cool\\w*|temper\\w*|heat\\w*|hot[estr]*\\b|cold[estr]*\\b)`
- `\\b(?:oceans?|seas?)[\\W_]*(?:level\\w*|temper\\w*|acid\\w*|ris[ei]\\w*|heat[ingswave]*|ph\\b)`
- `\\b(?:level\\w*|temper\\w*|acid\\w*|ris[ie]\\w*|heat[ingswave]*|ph\\b)\\w*[\\W_]*(?:oceans?|seas?)`
- `\\b(?:green[estr]?|altern\\w*|renewab\\w*|wind|solar|hydro(?:xid|ponic|phobi|fobi)\\w*|geo[\\W_]*ther\\w*|nuclear\\w*|nucular\\w*|clean[estr]*[\\W_]*energ\\w*`
- `\\b(?:green[estr]?|wind|solar|hydro\\w*|geother\\w*|clean[estr]*)[\\W_]*(?:tech\\w*|solut\\w*)`
- `\\bclean[estr][\\W_]*(?:tech\\w*|energ\\w*)`
- `\\bgreen[\\W_]*house\\w*[\\W_]*(?:gas\\w*|effect\\w*)`
- `\\bcarbon[\\W_]*(?:neutr|emi[st]|diox|foot)\\w*`
- `\\bco2\\b`
- `\\b(?:emit\\w*|produc\\w*|contrib\\w*)[\\W_]*carbon`
- `\\b(?:polar|arc?tic(?:al|le)\\w*|ant[\\W_]*arc?tic\\w*)[\\W_]*(?:ice\\w*|shelf\\w*|temper\\w*|warm\\w*|cool\\w*)`
- `\\bthu[mn][mn]?b[eu]rg\\b`
- `\\bipcc\\b`
- `\\b(?:montreal|kyoto)[\\W_]*proto\\w*\\b`
- `\\bice[\\W_]*(?:caps?\\b|shell[fves]*|sheets?)\\b`
- `\\bfossil[\\W_]*fuels?\\b`
- `\\b[dr]e[\\-]*forest\\w*`
- `\\brenewables\\b`
- `\\bwind[\\W_]*turbine[s]*\\b`
- `\\bparis[\\W_]*(?:accord|agreem)\\w*`

And all of the following regular expressions:

- ~~(?<!politicalw*)\bclimat~~
- ~~(?<!econom\w*)\bclimat~~
- ~~(?<!financ\w*)\bclimat~~
- ~~(?<!academ\w*)\bclimat~~
- ~~(?<!work\w*)\bclimat~~
- ~~(?<!social\w*)\bclimat~~
- ~~(?<!job\w*)\bclimat~~
- ~~(?<!employ[me]\w*)\bclimat~~
- ~~\bclimat\w*W*(?!control)~~

The former regular expressions include any posts with text tightly associated with climate change related topics, as well as bigrams such as “political climate” or “climate control” which refer to topics that are not specific to climate change while. The latter regular expressions ensure that any items that did match “climate” did so because climate appeared at least once not preceded by words such as “political,” “financial,” “job,” or “academic” and not followed by “control.”

3. Results

Climate Related NGram Properties
2005 to 2021, $n = 13.0 \times 10^6$

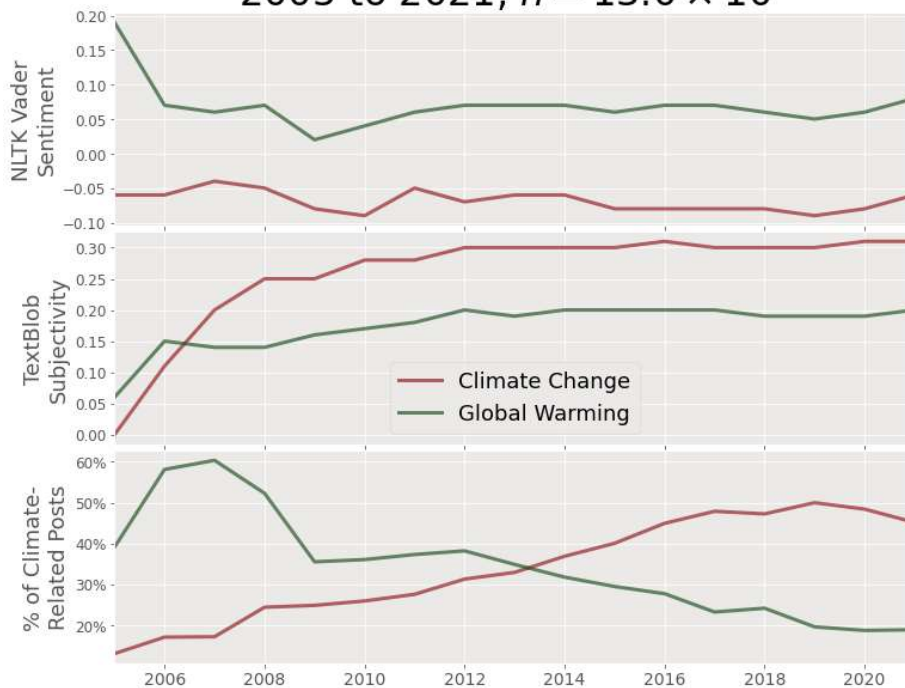


Figure 1: The sentiment, subjectivity, and percentage of posts related to climate

Figure 1 shows that the frequency of usage of the term "global warming" decreased over time being preferentially replaced with the term "climate change" (bottom) even though the sentiment associated with "climate change" remained more negative while the subjectivity for both terms increased over time (top) with "climate change" being associated with more subjective text than "global warming" since 2007 (middle).

Z komentarzem [A6]: Addressing Reviewer 1

Figure 1 shows the sentiment, subjectivity, and percentage of posts related to climate.

Percentage of All Posts Matching Climate-Change Related Terms from 2005 to 2021, $n = 11.5 \times 10^9$

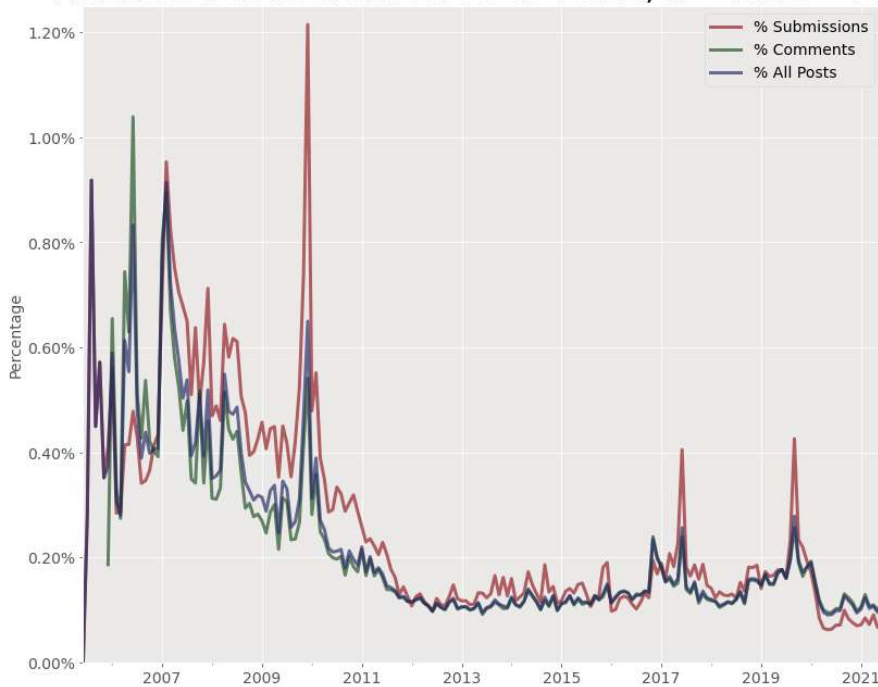


Figure 2: Percentage of all posts related to climate change

Figure 2 shows the relative frequency of climate-change-related topics in time as a percentage of discussions on Reddit. The discussion frequency has been trending down over the time period analyzed. Figure 2 shows the relative popularity of climate-change-related topics in time.

Z komentarzem [A7]: Addressing Reviewer 1

Mean NLTK Vader Composite Score by Subreddit for All Dates

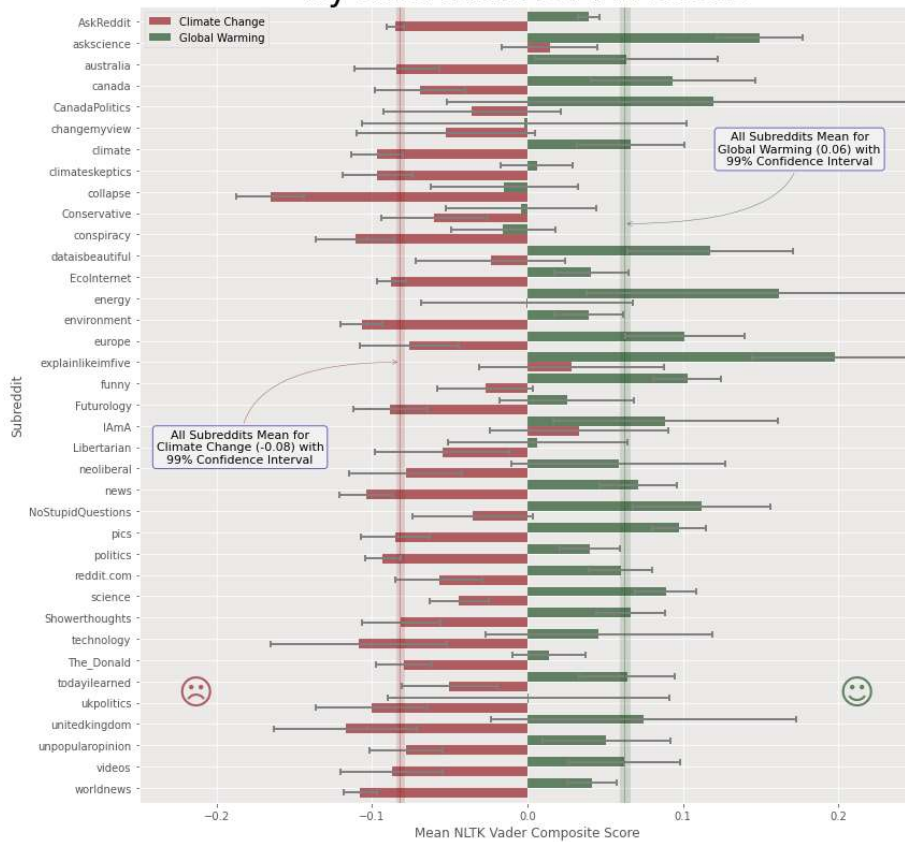


Figure 3: Mean NLTK Vader Composite Scores by Subreddit

Figure 3 depicts the mean NLTK Vader Sentiment Composite Score for climate-related posts containing the terms “climate change” or “global warming” across all subreddits, which are plotted as vertical lines with the shaded area representing the 99% confidence interval while the bars represent the mean for each of these for a given subreddit with the error bars showing the 99% confidence interval for that given subreddit and term. The subreddits (37 total) were selected by choosing the 50 highest volume subreddits with more than 100 climate-related posts.

Mean TextStat Difficult Words Score by Subreddit for All Dates

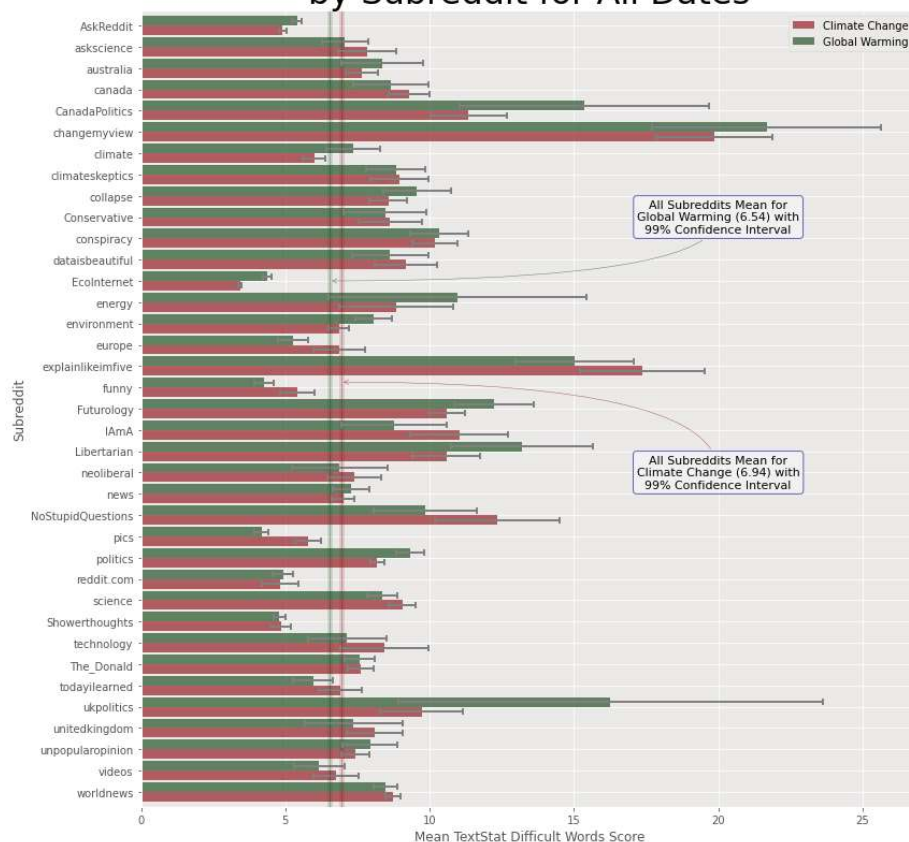


Figure 4: Mean TextStat Difficult Words Scores by Subreddit

In Figure 4, the mean TextStat Difficult Words score for all posts containing the terms “climate change” and “global warming” across all subreddits are plotted as vertical lines with the shaded area representing the 99% confidence interval while the bars represent the mean for each of these for a given subreddit with the error bars showing the 99% confidence interval for that given subreddit and term. The subreddits (37 total) were selected by choosing the 50 highest volume subreddits with more than 100 climate-related posts.

In the supplemental materials, we provide the raw analytical data, which includes calculated values calculated for a variety of text analyses, allowing for considerable additional data exploration.

Google Search Trend Scores from 2004 to 2022 for Climate Change and Global Warming 6-Month Centered Rolling Average

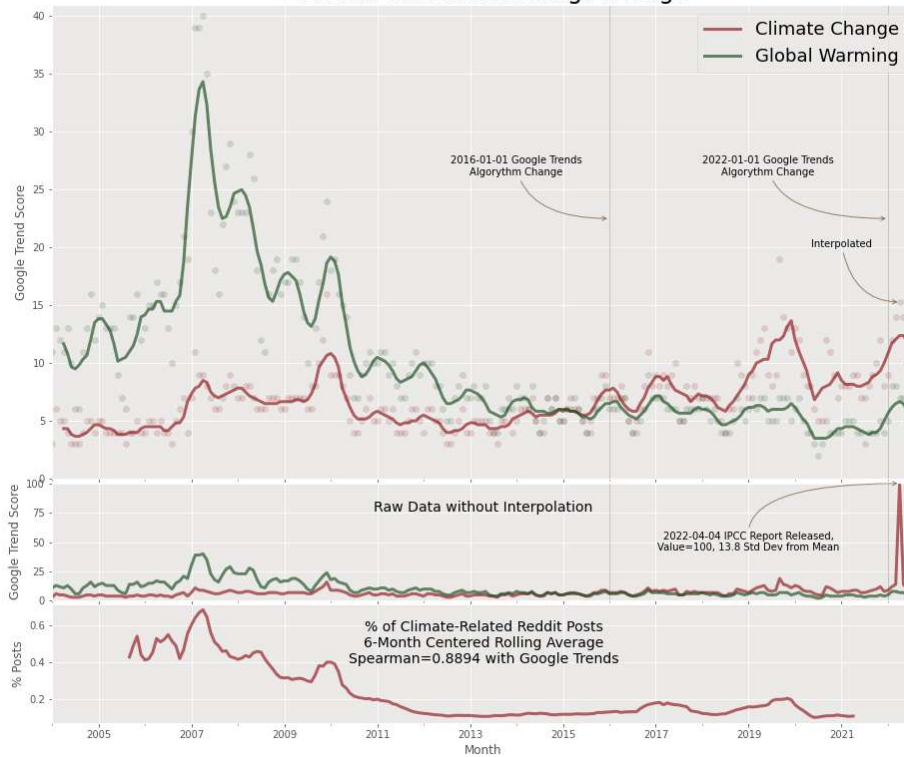


Figure 5: Google Search Trends

Figure 5 shows the Google search term trends for “climate change” and “global warming” (without quotes) for the 18 years from January 2004 to August 2022. Vertical lines represent the dates on which Google implemented algorithmic changes to how it tracks searches. In April 2022, the month during which the IPCC released its most recent and most alarming report, there was a large spike in searches reported; however, we are unable to ascertain if the two are related.

5. Discussion

Our first conclusion is that the shift from “global warming” to “climate change” indeed happened, but a decade later than the Republican initiative. Curiously, it was relatively abrupt, and around 2013, “climate change” overcame “global warming” in general discussions we analyzed (see Figure 1). It is also notable that Google search trends (Figure 5) show a similar switch in terminology occurring only two years later.

Our analysis found that, contrary to previous assumptions, “global warming” is associated with more positive sentiment polarity than “climate change.” Across all subreddits, the sentiment of posts using the term “global warming” was positive, while the sentiment of posts using the term “climate change” was negative. This result could indicate that those writing about “global warming” are less concerned about the climactic changes than those naming it “climate change,” which could be correlated with the slow uptake of new terminology. Interestingly, the subreddit “askscience,” which focuses on a scientific approach to the world, had the most positive sentiment overall for posts using either “climate change” or “global warming.”

The subjectivity of posts about “climate change” was much higher, meaning that people writing about it were expressing their views in a less authoritative tone, while people writing about “global warming” were phrasing their posts as if they were describing objective facts. This result could indicate that those writing about “climate change” are more conscious about separating opinions from facts and is, generally, not surprising. Still, this finding indicates that those who write about “global warming” may see their views as non-negotiable truths.

It is worth mentioning that during Donald Trump’s presidency, there was a small drop in both sentiment polarity and subjectivity - people were writing about “climate change” and “global warming” more negatively, and in language more typical for a description of truth rather than viewpoints. We do not have a good explanation for this phenomenon.

Moreover, it is ~~also~~ quite surprising that the discussions about “climate” in general have not increased significantly since 2009, and they remain at a stable percentage of overall posts. Only the last four years note an increase from about 20% to nearly 30%. They appear to be nearly perfectly correlated with “global warming” conversations - the more people talk about one, the less they do about the other (see Figure 2). Although we do not explore further the possible reasons for spikes in the number of conversations around 2010, 2018, and 2020, these present interesting opportunities for further investigation. For example, it may be that the spike around 2020 was driven by global news media covering the reduction in carbon emissions observed during the start of the COVID-19 pandemic.

Z komentarzem [A8]: Addressing reviewer #1

When the percentage of Reddit posts with climate-related topics is considered, we can even observe an evident decline. While 2005-2010 even up to 1.4% of topics were related to climate, after that period the numbers steadily decreased. It may be misleading, though: since we are measuring the proportions of topics, the overall growth of popularity of Reddit is likely largely contributing to the observed phenomenon. In other words, the absolute numbers of climate-related discussions have not decreased (in fact, they increased steadily), but the growth of discussions on other topics was more rapid. Nevertheless, we believe that this result is a mild warning of the fact that climate is becoming a less and less popular topic proportionally.

On a side note, we observe that, considering the word difficulty, the “ecointernet” subreddit had the least difficult words in sentences related to “climate change”, “global warming,” and “climate” in general. “Changemyview” subreddit, on the other hand, had the most difficult sentences - it may be that this subreddit, dedicated to confrontational discussions, relied more largely on sources and academic language. One surprising finding from our study is that “Explain to me like I’m 5” thread uses more complex words than other threads. In the supplement, we provide additional figures, illustrating this phenomenon.

As seen in Figure 3, the terms “climate change” appear to be associated very consistently with Vader sentiment scores that are less favorable than those of “global warming” irrespective of the subreddit, even when the subreddits vary greatly on the validity of climate-related claims (e.g., the Conservative, The_Donald, and conspiracy versus climate, askscience, and science), suggesting that the relatively more positive sentiment association with “global warming” may persist across political views and scientific literacy. Unsurprisingly, several subreddits with a stronger focus on educating people tended to use more positive tones (e.g., askscience, explainlikeimfive, IAmA). It is worth noting that this is contrary to the prediction of Bush-era political consultants (Burkeman, 2003).

We note that posts with the terms “climate change” have a small but very statistically significant increase in the use of difficult words when scores are aggregated across all subreddits, but the trend does not hold well within and across subreddits (see Figure 4). The subreddit dedicated to attempting to change the original poster’s (OP) views had the highest frequency of difficult words with respect to climate-related posts. It is interesting to note the considerable irony that the subreddit ostensibly meant to explain challenging topics using simple terminology, explainlikeimfive, had, on average, the second-highest frequency of use of difficult words. Additional research into the influence of the use of more sophisticated language on sentiment, the ability to change people’s opinions, or the ability to explain complex topics such as climate change may be warranted.

The most interesting findings are that the Google search trends (Figure 5) demonstrate a similar change in frequency of the terms from “climate change” in the searches before 2015 to “global warming” to those after. Although this occurred at a later date than that which we observe in the Reddit discussions circa 2013, it appears to be a consistent feature across two disparate Internet activities. Furthermore, the 6-month rolling averages for the sum of the two Google trend lines, which roughly approximates the overall climate-related searches over time, is strongly correlated with the percentage of Reddit climate-related discussions with a Spearman correlation coefficient of 0.89 further suggesting that the trends observed in Reddit discussions are reflective of broader trends across at least some Internet-based media.

In addition to highlighting novel findings associated with the terms evaluated with respect to sentiment, polarity, readability, and preferred nomenclature, our findings provide additional context to previous findings concerning climate related topics in online social media (Pearce et al. 2019; Fownes et al. 2018; Hautea et al. 2021; Oz et al. 2018). We would like to encourage those with access to relevant Facebook, Twitter/X, or other social media to compare those data with our findings.

6. Conclusions

Within North America, the global environmental change is highly politicized and ideologized in the public discourse~~Global environmental change is highly politicized and ideologized in the public discourse~~ (Stecula and Merkley, 2019; Chinn, Hart and Soroka, 2020). The mainstream media have been covering it systematically for years. Even though disproportionate attention has been given to contrarians rather than experts (Petersen, Vincent and Westerling, 2019), and even though extreme weather phenomena has been

Z komentarzem [A9]: Addressing reviewer #2

affecting the media attention more than scientific accounts (Pianta and Sisco, 2020), the overall media attention has been growing (Barkemeyer *et al.*, 2017; Layna and Altamirano, 2022). Moreover, even the more conservative outlets have started acknowledging the burning problem (Schmid-Petri *et al.*, 2017). It is important to note that climate related topics may not have the same level of politicization in all regions of the world (McCrigh et al. 2016).

In our article we have shown that the proportionate interest of the general public, as reflected in large-scale online discussions, has not been following the same path. Possibly because of exhausting the topic and being tired of continuous high alert state, other competing crises (Stoddart *et al.*, 2021), as well as simply other interesting topics, the discourse about climate change has taken less and less space proportionately.

Moreover, we show that discussions about “global warming” appear to use less sophisticated language and are more positive than those about “climate change.” This result suggests that perceptions of global environmental change manifest in language, which is a phenomenon that call for further exploration.

We also show that online discussions seem to precede Google searches in terms of language shift from “global warming” to “climate change”, which suggests that monitoring social media discussions may be a more immediate way of detecting changes in public attention and imaginary. especially given the interplay between dissemination of information in user generated content and more traditional mass media outlets which may not follow previously assumed patterns in other contexts (Porten-Cheé and Eilders 2015).

All of these results indicate that the window for mobilizing the general public's attention around global environmental change may be closing, and that the main debate around the topic among non-experts may be affected by the informal discussions on social media more than elsewhere. Although our analyses included only English posts, the majority of which were likely from North America, it is important to note that there may be significant differences in sentiment and polarity when looking at other regions of the world or other languages (McCrigh et al. 2016).

Further investigation into how discussions about climate change are shifting across various online platforms could provide a more comprehensive understanding of the evolution of public discourse on this critical issue.

Z komentarzem [A10]: Addressing Reviewer #2.

Sformatowano: Normalny, Wyjustowany

7. References

Al-Rfou, R. (2015) *Polyglot: A Massive Multilingual Natural Language Processing Pipeline*. search.proquest.com. Available at: <https://search.proquest.com/openview/a6748f01943c73c42dd77d55625097aa/1?pq-origsite=scholar&cbl=18750>.

Barkemeyer, R. *et al.* (2017) 'Media coverage of climate change: An international comparison', *Environment and Planning C: Politics and Space*, 35(6), pp. 1029–1054.

Bird, S., Klein, E. and Loper, E. (2009) *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. 'O'Reilly Media, Inc.'

- Broadbent, J. *et al.* (2016) 'Conflicting Climate Change Frames in a Global Field of Media Discourse', *Socius*, 2, p. 2378023116670660.
- Burkeman, O. (2003) 'Memo exposes Bush's new green strategy', *The guardian*. Available at: <https://amp.theguardian.com/environment/2003/mar/04/usnews.climatechange>.
- Chinn, S., Hart, P.S. and Soroka, S. (2020) 'Politicization and Polarization in Climate Change News Content, 1985-2017', *Science communication*, 42(1), pp. 112–129.
- Fariello, G., Jemielniak, D. and Sulkowski, A. (2021) 'Does Godwin's law (rule of Nazi analogies) apply in observable reality? An empirical study of selected words in 199 million Reddit posts', *New media & society*, p. 146144482110620.
- Ganczewski, G. and Jemielniak, D. (2022) 'Twitter is garbage: A Thick Big Data exploration of #zerowaste hashtag on Twitter in relation to packaging and food packaging materials', *Packaging Technology & Science* [Preprint]. Available at: <https://doi.org/10.1002/pts.2685>.
- Gaudette, T. *et al.* (2021) 'Upvoting extremism: Collective identity formation and the extreme right on Reddit', *New Media & Society*, 23(12), pp. 3491–3508.
- Hutto, C. and Gilbert, E. (2014) 'VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text', *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), pp. 216–225.
- Jun, S.-P., Yoo, H.S. and Choi, S. (2018) 'Ten years of research change using Google Trends: From the perspective of big data utilizations and applications', *Technological forecasting and social change*, 130, pp. 69–87.
- Layna, J. and Altamirano, L. (2022) 'Promises that Don't Work? COP26 and the Problems of Climate Change', *Engaging Science, Technology, and Society*, 8(3), pp. 118–129.
- Lineman, M. *et al.* (2015) 'Talking about Climate Change and Global Warming', *PloS one*, 10(9), p. e0138996.
- Neff, T. and Jemielniak, D. (2022) 'How do transnational public spheres emerge? Comparing news and social media networks during the Madrid climate talks', *New Media & Society*, p. 14614448221081426.
- Pearce, W. *et al.* (2019) 'The social media life of climate change: Platforms, publics, and future imaginaries', *Wiley interdisciplinary reviews. Climate change*, 10(2), p. e569.
- Petersen, A.M., Vincent, E.M. and Westerling, A.L. (2019) 'Discrepancy in scientific authority and media visibility of climate change scientists and contrarians', *Nature communications*, 10(1), p. 3502.
- Pianta, S. and Sisco, M.R. (2020) 'A hot topic in hot times: how media coverage of climate change is affected by temperature abnormalities', *Environmental research letters: ERL [Web site]*, 15(11), p. 114038.
- Plass, G.N. (1956) 'The carbon dioxide theory of climatic change', *Tell'Us*, 8(2), pp. 140–154.
- Schäfer, M.S. and Painter, J. (2021) 'Climate journalism in a changing media ecosystem: Assessing the production of climate change-related news around the world', *Wiley interdisciplinary reviews. Climate change*, 12(1). Available at: <https://doi.org/10.1002/wcc.675>.

Schmid-Petri, H. *et al.* (2017) 'A changing climate of skepticism: The factors shaping climate change coverage in the US press', *Public understanding of science* , 26(4), pp. 498–513.

Schuldt, J.P. and Roh, S. (2014) 'Media Frames and Cognitive Accessibility: What Do "Global Warming" and "Climate Change" Evoke in Partisan Minds?', *Environmental Communication*, 8(4), pp. 529–548.

Shehata, A. and Hopmann, D.N. (2012) 'Framing climate change', *Journalism Studies*, 13(2), pp. 175–192.

Sisco, M.R. and Weber, E.U. (2022) 'Local temperature anomalies increase climate policy interest and support: Analysis of internet searches and US congressional vote shares', *Global environmental change: human and policy dimensions*, 76, p. 102572.

Sonnett, J. (2022) 'Climate change risks and global warming dangers: a field analysis of online US news media', *Environmental Sociology*, 8(1), pp. 41–51.

Stecula, D.A. and Merkley, E. (2019) 'Framing Climate Change: Economics, Ideology, and Uncertainty in American News Media Content From 1988 to 2014', *Frontiers in Communication*. Available at: <https://doi.org/10.3389/fcomm.2019.00006>.

Stoddart, M.C.J. *et al.* (2021) 'Competing Crises? Media Coverage and Framing of Climate Change During the COVID-19 Pandemic', *Environmental Communication*, pp. 1–17.

Vu, H.T., Liu, Y. and Tran, D.V. (2019) 'Nationalizing a global phenomenon: A study of how the press in 45 countries and territories portrays climate change', *Global environmental change: human and policy dimensions*, 58, p. 101942.

Zehr, S.C. (2000) 'Public representations of scientific uncertainty about global climate change', *Public understanding of science* , 9(2), p. 85.

8. Appendices: Supplemental Data and MetaData

The data and the associated data used to produce the plots and findings is made freely available in a [Reddit-Climate-Data-2005-2021](#) shared folder with the following contents:

[Reddit-Climate-Data-2005-2021/](#)

- [Reddit General Climate Related Plots 2005-2021.ipynb](#) (2.2 MB)
The Jupyter Notebook compatible Google Colab Notebook used to create all the plots, and others, in this publication.
- [Headers Information.xlsx](#) (20KB)
A Microsoft Excel file containing information on the various headers (columns) found in the data files.
- [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.subreddit.csv](#) (77MB)
A (large) summary of all analyzed data grouped by subreddit.
- [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.subreddit_year.csv](#) (184MB)
A (large) summary of all analyzed data grouped by subreddit and year.
- [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.year.csv](#) (31KB)
A summary of all analyzed data grouped by year.

- [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.day_of_week.csv](#) (14KB)
A summary of all analyzed data grouped by the day of the week (0 = Monday, 6=Sunday).
- [RedditBoth-All-Climate-English-matching.csv](#) (10KB)
A summary of the number of submissions, comments, and both that matched climate related searches.
- [Unsorted-Top-25-NGrams-Analaysis-Results-Merged-All-Subreddits.csv](#) (23MB)
A summary of the (at most) top 25 NGrams analyzed unigrams, bi-grams, and 4-grams. Recall that “climate change” and “global warming” were masked as the unigrams “climate-change” and “global-warming” respectively for this analysis thereby enabling a comparison of the features of “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by “warming” vs. “global” not followed by “warming”.
- [Unsorted-Top-25000-NGrams-Analaysis-Results-Merged-All-Subreddits.csv](#) (51MB)
A summary of the (at most) top 25,000 NGrams analyzed unigrams, bi-grams, and 4-grams. Recall that “climate change” and “global warming” were masked as the unigrams “climate-change” and “global-warming” respectively for this analysis thereby enabling a comparison of the features of “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by “warming” vs. “global” not followed by “warming”.
- [All-Subreddits.search-term.unsorted-ngrams.all-time.csv.yearly.merged-top-250.xlsx](#) (6MB)
A summary in Microsoft Excel format with heat maps of the (at most) top 250 NGrams analyzed unigrams, bi-grams, and 4-grams. Recall that “climate change” and “global warming” were masked as the unigrams “climate-change” and “global-warming” respectively for this analysis thereby enabling a comparison of the features of “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by “warming” vs. “global” not followed by “warming”.
- [All-Subreddits.search-term.unsorted-ngrams.all-time.csv.yearly-used-columns.xlsx](#) (65KB)
A summary in Microsoft Excel format with heat maps containing only items used for some subset of analyses containing a subset of the (at most) top 250 NGrams analyzed unigrams, bi-grams, and 4-grams. Recall that “climate change” and “global warming” were masked as the unigrams “climate-change” and “global-warming” respectively for this analysis thereby enabling a comparison of the features of “climate” followed by “change” vs. “climate” not followed by “change” and “global” followed by “warming” vs. “global” not followed by “warming”.
- [Google-Trends-Global-Warming-and-Climate-Change-2004-2022.csv](#) (4KB)
The Google Trends data obtained at <https://trends.google.com/trends/explore?date=all&q=%2Fm%2F0d063v,Global%20Warming> on September 4, 2022
- [All-Climate-Related-Posts-Raw-Data-2005-2021.csv.zst](#) (5.54GB)
The complete raw data and text analysis data for all posts containing climate related matches.

Many thanks for the comments and suggestions for improvement. We are including a detailed reply to each of them below (reviewers' comments in magenta, our replies in blue, and changes in black), as well as a revision-tracked copy of the manuscript below.

The discussion and conclusion sections have been completely rewritten.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

This article conducts an analysis of discussions surrounding climate change and global warming from 2005 to 2021 through the examination of Reddit posts. The findings suggest that, overall, conversations pertaining to climate change tend to manifest a more negative sentiment compared to those about global warming. Furthermore, it is observed that around the year 2012, the term "climate change" began to be employed more frequently than "global warming." This study provides a detailed perspective on how public discourse on these critical environmental issues has evolved, reflecting shifts in public concerns and trends in the perception of the climate crisis.

1. Further detail on the Reddit platform should be discussed. It is not as globally recognized as platforms mentioned in the text (such as Twitter/ X). Who are its users? Could it be considered representative? Additionally, the differences with Twitter, which are only briefly mentioned, warrant further exploration.

Thank you for your observation. We have changed

Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends ([Fariello, Jemielniak and Sulkowski, 2021](#); [Gaudette et al., 2021](#)). In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.

To:

Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends ([Fariello, Jemielniak and Sulkowski, 2021](#); [Gaudette et al., 2021](#)). As of 2023, Reddit, with its unique structure centered around interest-based communities known as "subreddits," hosts over 850 million monthly active users (MAUs) (Curry, 2024) and 2.1 billion unique visitors (Tiano Bianchi, 2024), engaging in in-depth discussions and sharing a wide array of content. Notably, it is the 5th most visited social media platform globally (Dixon, 2024), with approximately 48% of visits (Tiago Bianchi, Feb 29. 2024) originating from the United States. Reddit represents approximately one third as much social media traffic as Twitter/X which serves as a hub for real-time updates and public discourse (Dixon, 2024). Although Facebook remains the dominant force with just over 3 billion monthly active users, catering to a broad demographic by facilitating personal connections and diverse content consumption of which approximately 24% are based in India and 12% are based in the United States (Dean, 2024), Reddit's model of specialized, user-moderated subreddits is particularly resonant among specific interest groups, steadily growing in its impact on public opinion and media trends.

In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.

2. The origin of the analyzed conversations should be clarified. Are they predominantly from the USA? This detail is not mentioned anywhere in the text.

We hope that the text above clarifies that the plurality of posts on Reddit are from the United States. We only consider posts in English; however, there is no way for us to determine with any precision the origin of posts since Reddit does not provide geolocation or other information.

3. Are the comments analyzed exclusively in English, or do they encompass multiple languages?

Yes. This is explicitly addressed in the methods section which we have revised slightly.

4. The discourse is noted to shift in 2002 (Bush administration as it is indicated in the text), yet the analyzed data originates from 2005. Implications should be addressed.

Although we do not address how this intentional shift in nomenclature influenced the discourse prior to 2005 since the dataset begins in late 2005, this would be an interesting area for further research.

5. The terms 'Reddit submissions' and 'Reddit comments' are utilized without explanation. For readers unfamiliar with Reddit, these concepts may be challenging to grasp.

Thank you for catching the lack of clarity. We have changed:

We obtained a large dataset ($n=11.486 \times 10^9$) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files).

To:

We obtained a large dataset ($n=11.486 \times 10^9$) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files). Submissions are the top-most post that initiates a discussion within a group known as a subreddit while comments are replies to either a submission or another comment.

6. The tools 'VADER' and 'TextBlob' are mentioned without description. Readers unfamiliar with these analysis tools are left without understanding of their application or significance. I know that VADER, for example, ranges from -1 to 1. But this kind of aspects do not appear in the text to give you more feedback about the content of the paper (just an example).

Thank you again for this observation. We have added the following text to section 2.1:

VADER (Valence Aware Dictionary and sEntiment Reasoner): A lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of a sentiment lexicon and a set of heuristic rules to estimate sentiment scores for textual data (Hutto & Gilbert, 2014). VADER's rule-based approach, which takes into account both the semantic meaning of words and the context in which they are used (including punctuation, capitalization, and modifiers), makes it highly effective for the nuanced understanding of sentiment in social media posts. This approach provides a robust framework for analyzing how sentiment regarding climate change and global warming is expressed in online discussions, contributing significantly to the study's findings on public engagement and perception related to these topics.

- **Sentiment Score:** VADER analyzes the text to calculate a compound score that indicates the overall sentiment. This compound score is a normalized, weighted composite score that ranges from -1 (extremely negative) to 1 (extremely positive). This scoring is beneficial in understanding the general sentiment of each post regarding climate change or global warming.

TextBlob: Provides a simple API for common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. In this study, we utilized TextBlob's sentiment analysis features to assess the polarity and subjectivity of the posts.

- **Polarity:** This refers to the emotional orientation of the words used in the text. Polarity is a float that lies between -1.0 and 1.0, where -1.0 indicates a highly negative sentiment, 0 indicates neutrality, and 1.0 indicates a highly positive sentiment. It essentially measures the degree to which the language used in a text leans towards a positive or negative emotion. This metric is crucial for understanding the general sentiment expressed in discussions about climate change and global warming on social media platforms.
- **Subjectivity:** This measures the amount of personal opinion and factual information contained in the text. The value ranges from 0 to 1, where 0 is very objective and 1 is highly subjective. Subjective text expresses personal feelings, views, or beliefs, whereas objective text provides factual information. Analyzing subjectivity helps in distinguishing between posts that are based on personal opinions and those based on objective reports, which is vital for evaluating how discussions on climate change are framed by the public.

In our analysis, these metrics enabled a deeper understanding of the emotional and subjective layers within the climate change discourse on Reddit, providing insights into how people emotionally engage with the topic and the extent to which discussions are opinion-driven or fact-based (Loria, 2020).

Loria, S. (n.d.). *TextBlob: Simplified Text Processing* — TextBlob 0.18.0.post0 documentation. Retrieved April 21, 2024, from <https://textblob.readthedocs.io/en/dev/>

7. It is uncertain whether the detailed description provided in the "selecting climate related posts" section is necessary with such a level of description.

Your point is well taken. We have revised Section 2.2 to be:

To ensure our analysis focused solely on content relevant to climate change, we developed a set of criteria to filter posts from the PushShift.io data repository. This involved using regular expressions to select posts that specifically mentioned climate change-related terms and exclude irrelevant contexts.

Methodology Summary: Our approach was twofold. First, we included posts that contained explicit references to climate-related topics, such as "global warming," "carbon emissions," and specific environmental policies (e.g., the Paris Accord). For example, one of our expressions

`\bglobal\\w*[\W_]*(?:warm\\w*|cool\\w*|temper\\w*|heat\\w*|hot[estr]*\b|cold[estr]*\b)` captures discussions around global temperature changes.

Second, we excluded posts where the term "climate" appeared in non-environmental contexts, such as "political climate" or "climate control," using negative lookbehind assertions in our regular expressions. For instance, the expression `(?<!political\\w*)\bclimat` ensures the word "climate" is not preceded by "political" or related modifiers.

Detailed List and Rationale: For readers interested in the complete list of regular expressions and their specific purposes, we have included these details in the Supplementary Materials. This includes a comprehensive breakdown of all filters used to accurately curate our dataset, ensuring that only relevant discussions were analyzed.

We hope that this simplification improves understandability and, given that we provide the explicit regular expressions in the supplemental information, we hope that there is ample information for others to reproduce this work.

8. In the results, Figures 1 and 2 are presented without immediate commentary, leaving them devoid of context.

We have expanded the information provided for Figures 1 and 2 as follows:

Change:

Figure 1 shows the sentiment, subjectivity, and percentage of posts related to climate.

To:

Figure 1 shows that the frequency of usage of the term "global warming" decreased over time being preferentially replaced with the term "climate change" (bottom) even though the sentiment associated with "climate change" remained more negative while the subjectivity for both terms increased over time (top) with "climate change" being associated with more subjective text than "global warming" since 2007 (middle).

Change:

Figure 2 shows the relative popularity of climate-change-related topics in time.

To:

Figure 2 shows the relative frequency of climate-change-related topics in time as a percentage of discussions on Reddit. The discussion frequency has been trending down over the time period analyzed.

9. The concept of 'TextBlob subjectivity' is mentioned, yet a more detailed explanation of the analysis conducted is warranted.

We hope that the expanded explanation that addresses your point number 6 which included additional sections explaining the TextBlob Polarity and Subjectivity values addresses this point as well.

10. For Figure 2, an inquiry into the factors behind the spikes in % submissions around the years 2010, 2017, and 2019 should be made.

We have added the following:

Although we do not explore further the possible reasons for spikes in the number of conversations around 2010, 2018, and 2020, these present interesting opportunities for further investigation. For example, it may be that the spike around 2020 was driven by global news media covering the reduction in carbon emissions observed during the start of the COVID-19 pandemic.

11. The term 'subreddits' is used in a context that may imply understanding, yet no definition or explanation is provided within the text.

We hope that the added explanation in the expanded text addressing your point number 1 addresses this as well.

12. Figures 3 and 4 are presented with minimal detailed commentary on the observed phenomena, primarily being described rather than analyzed in depth.

That is true. While the observed phenomena are interesting and worth reporting, we do not see an immediate explanation or interpretation we could offer. However, we believe that they provide an important context for the study. It is also important and relevant to show the differences between different subreddits, which is important for our discussion, as well as to allow further, more detailed studies to follow. We have added a sentence highlighting how this contrasts with the expectations of political consultants from 2002 ((Burkeman, 2003). We have additionally added a sentence in the paragraph addressing Figure 4 suggesting more research may be warranted.

13. The observations made in the discussions section appear to extend beyond merely analyzing the terminological shift from "global warming" to "climate change." However, these findings are not emphasized within the article itself.

We made our best effort to interpret the results of our analysis of climate-related discussions. We have tried to keep a reasonable balance between interpretive observations and just

reporting data. In this revision, we have completely revised the discussion section to address this comment.

14. I miss a statistical/econometric analysis beyond the a simple description

There are additional statistical and econometric analyses that could be done, and we would encourage others to do so; however, we hope that the strength of this paper is in the size of the dataset: we're not relying on a sample of the dataset, but rather analyze it all, which is particularly useful for new and emerging phenomena. The paper reports the key immediately relevant observation of terminological shift, and the word difficulty across subreddits, based on a study on an unprecedented scale. We agree that additional statistical and econometric analyses will be useful, however, we also believe that the current ones, we've provided, are sufficiently interesting for the larger audience.

My objective has been to highlight areas for improvement in providing clarity and context for readers, particularly those unfamiliar with the platforms and analytical tools mentioned.

We greatly appreciate the time and care you have taken to review our paper.

Reviewer #2 (Remarks to the Author):

Dear authors and editors,

thank you very much for giving me the chance to review your paper in which you checked in how far discussions on climate change on the internet platform Reddit changed over time. I am certain that your paper is of high relevance to the scientific community and offers novel insights on an important topic. I have not worked with Reddit posts and large language data sets myself, hence I feel unable to comment on your methodological proceedings.

Thank you kindly for your comments.

The comments I have on your paper revolve around your discussion section and conclusion predominantly. I have made specific comments within the manuscript which I will upload to this platform for greater clarity. Hence, my comments here are more general.

Thank you for taking the time to comment directly on the manuscript. We will address each of your comments below.

I feel that your discussion could use a more sustained engagement with previous research on the topic of social media discussions of climate change/ global warming. To me it was somewhat surprising to see a discussion section without references to other people's work. I believe there is much potential to enhance your valuable contribution if you could help the reader to better understand in how far your findings align/ differ from other research. I understand that the majority of work has been conducted on debates on X (former Twitter) or Facebook and not Reddit but still it would be useful to alert the reader to potential differences/ similarities.

Thank you for pointing us to additional research which may be of interest and relevance to those interested in our research. We have a large number of references, and rewritten the discussion section and the conclusions radically.

Since we are focused on English only posts which are predominantly from North America and we are unable to ascertain how many non North American originating posts there are, we opted not further address non North American contexts.

Comments from the PDF:

1. "Maybe you could rephrase this a bit, as your findings are much more detailed than a simple more/less answer that you claim to be the purpose of your study?"

Thank you for the suggestion. We have changed:

Purpose: Our purpose was to check if people discuss climate change more or less proportionally to other topics. We collected and analyzed 16 years and 11.5 billion Reddit posts using sentiment analysis.

To:

Purpose: The aim of this study was to examine the evolution of discourse surrounding 'climate change' and 'global warming' over 16 years within 11.5 billion Reddit posts. By employing sentiment analysis and other computational linguistics techniques, we sought to identify shifts in discussion frequency, sentiment polarity, and linguistic complexity in relation to these environmental topics. Our analysis not only reveals a change of terminology used in online environmental discourse and offers insights into public perception and engagement with these critical issues over time.

2. "This seems a bit generic; could you specify?" (Lines 49-50)

Agreed. We have deleted the comment. (Line 56)

3. "Check language?"

Thank you for catching this. We have removed the offending item.

4. comment: "?" regarding a sentence fragment (Lines 155-157).

We updated the methods section to be more concise and hopefully easier to understand.

5. "Interesting! Can you unpack and discuss this finding a bit more? Why 2013? Administrative change in leading countries? New IPCC insights? COP meetings?"

We tried to find an "easy" way to download the IPCC reports from 2005 on, but unfortunately, for many of them we could not find the complete report as one PDF or document to download, but only a site where one has to download each subsection by going to many different pages. An initial analysis of the IPCC reports sections we did download suggests that "global warming" is either never or almost never used so far. Thus, we believe that this is not immediately informative for our study – although may be useful in future research.

6. "I think the discussion could do with more in-depth debates on your findings and previous research. I am missing more engagement with other studies on social media debates and the topic of climate change. Such a contextualisation would help the reader to better evaluate and understand the significance of your findings."

It would be wonderful if we could analyze other online platforms; however, at this time we do not have access to a complete or substantial portion of other forums such as X (Twitter) or Facebook to perform a compelling analysis to answer this question. It is worth noting that, unlike previously, with current EULA as well as anti-scraping regulations and tools it is also impossible for all practical purposes.

We have expanded the discussion section

"Please give more details/ examples on what you mean with 'positive' and 'negative' here. It is not immediately clear to the reader."

We have added a detailed explanation of the TextBlob and VADER sentiment and polarity scores and categories.

8. "Or is the discussion shifting to other online platforms?"

As discussed above, we feel this would be an excellent line of investigation which may be hampered by the changes to the EULAs of providers such as Facebook and Twitter/X. We have added the following to the conclusion:

Further investigation into how discussions about climate change are shifting across various online platforms could provide a more comprehensive understanding of the evolution of public discourse on this critical issue.

9. "How important is the US presidency here? Is it possible to georeference the Reddit posts to get a better understanding of where people are located who are taking part in these discussions?"

Although interesting questions, one could attempt to determine the effect of the US presidency on discourse; however, it is not immediately apparent how we might best approach this question. We therefore leave it for future research. With respect to using geolocation information, but we cannot do the latter since Reddit does not provide that information.

Lack of peaks does not suggest a significant effect; however, may be interesting for further research.

10. "Can you define 'difficult words'?"

We have added an expanded explanation of "difficult words":

TextStat's Difficult Words analysis is a tool used to quantify the complexity of vocabulary in a text. It evaluates the readability of content by counting words not found on a predefined list of commonly understood words, typically based on the Dale-Chall word list ([Dale and Chall 1948](#)). This metric is particularly useful in assessing the accessibility of text to audiences with varying levels of literacy. In our study, we employed TextStat Difficult Words analysis to measure the linguistic complexity of discussions within Reddit posts concerning 'climate

change' and 'global warming.' By applying this analysis, we aimed to determine the sophistication of language that could potentially influence the comprehensibility and reach of the discourse on these environmental issues. The insights gained help to understand how the complexity of language in online discussions may affect public engagement and awareness.

11. "This is certainly true for the US and some other countries; however, not so very true for others (Sweden, Denmark...). Maybe re-phrase and make the statement somewhat less Anglo-centric?"

Point well taken. We have modified:

Global environmental change is highly politicized and ideologized in the public discourse (Stecula and Merkley, 2019; Chinn, Hart and Soroka, 2020).

To:

Within North America, the global environmental change is highly politicized and ideologized in the public discourse (Stecula and Merkley, 2019; Chinn, Hart and Soroka, 2020).

Also added:

It is important to note that climate related topics may not have the same level of politicization in all regions of the world ([McCright et al. 2016](#))

12. "I am not sure if I am convinced by this final conclusion. You have acknowledged before that the downward trend in Reddit discussions on cc is likely due to the increase in discussions on other topics. Hence, it might be a bit hasty to declare a 'closing window' here?"

Thank you for pointing this out. We have removed the claim.

Abstract

In the study of 16 years and 11.5 billion Reddit posts, we show that “climate change” replaced “global warming” in 2013 in frequency. We also show that, more recently, the climate has become a less frequent topic of online conversations and searches. Moreover, and contrary to previous findings, we show that “global warming” is increasingly associated with more positive sentiment than “climate change.” Additionally, we discover that discussions about “global warming” use less subjective language than “climate change.” Our findings show that global environmental change is a complicated topic, and while it is being discussed in the media more often, the engagement of the general public is difficult to maintain. This result brings a dire warning to policymakers.

1. Introduction

Global environmental change has been discussed in the media for many decades, with varied issue attention (Hase et al., 2021) and different framing from country to country (Vu, Liu and Tran, 2019). Climate journalism has emerged and evolved as a field (Schäfer and Painter, 2021).

“Global warming” and “climate change” are terms that are sometimes used interchangeably (Shehata and Hopmann, 2012; Broadbent *et al.*, 2016). Yet, they do have different semantic contexts and connotations: “Global warming” refers to a long-term increase in global temperatures, while “climate change” denotes changes in climate patterns, in particular involving the increased levels of carbon dioxide in the atmosphere (Lineman *et al.*, 2015). ~~Both terms are highly politicized, though, and for a good reason.~~

“Climate change” has been used in academic papers since the fifties (Plass, 1956), but it has been more widely popularized in the public discourse by Frank Luntz, a Republican communication strategist, who in 2002 advised the Bush administration to use it instead of “global warming,” as a less frightening term (Burkeman, 2003), and to make the perceived lack of scientific certainty a vital issue in the debate about the environment. ~~Although we do not address how this intentional shift in nomenclature influenced the discourse prior to 2005 since the dataset begins in late 2005, this would be an interesting area for further research.~~

~~It made a political sense:~~ even though for liberals, “global warming” and “climate change” are associated equally with rising temperatures and, overall, the heat impact on the planet, conservatives see “climate change” as more disconnected from it (Schuldt and Roh, 2014). It is also worth noting that the US media, in particular, are prone to relativization of the topic and frequently evoke scientific uncertainty and controversy regarding human-induced climate change (Shehata and Hopmann, 2012). Whether the Republican plan was successful or not, over the years, a noticeable shift in vocabulary from “global warming” to “climate change” in the popular press and academic papers alike occurred (Zehr, 2000; Sonnett, 2022).

The social perception of these terms on the Internet has been studied, e.g. by Google Trends (Jun, Yoo and Choi, 2018; Sisco and Weber, 2022). There have been also many

studies of the perception of climate change and environmental challenges on the social media, with a heavy bias towards Twitter (Lineman *et al.*, 2015; Pearce *et al.*, 2019; Ganczewski and Jemielniak, 2022), possibly because of the ease of data access.

However, it has not been a subject of a systematic review of large-scale online discussions. It is essential, though, as online discussion fora provide a platform for contrasting arguments and reasoning, in contrast to Twitter, which relies on short, linguistically-restricted statements, and in contrast to just Google Trends, which only reflects people's search interests. Online discussions are also a good proxy for general public interest and perceptions, and may help to understand the emergence of public spheres in climate-related topics (Neff and Jemielniak, 2022).

Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends (Fariello, Jemielniak and Sulkowski, 2021; Gaudette *et al.*, 2021). As of 2023, Reddit, with its unique structure centered around interest-based communities known as "subreddits," hosts over 850 million monthly active users (MAUs) (Curry, 2024) and 2.1 billion unique visitors (Tiano Bianchi, 2024), engaging in in-depth discussions and sharing a wide array of content. Notably, it is the 5th most visited social media platform globally (Dixon, 2024), with approximately 48% of visits (Tiago Bianchi, Feb 29, 2024) originating from the United States. Reddit represents approximately one third as much social media traffic as Twitter/X which serves as a hub for real-time updates and public discourse (Dixon, 2024). Although Facebook remains the dominant force with just over 3 billion monthly active users, catering to a broad demographic by facilitating personal connections and diverse content consumption of which approximately 24% are based in India and 12% are based in the United States (Dean, 2024), Reddit's model of specialized, user-moderated subreddits is particularly resonant among specific interest groups, steadily growing in its impact on public opinion and media trends.

In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.

~~Reddit is the largest and long-running platform for discussions in English, and using it allows for analyzing critical social changes and trends (Fariello, Jemielniak and Sulkowski, 2021; Gaudette *et al.*, 2021). In our study, we perform the first large-scale analysis of 11.5 billion posts in regard to climate change and global warming.~~

2. Material and methods

2.1 Data collection

We obtained a large dataset ($n=11.486 \times 10^9$) containing Reddit.com posts archived from the PushShift.io data repository (Baumgartner *et al.* 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files). Submissions are the top-most post that initiates a discussion within a group known as a subreddit while comments are replies to either a submission or another comment.~~We obtained a large dataset ($n=11.486 \times 10^9$) containing~~

~~Reddit.com posts archived from the PushShift.io data repository (Baumgartner et al. 2020) for the years 2005 through June 2021 for both Reddit Submissions (RS files) and Reddit Comments (RC files).~~ We uncompressed each monthly file, and removed duplicates. We then used a series of regular expressions to identify all posts related to climate change while ignoring all instances of political, economic, financial, academic, work, social, job, or employment climate, as well as terms such as “climate control” used to refer to topics not related to climate change while still retaining all posts that contained explicit references to climate change, global warming, and related climate specific topics such as references to the Paris Accord, the Kyoto Protocol, Greta Thunberg, and others. The complete list and order of expressions used is included in the supplemental materials.

This resulted in 15.341×10^6 posts likely to be related to climate change (0.134%). These posts were then converted from JSON format to a more data science friendly CSV format retaining all the relevant information on which we performed language detection using both the polyglot (Al-Rfou, 2015) and langdetect Python libraries, keeping only posts that were determined to be in English by both ($n=1.501 \times 10^6$, 97.9% of matched posts). We subsequently concatenated and lemmatized the title and body of all the remaining posts using standard NLTK library methods (nltk.stem.WordNetLemmatizer) and removed all English stop words (Bird, Klein and Loper, 2009). Each post was then analyzed using the VaderSentiment (Hutto and Gilbert, 2014), TextBlob sentiment, TextBlob naïve Bayes, and textstat python libraries among several other interesting metrics such as readability and language complexity scores provided in the supplemental data.

VADER (Valence Aware Dictionary and sEntiment Reasoner): A lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of a sentiment lexicon and a set of heuristic rules to estimate sentiment scores for textual data (Hutto & Gilbert, 2014).

VADER's rule-based approach, which takes into account both the semantic meaning of words and the context in which they are used (including punctuation, capitalization, and modifiers), makes it highly effective for the nuanced understanding of sentiment in social media posts. This approach provides a robust framework for analyzing how sentiment regarding climate change and global warming is expressed in online discussions, contributing significantly to the study's findings on public engagement and perception related to these topics.

- **Sentiment Score:** VADER analyzes the text to calculate a compound score that indicates the overall sentiment. This compound score is a normalized, weighted composite score that ranges from -1 (extremely negative) to 1 (extremely positive). This scoring is beneficial in understanding the general sentiment of each post regarding climate change or global warming.

TextBlob: Provides a simple API for common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. In this study, we utilized TextBlob's sentiment analysis features to assess the polarity and subjectivity of the posts.

- **Polarity:** This refers to the emotional orientation of the words used in the text. Polarity is a float that lies between -1.0 and 1.0, where -1.0 indicates a highly

negative sentiment, 0 indicates neutrality, and 1.0 indicates a highly positive sentiment. It essentially measures the degree to which the language used in a text leans towards a positive or negative emotion. This metric is crucial for understanding the general sentiment expressed in discussions about climate change and global warming on social media platforms.

- **Subjectivity:** This measures the amount of personal opinion and factual information contained in the text. The value ranges from 0 to 1, where 0 is very objective and 1 is highly subjective. Subjective text expresses personal feelings, views, or beliefs, whereas objective text provides factual information. Analyzing subjectivity helps in distinguishing between posts that are based on personal opinions and those based on objective reports, which is vital for evaluating how discussions on climate change are framed by the public.

In our analysis, these metrics enabled a deeper understanding of the emotional and subjective layers within the climate change discourse on Reddit, providing insights into how people emotionally engage with the topic and the extent to which discussions are opinion-driven or fact-based (Loria, 2020).

TextStat's Difficult Words analysis is a tool used to quantify the complexity of vocabulary in a text. It evaluates the readability of content by counting words not found on a predefined list of commonly understood words, typically based on the Dale-Chall word list ([Dale and Chall 1948](#)). This metric is particularly useful in assessing the accessibility of text to audiences with varying levels of literacy. In our study, we employed TextStat Difficult Words analysis to measure the linguistic complexity of discussions within Reddit posts concerning 'climate change' and 'global warming.' By applying this analysis, we aimed to determine the sophistication of language that could potentially influence the comprehensibility and reach of the discourse on these environmental issues. The insights gained help to understand how the complexity of language in online discussions may affect public engagement and awareness.

We then analyzed the unigram, bigrams, and trigrams and subsequently analyzing the occurrences of the “climate” unigram when not followed by “change” and the “climate change” bigram as well as the “global” unigram when not followed by “warming” and the “global warming” bigram. We accomplished this by replacing all permutations of occurrences of “climate change” with “climate-change” and “global warming” with “global-warming” after lemmatization and performing a unigram analysis. For each of the resultant unigrams, we analyzed sentiment, polarity, and readability subdivided by the top 40 subreddit, and looked at the relative trends for the unigrams across each year.

2.2 Selecting Climate Related Posts

To ensure our analysis focused solely on content relevant to climate change, we developed a set of criteria to filter posts from the PushShift.io data repository. This involved using

regular expressions to select posts that specifically mentioned climate change-related terms and exclude irrelevant contexts.

Methodology Summary: Our approach was twofold. First, we included posts that contained explicit references to climate-related topics, such as "global warming," "carbon emissions," and specific environmental policies (e.g., the Paris Accord). For example, one of our expressions

`\\bglobal\\w*[\\W_]*(?:warm\\w*|cool\\w*|temper\\w*|heat\\w*|hot[estr]*\\b|cold[estr]*\\b)`

captures discussions around global temperature changes.

Second, we excluded posts where the term "climate" appeared in non-environmental contexts, such as "political climate" or "climate control," using negative lookbehind assertions in our regular expressions. For instance, the expression `(?<\\bpolitical\\w*)\\bclimat` ensures the word "climate" is not preceded by "political" or related modifiers.

For readers interested in the complete list of regular expressions and their specific purposes, we have included these details in the Supplementary Materials. This includes a comprehensive breakdown of all filters used to accurately curate our dataset, ensuring that only relevant discussions were analyzed.

~~On the original JSONL formatted data downloaded from PushShift.io, we extract all posts whose title text and body text, when concatenated together, matched any one of the following regular expressions:~~

- ~~● \\bclimat(?:c|olog|arian|ard|is|mt|caust|sci)w*\\bclimistw*||~~
- ~~● \\b(?:globalw*|environw*|wea?therw*|oceans?|seas?|atmosw*|historiow*)[\\W_]*(?:warmw*|coolw*|temperw*|heatw*|hot[estr]*\\b|cold[estr]*\\b)||~~
- ~~● \\b(?:oceans?|seas?)[\\W_]*(?:levelw*|temperw*|acidw*|ris[oi]w*|heat[ingswave]*|ph\\b)||~~
- ~~● \\b(?:levelw*|temperw*|acidw*|ris[ic]w*|heat[ingswave]*|ph\\b)w*[\\W_]*(?:oceans?|seas?)||~~
- ~~● \\b(?:green[estr]?|alternw*|renewabw*|wind|solar|hydro(?:xide|ponic|phobi|fobi)w*|geo[\\W_]*therw*|nuolearw*|nuocularw*|clean[estr]*)[\\W_]*energ\\w*||~~
- ~~● \\b(?:green[estr]?|wind|solar|hydrow*|geotherw*|clean[estr]*)[\\W_]*(?:techw*|solu\\w*)||~~
- ~~● \\bclean[estr][\\W_]*(?:techw*|energ\\w*)||~~
- ~~● \\bgreen[\\W_]*housew*[\\W_]*(?:gasw*|effectw*)||~~
- ~~● \\bcarbon[\\W_]*(?:neutr|emi[st]|diox|foot)w*||~~
- ~~● \\bco2\\b||~~
- ~~● \\b(?:emitw*|producw*|contribw*)[\\W_]*carbon||~~
- ~~● \\b(?:polar|are?tic(?:al|e)w*|ant[\\W_]*are?ticw*)[\\W_]*(?:icew*|shelfw*|temperw*|warmw*|coolw*)||~~
- ~~● \\bthu[mn][mn]?b[eu]rg\\b||~~
- ~~● \\bipee\\b||~~
- ~~● \\b(?:montreal|kyoto)[\\W_]*protow*\\b||~~
- ~~● \\bice[\\W_]*(?:caps?\\b|shelfves*|sheets?)\\b||~~
- ~~● \\bfossil[\\W_]*fuels?\\b||~~
- ~~● \\b[dr]c[_]*forestw*||~~

- ~~\brenewables\b¶~~
- ~~\bwind[\W_]*turbine[s]*\b¶~~
- ~~\bparis[\W_]*(?:accord|agreement)\w*¶~~

~~And all of the following regular expressions:¶~~

- ~~(?<!\politico\w*)\belimat¶~~
- ~~(?<!\econom\w*)\belimat¶~~
- ~~(?<!\financ\w*)\belimat¶~~
- ~~(?<!\academ\w*)\belimat¶~~
- ~~(?<!\work\w*)\belimat¶~~
- ~~(?<!\social\w*)\belimat¶~~
- ~~(?<!\job\w*)\belimat¶~~
- ~~(?<!\employ[me]\w*)\belimat¶~~
- ~~\belimat\w*\W*(?!control)¶~~

~~The former regular expressions include any posts with text tightly associated with climate change related topics, as well as bigrams such as “political climate” or “climate control” which refer to topics that are not specific to climate change while. The latter regular expressions ensure that any items that did match “climate” did so because climate appeared at least once not preceded by words such as “political,” “financial,” “job,” or “academic” and not followed by “control.”~~

3. Results

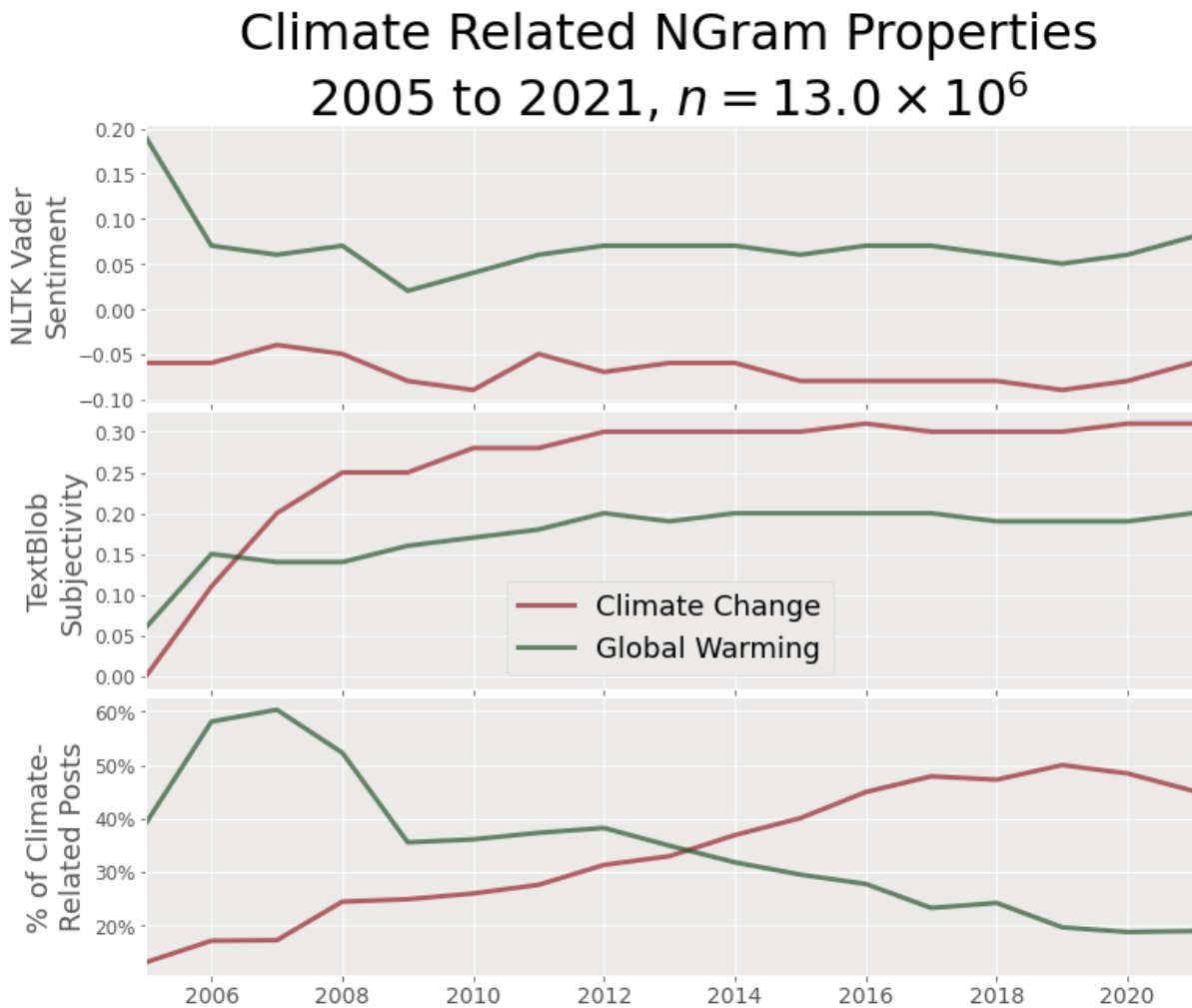


Figure 1: The sentiment, subjectivity, and percentage of posts related to climate

Figure 1 shows that the frequency of usage of the term “global warming” decreased over time being preferentially replaced with the term “climate change” (bottom) even though the sentiment associated with “climate change” remained more negative while the subjectivity for both terms increased over time (top) with “climate change” being associated with more subjective text than “global warming” since 2007 (middle).

~~Figure 1 shows the sentiment, subjectivity, and percentage of posts related to climate.~~

Percentage of All Posts Matching Climate-Change Related Terms from 2005 to 2021, $n = 11.5 \times 10^9$

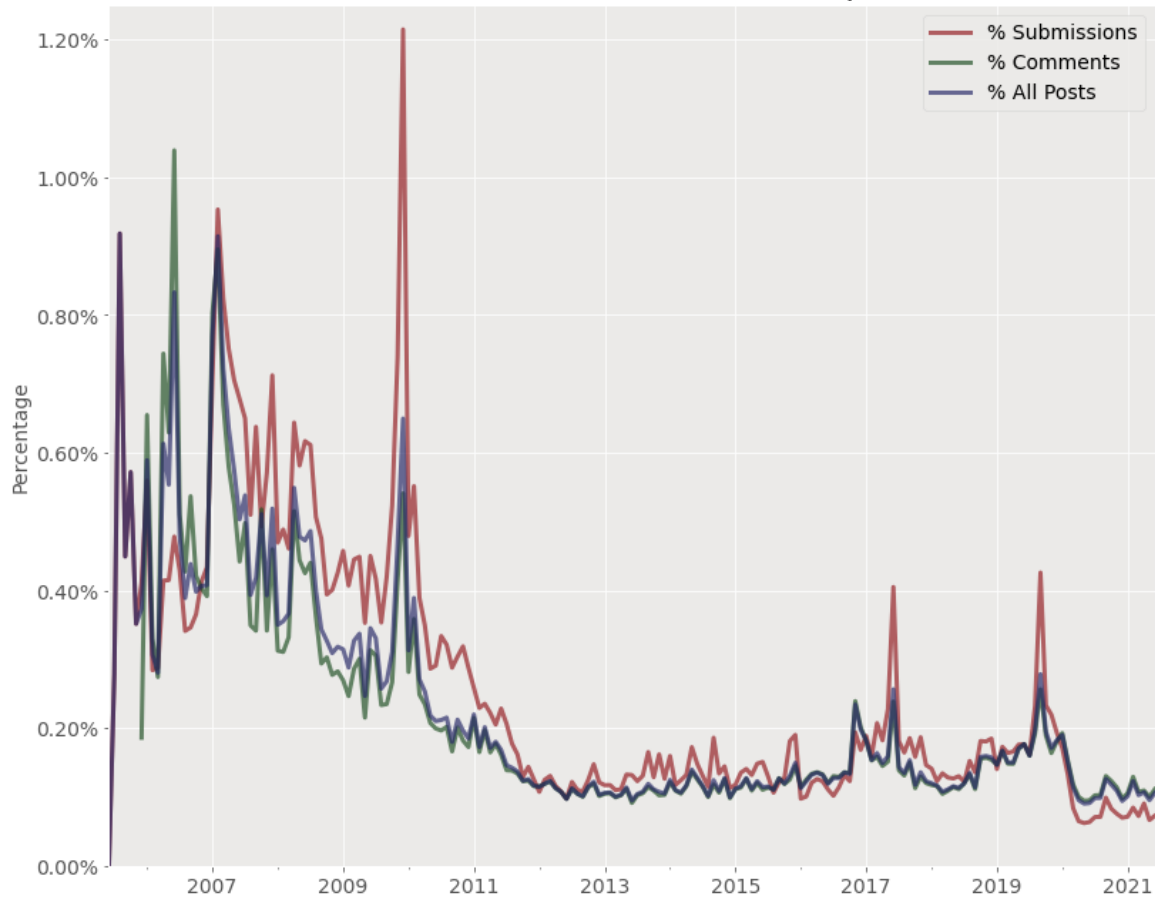


Figure 2: Percentage of all posts related to climate change

Figure 2 shows the relative frequency of climate-change-related topics in time as a percentage of discussions on Reddit. The discussion frequency has been trending down over the time period analyzed. ~~Figure 2 shows the relative popularity of climate-change-related topics in time.~~

Mean NLTK Vader Composite Score by Subreddit for All Dates

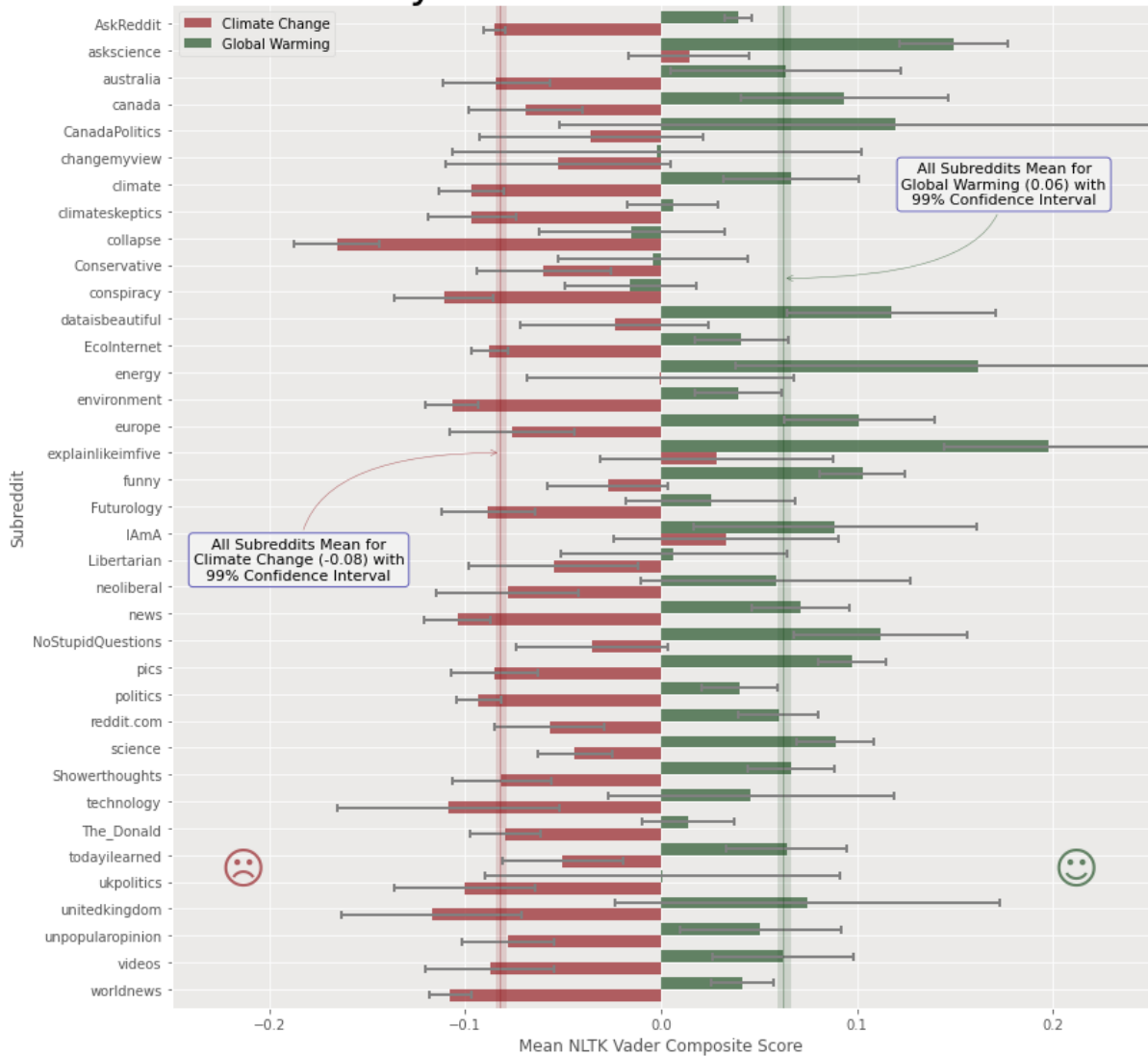


Figure 3: Mean NLTK Vader Composite Scores by Subreddit

Figure 3 depicts the mean NLTK Vader Sentiment Composite Score for climate-related posts containing the terms “climate change” or “global warming” across all subreddits, which are plotted as vertical lines with the shaded area representing the 99% confidence interval while the bars represent the mean for each of these for a given subreddit with the error bars showing the 99% confidence interval for that given subreddit and term. The subreddits (37 total) were selected by choosing the 50 highest volume subreddits with more than 100 climate-related posts.

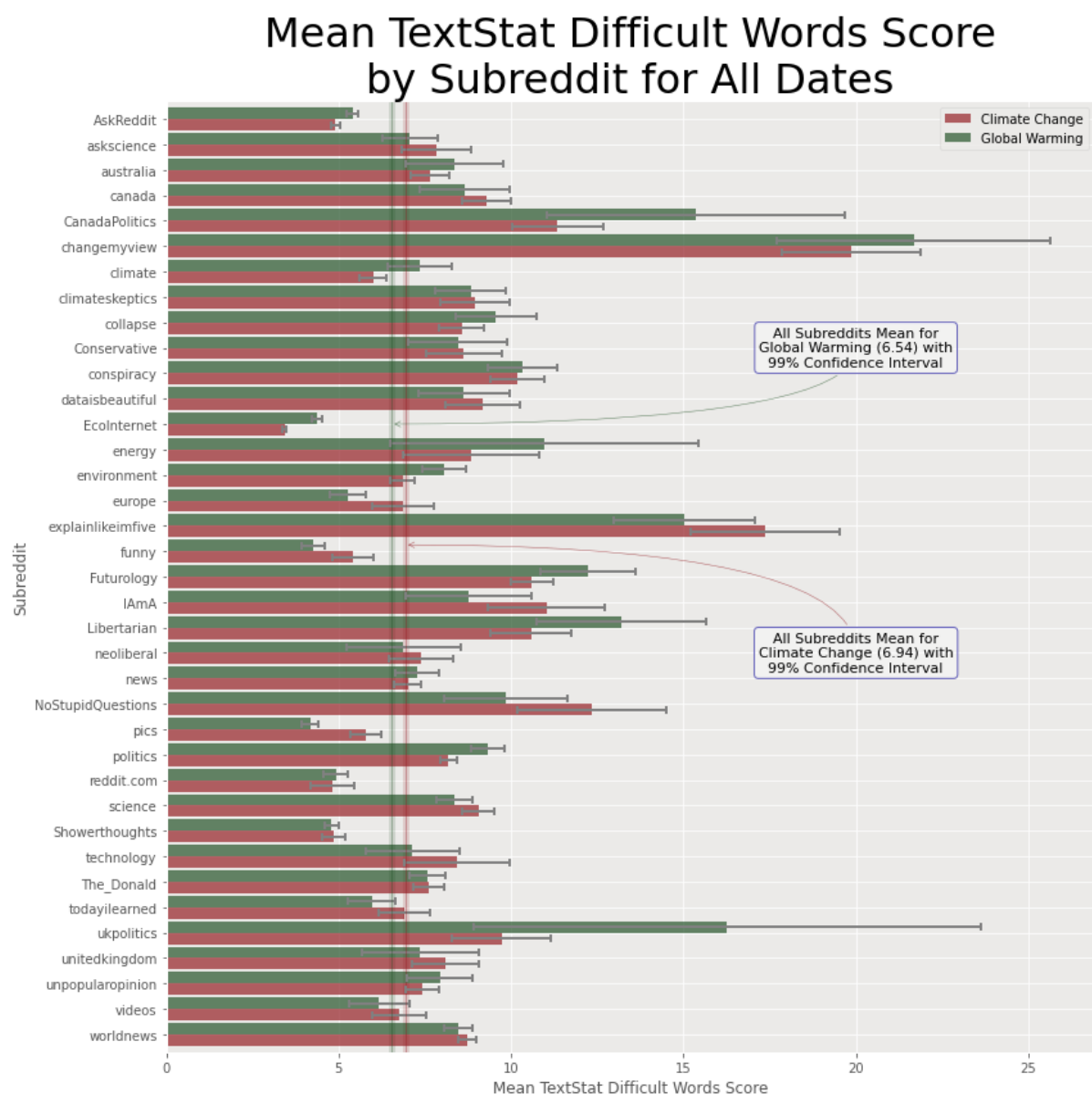


Figure 4: Mean TextStat Difficult Words Scores by Subreddit

In Figure 4, the mean TextStat Difficult Words score for all posts containing the terms “climate change” and “global warming” across all subreddits are plotted as vertical lines with the shaded area representing the 99% confidence interval while the bars represent the mean for each of these for a given subreddit with the error bars showing the 99% confidence interval for that given subreddit and term. The subreddits (37 total) were selected by choosing the 50 highest volume subreddits with more than 100 climate-related posts.

In the supplemental materials, we provide the raw analytical data, which includes calculated values calculated for a variety of text analyses, allowing for considerable additional data exploration.

Google Search Trend Scores from 2004 to 2022 for Climate Change and Global Warming 6-Month Centered Rolling Average

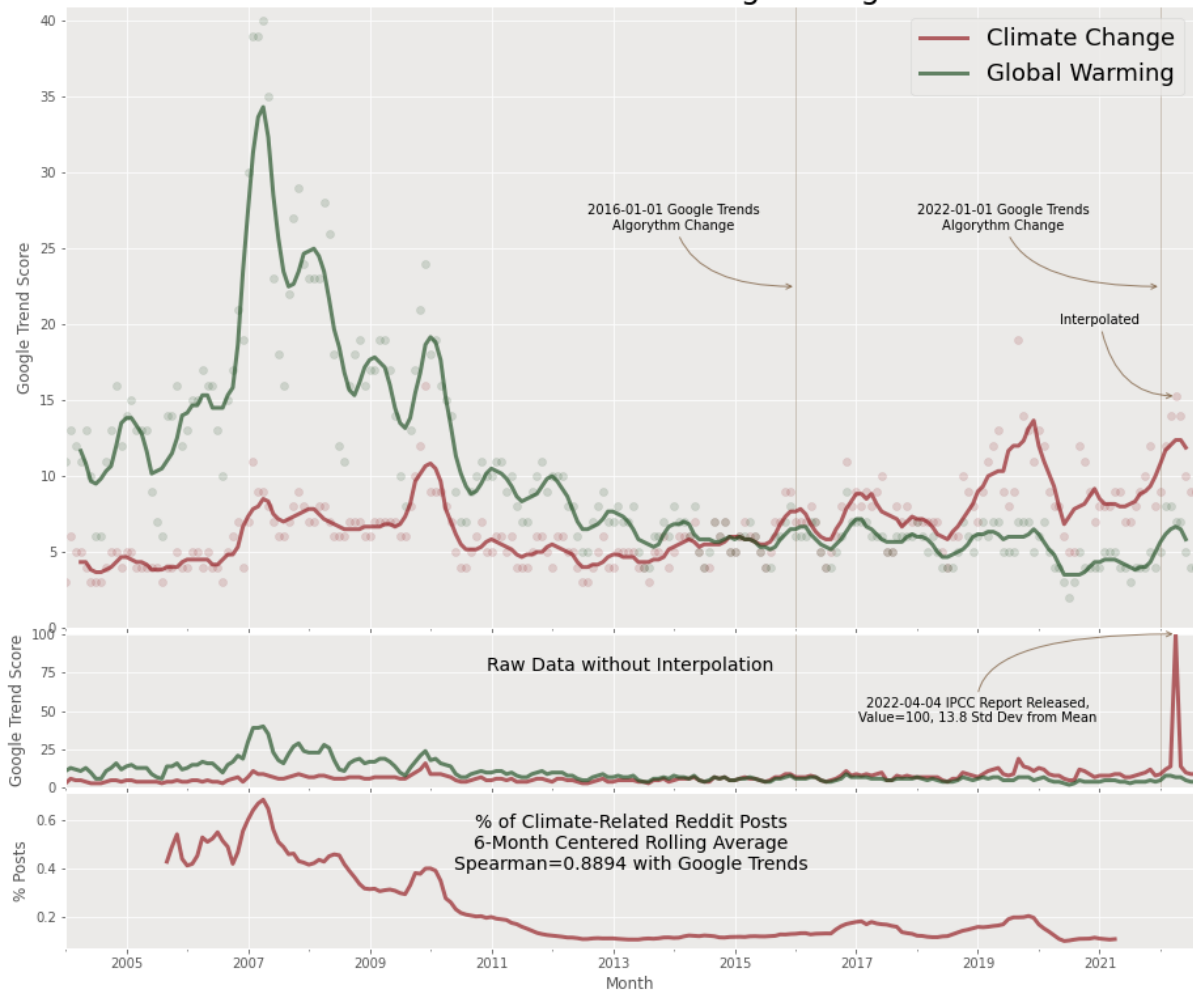


Figure 5: Google Search Trends

Figure 5 shows the Google search term trends for “climate change” and “global warming” (without quotes) for the 18 years from January 2004 to August 2022. Vertical lines represent the dates on which Google implemented algorithmic changes to how it tracks searches. In April 2022, the month during which the IPCC released its most recent and most alarming report, there was a large spike in searches reported; however, we are unable to ascertain if the two are related.

5. Discussion

Our analysis confirms the lexical transition from “global warming” to “climate change” indeed took place, albeit a decade later than the Republican initiative purportedly instigating this shift. This change was abrupt, with “climate change” becoming the dominant term in the discussions analyzed around 2013, as illustrated in Figure 1. This shift in terminology is further corroborated by Google search trends that reveal a parallel change occurring around

two years later, as shown in Figure 5. These findings suggest a broad adoption of the term "climate change" in both digital discourse and search behaviors during this period, reflecting broader societal and political shifts.

Furthermore, extending our investigation into Google Trends data from 2004 to 2021, well beyond those previously analyzed in studies focusing on searches before 2014 (Leiserowitz et al., 2014; Lineman et al., 2015), we found a similar shift in terminology from "climate change" to "global warming" around 2015 beyond Reddit. This alignment between Reddit discussions and Google search trends, despite the slight difference in timing, supports the consistency of this lexical evolution across various Internet platforms. Additionally, the six-month rolling averages for the sum of the two Google trend lines, which approximately represent the overall climate-related searches over time, show a strong correlation (Spearman correlation coefficient of 0.89) with the percentage of Reddit discussions about climate-related topics. This correlation further suggests that the trends observed in Reddit discussions are reflective of broader trends across Internet-based media, underscoring the pervasive shift in how climate issues are discussed online.

Although some prior studies viewed the terms somewhat interchangeably in traditional mass media from 1995-2004 (Boykoff, 2008), more recent studies suggest important differences. For example, one study found that the term "global warming" elicited lower levels of belief in the phenomenon, especially among Republicans, while the term "climate change" elicited higher levels of belief, particularly among the same group (Schuldt et al., 2011). Another study found that "red states" (i.e. more conservative states in the United States) tended to prefer the term "global warming" while "blue states" (less conservative) tended to prefer the term "climate change" (Jang & Hart, 2015). More recently, one study found that on twitter "global warming" was more closely associated with human causation and conveyed more of an immediate and direct impact, whereas "climate change" appeared to be a broader term that could be framed as a natural phenomenon, potentially downplaying the necessity for human intervention (Fownes et al., 2018). Another study found that tweets mentioning "climate change" were more likely to be associated with environmental and political content, whereas tweets using "global warming" were more likely to be linked to weather (including heatwaves) and energy, and the authors suggested that terminology may influence public perception and engagement (Pearce et al., 2019). A 2014 survey of 1,657 adults found that people reported using and hearing "global warming" far more often than "climate change" - 45% of the time vs. 12% and 35% vs. 16% respectively (Leiserowitz et al., 2014); although the term preferences are consistent with our study's observed shift in Reddit and Google Search Trends around that time, the significant difference between the perceived frequencies and those observed suggests there may be important underlying factors that require further investigation.

The analysis of the TextBlob subjectivity scores revealed a significantly higher level of subjectivity in posts pertaining to "climate change," (Figure 1) suggesting that individuals discussing this term tend to express their perspectives in a less objective manner. Conversely, discussions labeled under "global warming" were characterized by a tone suggesting objective factuality. This distinction may imply that contributors discussing "climate change" are potentially more vigilant in differentiating between opinion and fact. Such an observation is aligned with expectations and underscores a critical discourse dynamic: it appears that those referring to "global warming" may be more likely to perceive

their viewpoints as unequivocal truths. This distinction in discourse styles not only illuminates differing perceptions but also potentially influences the public and policy discourse surrounding climate issues. In short, the choice of terms may significantly impact how the underlying phenomenon has been received.

Our analysis challenges prevailing assumptions by demonstrating that the term "global warming" is associated with more positive sentiment compared to "climate change." An average positive sentiment was consistently observed over all time periods analyzed for "global warming," whereas "climate change" corresponded with predominantly negative scores (Figure 1). This disparity may suggest that individuals discussing "global warming" are less apprehensive about climatic changes than those using the term "climate change," possibly reflecting a delay in adopting the newer terminology. Notably, the subreddit "askscience," known for its scientific rigor, displayed the most positive sentiments for both terms (found in supplemental data) and had some of the highest composite scores (Figure 3). Contrasting with a 2015 study (Lineman et al., 2015), which utilized Semantria® to analyze Twitter sentiments over two months, our broader assessment over 15 years using VADER sentiment scores uncovered an opposing trend, aligning more closely with sentiment self-report survey results from 2014 (Leiserowitz et al., 2014). This incongruence underscores the need for further exploration into the dynamics shaping sentiment. It is notable that during Donald Trump's presidency, there was a modest decrease in both sentiment polarity and subjectivity within discussions of "climate change" and "global warming." Contributions during this period trended towards more negative tones and were phrased more as declarations of fact than expressions of personal viewpoints. The underlying reasons for this shift in discourse style remain unclear and warrant further investigation.

As illustrated in Figure 3, the term "climate change" consistently registers less favorable Vader sentiment scores compared to "global warming" across various subreddits, regardless of the spectrum of views on climate validity presented by these forums (such as The Conservative, The_Donald, and conspiracy versus climate, askscience, and science). This pattern indicates that the more positive sentiment typically associated with "global warming" transcends political and scientific divides. Notably, subreddits dedicated to educational content, including askscience, explainlikeimfive, and IAmA, tend to adopt more positive tones in their discussions. This consistency suggests a broad, enduring perception gap between the two terms.

We observed that posts containing the terms "climate change" exhibit a small but statistically significant increase in the use of complex vocabulary across all subreddits. However, this trend does not consistently manifest within individual subreddits or across different forums (refer to Figure 4). The subreddit aimed at changing the original poster's (OP) views displayed the highest usage of complex words in discussions related to climate topics. Paradoxically, the subreddit "explainlikeimfive," which is intended to simplify complex subjects for easier understanding, actually had the second-highest average use of complex vocabulary in its posts.

When considering the proportion of Reddit posts addressing climate-related topics, a noticeable decline is observed. From 2005 to 2010, up to 1.4% of discussions were climate-related. However, this percentage has since shown a steady decline, ultimately

averaging around 0.1% of posts by 2021 (Figure 2). This apparent decrease may be somewhat misleading due to the overall growth in Reddit's popularity, which could disproportionately affect the perceived share of climate discussions. In reality, while the absolute number of climate-related discussions has continued to increase, the expansion of conversations on other topics has outpaced them. Despite the actual increase in climate discussions, this trend suggests a relative decrease in engagement with climate issues within the Reddit community, serving as a subtle indicator that the topic may be diminishing in proportional popularity.

As an additional observation, our analysis of word difficulty revealed that the "ecointernet" subreddit utilized the least complex language in discussions pertaining to "climate change," "global warming," and climate topics in general. Conversely, the "Changemyview" subreddit, which often hosts debates and confrontational dialogues, exhibited the most complex sentence structures, likely due to a reliance on academic sources and technical language. Surprisingly, the "Explain to Me Like I'm 5" thread, which aims to simplify concepts, employed more complex vocabulary than other forums. Additional data and visualizations supporting this finding are provided in the supplementary materials.

Our findings deepen the existing literature by offering a comprehensive analysis of Reddit discussions, contrasted against broader search term trends, thereby enhancing our understanding of the evolution of climate discourse across diverse online platforms. ~~Our first conclusion is that the shift from "global warming" to "climate change" indeed happened, but a decade later than the Republican initiative. Curiously, it was relatively abrupt, and around 2013, "climate change" overcame "global warming" in general discussions we analyzed (see Figure 1). It is also notable that Google search trends (Figure 5) show a similar switch in terminology occurring only two years later. ¶¶~~

~~Our analysis found that, contrary to previous assumptions, "global warming" is associated with more positive sentiment polarity than "climate change." Across all subreddits, the sentiment of posts using the term "global warming" was positive, while the sentiment of posts using the term "climate change" was negative. This result could indicate that those writing about "global warming" are less concerned about the climactic changes than those naming it "climate change," which could be correlated with the slow uptake of new terminology. Interestingly, the subreddit "askscience," which focuses on a scientific approach to the world, had the most positive sentiment overall for posts using either "climate change" or "global warming." ¶¶~~

~~The subjectivity of posts about "climate change" was much higher, meaning that people writing about it were expressing their views in a less authoritative tone, while people writing about "global warming" were phrasing their posts as if they were describing objective facts. This result could indicate that those writing about "climate change" are more conscious about separating opinions from facts and is, generally, not surprising. Still, this finding indicates that those who write about "global warming" may see their views as non-negotiable truths. ¶¶~~

~~It is worth mentioning that during Donald Trump's presidency, there was a small drop in both sentiment polarity and subjectivity—people were writing about "climate change" and "global warming" more negatively, and in language more typical for a description of truth rather than viewpoints. We do not have a good explanation for this phenomenon. ¶¶~~

Moreover, it is also quite surprising that the discussions about “climate” in general have not increased significantly since 2009, and they remain at a stable percentage of overall posts. Only the last four years note an increase from about 20% to nearly 30%. They appear to be nearly perfectly correlated with “global warming” conversations—the more people talk about one, the less they do about the other (see Figure 2). Although we do not explore further the possible reasons for spikes in the number of conversations around 2010, 2018, and 2020, these present interesting opportunities for further investigation. For example, it may be that the spike around 2020 was driven by global news media covering the reduction in carbon emissions observed during the start of the COVID-19 pandemic.¶¶

When the percentage of Reddit posts with climate-related topics is considered, we can even observe an evident decline. While 2005-2010 even up to 1.4% of topics were related to climate, after that period the numbers steadily decreased. It may be misleading, though: since we are measuring the proportions of topics, the overall growth of popularity of Reddit is likely largely contributing to the observed phenomenon. In other words, the absolute numbers of climate-related discussions have not decreased (in fact, they increased steadily), but the growth of discussions on other topics was more rapid. Nevertheless, we believe that this result is a mild warning of the fact that climate is becoming a less and less popular topic proportionally.¶¶

On a side note, we observe that, considering the word difficulty, the “ecointernet” subreddit had the least difficult words in sentences related to “climate change”, “global warming,” and “climate” in general. “Changemyview” subreddit, on the other hand, had the most difficult sentences—it may be that this subreddit, dedicated to confrontational discussions, relied more largely on sources and academic language. One surprising finding from our study is that “Explain to me like I’m 5” thread uses more complex words than other threads. In the supplement, we provide additional figures, illustrating this phenomenon.¶¶

As seen in Figure 3, the terms “climate change” appear to be associated very consistently with VADERVader sentiment scores that are less favorable than those of “global warming” irrespective of the subreddit, even when the subreddits vary greatly on the validity of climate-related claims (e.g., the Conservative, The_Donald, and conspiracy versus climate, askscience, and science), suggesting that the relatively more positive sentiment association with “global warming” may persist across political views and scientific literacy. Unsurprisingly, several subreddits with a stronger focus on educating people tended to use more positive tones (e.g., askscience, explainlikeimfive, IAmA). It is worth noting that this is contrary to the prediction of Bush-era political consultants (Burkeman, 2003).¶¶

We note that posts with the terms “climate change” have a small but very statistically significant increase in the use of difficult words when scores are aggregated across all subreddits, but the trend does not hold well within and across subreddits (see Figure 4). The subreddit dedicated to attempting to change the original poster’s (OP) views had the highest frequency of difficult words with respect to climate-related posts. It is interesting to note the considerable irony that the subreddit ostensibly meant to explain challenging topics using simple terminology, explainlikeimfive, had, on average, the second-highest frequency of use of difficult words. Additional research into the influence of the use of more sophisticated language on sentiment, the ability to change people’s opinions, or the ability to explain complex topics such as climate change may be warranted.¶¶

The most interesting findings are that the Google search trends (Figure 5) demonstrate a similar change in frequency of the terms from “climate change” in the searches before 2015 to “global warming” to those after. Although this occurred at a later date than that which we observe in the Reddit discussions circa 2013, it appears to be a consistent feature across two disparate Internet activities. Furthermore, the 6-month rolling averages for the sum of the two Google trend lines, which roughly approximates the overall climate-related searches over time, is strongly correlated with the percentage of Reddit climate-related discussions with a Spearman correlation coefficient of 0.89 further suggesting that the trends observed in Reddit discussions are reflective of broader trends across at least some Internet-based media.¶

In addition to highlighting novel findings associated with the terms evaluated with respect to sentiment, polarity, readability, and preferred nomenclature, our findings provide additional context to previous findings concerning climate-related topics in online social media (Pearce et al. 2019; Fownes et al. 2018; Hautea et al. 2021; Oz et al. 2018). We would like to encourage those with access to relevant Facebook, Twitter/X, or other social media to compare those data with our findings.

6. Conclusions

Global environmental change is a highly politicized and ideologically charged topic within public discourse (Chinn et al., 2020; Stecula & Merkley, 2019). The mainstream media have been systematically covering this issue for years, often disproportionately highlighting contrarian views over expert opinions (Petersen et al., 2020). Despite extreme weather events drawing significant media attention (Pianta & Sisco, 2020), overall media coverage on climate change has grown (Barkemeyer et al., 2017; Layna & Altamirano, 2022). Moreover, even the more conservative outlets have started acknowledging the burning problem (Schmid-Petri et al., 2017). It is important to note that climate-related topics may not have the same level of politicization in all regions of the world (McCright et al., 2016).

Unlike the growing media attention, our analysis of large-scale online discussions reveals a different trend. Over 16 years of Reddit posts show a decreasing proportional interest in climate-related topics. While the absolute number of climate discussions has increased, the expansion of other topics on Reddit has outpaced it, suggesting a relative decline in engagement with climate issues. This trend might be attributed to various factors, including topic fatigue, competition from other crises (Stoddart et al., 2021), and the emergence of other engaging topics.

Our study highlights significant linguistic and sentiment differences between the terms “global warming” and “climate change.” Discussions about “global warming” tend to use less sophisticated language and exhibit more positive sentiment compared to those about “climate change.” This finding suggests that the term “global warming” is perceived less negatively and is communicated in a more accessible manner. This linguistic distinction is crucial as it can influence public perception and engagement with the topic.

Furthermore, our analysis indicates that online discussions precede Google searches in reflecting the language shift from "global warming" to "climate change." This suggests that social media platforms like Reddit can serve as early indicators of changing public attention and discourse. Monitoring these platforms may provide more immediate insights into shifts in public interest and sentiment than traditional search trend analyses.

Our findings have significant implications for policymakers, communicators, and climate advocates. The decreasing proportional interest in climate discussions on Reddit highlights the challenge of maintaining public engagement on this critical issue. The language used to discuss climate change can affect how the topic is perceived and engaged with by the public. Therefore, strategic communication efforts should consider these linguistic nuances to effectively mobilize public interest and action.

Moreover, our study underscores the importance of addressing the potential gap between media coverage and public engagement. While the media continues to amplify climate issues, translating this attention into sustained public concern and action remains a challenge. Policymakers and communicators must develop innovative strategies to reinvigorate public interest and participation in climate discussions, ensuring that the urgency of global environmental change is not lost amid other competing issues.

In conclusion, the window for mobilizing public attention around global environmental change may be closing. The observed trends suggest that informal discussions on social media platforms significantly influence the broader debate among non-experts. To address the climate crisis effectively, it may be imperative to leverage these platforms for proactive and engaging communication strategies. Understanding and harnessing the dynamics of online discourse can play a crucial role in shaping public perception and driving collective action towards mitigating global environmental change.

~~Within North America, the global environmental change is highly politicized and ideologized in the public discourse. Global environmental change is highly politicized and ideologized in the public discourse (Stecula and Merkley, 2019; Ghinn, Hart and Soroka, 2020). The mainstream media have been covering it systematically for years. Even though disproportionate attention has been given to contrarians rather than experts (Petersen, Vincent and Westerling, 2019), and even though extreme weather phenomena has been affecting the media attention more than scientific accounts (Pianta and Sisco, 2020), the overall media attention has been growing (Barkemeyer et al., 2017; Layna and Altamirano, 2022). Moreover, even the more conservative outlets have started acknowledging the burning problem (Schmid Petri et al., 2017). It is important to note that climate related topics may not have the same level of politicization in all regions of the world (McCrigh et al., 2016).~~

~~In our article we have shown that the proportionate interest of the general public, as reflected in large scale online discussions, has not been following the same path. Possibly because of exhausting the topic and being tired of continuous high alert state, other competing crises (Stoddart et al., 2021), as well as simply other interesting topics, the discourse about climate change has taken less and less space proportionately.~~

~~Moreover, we show that discussions about "global warming" appear to use less sophisticated language and are more positive than those about "climate change." This result suggests that~~

perceptions of global environmental change manifest in language, which is a phenomenon that call for further exploration.¶¶

We also show that online discussions seem to precede Google searches in terms of language shift from “global warming” to “climate change”, which suggests that monitoring social media discussions may be a more immediate way of detecting changes in public attention and imaginary, especially given the interplay between dissemination of information in user generated content and more traditional mass media outlets which may not follow previously assumed patterns in other contexts (Porten Gheé and Eilders 2015)..¶¶

All of these results indicate that the window for mobilizing the general public’s attention around global environmental change may be closing, and that the main debate around the topic among non-experts may be affected by the informal discussions on social media more than elsewhere. Although our analyses included only English posts, the majority of which were likely from North America, it is important to note that there may be significant differences in sentiment and polarity when looking at other regions of the world or other languages (McGrigh et al. 2016).

Further investigation into how discussions about climate change are shifting across various online platforms could provide a more comprehensive understanding of the evolution of public discourse on this critical issue.¶¶

7. References

- Al-Rfou, Rami. 2015. “Polyglot: A Massive Multilingual Natural Language Processing Pipeline.” search.proquest.com.
<https://search.proquest.com/openview/a6748f01943c73c42dd77d55625097aa/1?pq-origsite=gscholar&cbl=18750>.
- Barkemeyer, Ralf, Frank Figge, Andreas Hoepner, Diane Holt, Johannes Marcelus Kraak, and Pei-Shan Yu. 2017. “Media Coverage of Climate Change: An International Comparison.” *Environment and Planning C: Politics and Space* 35 (6): 1029–54.
- Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. “O’Reilly Media, Inc.”
- Boykoff, Maxwell T. 2008. “Lost in Translation? United States Television News Coverage of Anthropogenic Climate Change, 1995–2004.” *Climatic Change* 86 (1): 1–11.
- Broadbent, Jeffrey, John Sonnett, Iosef Botetzagias, Marcus Carson, Anabela Carvalho, Yu-Ju Chien, Christopher Edling, et al. 2016. “Conflicting Climate Change Frames in a Global Field of Media Discourse.” *Socius* 2 (January): 2378023116670660.
- Burkeman, Olive. 2003. “Memo Exposes Bush’s New Green Strategy.” *The Guardian*, 2003.
<https://amp.theguardian.com/environment/2003/mar/04/usnews.climatechange>.
- Chinn, Sedona, P. Sol Hart, and Stuart Soroka. 2020. “Politicization and Polarization in Climate Change News Content, 1985–2017.” *Science Communication* 42 (1): 112–29.
- Dale, Edgar, and Jeanne Sternlicht Chall. 1948. *A Formula for Predicting Readability*. Bureau of Educational Research, Ohio State University.

- Fariello, Gabriele, Dariusz Jemielniak, and Adam Sulkowski. 2021. "Does Godwin's Law (rule of Nazi Analogies) Apply in Observable Reality? An Empirical Study of Selected Words in 199 Million Reddit Posts." *New Media & Society*, December, 146144482110620.
- Fownes, Jennifer R., Chao Yu, and Drew B. Margolin. 2018. "Twitter and Climate Change." *Sociology Compass* 12 (6): e12587.
- Ganczewski, Grzegorz, and Dariusz Jemielniak. 2022. "Twitter Is Garbage: A Thick Big Data Exploration of #zerowaste Hashtag on Twitter in Relation to Packaging and Food Packaging Materials." *Packaging Technology & Science*, August. <https://doi.org/10.1002/pts.2685>.
- Gaudette, Tiana, Ryan Scrivens, Garth Davies, and Richard Frank. 2021. "Upvoting Extremism: Collective Identity Formation and the Extreme Right on Reddit." *New Media & Society* 23 (12): 3491–3508.
- Hutto, C., and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8 (1): 216–25.
- Jang, S. Mo, and P. Sol Hart. 2015. "Polarized Frames on 'climate Change' and 'global Warming' across Countries and States: Evidence from Twitter Big Data." *Global Environmental Change: Human and Policy Dimensions* 32 (May): 11–17.
- Jun, Seung-Pyo, Hyoung Sun Yoo, and San Choi. 2018. "Ten Years of Research Change Using Google Trends: From the Perspective of Big Data Utilizations and Applications." *Technological Forecasting and Social Change* 130 (May): 69–87.
- Layna, Juan, and Leandro Altamirano. 2022. "Promises That Don't Work? COP26 and the Problems of Climate Change." *Engaging Science, Technology, and Society*, December. <https://doi.org/10.17351/ests2022.1377>.
- Leiserowitz, Anthony, Geoff Feinberg, Seth Rosenthal, Nicholas Smith, Ashley Anderson, Connie Roser-Renouf, and Edward Maibach. 2014. "What's in a Name? Global Warming vs. Climate Change." *Yale Project on Climate Change Communication*.
- Lineman, Maurice, Yuno Do, Ji Yoon Kim, and Gea-Jae Joo. 2015. "Talking about Climate Change and Global Warming." *PloS One* 10 (9): e0138996.
- McCright, Aaron M., Riley E. Dunlap, and Sandra T. Marquart-Pyatt. 2016. "Political Ideology and Views about Climate Change in the European Union." *Environmental Politics* 25 (2): 338–58.
- Neff, Timothy, and Dariusz Jemielniak. 2022. "How Do Transnational Public Spheres Emerge? Comparing News and Social Media Networks during the Madrid Climate Talks." *New Media & Society*, March, 14614448221081426.
- Pearce, Warren, Sabine Niederer, Suay Melisa Özkula, and Natalia Sánchez Querubín. 2019. "The Social Media Life of Climate Change: Platforms, Publics, and Future Imaginaries." *Wiley Interdisciplinary Reviews. Climate Change* 10 (2): e569.
- Petersen, Alexander Michael, Emmanuel M. Vincent, and Anthony Leroy Westerling. 2020. "Discrepancy in Scientific Authority and Media Visibility of Climate Change Scientists and Contrarians." *Nature Communications* 11 (1): 553.
- Pianta, Silvia, and Matthew R. Sisco. 2020. "A Hot Topic in Hot Times: How Media Coverage of Climate Change Is Affected by Temperature Abnormalities." *Environmental Research*

Letters: ERL [Web Site] 15 (11): 114038.

Plass, Gilbert N. 1956. "The Carbon Dioxide Theory of Climatic Change." *Tell'Us* 8 (2): 140–54.

Schäfer, Mike S., and James Painter. 2021. "Climate Journalism in a Changing Media Ecosystem: Assessing the Production of Climate Change-related News around the World." *Wiley Interdisciplinary Reviews. Climate Change* 12 (1). <https://doi.org/10.1002/wcc.675>.

Schmid-Petri, Hannah, Silke Adam, Ivo Schmucki, and Thomas Häussler. 2017. "A Changing Climate of Skepticism: The Factors Shaping Climate Change Coverage in the US Press." *Public Understanding of Science* 26 (4): 498–513.

Schuldt, Jonathon P., Sara H. Konrath, and Norbert Schwarz. 2011. "'Global Warming' or 'climate Change'? Whether the Planet Is Warming Depends on Question Wording." *Public Opinion Quarterly* 75 (1): 115–24.

Schuldt, Jonathon P., and Sungjong Roh. 2014. "Media Frames and Cognitive Accessibility: What Do 'Global Warming' and 'Climate Change' Evoke in Partisan Minds?" *Environmental Communication* 8 (4): 529–48.

Shehata, Adam, and David Nicolas Hopmann. 2012. "Framing Climate Change." *Journalism Studies* 13 (2): 175–92.

Sisco, Matthew R., and Elke U. Weber. 2022. "Local Temperature Anomalies Increase Climate Policy Interest and Support: Analysis of Internet Searches and US Congressional Vote Shares." *Global Environmental Change: Human and Policy Dimensions* 76 (September): 102572.

Sonnett, John. 2022. "Climate Change Risks and Global Warming Dangers: A Field Analysis of Online US News Media." *Environmental Sociology* 8 (1): 41–51.

Stecula, Dominik A., and Eric Merkley. 2019. "Framing Climate Change: Economics, Ideology, and Uncertainty in American News Media Content From 1988 to 2014." *Frontiers in Communication*. <https://doi.org/10.3389/fcomm.2019.00006>.

Stoddart, Mark C. J., Howard Ramos, Karen Foster, and Tuomas Ylä-Anttila. 2021. "Competing Crises? Media Coverage and Framing of Climate Change During the COVID-19 Pandemic." *Environmental Communication*, August, 1–17.

Vu, Hong Tien, Yuchen Liu, and Duc Vinh Tran. 2019. "Nationalizing a Global Phenomenon: A Study of How the Press in 45 Countries and Territories Portrays Climate Change." *Global Environmental Change: Human and Policy Dimensions* 58 (September): 101942.

Zehr, Stephen C. 2000. "Public Representations of Scientific Uncertainty about Global Climate Change." *Public Understanding of Science* 9 (2): 85. Al-Rfou, R. (2015) Polyglot: A Massive Multilingual Natural Language Processing Pipeline. [search.proquest.com. Available at: https://search.proquest.com/openview/a6748f01943e73e42dd77d55625097aa/1?pq-origsite=scholar&cbl=18750.](https://search.proquest.com/openview/a6748f01943e73e42dd77d55625097aa/1?pq-origsite=scholar&cbl=18750)

Barkemeyer, R. et al. (2017) 'Media coverage of climate change: An international comparison', *Environment and Planning C: Politics and Space*, 35(6), pp. 1029–1054.

Bird, S., Klein, E. and Loper, E. (2009) *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.

~~Broadbent, J. et al. (2016) 'Conflicting Climate Change Frames in a Global Field of Media Discourse', *Socius*, 2, p. 2378023416670660.¶¶~~

~~Burkeman, O. (2003) 'Memo exposes Bush's new green strategy', *The guardian*. Available at: <https://amp.theguardian.com/environment/2003/mar/04/usnews.climatechange>.¶¶~~

~~Chinn, S., Hart, P.S. and Soroka, S. (2020) 'Politicoization and Polarization in Climate Change News Content, 1985-2017', *Science communication*, 42(1), pp. 112-129.¶¶~~

~~Fariello, G., Jemielniak, D. and Sulkowski, A. (2021) 'Does Godwin's law (rule of Nazi analogies) apply in observable reality? An empirical study of selected words in 199 million Reddit posts', *New media & society*, p. 146144482110620.¶¶~~

~~Ganczewski, G. and Jemielniak, D. (2022) 'Twitter is garbage: A Thick Big Data exploration of #zerowaste hashtag on Twitter in relation to packaging and food packaging materials', *Packaging Technology & Science* [Preprint]. Available at: <https://doi.org/10.1002/pts.2685>.¶¶~~

~~Gaudette, T. et al. (2021) 'Upvoting extremism: Collective identity formation and the extreme right on Reddit', *New Media & Society*, 23(12), pp. 3491-3508.¶¶~~

~~Hutto, C. and Gilbert, E. (2014) 'VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text', *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), pp. 216-225.¶¶~~

~~Jun, S. P., Yoo, H.S. and Choi, S. (2018) 'Ten years of research change using Google Trends: From the perspective of big data utilizations and applications', *Technological forecasting and social change*, 130, pp. 69-87.¶¶~~

~~Layna, J. and Altamirano, L. (2022) 'Promises that Don't Work? COP26 and the Problems of Climate Change', *Engaging Science, Technology, and Society*, 8(3), pp. 118-129.¶¶~~

~~Lineman, M. et al. (2015) 'Talking about Climate Change and Global Warming', *PloS one*, 40(9), p. e0138996.¶¶~~

~~Neff, T. and Jemielniak, D. (2022) 'How do transnational public spheres emerge? Comparing news and social media networks during the Madrid climate talks', *New Media & Society*, p. 14614448221081426.¶¶~~

~~Pearce, W. et al. (2019) 'The social media life of climate change: Platforms, publics, and future imaginaries', *Wiley interdisciplinary reviews. Climate change*, 10(2), p. e569.¶¶~~

~~Petersen, A.M., Vincent, E.M. and Westerling, A.L. (2019) 'Discrepancy in scientific authority and media visibility of climate change scientists and contrarians', *Nature communications*, 40(1), p. 3502.¶¶~~

~~Pianta, S. and Sisco, M.R. (2020) 'A hot topic in hot times: how media coverage of climate change is affected by temperature abnormalities', *Environmental research letters: ERL* [Web site], 15(11), p. 114038.¶¶~~

~~Plass, G.N. (1956) 'The carbon dioxide theory of climatic change', *Tell'Us*, 8(2), pp. 140-154.¶¶~~

~~Schäfer, M.S. and Painter, J. (2021) 'Climate journalism in a changing media ecosystem: Assessing the production of climate change-related news around the world', *Wiley interdisciplinary reviews. Climate change*, 12(1). Available at: <https://doi.org/10.1002/woc.675>.¶¶~~

~~Schmid-Petri, H. et al. (2017) 'A changing climate of skepticism: The factors shaping climate~~

~~change coverage in the US press', Public understanding of science , 26(4), pp. 498–513.¶~~

~~Schuldt, J.P. and Roh, S. (2014) 'Media Frames and Cognitive Accessibility: What Do "Global Warming" and "Climate Change" Evoke in Partisan Minds?', Environmental Communication, 8(4), pp. 529–548.¶~~

~~Shehata, A. and Hopmann, D.N. (2012) 'Framing climate change', Journalism Studies, 13(2), pp. 175–192.¶~~

~~Sisco, M.R. and Weber, E.U. (2022) 'Local temperature anomalies increase climate policy interest and support: Analysis of internet searches and US congressional vote shares', Global environmental change: human and policy dimensions, 76, p. 102572.¶~~

~~Sonnett, J. (2022) 'Climate change risks and global warming dangers: a field analysis of online US news media', Environmental Sociology, 8(1), pp. 41–51.¶~~

~~Stecula, D.A. and Morkley, E. (2019) 'Framing Climate Change: Economics, Ideology, and Uncertainty in American News Media Content From 1988 to 2014', Frontiers in Communication. Available at: <https://doi.org/10.3389/fcomm.2019.00006>.¶~~

~~Stoddart, M.C.J. et al. (2021) 'Competing Crises? Media Coverage and Framing of Climate Change During the COVID-19 Pandemic', Environmental Communication, pp. 1–17.¶~~

~~Vu, H.T., Liu, Y. and Tran, D.V. (2019) 'Nationalizing a global phenomenon: A study of how the press in 45 countries and territories portrays climate change', Global environmental change: human and policy dimensions, 58, p. 101942.¶~~

~~Zehr, S.C. (2000) 'Public representations of scientific uncertainty about global climate change', Public understanding of science , 9(2), p. 85.~~

8. Appendices: Supplemental Data and MetaData

The data and the associated data used to produce the plots and findings is made freely available in a [Reddit-Climate-Data-2005-2021](#) shared folder with the following contents:

[Reddit-Climate-Data-2005-2021/](#)

- [Reddit General Climate Related Plots 2005-2021.ipynb](#) (2.2 MB)
The Jupyter Notebook compatible Google Colab Notebook used to create all the plots, and others, in this publication.
- [Headers Information.xlsx](#) (20KB)
A Microsoft Excel file containing information on the various headers (columns) found in the data files.
- [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.subreddit.csv](#) (77MB)
A (large) summary of all analyzed data grouped by subreddit.
- [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.subreddit_year.csv](#) (184MB)
A (large) summary of all analyzed data grouped by subreddit and year.
- [RedditBoth-All-Climate-English-both_text-reduced.grouped-by.year.csv](#) (31KB)
A summary of all analyzed data grouped by year.