Exploring Topics and Political biases in News Outlet Climate Messaging

Lambert Francis, Patrick Janulewicz

Abstract

With the relevance of climate change in today's society, it is natural to wonder how climate-related information is being covered. In this paper, we examine the impact of political leaning on mainstream media coverage of climate-related news articles. To do this, climate articles from various different outlets were gathered using a keyword search over a balanced set of thousands of articles released between 2017 and 2019, inclusively. The data was then analyzed in a multifaceted manner. Firstly, manual open coding was performed to determine article categories. These categories were then studied using TF-IDF, sentiment analysis, and a trained climate stance classifier. Our main contributions include a climate-relevant subset of news articles from 2017-2019, a breakdown of climate change topics discussed in the news, and an analysis of how climate messaging differs between news outlets with different political backgrounds.

Introduction

The topic of climate change quickly rose to prominence in the 1980s as the scientific community began to recognize the grave threats it poses. The term has since become ubiquitous, being well known by scientists and the general public alike. Messaging of climate-related information can be found in countless aspects of people's lives, including science communication, marketing, and media. In this paper, we will focus on the third; in particular, we will examine climate messaging and communication in mainstream media. We examine the general sentiment of various outlets on climate change and whether their own political leanings have an effect on the tone of their communication. Furthermore, we will examine the coverage of these outlets on various topics and consider whether this reinforces or casts doubt on their impartiality.

We present a multifaceted approach to analyze climate messaging in mainstream media. In this paper, we consider multiple techniques, each varying in their level of sophistication and complexity. On the more traditional side, human annotation was used to classify a subset of articles by topic, and the highest scoring words by TF-IDF were determined. A more intermediate approach uses a latent Dirichlet allocation (LDA) to categorize the full dataset and sentiment analysis to study the overall tone per topic. Finally, we use a BERT classifier developed by Luo, Card, and Jurafsky [6] which is specifically trained for climate-related stance detection. These different approaches can then be compared and contrasted to find their agreement, disagreement, and relative effectiveness.

While this project takes an in-depth look at climate messaging in mainstream media, it is is important to acknowledge its constraints. One limitation of this project is the fact that the dataset covers a three-year period from 2017 to 2019, inclusively. The data was taken from the POLUSA dataset [4]; while this data is plentiful, balanced, and organized, it covers only a small period of time. Climate-related news in mainstream media has been covered for decades. We therefore do not claim to provide an analysis of the entire history of climate change.

The POLUSA dataset is designed to represent "the online media landscape as perceived by an average US news consumer". It is balanced with respect to outlet popularity and publication date. The patterns obeyed by the articles therefore obey the popularity of the outlet rather than the popularity of the articles themselves.

Moreover, we only consider articles from U.S. outlets written in English. While the outlets provided in this paper are among the largest and most influential in the world, we do not extend the analysis to other countries or languages.

We also limit our analysis to mainstream outlets. Although the selected outlets certainly have their biases, each source has a sufficiently large and established network of followers. We do not consider smaller scale or fringe networks in this analysis.

Relevant Work

Various studies have gathered and analyzed climate-related data in the hopes of better understanding opinions on the matter. In the past, a natural way of acquiring this data was through television and daily print media [2]. In more recent times, however, the presence of the internet has become far more dominant, and most articles can be obtained digitally. Previous large-scale studies such as [1] have examined such news articles and established links between topics and outlet bias. Other studies have considered the difference between mainstream media and social media [7]. Finally, and perhaps the most prominent related work to this study is the work of Luo, Card, and Jurafsky [6]. The paper aims to study the news articles' stance on climate change. The study introduces the Global Warming Stance Dataset (GWSD), which contains two thousand climate related sentences and classifies them based on whether they are dismissive, neutral, or agreeing of climate change. This dataset was then used to train a BERT classifier which has an accuracy competitive with human performance. This trained model is used to classify various articles in the dataset and sort them based on their position toward climate change. In this paper, we aim to expand on the model's use and also explore alternative ways of performing analysis of climate-related news articles.

Dataset Creation

Source Data

Construction of the climate-relevant dataset begins with the POLUSA dataset [4]. This dataset contains 0.9M articles spanning from January 2017 to August 2019 and consists of 18 different publishers. It designed to represent the view of "the online media landscape as perceived by an average US news consumer". To achieve this, the authors balanced the dataset in terms of outlet popularity and publication date. POLUSA labels outlets with a political leaning of *LEFT, CENTER, RIGHT,* or *UNDEFINED*. These political scores are drawn from 8 aggregate sources. More information about difference source types, and disagreement thresholds can be found in the POLUSA paper.

Keyword Selection

In order to trim the POLUSA dataset down to only climate-relevant articles many different keywords were considered for filtering by the article body, for example, *climate change, global warming, sea level rise, disaster & climate, deforestation, carbon emission & climate, habitat & climate, paris agreement, greta thunberg, renewable & climate, climate march, polar bear & climate, carbon dioxide & climate, carbon tax*, where an ampersand represents the typical "and" boolean operator. The motivation behind the exploration of these keywords was by reviewing major climate events, well-accepted impacts of climate change, and keywords related to 'climate change' on google analytics. These keywords combined offered an increase in dataset size of 13% compared to filtering using only *climate change, global warming, greenhouse gas,* and *greenhouse gases*. Given that many other articles use only these simple keywords and the increase in false positives from additional keywords, the final keywords were chosen to be those three to better enable comparison and more easily continue filtering [2] . We note that this choice potentially excludes fringe articles that talk about effects of climate change without directly mentioning it, that perhaps include unique perspectives. After applying these keyword filters and analyzing a random sample of 30 articles, 18 articles were deemed relevant to climate change; the major offenders were typically political articles with offhand mentions of climate change as part of a candidate's or politician's policy.

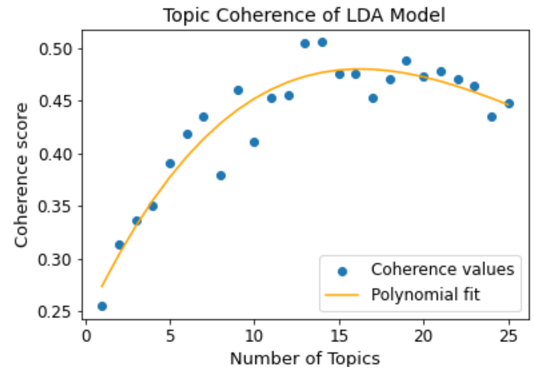

Figure 1: Coherence values as a function of the number of topics. Note that the lowest possible number of topics before the plateau is considered optimal.

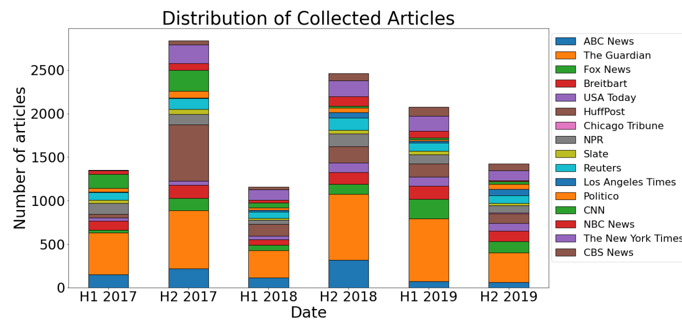

Figure 2: Distribution of collected articles after filtering. Data is divided into 6 month periods between the first half of 2017 and the second half of 2019.

Topic Modeling

To further increase the proportion of climate-change relevant articles, experiments were performed using Facebook's bart-large-mnli transformer model for zero shot text classification [10]. Using possible topics of *climate impact* and *climate change*, the headlines of 30 climate-relevant labeled (Y/N) articles were scored. Using a threshold score of 0.25 resulted in 26/30 of the articles being correctly labeled, with 3 false negatives, 1 false positive. After performing the keyword filter and topic modeling with the aforementioned threshold, we are left with 12,430 articles (1.37% of the original dataset).

Methods

Annotation

In order to understand the topics around climate change being discussed in the media, two independent open codings were performed over 50 randomly selected article bodies by the authors of this paper. After discussion, the two codings were merged. The coding was then applied to another 50 randomly selected article bodies in another iteration to refine the topics.

In addition to the manual open coding, a Latent Dirichlet allocation was performed using the Gensim model [8]. To do this, the body of each article was taken and cleaned by removing stopwords and unwanted characters, as well as lemmatizing. The text was also split up into individual words, bigrams, and trigrams, and the corpus was subsequently created. To decide the number of categories for the LDA, the coherence values were computed and plotted in figure 1. While the coherence score reached a maximum at around 15 topics, human judgment found this number to be too large. The number with the highest coherence that was not deemed to overfit was approximately eight. Table 6 shows the keywords obtain from the LDA and their inferred topics.

In order to feasibly code a significant number of articles and thus have sufficient articles to compute meaningful statistics for each topic, we decide to annotate headlines instead of article bodies. Testing headline annotation on the second sample of 50 articles from before, we find that 3 articles are incorrectly classified based off the headline and 9 are uncertain between multiple topics. During the final annotation, if the headline appeared ambiguous the lead sentence was referenced and if also found ambiguous then the body. If the body was needed to classify the article then it was also tagged with a 'B' label so that it could be ignored for analyses involving solely headlines. Classifying the articles was done carefully, prioritizing correct labels over quantity of labels. If the annotator was unsure about the category even after reading the article body, it was tagged as 'Other'.

The resulting topic names as well as an example for each with justification can be found in Table 7 in the appendix. More general guidelines on topic criteria are given below. The criteria used to code topics is not a direct correspondence to the label definition. For instance, a politics coded headline might focus on conflict over climate change between two celebrities with no political figure involved.

Some overlaps between the topics are inevitable. For instance, a headline discussing how solar energy is reducing emissions is ambiguous; It fits both *Mitigation* and *Energy*. In cases such as these, both labels are assigned to the headline. Multi-labeled headlines were less than 10% of the labeled sample. The above examples are not written in stone as belonging to their respective category. A headline on the Paris Accord might make no mention of international affairs and instead focus solely on its mitigation aspects and would be coded as mitigation.

Analysis

To prepare the data for feature extraction, a standard text cleaning was performed. This involved the downcasing of all text data in addition to removing punctuation. An additional analysis into advertisement removal was performed for each outlet, however, much pollution remained in the text data even after applying those filters. This pollution greatly impacts computing standard TF-IDF scores across political leanings as one outlet will have unique formats.

Topic	Description
Politics	Conflict between two (usually famous) parties in the context of climate change.
Mitigation	Specifics of a mitigation plan or technology (Solar, Recycling initiative, Emission cuts, ...).
Call to Action	Argument or appeal to act in some regard. (Leader paints climate change as crisis demanding action, Youth protests, ...)
Energy	Energy technology, country energy transitions, energy companies (Report on solar, coal usage discussion, BP report).
Environmental Impact	Scientific reports, climate change related effects on environment, often long term consequences which don't immediately impact humans.
International	Dialogue between nations (G20, Paris Accord).
Human Impact	How humans are being impacted, often short term (air pollution death rates increase, ethnic populations displaced due to sea level rise, ...).
Climate Denial	Discussion on belief in climate change, either by climate change proponents or deniers.
Public Perception	Discussion of general public's attitude towards climate change (Survey results, individual responsibility).
Other	Relevant to climate change but does not fit the typology.

Table 1: General coding guidelines for different categories. Specific examples can be found in Table 7 in the appendix.

To mitigate this, alternative TF-IDF metrics were explored [2] . The metric settled upon allows words to have scores proportional to their term frequency, even if they are not unique across all documents. Also, an additional class frequency term is included to bias scores towards words with greater presence throughout a class. This alternative metric resulted words much more related to climate change such as 'paris', 'epa', and 'emissions'.

Results

Overall

The transition to the climate-related dataset creates a significant shift in outlet distribution. The proportions change as follows: the number of Left, Center, Right, and Undefined outlets change from 0.31 → 0.53, 0.27 → 0.22, 0.16 → 0.12, 0.26 → 0.14, respectively. The differences between are shown in Figure 3.This shows that filtering for climate-based articles drastically increases the proportion of articles from Left leaning outlets.

The top 5 TF-IDF scores on article headlines for each outlet can be found in Table 4. Unique words in the top 5 include 'Global' and 'Energy', both from the left. Of these,

Code	Topic
P	Politics
O	Other
M	Mitigation
C	Call to Action
E	Energy
EE	Environmental Impact
I	International
H	Human Impact
D	Climate Denial
PP	Public Perception

Table 2: Open coding topics and their respective codes.

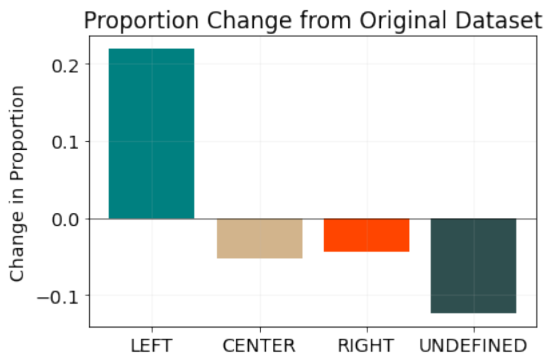

Figure 3: Change in political leaning article proportion after applying the climate filtering steps discussed above.

the other leanings all contained 'Global' in the top 10 but only 'Energy' in the top 20.

Applying the global warming stance detection model of Luo et al. to article headlines over each political leaning and averaging yields the following scores: 0.048, 0.044, 0.009, and 0.006 for center, left, right, and undefined leanings, respectively. A score for an individual headline ranges from -1 to 1 for doubting and affirming climate change. The low average scores suggest that most articles are neutral. However, center and left leaning news does appear to have more articles with stances affirming climate change than right and undefined leaning ones. Applying to article bodies instead results in a Left > Center > Undefined > Right trend, with respective average stance scores of 0.105, 0.070, 0.036, -0.014. We note that the scores have greatly increased, but a similar trend remains with left and center having notably higher scores relative to right and undefined.

Topics

The topic distribution can be found in Figure 5. Of the coding topics the top 5 form 66% of the dataset. These topics are Politics, Mitigation, Call to Action, Energy, and Environmental Impact.

The landscape of topic distribution changes dramatically within different political leanings, as shown in 4. Significant discrepancies in topic coverage can be seen via rows with dramatic shifts in color. For instance, Right leaning outlets report much more on politics than other topics. Other notable trends include: a lack of mitigation reporting in Right outlets, less calls to action in Center outlets, and a lack of energy for Right leaning.

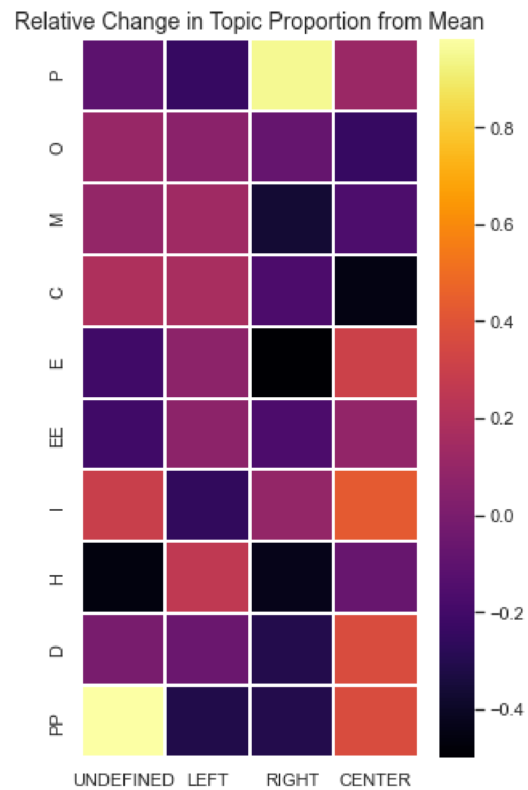

Figure 4: Relative percent change in topic proportion from the proportion of the entire dataset. A value of 1 implies that the proportion of that topic in that leaning is twice that of the entire dataset.

As previously mentioned, a LDA was used to cluster like categories. The GWSD-trained Bert model was then applied to the topics and each corresponding outlet. A heatmap of this data can be found in Figure 6 in the appendix. The map displays the average score of each outlet and each topic. Note that an acknowledgement of climate

Leaning	Positive	Neutral	Negative	Score
Left	0.152	0.765	0.084	0.044
Center	0.117	0.846	0.037	0.047
Right	0.111	0.781	0.107	0.009
Undefined	0.082	0.873	0.045	0.006

Table 3: Proportion of articles in each leaning that were classified as positive, neutral, or negative using the GWSD classifier for each leaning. An additional aggregate score is included $Score = P_{+} + P_0 + P_{-}$, where P_i is the proportion of articles classified as i.

Left	Center	Right	Undefined
Trump	Trump	Trump	Trump
Global	California	GND	Paris
Emissions	Paris	AOC	GND
Coal	Coal	California	EPA
Energy	EPA	EPA	Emissions

Table 4: Top 5 relevant TF-IDF for each political leaning. A modified TF-IDF is used, allowing scores to not vanish with IDF scores of 0 [9]. Some words deemed noise are removed from consideration, such as 'delingpole' a climate journalist who authors many articles in the dataset. 'GND' is an abbreviation for 'Green New Deal'.

change corresponds to a score of +1, a dismissal of climate change corresponds to a score of -1, and neutrality corresponds to a score of 0.

To compare and contrast with the GWSD-trained model, sentiment analysis was also performed on article headlines using TextBlob's polarity score [5]. This resulted in Figure 7, which can be found in the appendix.

Topic TF-IDF'S

The top 3 modified TF-IDF words for each topic can be found in Table 5. Some additional thoughts on the TF-IDF words for certain topics are included below.

Environmental Impact Environmental impact contained less sensational words, ones more consistent with its removal from immediate effects on human lives, such as 'warming', 'scientists', 'arctic', and 'record'. Interestingly, there is a lack of words relating to emissions despite anecdotally annotating many articles in this vein. This suggests that emissions are commonly discussed throughout the other topics; In fact we will see they are prevalent when discussing mitigation and energy.

Energy Discussion of energy was fairly dominated by fossil fuels. The top words were 'coal', 'oil', 'fossil fuels', 'emissions', and 'trump'. Given recent hype over renewable energies it is surprising to see no mention of solar or

Topic	Word 1	Word 2	Word 3
P	Trump	GND	AOC
M	GND	emissions	tax
C	action	rebellion	youth
E	coal	oil	Trump
EE	scientists	says	heat
I	Trump	Paris	talks
H	study	risk	oil
D	news	science	real
PP	americans	poll	policies

Table 5: Top 3 relevant TF-IDF for each topic. A modified TF-IDF is used, allowing scores to not vanish with IDF scores of 0 [9]. Some words deemed noise are removed from consideration, such as 'just'. 'GND' is an abbreviation for 'Green New Deal'.

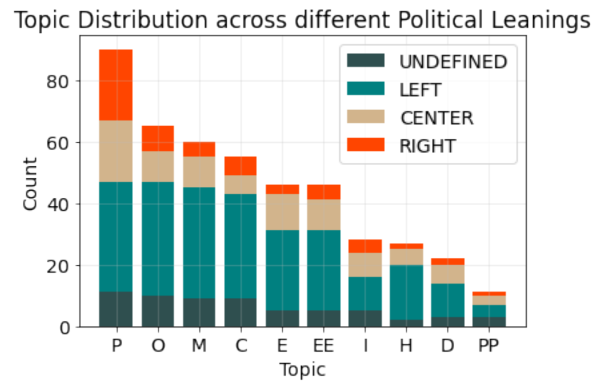

Figure 5: Topic distributions across different political leanings. Please refer to Table 2 for the associated topics.

wind power; One potential reason for this is that articles are sampled from 2017 to 2019, entirely during Trump's presidency of which a major policy was fossil fuels. It is not clear whether sampling another time period would result in greater discussion over renewable or other energy topics and is an avenue that could be investigated in future works.

Politics Two figures dominated the discussion in Politics, Alexandra Ocasio-Cortez as well as President Trump were both among the top mentions. Both of these figures are strong symbols of their respective political associations. Additional words included 'EPA' (U.S. Environmental Protection Agency), 'California', and 'Democrats'.

Discussion

Below we focus on answering two questions. What are the main topics discussed by the media in the context of climate change, and what effect, if any, does political leaning have on article messaging. We focus the discussion of individual topic trends on the more populous topics.

Topics

From the topic distributions shown in Figure 5, we see that coverage of climate change is largely skewed towards categories involving high profile figures or areas of conflict. The politics and call to action categories are both examples of this and collectively form 32.2% of the labeled dataset. This result supports previous work showing that journalistic norms cause a bias in climate reporting, specifically by 'adherence to ... personalization, dramatization ...' [2].

Discussion of actual climate change impact is comprised largely of the Environmental Impact and Human Impact categories. Together they account for 16.2% of the dataset. This is smaller than one might expect considering the huge scope of discussion that is contained in climate mitigation and climate impact. This further demonstrates that when climate change is discussed in the media, it is primarily as the side actor of another main event. For example, a report on youth protesting over climate change focuses on the protest itself rather than climate change mitigation or impact. One motivation of the separation of environmental and human impacts of climate change was to discover whether the media favored covering the shorter or longer term effects associated with the two categories (See 1. From the topic distributions we see that discussion on generally longer-term more environmental effects is more common. This is at odds with the journalistic preference for dramatization. A explanation reconciliation for this discrepancy is that human-related impacts simply aren't attributed to climate change. For example, an article on floods in neighborhoods causing relocations might not attribute this to climate change and so would not be picked up in our dataset. Another explanation could be that there are simply not many human-related impacts or that those impacts do not manifest in a large amount of events that would warrant an article.

Some notable trends arose in the TF-IDF words across different topics. One commonality was 'Trump' referencing the at the time President. He appeared in 5 of the top 5 TF-IDF results by topic. The two categories focused on impact included the words 'study' and 'scientists', 'say', suggesting that media messaging here is focused more on scientific communication. Mitigation was composed primarily of financial terms, suggesting that news media focuses coverage on the economic aspects of mitigation plans, this is one of the categories where Trump was not mentioned, and also a category with less right leaning articles. This suggests that mitigation is not a topic of focus for the right political ideology. 'emissions' and words relating to fossil fuels were also discussed in the mitigation and energy topics. Surprisingly, there was no reference to any forms of green energy. In fact, the closest mention to green energy in any of the categories was through the mention of the green new deal.

Political Effect

TF-IDF outcomes using political leanings as classes yielded no glaring abnormalities; The largest difference being the left's greater use of the word 'energy' than other outlets. A much greater shift was the political leaning balance of articles when filtering to climate related articles. One possible reason for this is that left leaning articles tend to actually write more about climate change. Given that climate change is a cause often championed by liberal parties, this seems like a reasonable conclusion. However, it is also possible a bias was introduced in the keyword filtering. One scenario could be that other political leanings prefer using 'global warming' as a key word and deliberately avoided using 'climate change' in their article bodies. One might expect this from far right outlets as 'global warming' paints a less scary picture of climate change and this ideology is linked to climate-skepticism [3].

Topic coverage also shifts among different political leanings as shown in Figure 4. Most notably, right leaning articles have dramatically more articles in the politics category. This indicates that right leaning outlets cover conflict between parties more often. These articles also tended to report less on mitigation, which could be a potential association with climate-skepticism. Additionally, discussion of energy was lacking among right leaning articles, this is surprising, given a main talking point through the Trump administration was coal and fossil fuels. Center articles had a distinct lack of articles identifying as a call to action, a possible result of a more middle of the road ideology. Left articles were fairly balanced.

The global warming stance detector discussed in the Relevant Works section showed that undefined and right leaning articles had a lower occurrence of affirming stances than left and center leaning articles on headlines. Though all scores were very small, and hence no outlet took a strong stance on climate change, the shift in scores was notable. This shift suggests that a higher proportion of articles from left and center leaning outlets are favorable to climate change than right and undefined leaning ones. We note that left leaning articles had the highest proportion of a positive stance.

Conclusion

In this paper, we investigated what topics news outlets discussed in the context of climate change, and how their messaging differed based on their political alignment. A dataset of 12,430 articles was created using a series of keyword and topic modeling filters from an original dataset of 0.9M articles. A subset of 500 of these articles were labeled by headline on topics developed through an open coding process.

Using top TF-IDF words grouped by topic in addition to topic proportions themselves, we explore how the media covers different parts of climate change and find that their main focus is on conflict or statements made by notable figures. Topics are also discussed in the context of larger macro-themes such as mitigation and impact. We find that Trump is a dominant part of the conversation, appearing in the top 5 TF-IDF scores for every leaning and in 5 of the topics. Other interesting topics or lack thereof are mentioned in the discussion.

Our political analysis showed no great difference in the discussion of topics when they were discussed in articles from different leanings. However, the quantity and ratios of topics discussed had significant shifts with political leaning; With most notable trends of climate change in general being covered far more by left leaning outlets, and with largely different topic distributions in right leaning articles compared to the overall trends. One caveat to this is that right and undefined leaning articles had lower tendencies to affirm climate change, which might suggest a bias not represented in analyzing top TF-IDF words. This is a potential avenue for further study.

To summarize, our main contributions are

1. an analysis on how climate messaging differs with political leaning, if at all.
2. a subset of the POLUSA dataset containing climate-relevant articles, and an additional smaller subset of manually annotated articles.
3. a discussion of possible sub-topics of the more general topics found in the open coding, indicating potential

avenues for future research.

References

[1] Jeremiah Bohr. ““Reporting on climate change: A computational analysis of U.S. newspapers and sources of bias, 1997–2017””. In: *Global Environmental Change* 61 (2020), p. 102038. ISSN: 0959-3780. DOI: https://doi.org/10.1016/j.gloenvcha.2020.102038. URL: https://www.sciencedirect.com/science/article/pii/S095937801931088X.

[2] Maxwell T. Boykoff and Jules M. Boykoff. “Climate change and journalistic norms: A case-study of US mass-media coverage”. In: *Geoforum* 38.6 (2007). Theme Issue: Geographies of Generosity, pp. 1190–1204. ISSN: 0016-7185. DOI: https://doi.org/10.1016/j.geoforum.2007.01.008. URL: https://www.sciencedirect.com/science/article/pii/S0016718507000188.

[3] Bernhard Forchtner, Andreas Kroneder, and David Wetzel. “Being Skeptical? Exploring Far-Right Climate-Change Communication in Germany”. In: *Environmental Communication* 12.5 (2018), pp. 589–604. DOI: 10.1080/17524032.2018.1470546. eprint: https://doi.org/10.1080/17524032.2018.1470546. URL: https://doi.org/10.1080/17524032.2018.1470546.

[4] Lukas Gebhard and Felix Hamborg. “The POLUSA Dataset: 0.9M Political News Articles Balanced by Time and Outlet Popularity”. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. ACM, 2020. DOI: 10.1145/3383583.3398567. URL: https://doi.org/10.1145\%2F3383583.3398567.

[5] Steven Loria. “textblob Documentation”. In: *Release 0.15* 2 (2018).

[6] Yiwei Luo, Dallas Card, and Dan Jurafsky. “Detecting Stance in Media on Global Warming”. In: (2020). DOI: 10.48550/ARXIV.2010.15149. URL: https://arxiv.org/abs/2010.15149.

[7] Alexandra Olteanú et al. “Comparing Events Coverage in Online News and Social Media: The Case of Climate Change”. In: May 2015.

[8] Radim Rehurek and Petr Sojka. “Gensim–python framework for vector space modelling”. In: *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic* 3.2 (2011).

[9] Rajendra Kumar Roul, Jajati Keshari Sahoo, and Kushagr Arora. “Modified TF-IDF Term Weighting Strategies for Text Categorization”. In: *2017 14th IEEE India Council International Conference (INDICON)*. 2017, pp. 1–6. DOI: 10.1109/INDICON.2017.8487593.

[10] Thomas Wolf et al. *HuggingFace's Transformers: State-of-the-art Natural Language Processing*. 2019. DOI: 10.48550/ARXIV.1910.03771. URL: https://arxiv.org/abs/1910.03771.

Appendix

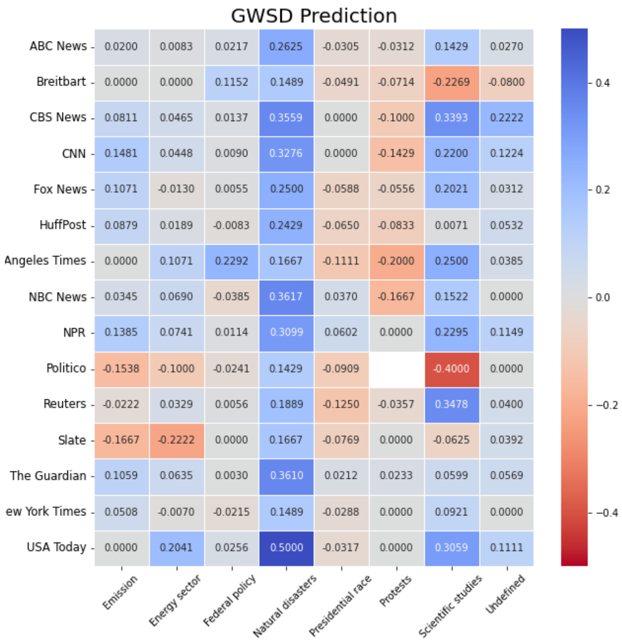

Figure 6: Agreement heatmap per LDA topic of each outlet according to the trained Bert classifier. Each value is calculated by computing the average score of each headline, where -1 is dismissive, 0 is neutral, is 1 is acknowledging of climate change. Note that The Chicago Tribune was ommitted due to a lack of data for many topics.

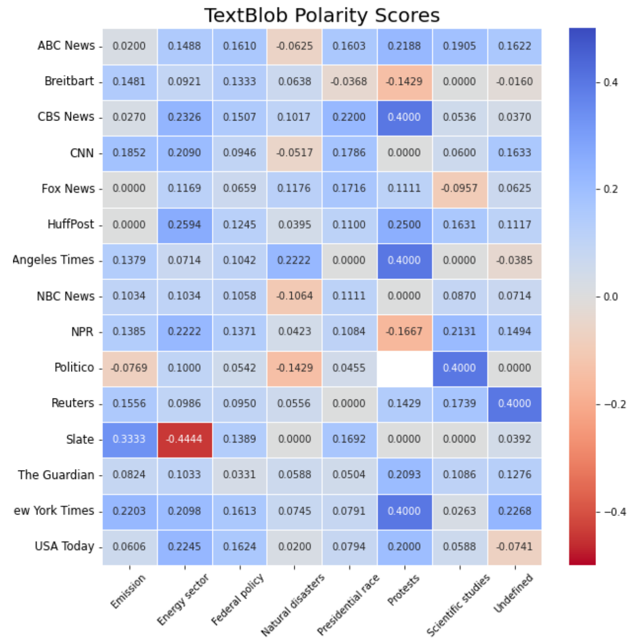

Figure 7: Polarity heatmap per LDA topic of each outlet according to TextBlob. Values are computed by computing the average score of each headline, where -1 is negative, 0 is neutral, and 1 is positive. Note that The Chicago Tribune was ommitted due to a lack of data for many topics.

Top scoring words	Inferred topic
emission, energy, carbon, government, coal, power, renewable	Energy sector
state, protest, group, activist, oil, bill, emergency, protester	Protests
car, gas, city, fuel, amazon, pollution	Emission
climate, year, report, scientist, study	Scientific studies
people, change, climate, take, make, time	Undefined
climate, plan, administration, issue, country, president, trump, federal	Federal poliicy
water, fire, year, land, risk, flood	Natural disaster
candidate, campaign, democratic, debate, party, election, vote	Presidential race

Table 6: List of keywords in each LDA cluster and their inferred meaning.

Topic	Examples	Headline	Coding	Justification
Politics	Attacking AOC	ocasio cortez contradicts herself on role of government in massive and unprecedented 'green new deal'	P	Headline focuses on AOC, not the actual content relevant to climate change (green new deal)
Mitigation	Investors urge... to cut emissions Mitigation Technique	turning carbon dioxide into rock forever	M	Discussion of method to mitigate emissions
Call To Action	Obama promotes action Pope addresses oil companies	barack obama on food and climate change 'we can still act and it won't be too late'	C	Appeal for action to be taken
Energy	Energy Transition BP move to renewables Limiting coal production	score one for corn in battle over biofuel a rare setback for big oil	E	Focus is on green energy and fossil fuels
Environmental Impact	Melting glaciers Habitat loss IPCC Report	remote hawaiian island wiped off the map	EE	A long-term environmental effect
International	Discussion between two countries on climate	german official to hold talks with turkey on climate issue	I	Focus is on two nations interacting, not particular subject of climate change
Human Impact	Disruption to daily services Loss of homes	london tubes schools and homes 'face climate change chaos'	H	Impact on daily human life.
Climate Denial	Statement on GW Existence claim of emissions not relevant to climate change	trump pick for top environmental post called belief in global warming a 'kind of paganism'	D	Headline focuses is not on Trump's pick but on their denial of climate change.
Public Perception	Views of coal demographic Poll Results	epa heads to coal country to hear views on an obama climate rule	PP	Stance of a demographic on Climate Change
Other	Ethical Controversy	epa chief faces mounting scrutiny for ethics violations	O	No clear fit

Table 7: Resulting Typology of an open coding over 50 climate related articles. Summaries of real examples are given as well as one entire headline and justification for each topic.