

A Topic Modelling Approach for Developing a Typology of Climate Change Misinformation on Twitter

Eleanor Dixon-Brandt¹

¹ King's College London, United Kingdom
eleanor.k.dixon-brandt@kcl.ac.uk

Abstract. From a dataset of 237,937 climate change-relevant tweets compiled from 2021 to 2023, this study uses manual analysis, Latent Dirichlet Allocation (LDA) topic modelling, and verification methods to map and model the topics of discussion across climate change discourses on Twitter. The methods were applied to test and control profiles, where the former focused on climate change misinformation and the latter on climate change information. From the LDA results, nearly one third of tweets in the test profiles were classified as 'Questioning the Science', which refers to challenging and interrogating the science behind climate change, while for the control profiles, 'General Climate Change' was the most frequently occurring topic. From this study, various efforts to challenge climate change misinformation can be developed and applied, as the relative emphasis on themes and topics has been modelled.

Keywords: Topic Modelling, Latent Dirichlet Allocation, Natural Language Processing, Machine Learning, Climate Change, Misinformation, Twitter.

1 Introduction

Currently, one of the greatest risks to society, the environment, and the global economy is climate change, defined as "a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods" [1]. Impacts can include increased extreme events, such as forest fires, flooding, and tropical storms, and changes in ocean and atmospheric temperatures [2]. Understandings of the risks inherent in such events are confounded by inaccurate, inconsistent, and manipulated information online, which undermines climate science. This includes misinformation, disinformation, and fake news, all of which are concerned with the communication of false or misleading messages [3, 4]. Though most individuals in the USA and Britain, 73% and 68%, respectively, accept that the climate is changing, a present and vocal group nevertheless challenges the scientific consensus and hinder public action to mitigate and adapt to climate change [5, 6]. To address climate change misinformation effectively, an understanding of the topics discussed in online communities is required. This study uses manual analysis, Latent Dirichlet Allocation (LDA) topic modelling, and verification methods to identify and classify the topics of discussion across climate change discourses on Twitter.

2 Methods

The methods followed three stages (Fig. 1). First, a manual content analysis was conducted which consisted of labelling a random subsample of 360 climate change misinformation and climate change information-relevant tweets. Then, LDA topic modelling, an unsupervised machine learning algorithm which identifies the relative emphasis of themes and topics in a corpus, was applied to the complete dataset. The LDA topic model was trained to output 30 topics from the dataset and these topics were subsequently labelled. Finally, a blind-manual analysis was carried out, of a different subsample of tweets to which the LDA topic model had assigned the highest probability of belonging to a single topic. This last method assessed the accuracy of the LDA topic modelling process.

2.1 Data

The dataset contained 115,978 tweets for the control profiles and 121,959 tweets for the test profiles. The total number of tweets was therefore 237,937 tweets. Control and test profiles were identified with various search and bias terms and were filtered based on criteria including Twitter profile name, number of followers, verification status, ‘about’ description, history, activity level, and the number of links to denial websites or blogs (Fig. 1). The tweets were collected between 2021 and 2023 through the Twitter full-archive search application programming interface (API) and all were in English.

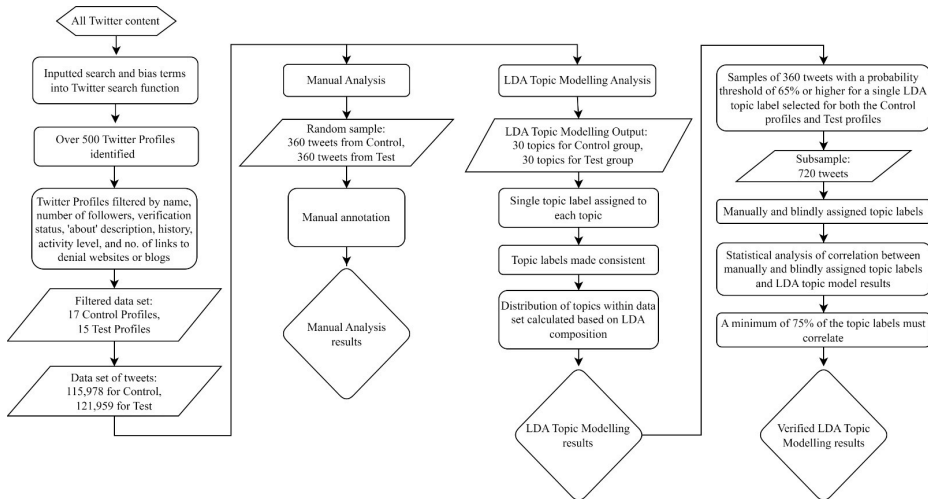


Fig. 1. Twitter Data Retrieval and Analysis Flowchart

3 Results

3.1 Manual Content Analysis Results, Twitter, 2021 – 2023

The manual analysis results (Table 1) indicate ‘Action Campaigns’ and ‘Questioning the Science’ are the most frequently occurring topics for the control and test profiles at 15.00% and 19.72%, respectively. The top five topics for the control profiles consist of ‘Action Campaigns’, ‘General Climate Change’, ‘Social Media’, ‘Media’, and ‘Politics’. These five topics occur, in total, 54.17% of the time. For the test profiles, the top five topics are ‘Questioning the Science’, ‘Politics’, ‘Not Applicable’, ‘Denial Promotion’, and ‘Covid / Health’. These topics occur 73.05% of the time. This distribution indicates that the control profiles contained more diverse topics. Additionally, ‘Covid / Health’ appears in the test profiles 7.22% of the time, while in the control profiles, this is only 1.94%.

Table 1. Manual Content Analysis Results, Twitter, 2021-2023.

Topic - Control	Prevalence (n=360)	Topic - Test	Prevalence (n=360)
Action Campaigns	15.00%	Questioning the Science	19.72%
General Climate Change	11.94%	Politics	16.11%
Social Media	11.67%	NA	15.83%
Media	7.78%	Denial Promotion	14.17%
Politics	7.78%	Covid / Health	7.22%
Event Promotion	7.50%	Energy	6.11%
Energy	7.22%	Media	3.89%
Extreme Events	7.22%	Action Campaigns	3.06%
Anti Denialism	5.56%	Economy / Money	2.78%
NA	4.17%	Other	11.15%
Economy / Money	3.06%		
Electricity Related Solutions	3.06%		
Other	8.06%		

3.2 LDA Topic Modelling Results, Twitter, 2021 – 2023

The LDA topic model results below (Table 2) indicate that the most frequently occurring topic for the control profiles was ‘General Climate Change’, at 16.81%, while, for the test profiles, this was ‘Questioning the Science’, at 29.99%. In the control profiles, the second most frequently occurring topic was labelled ‘Not Applicable’, meaning the topic contained terms which did not indicate a distinctive topic. The second most frequently occurring topic for the test profiles was labelled ‘Politics’, with a frequency of 16.96%. There are some common topics between the control and test profiles, such as ‘Economy’, ‘Electric Vehicle Solutions’, and ‘Event Promotion’, amongst others. Except for ‘Politics’, all common topics across the control and test datasets have a

frequency of less than 10.00%. The LDA topic model indicates that there are the opposing topics ‘Action Campaigns’ and ‘Anti Action Campaigns’ within the control and test profiles, with a frequency of 12.85% and 3.30%, respectively.

Table 2. LDA Topic Modelling Results, Twitter, 2021-2023

Topic - Control	Prevalence (n=115978)	Topic - Test	Prevalence (n=121959)
General Climate Change	16.81%	Questioning the Science	29.99%
NA	16.15%	Politics	16.96%
Action Campaigns	12.85%	Denial Promotion	9.85%
Impacts of Climate Change	10.26%	Extreme Events	9.79%
Energy	9.84%	Economy	6.83%
Electric Vehicle Solutions	6.69%	Emissions	6.61%
Economy	6.67%	Energy	3.61%
Environments	6.35%	Electric Vehicle Solutions	3.35%
Anti Denialism	3.91%	Covid / Health	3.34%
Politics	3.63%	Anti Action Campaigns	3.30%
Event Promotion	3.43%	Event Promotion	3.22%
International Agreements	3.40%	Sea Levels and Temperature Data	3.13%

3.3 Verification Analysis and Results

The verification process consisted of manually labelling 720 tweets to which the LDA topic model had assigned the highest probability of belonging to a single topic (Fig. 1). For the control profiles, 360 tweets were used with a probability greater than 67.77% of belonging to a single topic, while for the test profiles the 360 tweets had probabilities greater than 69.29%. All tweets were labelled blindly and then correlated with the highest probability topic for the tweet produced by the model. The accuracy of the labels was 80.00% for the control profiles and 89.70% for the test profiles. The model is consequently able to provide an accurate typological assessment for this climate change (misinformation) dataset.

4 Characterizing Latent Dirichlet Allocation Topics

The ‘Questioning the Science’ topic occupies nearly one third of all topics within the test profiles and includes discussion of climate scientists, scientific methodologies, and the consensus amongst scientists about the occurrence of climate change. This topic discusses the objectivity and integrity of scientists, as seen in some of the terms used, such as ‘corruption’, ‘deception’, ‘scam’, ‘trust’, ‘honest’, and ‘propaganda’. These words have a high co-occurrence and are associated with the values and principles of scientists. This closely relates to other studies, such as that conducted by Kovaka, who explained that some believe scientists “must be disinterested truth-seekers, free from the influence of bias or wishful thinking” [7, p. 2366]. The ‘messiness’ behind scientific

methodologies is also a subtopic within the ‘Questioning the Science’ topic. This is apparent in the use of terms such as, ‘models’, ‘failed’, ‘predict’, ‘claims’, ‘false’, ‘published’, ‘actual’, ‘analysis’, ‘computer’, and ‘reports’. Test profiles discuss the chaotic nature of scientific inquiry, pointing out that it is “not [as] rigidly organized or [as] modular as a scientific worldview seeks to present it” [8, p. 646].

The topic ‘Denial Promotion’ refers to misinformation which is being disseminated but is not explicitly questioning or undermining climate change science. This includes alternative explanations for climate change, such as grand solar minimum and natural variations in the Earth’s climate, as well as rhetorical questions such as ‘Is wind energy killing our ducks?’ and ‘what do wind turbines do when the wind doesn’t blow?’. It also includes the promotion of known denial websites, such as ‘www.wattsup-withthat.com’ and ‘www.climate-skeptic.com’. Additionally, the topic discusses challenges to users’ arguments by explaining that their comments have been previously ‘debunked’ and emphasizes that even if climate change is happening, it does not call for narratives of alarmism, crisis, and doomsday. These modelled typologies are distinct from the typologies which question climate science. ‘Denial Promotion’ discourses are not challenging or undermining the scientific institutions, various mechanisms, guiding ethe, and community practices to which scientists are committed [9].

The topic of ‘Politics’ appears in both the test (16.96%) and control (3.63%) profiles. For the control profiles, the topic includes discussing American, Canadian, and British politics and encourages users to vote. Additionally, the topic includes the role of caucuses and legislation, the need for leadership, and Donald Trump’s role in removing the USA from the Paris Agreement in 2020. In contrast, the test profiles focus on anti-left wing / anti-liberal / anti-democrat policies, with references to communism. This includes discussing how such policies might impact jobs, livelihoods, and freedom of speech. The high frequency of the ‘Politics’ topic, its anti-left focus, and the subtopics within the ‘Questioning the Science’ topic all indicate a political interpretation of (climate) science. There is a suggestion that science, like politics, is a ‘system’ which can be corrupted, rejected, and stripped of authority, sentiments which consequently enable some to reject science as illegitimate [10]. The control profiles focus on how politics, politicians and policies can benefit individuals and the relationship between such parties, while the test profiles discuss the negative impacts politics, politicians and policies might have on individuals, ideas which are rooted in anti-left-wing sentiments.

‘Action Campaigns’ and ‘Anti Action Campaigns’ are both topics which appear in the control and test profiles with a 12.85% and 3.30% frequency, respectively. The control profiles urgently call users to action, emphasizing the need to challenge and address climate change. This includes discussions about moral responsibility, youth activism / campaigns, carbon footprints, recycling methods, and reducing consumption. The test profiles, in comparison, refer to the inconvenience of action campaigns and express anti-activist sentiments. There is a particular focus on specific campaigns, such as anti-Insulate Britain and anti-Extinction Rebellion, or individuals, such as anti-Greta Thunberg. There are also subtopics related to green policies and how they are costly, socialist, and radical. These noteworthy differences provide an indication of the desired outcomes. The control profiles encourage users to participate in challenging climate change, while the test profiles focus on drawing attention to the failures of activism.

5 Conclusion, Limitations, and Future Research

In conclusion, this study uses three methods to model the most distinctive topics on climate change information and climate change misinformation profiles on Twitter. The test profiles focus on scientists, scientific methodologies, and the consensus, and suggest that all are corrupt in some form. The frequency of the ‘Politics’ topic indicates that science is viewed as a ‘system’, not a ‘method’, which can be rejected, similar to the rejection of left-wing ideologies. The control profiles, in contrast, discuss climate change in general terms and focus on calling users to action.

One of the central components of the methodology involved using various search and bias terms to identify the Twitter profiles. Future research could include other terms, such as ‘climate change falsehoods’ or ‘global warming lies’, and these may potentially produce a different set of profiles and datasets. Additionally, more combinations of the search and bias terms could be explored. Twitter remains an important site for information and misinformation dissemination, and modelling topics can provide a framework for challenging misinformation. Future research should model climate change misinformation topics on alternative platforms.

References

1. United Nations Framework Convention on Climate Change, ARTICLE 1, DEFINITIONS*, <https://unfccc.int/resource/ccsites/zimbab/conven/text/art01.htm>, last accessed 2023/01/11.
2. Tang, K.H.D.: Climate change and its impacts on mental wellbeing. *Global Academic Journal of Humanities and Social Science*, 3(4), 144-151 (2021).
3. Shu, K., Wang, S., Lee, D., Liu, H.: Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements. In: Shu, K., Wang, S., Lee, D., Liu, H. (eds.) *Disinformation, Misinformation, and Fake News in Social Media*. Lecture Notes in Social Networks. Springer, Cham (2020).
4. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-36 (2017).
5. Ballew, M.T., Leiserowitz, A., Roser-Renouf, C., Rosenthal, S.A., Kotcher, J.E., Marlon, J.R., Lyon, E., Goldberg, M.H., Maibach, E.W.: Climate change in the American mind: Data, tools, and trends. *Environment: Science and Policy for Sustainable Development*, 61(3), 4-18 (2019).
6. Belief in climate change, <https://yougov.co.uk/topics/politics/trackers/belief-in-climate-change>, last accessed 2023/01/11.
7. Kovaka, K.: Climate change denial and beliefs about science. *Synthese*, 198(3), 2355-2374 (2021).
8. Nichols, M.D., Petzold, A.M.: A crisis of authority in scientific discourse. *Cultural Studies of Science Education*, 16(2), 643-650 (2021).
9. Blancke, S., Boudry, M.: “Trust Me, I’m a Scientist” How Philosophy of Science Can Help Explain Why Science Deserves Primacy in Dealing with Societal Problems. *Science & Education*, 31(5), 1141-1154 (2022).
10. Fischer, F.: Post-Truth Populism and Scientific Expertise: Climate and Covid Policies from Trump to Biden. *International Review of Public Policy*, 4(1), 115-122 (2022).