Proceedings of the Nineteenth International AAAI Conference on Web and Social Media (ICWSM 2025)

Mapping the Climate Change Landscape on TikTok

Alessia Galdeman, Luca Maria Aiello

IT University of Copenhagen
gald@itu.dk, luai@itu.dk

Abstract

Social media platforms shape climate action discourse. Mapping these online conversations is essential for effective communication strategies. TikTok's climate discussions are particularly relevant given its young, climate-concerned audience. In this work, we collect the first TikTok dataset on climate topics. We collected 590K videos from 14K creators along with their follower networks. By applying topic modeling to the video descriptions, we map the topics discussed on the platform on a climate taxonomy that we construct by consolidating existing categorizations. Results show TikTok creators primarily approach climate through the angle of lifestyle and dietary choices. By examining semantic connections between topics, we identified non-climate "gateway" topics that could draw new audiences into climate discussions.

1 Introduction

Social media platforms are vital spaces for global climate change dialogue (Pearce et al. 2019). While enabling broader participation, understanding their impact requires analyzing emerging platforms and engagement patterns. Traditional research has primarily focused on text-based platforms such as Twitter and Reddit (Fownes, Yu, and Margolin 2018; Effrosynidis et al. 2022; Treen et al. 2022), but contemporary social media landscapes are increasingly dominated by short-form video content, particularly among younger audiences (Hautea et al. 2021). Content creators on these platforms serve as influential information disseminators and opinion leaders, wielding significant power in shaping climate discourse and potentially bridging polarized viewpoints (Falkenberg et al. 2022). However, the characteristics and dynamics of modern platforms like TikTok remain relatively unexplored in the literature (Corso, Pierri, and De Francisci Morales 2024), particularly regarding how broader topics are related to the climate change discussions. While previous studies have examined climate content on YouTube and its role in activist communities (Shapiro and Park 2018; Uldam and Askanius 2013), research on TikTok has been limited to content analysis and linguistic features (Hautea et al. 2021; Basch, Yalamanchili, and Fera 2022; Nguyen 2023) or to small-scale quantitative studies on climate video creators (Pera and Aiello 2024). As a result, there remains a crucial gap in understanding how climate discussions are shaped across modern platforms and identifying topics that act as bridges towards climate-related topics. This work addresses these gaps through two main contributions. First, we create a dataset containing 590361 climate-related TikTok videos and the social network between their $14K$ creators. We map these videos onto a taxonomy of climate topics that we obtain by combining existing literary sources. Second, we explore how different topics are related to one another, identifying non-climate topics that are semantically close to the climate discussion, thus having the potential to introduce new audiences to the climate discourse. We addressed two main research questions:

RQ1: *What are the climate topics discussed on TikTok and in what proportion? Do they cover known taxonomies?*

RQ2: *What non-climate topics on TikTok are potential 'gateway' topics for people to start engaging with climate topics?*

2 Taxonomy of Climate Topics

To create a taxonomy of climate change topics, we aggregated categorizations proposed by existing studies (Basch, Yalamanchili, and Fera 2022; Dahal, Kumar, and Li 2019; Grouverman et al. 2019; Pera and Aiello 2024; Pupneja et al. 2023). These are empirical studies on either Twitter or TikTok. They all provide topic lists extracted from their specific video collections, which often focus on particular subtopics and therefore do not exhaustively capture all the facets of the climate change discourse. To the best of our knowledge, this work represents the first attempt to create a data-driven taxonomy of climate change topics on TikTok, informed by both empirical evidence and existing literature. We first collected all topic labels from the referenced studies (51 in total). We then manually merged near-duplicates and grouped the labels into 10 semantically coherent top-level categories. The resulting taxonomy is visualized in Figure 1.

3 Data Collection

We collect all the data through the TikTok Research API (TikTok 2025), and make the dataset public[1]. First, we iden-

[1]Dataset's DOI: 10.5281/zenodo.15131067

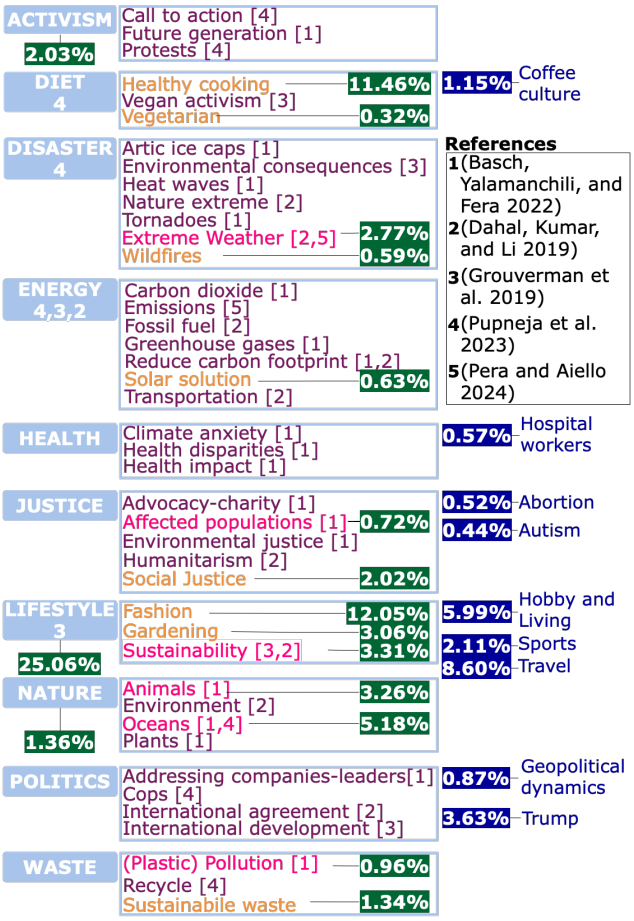

Figure 1: **Taxonomy of climate-related topics.** Labels are color-coded based on their occurrence: purple indicates presence only in previous works, orange shows topics found exclusively in our data collection, and magenta represents topics present in both. Non-climate topics are shown in blue. Superscript numbers on purple and magenta topics indicate the work where these topics were originally identified.

tify the set of hashtags that are used on TikTok to tag climate-related content. To do so, we start by filtering 6 existing lists (GreenPeace 2023; Wired 2021; PalauProject 2022; NYTimes 2023; HarvardPique 2024; EcoTokCollective 2024) of climate-change centered accounts. From these lists, we obtain 49 *seed users*, highly active on TikTok and specialized in posting videos related to climate change. We deliberately exclude news outlets from this list as they are not climate change-focused accounts. To mitigate the potential bias of this initial selection, we augment the seed set with a snowball expansion (Goodman 1961). From the $16,533$ accounts followed by our seed users, we selected the top 0.5% based on the number of follows from seed users. We then manually reviewed these accounts for topical relevance, adding the climate-dedicated profiles to our seed set. This step yields 56 new seed users, for a total of 105. We extract the metadata of all $17,338$ public videos posted by seed users, capturing their complete posting history from their first video on the platform until the time of data collection (October 2024). We manually review the relevant hashtags in the 99^{th} percentile of the frequency distribution and excluded irrelevant hashtags such as engagement tags (e.g., `#viral`). Since our aim is to capture also the broader topical context around climate change discussions, we keep both climate-related keywords, as well as broader contextual hashtags that are still relevant to our taxonomical categories (e.g., `#politics`). The final list contains 193 hashtags, which we manually assign to the most fitting taxonomical branches (Table 1 in Appendix).

We query the API to collect videos for each hashtag separately. To retrieve a comparable number of videos across the ten taxonomical categories, we define a per-hashtag video quota q_h that is inversely proportional to the number of hashtags in its category, namely: $q_h = \frac{37*100}{T(h)}$ where $T(h)$ is the number of hashtags in the taxonomical category of hashtag h and 37 is the maximum value of $T(h)$. We collected videos in two-week intervals for each hashtag, from 16 October 2021 to 15 October 2022. This timeframe encompasses significant climate-related events including COP26, major climate protests, and extreme weather events. In total, we collected 1,326,684 videos from 199,968 users.

Finally, we construct the follower network among users. We first restrict our sample to the $32,765$ users with a minimum activity level of two videos posted within the timeframe of interest. We then collect the $536,585$ follower connections outgoing from these creators, and retained the links whose both endpoints were in our user set. After discarding singleton nodes, we were left with a network of 13867 *creators* connected by $161,891$ links.

4 Climate Topics on TikTok

Hashtags are only broadly indicative of a video's theme. To obtain a more detailed description of the topics discussed in the videos we gather, we augment the video metadata with the video descriptions. To do that, we focus on the 590361 video English-language descriptions longer than three words after the removal of stopwords and non-English words. For each video, we create a document by concatenating its description text and hashtags. To these documents, we apply a BERTopic pipeline (Grootendorst 2022) that i) generates sentence-BERT embeddings for each document (Reimers 2019); ii) reduces the embedding dimensionality with UMAP (McInnes, Healy, and Melville 2018); and applies HDBSCAN clustering to group descriptions into topics (Campello, Moulavi, and Sander 2013). The optimal parameters for this pipeline were determined through Randomized Cross-Validation search, maximizing the relative validity measure (Moulavi et al. 2014). The hyperparameters corresponding to the best combination are: `min_cluster_size` $= 1000$, `min_samples` $= 500$, `cluster_selection_epsilon` $= 0.05$. This resulted in 45 distinct clusters. We label the topics using the support of a Large Language Model (LLM). We randomly sample 20% of the documents from each cluster and parse each of them with the `gpt-4o-mini` model (OpenAI API (OpenAIResearch 2024)) to assign a single short video label. In

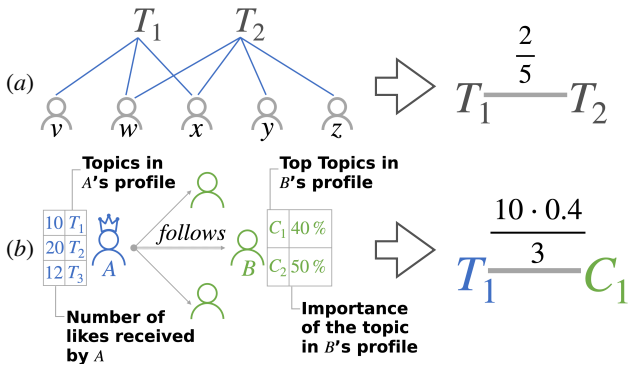

Figure 2: (a) **Projection of the user-topic bipartite network.** Topic T_1 is discussed with high frequency by 3 users (v, w, z), while topic T_2 by 4 users (w, x, y, z). In the projected graph on the topic layer, the edge (T_1, T_2) has $w = \frac{2}{5}$. (b) **Climate and non-climate bridge topic pairs.** Non-climate user A has received 10 likes on topic T_1. This is divided by its 3 neighbors and weighted by 40% as it represents the strength of topic C_1 for climate-user B. So, the weight of the edge (T_1, C_1) concerning the couple of users (A, B) is $\frac{10 \cdot 0.4}{3}$.

a second stage, we ask the same model to consolidate the individual topics into a single, coherent cluster label. Both the individual document labeling and the consolidation steps were performed independently on all 45 clusters (prompts in Appendix). Finally, we manually checked and grouped similar topics whose labels have high semantic similarity (e.g., *Sustainable fashion* and *Thrift fashion* under *Fashion*), which results in 26 topics, 17 of which are related to climate.

In Figure 1, climate-related subtopics, along with their volume percentage, are shown in one of the three colors: *(i)* purple, if they appear only from a theoretical point of view; *(ii)* orange, if they appear only in our data collection; *(iii)* magenta, if they appear in both cases. Non-climate topics are colored in blue and positioned around their parent topic to complete the taxonomical structure. TikTok content covers only a few taxonomical categories, with most videos focusing on climate topics related to either *Lifestyle*, *Diet*, or *Nature*.

5 Boundaries of Climate Discussion

We explore the interface between climate and non-climate content on TikTok, with the objective of charting potential gateway topics to introduce new users to the climate discussion. To do so, we use two complementary analytical perspectives. The first estimates the similarity between topics by leveraging the co-occurrence of topics in the video production of TikTok creators (*topic perspective*). The latter looks at follower links from creators who do not typically discuss climate-related topics to climate creators (*user perspective*).

To implement **topic perspective**, we construct a bipartite network that connects users to their respective topics. An edge from user u to topic t has a weight w representing the proportion of the user's video production dedicated to that particular topic. To focus on strong topic associations, we retain only links with weights in the top quartile of the distribution. We used the Jaccard index to project this bipartite network onto the topic dimension. The resulting network is an undirected graph where the edge connecting topics T_1 and T_2 is weighted by the ratio of the number of creators who have uploaded videos in both T_1 and T_2 over the number of creators who have uploaded videos on either T_1 or T_2 (Figure 2a).

To implement **user perspective**, we first profile creators to distinguish climate creators from non-climate creators, and then look at the following links between them. We characterize creators with a 26-dimensional profile vector that encodes the fraction of videos they published on each topic. We then compute a *greenness* score for each user, defined as the percentage of their published videos that belong to climate-related topics. We use the follow network between creators and our established greenness metric to evaluate how climate-related videos might gain new viewers through the recommendations of influential creators whose production is not focused on climate, but who follow climate-related creators that they could advertise. We stick to a conservative scenario where we look at follower links from creators with no video on climate (0% *greenness*) to those with full focus on climate topics (100% *greenness*). We consider only non-climate top influencers above the 75th percentile of the distribution of number of likes received.

For each edge connecting a non-climate node to a climate node, we create the respective topic pairs, selecting all non-climate topics for the first node and the most discussed topics for the climate node (i.e., covering at least 62.5% of the user profile, corresponding to the 90th percentile of the coverage distribution). For a non-climate node A and climate node B, the weight of a topic pair (T_1, C_1) is calculated as the number of likes that A received on topic T_1 (a proxy of A's audience reach on topic T_1), divided by A's out-degree (i.e. number of people that user A is following, to simulate an equal probability of A to advertise content of any climate-related accounts among its contacts), and multiplied by the importance of C_1 in B's profile (to model the likelihood that a user would be exposed to topic C_1 when landing on B's page). Formally: $w(T_1, C_1)_{A,B} = \frac{L(T_1)_A \cdot p(C_1)_B}{degree(A)}$, where $p(C_1)_B$ is the percentage of B's videos related to topic C_1 (see example in Figure 2b). The final weight for each topic pair is the sum of weights across all user pairs.

The relationships between topics are shown in Figure 3. From the topic perspective (Figure 3a), we observe stronger relationships among climate-related topics, especially among *lifestyle* ones. Other notable high-value pairs include *gardening* with *healthy cooking*, and *social justice* with *climate activism*. Concerning pairs between climate and non-climate topics, *travel* emerges as a significant bridge towards climate-related content, showing substantial overlap with *lifestyle*, *extreme weather*, *oceans*, and *animals*. Additionally, noteworthy patterns emerge around *Trump*-related content, which shows strong co-occurrence with both *lifestyle* topics and *social justice* discussions, indicating the intersection of political discourse with both ev-

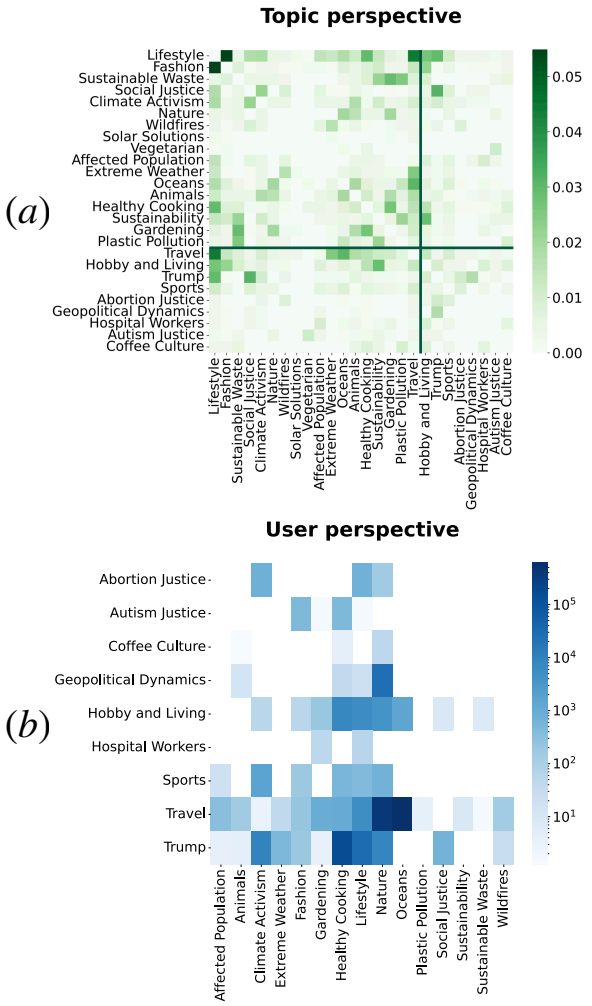

Figure 3: **Weights on topic pairs.** (a) *Topic perspective.* The heatmap reflects the proportion of shared userbase. (b) *User perspective.* Rows correspond to non-climate topics while columns collect climate-related topics. The weights represent the potential visibility that non-climate topics could bring to climate-related topics.

eryday life and environmental advocacy.

From the user perspective (Figure 3b), we see that influential creators focusing on *travel*, *Trump*, and *hobby and living*-related content show the highest potential for directing audience attention toward climate topics. Among these, *travel*-focused creators demonstrate the strongest potential for engaging their audiences with climate-related content, particularly with topics that share semantic proximity to travel, such as *oceans*, *nature*, *animals*, and *healthy cooking*. *Hobby and living*-oriented creators show promising pathways for influencing engagement with climate-related content such as *healthy cooking*, *lifestyle*, *nature* and *oceans*. Finally, creators focusing on *Trump*-related content exhibit unexpected potential for directing audiences toward *healthy cooking*, *lifestyle* and *nature* topics, as well as *climate activism* and *social justice*.

6 Discussion and Conclusion

The scope of this study is limited by the constraints that the TikTok APIs put on the volume of information that we could gather during three months of data collection. Additionally, our taxonomy is guided by empirical studies, mainly conducted on Twitter, which together with the manual categorization phase could have influenced the final classification structure; future work could expand and refine our taxonomy. Nevertheless, our work has produced the first public dataset to study the climate discourse on TikTok, while advancing our understanding of climate communication on the platform in two ways. First, we provide a comprehensive mapping of climate-related topics grounded in previous works, and complement it with empirical data to estimate the public engagement on these topics on TikTok. Surprisingly, we find that many topics that are relevant to climate action are not prominently featured on the platform. Second, we identify "gateway" topics (e.g., certain themes on politics and travel) with the potential for engaging users who might not primarily seek climate-related content, therefore broadening the reach of climate communicators. This work opens several promising avenues for future research, including the study of the coordination patterns among content creators, and the analysis of the potential impact of endorsement campaigns by non-climate influencers on climate-related topic engagement. Such studies could provide valuable insights into effective strategies for expanding climate awareness and engagement across diverse audience segments on social media platforms. Importantly, our findings can serve as practical guidelines for campaigns aimed at engaging young audiences in topics related to climate change and climate action.

7 Related Work

While TikTok has emerged as an influential social media platform, academic research examining its content remains limited, with most previous work on this topic focusing on Twitter (Dahal, Kumar, and Li 2019; Grouverman et al. 2019). Existing studies on TikTok have primarily focused on content analysis across select domains, including educational content (O'Donnell, Jerin, and Mu 2023), autism awareness (Gilmore et al. 2024), and preliminary investigations into climate change communication (Basch, Yalamanchili, and Fera 2022; Hautea et al. 2021; Nguyen 2023). Some studies have looked at TikTok usage through surveys (Omar and Dequan 2020). Basch, Yalamanchili, and Fera (2022) conducted one of the first analyses of videos tagged with `#climatechange` on TikTok, while (Pera and Aiello 2024)) provided early comparative insights through a cross-platform analysis of climate discourse on TikTok and YouTube on a small set of creators. In this work we looked at the climate discussion from the semantic perspective, using topic modeling. However, previous studies have approached the study of climate discourse from alternative angles including arguments for climate delay (Lamb et al. 2020), narratives around climate change (Fløttum and Gjerstad 2017), and historical perspectives on the opinions and goals around climate and the climate transition (Rahman 2013).

Ethical Statement

This study engages with user-generated content on TikTok, where content moderation is platform-dependent. While our analysis maps the landscape of climate change discourse, we acknowledge that some content may express views in ways that could be problematic or counterproductive to constructive collective action. Our broad analytical approach focuses on identifying and understanding the topical structure of climate-related discussions, rather than evaluating individual content for accuracy or potential harm. We recognize the inherent challenges of studying platforms with inconsistent content regulation and the importance of considering their ethical implications. Furthermore, we acknowledge that the methodological framework presented here could potentially be repurposed for harmful objectives, such as the strategic dissemination of misinformation. While mindful of these risks, we maintain that the potential benefits of this research for addressing climate change challenges outweigh these concerns. Nevertheless, we are dedicated to promoting responsible implementation of our methods/dataset and actively discouraging their misuse.

Acknowledgments

We acknowledge the support from the Carlsberg Foundation through the COCOONS project (CF21-0432).

References

Basch, C. H.; Yalamanchili, B.; and Fera, J. 2022. #Climate Change on TikTok: A Content Analysis of Videos. *Journal of Community Health*, 47(1): 163–167.

Campello, R. J.; Moulavi, D.; and Sander, J. 2013. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, 160–172. Springer.

Corso, F.; Pierri, F.; and De Francisci Morales, G. 2024. What we can learn from TikTok through its Research API. In *16th ACM Web Science Conference*, 110–114.

Dahal, B.; Kumar, S. A. P.; and Li, Z. 2019. Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, 9(1): 24.

EcoTokCollective. 2024. EcoTokCollective - Our Team. https://www.ecotokcollective.com/our-team.

Effrosynidis, D.; Karasakalidis, A. I.; Sylaios, G.; and Arampatzis, A. 2022. The climate change Twitter dataset. *Expert Systems with Applications*, 204: 117541.

Falkenberg, M.; et al. 2022. Growing polarization around climate change on social media. *Nature Climate Change*, 12(12): 1114–1121.

Fløttum, K.; and Gjerstad, Ø. 2017. Narratives in climate change discourse. *Wiley Interdisciplinary Reviews: Climate Change*, 8(1): e429.

Fownes, J. R.; Yu, C.; and Margolin, D. B. 2018. Twitter and climate change. *Sociology Compass*, 12(6): e12587.

Gilmore, D.; Radford, D.; Haas, M. K.; Shields, M.; Bishop, L.; and Hand, B. 2024. Building community and identity online: a content analysis of highly viewed# autism TikTok videos. *Autism in Adulthood*, 6(1): 95–105.

Goodman, L. A. 1961. Snowball sampling. *The annals of mathematical statistics*, 148–170.

GreenPeace. 2023. Climate Voices on Social Media of 2023. https://www.greenpeace.org/international/story/64559/climate-voices-on-social-media-2023/.

Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Grouverman, A.; Kollanyi, B.; Howard, P.; Barash, V.; and Lederer, T. 2019. Climate change consensus and skepticism: mapping climate change dialogue on Twitter and Facebook.

HarvardPique. 2024. Climate Creators to Watch 2024 Website. https://climatecreatorstowatch.com/.

Hautea, S.; Parks, P.; Takahashi, B.; and Zeng, J. 2021. Showing they care (or don't): Affective publics and ambivalent climate activism on TikTok. *Social media+ society*, 7(2): 20563051211012344.

Lamb, W. F.; Mattioli, G.; Levi, S.; Roberts, J. T.; Capstick, S.; Creutzig, F.; Minx, J. C.; Müller-Hansen, F.; Culhane, T.; and Steinberger, J. K. 2020. Discourses of climate delay. *Global Sustainability*, 3: e17.

McInnes, L.; Healy, J.; and Melville, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Moulavi, D.; Jaskowiak, P. A.; Campello, R. J.; Zimek, A.; and Sander, J. 2014. Density-based clustering validation. In *SIAM international conference on data mining*, 839–847.

Nguyen, H. 2023. TikTok as learning analytics data: Framing climate change and data practices. In *LAK23: 13th International Learning Analytics and Knowledge Conference*.

NYTimes. 2023. With TikTok and Lawsuits, Gen Z Takes on Climate Change. https://www.nytimes.com/2023/08/19/climate/young-climate-activists.html.

Omar, B.; and Dequan, W. 2020. Watch, share or create: The influence of personality traits and user motivation on TikTok mobile video usage.

OpenAIResearch. 2024. OpenAI Terms of Use. https://openai.com/policies/usage-policies/.

O'Donnell, N.; Jerin, S. I.; and Mu, D. 2023. Using TikTok to educate, influence, or inspire? A content analysis of health-related EduTok videos. *Journal of Health Communication*, 28(8): 539–551.

PalauProject. 2022. Environmental Influencers Using TikTok To Fight Climate Change 2022. https://www.palauproject.com/blog-post/environmental-influencers-tiktok-climate-change.

Pearce, W.; Niederer, S.; Özkula, S. M.; and Sánchez Querubín, N. 2019. The social media life of climate change: Platforms, publics, and future imaginaries. *Wiley interdisciplinary reviews: Climate change*.

Pera, A.; and Aiello, L. M. 2024. Shifting Climates: Climate Change Communication from YouTube to TikTok. In *ACM Web Science Conference*, 376–381. Stuttgart Germany: ACM. ISBN 9798400703348.

Pupneja, Y.; Zou, J.; Lévy, S.; and Huang, S. 2023. Understanding Opinions Towards Climate Change on Social Media. ArXiv:2312.01217.

Rahman, M. I.-u. 2013. Climate change: A theoretical review. *Interdisciplinary Description of Complex Systems: INDECS*, 11(1): 1–13.

Reimers, N. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084*.

Shapiro, M. A.; and Park, H. W. 2018. Climate change and YouTube: Deliberation potential in post-video discussions. *Environmental Communication*, 12(1): 115–131.

TikTok. 2025. TikTok for Developers. https://developers.tiktok.com/doc/overview.

Treen, K.; Williams, H.; O'Neill, S.; and Coan, T. G. 2022. Discussion of climate change on Reddit: Polarized discourse or deliberative debate? *Environmental Communication*, 16(5): 680–698.

Uldam, J.; and Askanius, T. 2013. Online civic cultures? Debating climate change activism on YouTube. *International Journal of Communication*, 7: 20.

Wired. 2021. Meet the Climate Change Activists of TikTok. https://www.wired.com/story/climate-change-tiktok-science-communication/.

Paper Checklist

- For most authors...
  1. Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes.
  2. Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes.
  3. Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes.
  4. Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we discuss potential biases of our data in the "Discussion and conclusion" section.
  5. Did you describe the limitations of your work? Yes, limitations are presented and discussed in the "Discussion and conclusion" section.
  6. Did you discuss any potential negative societal impacts of your work? Yes, we address negative societal impact in the "Ethical considerations" section.
  7. Did you discuss any potential misuse of your work? Yes, we discuss potential misuse in the "'Ethical considerations" section.
  8. Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, the anonymized data is shared and documented in the anonymised url.
  9. Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
- Additionally, if your study involves hypotheses testing...
  1. Did you clearly state the assumptions underlying all theoretical results? Yes.
  2. Have you provided justifications for all theoretical results? Yes.
  3. Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
  4. Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes.
  5. Did you address potential biases or limitations in your theoretical framework? Yes, these are addressed in the "Discussion and conclusion" section.
  6. Have you related your theoretical results to the existing literature in social science? Yes, results are related to existing literature in two sections: "Climate Topics on TikTok" and "Related Works".
  7. Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes, it is discussed in the "Discussion and conclusion" section.
- Additionally, if you are including theoretical proofs...
  1. Did you state the full set of assumptions of all theoretical results? NA.
  2. Did you include complete proofs of all theoretical results? NA.
- Additionally, if you ran machine learning experiments...
  1. Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes.
  2. Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, HDBSCAN clustering details are reported in the "Climate Topics on TikTok" section.
  3. Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
  4. Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
  5. Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes.
  6. Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes, discussed in the "Climate Topics on TikTok" section.
- Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
  1. If your work uses existing assets, did you cite the creators? Yes, we used the tiktok and OpenAI API.

2. Did you mention the license of the assets? Yes. we provide the links to the TikTok terms in "Data Collection" and the ones by OpenAI in "Climate Topics on TikTok"
3. Did you include any new assets in the supplemental material or as a URL? Yes, anonym.
4. Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA.
5. Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA, the identity of creators has no role in the study, as explained in the Datasheet
6. If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? Yes, as cited in the "Ethical considerations" section.
7. If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? Yes, in the GitHub repository.

- Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
  1. Did you include the full text of instructions given to participants and screenshots? NA.
  2. Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA.
  3. Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA.
  4. Did you discuss how data is stored, shared, and deidentified? NA.

Appendix

Hashtags for video collection

TOPIC	N. HASHTAGS	HASHTAGS LIST
ACTIVISM	6	climateaction, greenwashing, earthmonth, savetheearth, genzforchange, earthday
CLIMATE	9	climatechange, climatecrisis, environmentalist, climate, globalwarming, climativity , eco, environment, climatesolutions
DIET	13	plantbased, recipe, recipes, food, foodie, homemadekitchenrestock, easyrecipes, vegan, cooking, kitchenrestock, foodtiktok, vegetarian, sourdough
DISASTER	32	climateemergency, hurricane, weather, tornado, floods, tornadoes, storm, cyclone, flood, rain, flooding, disaster, naturaldisaster, fire, extremeweather, wildfire, storms, severeweather, wild, heatwave, earthquake, stormchasing, extreme, tornadowarning, hurricanes, heatwaves, hurricaneseason, emergency, water, thunderstorm, blizzard, disasters
ENERGY	6	stopwillow, cleanenergy, renewableenergy, fossilfuels, pollution, bigoil
HEALTH	8	goodnews, livebrightly, health, spreadlove, hope core, climateoptimism, goodclimatenews, todaysgoodnews
JUSTICE	20	community, climatejustice, environmentaljustice, indigenous, nativetiktok, indigenoustiktok, native, socialjustice, nativeamerican, justice, humanrights, brown, browntiktok, muslims, indigenousrights, environmentalracism, navajo, blacktiktok, muslimtiktok, racialjustice
LIFESTYLE	37	sustainability, sustainable, sustainableliving, ecotok, ecohome, minimalist, slowliving, thrifted, sustainablestyle, diy, sustainablelifestyle, sustainabilitytiktok, fastfashion, travel, vintage, thrift, thriftedstyle, environmentallyfriendly, ecofriendly, sustainablefashion, sustainabletravel, ecofriendlyliving, minimalism, consciousconsumer, slowfashion, secondhand, thrifting, homemade, gardening, garden, sundayreset, ecotravel, vanlife, sustainableswaps, thriftstorefinds, ecofriendlyproducts, simpleliving
NATURE	17	nature, beach, ocean, oceanconservation, animals, sand, earth, wildlife, amazon rainforest, marinebiology, savetheocean, marinebio, marinescience, bees, marineconservation, oceanlife, conservation
SCIENCE	8	learnontiktok, acespace, biology, education, science, research, womeninstem, moreyouknow
POLITICS	15	cop27, politics, cop28, cop, democrat, politicstiktok, conservative, cop28dubai, liberal, democrats, republicans, political, politicaltiktok, cop15, cop28_uae
WASTE	22	lowwaste, plasticpollution, trash, litter, compost, plastic, upcycle, reuse, upcycling, bandisposablevapes, foodwastetip, beachcleanup, zerowaste, plasticfree, cleanup, foodwastesolution, recycle, foodwasteprevention, trashboy, recycling, reusable, zerowasteliving

Table 1: Hashtag list used for the video collection, with the relative macrotopic.

Prompts

Prompt 1: "`I have a topic that contains the following document:` **`doc.`** `Based on the above information, can you give a short label of the topic (max 2 words)? Make sure you to only return the label and nothing more.`"
where **`doc`** is one single video, specifically the concatenation of description and hashtag list after cleaning (stopwords and punctuation removal, filtered out words not in the English corpus)
Prompt 2: "`I have topic that is described by the following keywords:` **`subtopics`**`.Based on the above information, can you give a short label of the topic (max 2 words)? Make sure you to only return the label and nothing more.`".
where **`subtopics`** refers to the concatenation of all subtopic extracted for the sampled document in the cluster.