Analysis of Climate Change Misleading Information in TikTok

Clara Baltasar[1], Sergio D'Antonio Maceiras[1], Alejandro Martín[1] and David Camacho[1]

[1]*Department of Computer System Engineering, Universidad Politécnica de Madrid, Calle de Alan Turing, 28031 Madrid, Spain*

Abstract

In today's digital landscape, social media platforms have become important areas for disseminating information, ranging from legitimate discourse to misinformation, especially on critical topics such as Climate Change. This study employs claim detection and clustering techniques to analyze misleading information within an initial dataset of videos. Initially, the study identified 5,352 videos out of a total of 8,151 that warranted further investigation. Utilizing clustering methods, it was discovered that the prevalence of misinformation was surprisingly lower than anticipated. Most of the clusters showcases content that promotes sustainability and raises environmental awareness, strengthened by corroborated information of fact-checking agency EFE Verifica. Conversely, there are two clusters that focuses on videos propagating misinformation, conspiracy theories, active discussion and debate, highlighting the necessity of consuming media with caution. Looking ahead, combating misinformation necessitates the enhancement of digital literacy and the cultivation of critical thinking skills. This research aims to leverage technology and verified information from credible organizations to identify, analyze, and mitigate the influence of misleading content on social media, thus better understanding its dynamics and reducing its adverse impacts.

Keywords

Climate change, TikTok videos, Misleading information, Environmental issues

1. Introduction

The spread of climate change misinformation on TikTok has become a serious concern for both environmental scientists and social media analysts. Understanding the dynamics and impact of this phenomenon requires in-depth analysis of the most viral content on the platform Viral content on TikTok often has several key characteristics: visually appealing, emotional and easy to understand. In the context of climate change misinformation, these elements are often used to attract audiences and spread false narratives. For example, many viral videos contain dramatic images of natural disasters or melting ice caps combined with misleading or inaccurate explanations. These videos often contain sensational and apocalyptic predictions, which can evoke strong emotional responses and encourage users to share the content widely.

A perfect example is videos falsely claiming that climate change is a fraud committed by governments or corporations for financial gain. These clips often feature conspiracy theory imagery, which is particularly appealing to viewers who are already skeptical of mainstream science. Additionally, such content creators often pretend to be tipsters, exploiting the feeling of inside information to increase their authority and attract more views.

TikTok's viral misinformation patterns are influenced by the platform's algorithm, which prioritizes content that generates high levels of engagement. Videos with lots of likes, comments and shares appear more frequently on the For You page, which exponentially increases their reach. This creates a feedback loop in which misinformation continuously reaches new users, thereby increasing its spread.

Collaboration features also play a key role in the spread of misinformation. These features allow users to directly interact with existing videos by adding their own comments or creating reaction videos.

Proceedings of the 1st Workshop on COuntering Disinformation with Artificial Intelligence (CODAI), co-located with the 27th European Conference on Artificial Intelligence (ECAI), pages 54–61, October 20, 2024, Santiago de Compostela, Spain
cm.baltasar@upm.es (C. Baltasar); sergio.dantonio@upm.es (S. D. Maceiras); alejandro.martin@upm.es (A. Martín); david.camacho@upm.es (D. Camacho)

Not only does this increase the visibility of the original content, but it also creates a sense of community and brainstorming around misinformation, which further contributes to its virality.

The widespread dissemination of climate change misinformation on TikTok has significant implications for public perception. Studies such as [1] have shown that repeated exposure to misinformation can lead to the formation of false beliefs and increased skepticism towards scientific consensus. On TikTok, where the user base skews younger, the impact is particularly concerning. Young users are still developing their understanding of complex issues like climate change and their perceptions can be heavily influenced by the content they consume on social media. Furthermore, viral misinformation can undermine efforts to promote accurate scientific information.

In this research, we undertake an analysis of misinformation around climate change in TikTok. We filtered 5,352 from a total of 8,151 videos related to conversations around climate. Then, we extracted information from the videos, such as keywords and the transcription. Then we applied clustering algorithms to identify different subtopics and conversations, which allowed us to identify groups promoting negationist theories about climate change.

The rest of this article is organised as follows. Section II presents an analysis of the state-of-the-art literature, Section III describes the methodology, Section IV the results, Section V analyses the presence of hoaxes and finally, Section VI presents a series of conclusions.

2. State of the art

In the digital age, social media platforms have become a prime battleground for the dissemination of information and the spread of ideas, both legitimate and otherwise. One such arena is the discussion surrounding climate change, where the impact of social media, AI and algorithmic systems has become a growing concern [2]. The proliferation of misinformation and disinformation on social media platforms has garnered significant attention from researchers, especially concerning critical issues like climate change. This section reviews some relevant contributions in this area, identifying the main trends and findings in the existing literature.

Several studies as [3], agree that fake news is an old concept; it has existed and will exist as long as publishers continue to use misleading information to promote their interests and this has been happening since before the Internet even existed. Nowadays, as [4] points out, misleading information can come in various forms, such as fake new, disinformation or misinformation, which are easily spread through social media.

Many studies have investigated the spread of misinformation on social media. [5] examined the mechanisms through which misinformation is disseminated on platforms like Twitter. They presented a language model who detects fake news spreaders on Twitter. In the same way, other studies as [6] examines the stability and evolution of network structure and discussion topics among a group of prominent climate change deniers. The findings reveal that while the climate change denial network remains stable in terms of size and core group composition, sub-groups continuously emerge and dissolve.

[7] explored the dynamics of misleading information on environmental issues on TikTok, a relatively new and rapidly growing social media platform. This study found that TikTok's short-form video content and highly engaging user interface create a fertile ground for the viral spread of both accurate and inaccurate information. This study, also emphasized the importance of science education in addressing misleading information.

Also, the purpose of the study [8], was to describe content related to climate change on TikTok. Their findings indicate that climate change is being represented on TikTok as a legitimate and anxiety provoking issue. Although only a few videos included in their sample are disinformation, these garnered millions of views. Therefore, they concluded that the presence of credible professionals is essential on platforms like TikTok to increase the chances that messages are as complete as time allows, while also being scientifically sound.

3. Architecture development

This section describes the methodology employed to extract, filter and analyze all extracted information from the social network used in this research: TikTok.

3.1. Data collection

The initial phase of our research involved the systematic collection of data using TikTok's research API designed for developers [9].

We meticulously extracted metadata from videos associated with a selection of environmentally relevant hashtags. These hashtags included #climatechange, #ecofriendly, #sustainability and #ecotok.

Upon securing the initial dataset, we conducted a comprehensive analysis to identify the most frequently used hashtags within these videos. This analysis led to the discovery of additional pertinent hashtags such as #zerowaste, #naturetok, #globalwarming, #climatecrisis, #savetheplanet, #ecology, #plasticfree, #sustainable, #savetheworld, #recycle, #recycling, #upcycling, #saveourplanet, #upcycle, #bethechange, #environment, #climateaction, #climateemergency, #climate, #plasticpollution, #plasticwaste, #savetheocean, #saveouroceans and #eco. This iterative process was repeated meticulously until we had compiled an exhaustive list of the most prevalent hashtags in our dataset, which are those shown above.

Subsequent to the environmental hashtag analysis, we turned our focus to the investigation of misinformation related to climate change. We identified and collected metadata associated with hashtags that propagate misinformation, such as #climatelies, #climatehoaxx, #climatehoaks, #climatelie, #climatehoax, #globalwarmingisfake, #globalwarminghoax, #globalwarmingisahoax, #carbonkleptomania, #globalcooling, #climatechangehoax, #noclimateemergency, #climatescam, #weathermanipulation, #stopglobalwarming and #globalwarmingsucks. And just like before, we repeated this iterative process until we had compiled an exhaustive list of the most prevalent hashtags that propagate misinformation in our dataset, which are those shown above.

Through this rigorous process, we accumulated a total of 8,151 video metadata entries, covering the period from January 2020 to June 2024.

3.2. Data characterization in TikTok

Once all videos were retrieved, the second step was to extract the audio channel, converting from .mp4 to format .mp3, and then we used the Whisper model [10] which is a pre-trained model for automatic speech recognition and speech translation, to get the transcription from the audio of the videos. We obtained a total of 6,998 text transcriptions and we started to work with that.

The next step was to analyze the transcriptions. First, we divided the transcriptions into sentences and two pre-trained models were loaded: the SentenceTransformer model [11], for creating dense vector representations of sentences, and the KeyBERT model [12], for extracting keywords from the transcriptions.

Our aim was to do clustering to all transcriptions and to obtain the most relevant keywords of each cluster. For this, to find the optimal number of clusters for K-Means, the elbow method and silhouette scores were used, and it was found that this number was ten clusters.

Then, we applied K-Means clustering to the embeddings and when we analyzed the most relevant keywords of each cluster we obtained some keywords that didn't make sense on the issue of climate change, like "foryourpage", "fyp" or "for watching", so we established them as stopwords but the results obtained were not much better. After trying several things we realized that many of the videos only had songs or phrases that were not very relevant as sound, so we decided to leave the transcriptions aside.

The second option was to use the "video_description" that we obtained when we extracted metadata from videos using TikTok's research API designed for developers [9], which owns the description of the video, and apply SentenceTransformer [11] and KeyBERT model [12] to this descriptions.

As we mentioned before, we accumulated a total of 8,151 video metadata entries, covering the period from January 2020 to June 2024, so the data collection we used in the end was larger than expected.

To find the optimal number of clusters for K-Means, the elbow method and silhouette scores were used, and it was found that this number was seven clusters. Then, we applied K-Means clustering to the descriptions of the videos and we obtained seven clusters well-defined, which are shown in the section 4.

4. Results

The image below, Fig. 1, represents the clustering of descriptions from 8,151 TikTok videos. We employed K-means clustering and Principal Component Analysis (PCA) to visualize and understand the thematic grouping of these transcriptions. The scatter plot shows the distribution of seven clusters, differentiated by colors. This section provides a detailed analysis of these clusters, discussing the most relevant keywords, potential misinformation and aggregate statistics for likes, views and comments.

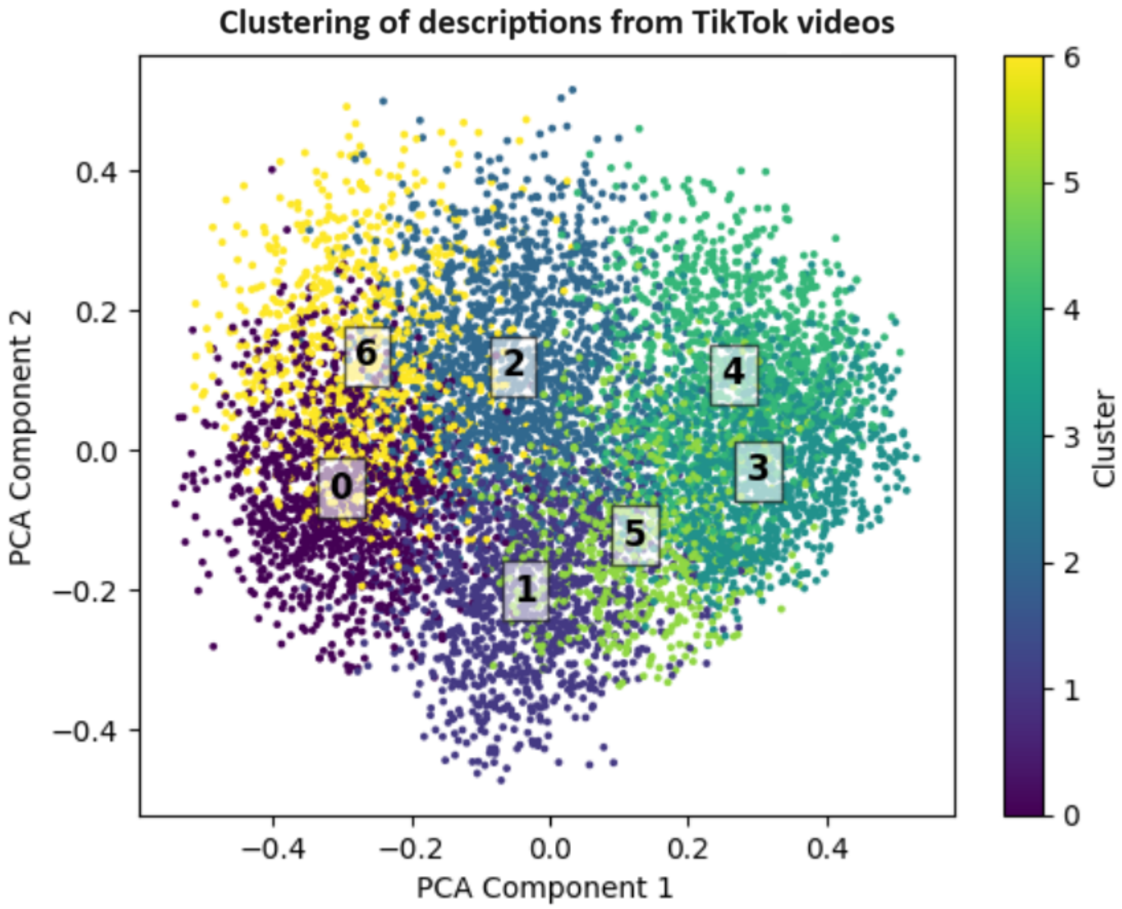

Figure 1: Clustering of descriptions from TikTok videos.

The clustering algorithm identified seven distinct clusters among the videos descriptions. Each cluster represents a group of videos with similar content based on their descriptions. The cluster analysis is presented below:

- **Group 0**. This cluster, represented in the scatter plot by the color purple, contains 1,260 videos. This cluster is characterized by prominent keywords like "sewing", "upcycling", "thrift", "thriftflip" and "fashion". This suggests a focus on sewing techniques, clothing upcycling and sustainable fashion practices.
- **Group 1**. It is identified in the scatter plot by the color dark blue, comprises 1,300 videos. This cluster is characterized by content related to TikTok, duets, life hacks and fashion. Common

keywords include "duet with", "duet", "lifehacks" and "tiktok". This indicates a focus on social interaction and entertainment.

- **Group 2**. It is represented in the scatter plot by the color blue, contains 1,665 videos. This cluster is distinguished by its focus on sustainability, environment and ecological awareness. Keywords include "sustainability", "environment", "eco", "nature" and "ecofriendly". This reflects a strong interest in sustainable practices and environmental protection.
- **Group 3**. Identified in the scatter plot by the color turquoise, comprises 1,243 videos. This cluster is characterized by discussions on weather manipulation, chemtrails and climate-related conspiracy theories. Keywords include "weathermanipulation", "chemtrails", "geoengineering", "weathermodification" and "conspiracytiktok".
- **Group 4**. This cluster, represented in the scatter plot by the color green, contains 1,139 videos. This cluster focuses on climate change, global warming, climate action and its controversies. Common keywords include “climatechange”, “globalwarming”, “climatecrisis”, “climateaction” and “gretathunberg”. Also included are terms that deny or question climate change (“climatelies”, “globalwarmingisfake” and “climatehoaxx”). The activist “gretathunberg” is mentioned, suggesting discussions about her influence.
- **Group 5**. This cluster, identified in the scatter plot by the color light green, comprises 636 videos. This cluster is characterized by its focus on climate action, saving the planet and environmental awareness. Keywords include “stopglobalwarming”, “savetheworld”, “savetheplanet” and “saveouroceans”.
- **Group 6**. This last cluster, represented in the scatter plot by the color yellow, contains 908 videos. This cluster is characterized by content related to recycling, waste reduction and sustainable practices. Keywords include “recycle“, “recycling”, “zero waste”, “reuse” and “crafts”.

4.1. Keyword Frequency, videos count, likes, views and comments Analysis

To provide the big picture of the results, Table 1 lists the most frequent keywords for each cluster, The total number of videos in each cluster and the total likes, views and comments for videos in each cluster. This data highlights the distribution of videos across the seven thematic clusters and provides insight into the engagement levels of videos in different thematic groups.

An analysis of the keywords and content themes across clusters reveals distinct focal areas and potential areas of misinformation. Clusters like Cluster 0 and Cluster 6 emphasize practical advice and advocacy for environmental issues, showcasing keywords related to sewing, upcycling, recycling and sustainable living. For instance, Cluster 0 includes terms like “sewing”, “upcycling” and “thrift” while Cluster 6 is characterized by “recycle”, “recycling” and “plasticfree”. These clusters tend to present more straightforward and educational content, making them less likely to contain misinformation.

In contrast, Cluster 3 stands out due to the prevalence of terms such as "weathermanipulation", "chemtrails" and "geoengineering", pointing to content related to conspiracy theories. This cluster is more likely to mix factual information with misleading claims, which can negatively influence public perception.

Also Cluster 4, despite containing terms as "climatechange", "globalwarming", "climatecrisis" and "climateaction", it also contains terms as "climatelies", "globalwarmingisfake" and "climatehoaxx; which means it focuses on climate change, global warming, climate action and its controversies.

The focus on conspiracy theories contrasts with the more practical and educational nature of the content in Clusters 0 and 6.

When considering the total likes, views and comments of these clusters, significant differences emerge. Cluster 5, for instance, which includes keywords like "stopglobalwarming" and "savetheworld" has the highest number of likes (3,053,581). This suggests that videos in this cluster receive higher acceptance and engagement from users, possibly due to the urgency and global appeal of climate change messaging.

Despite having fewer likes, Cluster 6 has the highest number of views (31,976,184). This indicates that videos in this cluster may be viewed more frequently, possibly due to a growing interest in recycling and

Cluster	Keywords and Frequency	Videos Count	Likes	Views	Comments
0	sewing (37) upcycling (24) thrift (24) thriftflip (20) fashion (19)	1,260	1,160,419	11,666,224	15,321
1	duet with (59) duet (47) bethechange (28) tiktok (25) lifehacks inspiration (19)	1,300	839,619	6,303,510	40,536
2	sustainability (63) environment (40) sustainable (40) ecology (37) eco (27)	1,665	1,778,118	18,963,898	39,140
3	weathermanipulation (338) chemtrails (136) weathermodification (69) weathermanipulation lexky (60) coveringthesun weathermanipulation (51)	1,243	681,355	3,707,237	57,980
4	climatechange (238) globalwarming (116) global warming (67) climate change (47) climatecrisis (36)	1,139	1,511,484	10,797,769	48,182
5	stopglobalwarming (162) savetheworld (87) savetheplanet (32) globalwarming (29) saveouroceans (18)	636	3,053,581	17,783,375	52,676
6	recycle (126) recycling (118) recycled (27) plasticfree (18) waste (15)	908	1,621,028	31,976,184	17,701

Table 1
Clustering analysis results

zero-waste topics. The higher view count suggests that the audience for recycling content is substantial, even if individual engagement through likes is lower.

Cluster 1, by the other hand, characterized by keywords such as "duet with", "duet" and "bethechange" highlights a significant trend in social media engagement. With 1,300 videos, Cluster 1 focuses on collaborative content, challenges and inspirational messages.

Cluster 3, which includes conspiracy-related keywords, has the highest number of comments (57,980), and Cluster 4 which also mentions "gretathunberg", has the third higher number of comments (48,182), indicating active discussion and debate. This could be due to the controversial nature of the topics, which often elicit strong opinions and engagement from viewers, as we saw in 1. In comparison, Cluster 2, which focuses on sustainability and ecology, has a lower number of comments (39,140) but still shows significant engagement, reflecting interest in environmental issues.

5. Disinformation in TikTok

In the digital age, the spread of misinformation, which is often mistakenly believed to be reliable, is a serious problem. As claim detection technology advances, the models for the detection of potentially misleading content have become more accurate. We used an automated claim detection model from Huggingface[1] flagged a significant number of videos, suggesting they were worthy of fact-checking. Specifically, 5,352 videos were initially identified as requiring review due to concerns of possible misrepresentation.

However, after further analysis using clustering techniques, we discovered an interesting finding: the actual number of videos related to disinformation and fake news was significantly lower than the original estimate, as we saw in 4.1. This difference highlights the importance of using more diverse methods to distinguish reviewable content from non-reviewable content.

Using a clustering algorithm allows us to divide tagged videos into distinct groups based on their subject content and inferred features. It's worth noting that we've divided these videos into seven groups, each revealing unique interaction patterns and thematic focus.

Group 0 emerged as a hub for sutainable fashion practices, clothing upcycling and sewing techniques. Our validation process, including consultation of correborated information of fact-checking agency EFE Verifica [13], confirmed the absence, mostly, of misinformation associated with these keywords. Instead, it underscored a commitment to promoting responsible consumption practices. Group 1 indicates a focus o social interaction, practical tips and entertaiment, which, for the most part, does not contain misinformation

Group 2 is distinguished by its focus con sustainability, environment and ecological awareness. It was underscored a commitment to promoting responsible sustainability and ecological discourse. Group 3 is characterized by discussions on weather manipulation, chemtrails and climate-related conspiracy theories. Despite not all videos in this cluster perpetuating misinformation, those that did sparked intense debates and polarized discussions among viewers. Group 4 focuses on climate change, global warming, climate actions and its controversies. Although many of the videos are based on climate actions, there are also some other certain controversies and debates.

Group 5 is characterized by its focus on climate action, saving the planet and environmental awareness. Our validation process, confirmed the absence, mostly, of misinformation associated with these keywords. Group 6 is characterized by content related to recycling, waste reduction and sustainable practices. The commitment to promoting responsible sustainability and recycling practices was highlighted.

6. Conclusion

As we have seen, in the digital era, social media platforms have become critical battlegrounds where misinformation about critical issues like climate change proliferates. Our study utilized claim detection and clustering techniques to analyze a substantial dataset of videos. Initially flagging 5,352 videos for potential misinformation, we found through clustering that the actual prevalence of misinformation, particularly in Cluster 3 and some in Cluster 4, despite existing, was lower than expected.

The rest of clusters predominantly featured content promoting sustainability and environmental awareness, corroborated by information of fact-checking agency EFE Verifica [13]. In contrast, Cluster 3 and 4 contained videos with keywords associated with misinformation and conspiracy theories. Collaboration with multilingual fact-checking platforms underscored the need for vigilant media consumption practices.

In summary, our study contributes to understanding and addressing the challenges posed by misinformation in digital media, emphasizing the importance of informed media consumption and collaborative efforts in safeguarding information integrity.

[1]https://huggingface.co/Nithiwat/xlm-roberta-base_claim-detection

Acknowledgements

This work has been funded by Grant PLEC2021-007681 (XAI-DisInfodemics), by the “European Union” or by the “European Union NextGenerationEU/PRTR”, by grant PCI2022-134990-2 (MARTINI) of the CHISTERA IV Cofund 2021 program, funded by MCIN/AEI/10.13039/501100011033 and by the “European Union NextGenerationEU/PRTR”, by Calouste Gulbenkian Foundation, under the project MuseAI - Detecting and matching suspicious claims with AI and by European Comission under IBERIFIER Plus - Iberian Digital Media Observatory (DIGITAL-2023-DEPLOY- 04-EDMO-HUBS 101158511).

References

[1] U. K. H. Ecker, S. Lewandowsky, J. Cook, et al., The psychological drivers of misinformation belief and its resistance to correction, Nature Reviews Psychology 1 (2022) 13–29. URL: https://doi.org/10.1038/s44159-021-00006-y. doi:10.1038/s44159-021-00006-y.

[2] V. Galaz, H. Metzler, S. Daume, A. Olsson, B. Lindström, A. Marklund, Ai could create a perfect storm of climate misinformation, 2023. arXiv:2306.12807.

[3] M. Aldwairi, A. Alwahedi, Detecting fake news in social media networks, Procedia Computer Science 141 (2018) 215–222. URL: https://www.sciencedirect.com/science/article/pii/S1877050918318210. doi:https://doi.org/10.1016/j.procs.2018.10.171, the 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018) / The 8th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2018) / Affiliated Workshops.

[4] P. N. Petratos, Misinformation, disinformation, and fake news: Cyber risks to business, Business Horizons 64 (2021) 763–774. URL: https://www.sciencedirect.com/science/article/pii/S000768132100135X. doi:https://doi.org/10.1016/j.bushor.2021.07.012, cIBER SPECIAL ISSUE: CYBERSECURITY IN CRISIS.

[5] J. Huertas-Tato, A. Martín, D. Camacho, Bertuit: Understanding spanish language in twitter with transformers, Expert Systems 40 (2023) e13404. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.13404. doi:https://doi.org/10.1111/exsy.13404. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.13404.

[6] A. Yang, Exploring the network and topic stability in climate change deniers’ disinformation network: A longitudinal study (2024).

[7] B. Braga Pereira, S. Ha, Environmental issues on tiktok: Topics and claims of misleading information, Journal of Baltic Science Education 23 (2024) 131–150. doi:10.33225/jbse/24.23.131.

[8] F. J. Basch CH, Yalamanchili B, #climate change on tiktok: A content analysis of videos, J Community Health 47(1) (2022) 163–167. doi:10.1007/s10900-021-01031-x.

[9] T. Company, Tiktok api, 2024. URL: https://developers.tiktok.com/.

[10] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, Robust speech recognition via large-scale weak supervision, 2022. URL: https://arxiv.org/abs/2212.04356. doi:10.48550/ARXIV.2212.04356.

[11] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019. URL: https://arxiv.org/abs/1908.10084.

[12] M. Grootendorst, Keybert: Minimal keyword extraction with bert., 2020. URL: https://doi.org/10.5281/zenodo.4461265. doi:10.5281/zenodo.4461265.

[13] Agencia EFE, EFE Verifica, ???? URL: https://efe.com/.