



Unveiling Bias in YouTube Shorts: Analyzing Thumbnail Recommendations and Topic Dynamics

Mert Can Cakmak^(✉), Nitin Agarwal, Selimhan Dagtas, and Diwash Poudel

COSMOS Research Center, University of Arkansas - Little Rock,
Little Rock, AR, USA

{mccakmak, nagarwal, sedagtas, dpoudel}@ualr.edu

Abstract. In the dynamic landscape of digital media, YouTube Shorts have emerged as a popular format, captivating users with their brief and engaging content. However, the recommendation algorithms driving these videos often exhibit biases that influence which thumbnails are prominently displayed. This study delves into the biases present in YouTube's recommendation algorithms, focusing on the thumbnails of YouTube Shorts, which play a crucial role in attracting viewers. Thumbnails, as powerful visual elements, significantly impact user decisions and engagement. By utilizing advanced topic modeling and content generation techniques, we analyzed a substantial dataset of YouTube Shorts' thumbnails. Our analysis, employing generative AI and BERTopic models, reveals notable shifts in topic distribution across recommendation cycles, it highlights a preference for certain types of content. These biases not only affect content visibility but also steer user engagement towards popular, yet potentially less diverse, topics. The findings of this study enhance the understanding of algorithmic biases in digital platforms and aim to promote more equitable and transparent content recommendation practices.

Keywords: YouTube Shorts · Thumbnails · Recommender System · Algorithmic Bias · Generative AI · Topic Modelling · Topic Clustering

1 Introduction

In the contemporary digital age, humans are influenced by external digital stimuli, particularly through recommendation algorithms that impact emotions, thoughts, and actions [1, 2]. On YouTube, thumbnails are powerful visual attractors that significantly affect a user's decision to view a video, making them crucial for capturing interest and guiding subsequent actions. The trend towards shorter video formats, such as YouTube Shorts, has reshaped content consumption, catering to fast-paced lifestyles seeking brief, engaging content.

This research explores biases in recommendation algorithms as they pertain to YouTube Shorts' thumbnails. By examining how these visual elements are

recommended and disseminated, the study aims to uncover patterns of bias. Specifically, it addresses the following research questions:

- **RQ1:** How does the topical content of YouTube Shorts’ thumbnails change over time through recommendations?
- **RQ2:** What types of topics are more and less frequently recommended for YouTube Shorts after multiple recommendation cycles?
- **RQ3:** How does the content depicted in these thumbnails, as recommended by YouTube’s algorithm, perpetuate biases on the platform?

To answer these questions, up-to-date topic modeling and content generation techniques were utilized. This research aims to enhance the understanding of the effects of recommendation algorithms on thumbnail content and their implications for content diversity and user engagement.

This paper is structured as follows: Sect. 2 reviews key concepts and relevant literature. Section 3 details our data collection and analytical methods. Section 4 presents findings on biases in YouTube Shorts’ recommendations with graphical analyses. Finally, Sect. 5 summarizes key insights and discusses implications for future research and practical applications.

2 Literature Review

This literature review provides an overview of key studies relevant to various aspects of our research.

2.1 South China Sea Dispute

The South China Sea (SCS) is a critical geopolitical region with significant attention due to overlapping territorial claims and its strategic importance for maritime routes and natural resources.

The research by [3] highlights the dual factors of natural resources and freedom of navigation, emphasizing the SCS’s abundant resources like oil, natural gas, and rich fishing grounds. They also stress the SCS’s importance as a vital trade route. The work by [4] examines China’s strategic approach, detailing efforts to consolidate claims and expand influence through diplomatic, economic, and military measures. The study by [5] analyzes China’s assertiveness from 1970 to 2015, identifying key turning points and highlighting the cumulative effects of China’s actions.

These studies underscore the SCS’s impact on regional stability, international maritime law, and the balance of power in the Asia-Pacific region.

2.2 YouTube Shorts and Thumbnails

YouTube Shorts, introduced to meet the demand for short-form content, have quickly become a dominant format, particularly in entertainment categories [6].

These videos attract higher engagement metrics compared to regular videos (RVs) but pose monetization challenges due to fewer advertisement opportunities, requiring new revenue strategies for creators [7]. The popularity of Shorts reflects changes in viewer behavior, aligning with shorter attention spans [8].

Thumbnails on YouTube play a crucial role in attracting viewers and influencing video selection, directly impacting click-through rates (CTR) and engagement metrics. Visually appealing thumbnails with high view counts are more likely to be selected [9]. However, clickbait thumbnails can lead to viewer dissatisfaction when the content does not meet expectations [10, 11]. Technological advancements like Optical Character Recognition (OCR) help detect and avoid clickbait, ensuring accurate representation of video content [12]. Thumbnails also influence algorithmic recommendations, as higher CTRs lead to more favorable placements in user feeds.

2.3 Recommendation Bias

Recommender systems significantly influence content consumption, often embedding biases and creating filter bubbles. Bias can arise from user preferences, algorithmic design, and training data. The study by [13] highlights how recommendation algorithms shape public discourse by promoting emotionally charged content [14]. Another work by [15] discusses content drift towards homogeneous and radical themes, emphasizing the importance of monitoring these shifts. Audits by [16–18] reveal the promotion of biased content, underlining the need for interventions. The work by [19] found that YouTube’s algorithm promotes content with specific emotional tones, affecting user engagement. Cross-topical analysis by [20] reveals biases in diverse contexts and highlight content drift risks. Lastly, the research by [21] found that YouTube’s algorithm can lead users, especially those with right-leaning ideologies, towards radical content [22]. Thus, addressing recommender bias and drift is crucial for a well-informed public.

To the best of our knowledge, no previous studies have specifically investigated biases in YouTube Shorts and their thumbnails. Therefore, our research offers a unique and novel contribution to the understanding of algorithmic biases in digital media platforms.

3 Methodology

This section details the methods used for data collection, topic modeling, and analysis of YouTube Shorts’ thumbnails to investigate recommendation biases.

3.1 Data Collection

To initiate data collection, we held workshops with experts to generate relevant keywords for our search, targeting YouTube Shorts videos.

Due to the YouTube Data API’s limitations with Shorts, we used APIFY’s YouTube Scraper [23] to collect 1,210 unique video IDs. Finding this insufficient, we employed a snowball method to generate additional keywords, using the YouTube Data API and transcriptions from previous research [24, 25].

The keywords for the South China Sea Dispute covered legal rulings, geopolitical tensions, and economic interests, enabling us to collect 2,094 unique video IDs for a detailed analysis of the conflict.

To measure bias in YouTube Shorts recommendations, we developed a custom scraping method.

Using collected video IDs as seed videos, we ensured a neutral user profile with fresh WebDriver instances. Automated with Selenium, our script scrolled through recommendations (depth) to a depth of 50, then started a new session for the next video.

We collected 104,700 videos, and after filtering out unavailable ones, our final dataset included 100,300 videos with their thumbnail images, obtained via the YouTube Data API.

3.2 Caption Generation

To investigate thumbnail images in detail and understand their context, we generated captions that describe the contents and events depicted in the images.

For this, we used GPT-4 Turbo, an enhanced version of OpenAI’s GPT-4 language model. This model is optimized for speed and cost-effectiveness while maintaining similar capabilities and performance to the standard GPT-4, making it suitable for applications requiring quick responses and scalability. GPT-4 Turbo excels in natural language understanding and generation, supporting tasks from text completion and translation to content creation and conversational AI [26].

3.3 Topic Modelling

To understand the thumbnails’ captions, we classified them into topics to track their evolution through recommendations.

We used GPT-4o, a refined and efficient version of GPT-4 designed for faster performance and lower costs while maintaining advanced natural language capabilities [26]. This model generated two types of topics general topics and categorized topics with specific constraints.

We also used BERTopic [27], a technique leveraging BERT embeddings to capture semantic similarities, resulting in coherent and interpretable topics. A fine-tuned version pre-trained on approximately 1,000,000 Wikipedia pages [28] was utilized, identifying 2,377 distinct topics. This robust framework effectively analyzed the thematic content of videos.

3.4 Clustering Topics

To analyze the topics discussed in Sect. 3.3, we clustered them due to their large number.

We filtered out non-informative topics like ‘Photograph(s)’, ‘Thumbnail(s)’, ‘Image(s)’, and ‘Video(s)’.

BERT embeddings [29] were generated to capture semantic meaning through dense vector representations, considering both preceding and following contexts for accuracy.

These embeddings were reduced to two dimensions using t-SNE for visualization [30], which reveals intricate local patterns effectively.

The reduced features were clustered using the OPTICS algorithm [31], which handles varying data densities and is more flexible than other methods.

4 Results

This section presents our findings on biases in YouTube Shorts' recommendations, supported by graphical analyses of topic shifts and distributions.

4.1 Clustered General Topics with GPT

The GPT model generated 2,314 unique general topics. To visualize the topic clusters, we plotted them in a 2D space at various depths using t-SNE components for dimensionality reduction. We visualized the top 50 high-frequency topics. The legend distinguishes clusters by color and shows noise in gray.

As shown in Fig. 1, representing depth 0, topics are clustered together. Cluster 0 includes terms like politics, history, diplomacy, war, and military, indicating political themes. Clusters 3, 4, and 5 contain terms like ships, aircraft, and fishing, relating to aircraft carriers and economic perspectives at sea. Cluster 7 includes terms like geopolitics, map, and Philippines, highlighting the geographic perspective of the topic. Other clusters depict activities such as broadcasts, meetings, and presentations, with some, like Cluster 1, showing animated explanations of the topic. Initial depth videos were highly relevant to the investigated topic.

At depth 5, as displayed in Fig. 2, the topics are vastly different from the initial ones. Cluster 0 shows crafting topics, Cluster 1 covers machines and robotics, Cluster 3 is mostly about gaming, Cluster 4 includes dance, gym, and martial arts, Cluster 7 features child and dog-related terms, and Cluster 8 contains memes. The original topics have almost completely faded, with many new topics emerging.

We did not include all depths here because topics drifted early, and space constraints limited our illustrations to these two depths. More depths will be shown in the upcoming result subsections.

4.2 Categorized Topics with GPT

In this section, we investigated the categorized topics mentioned earlier in Sect. 3.3. Using the GPT model, we generated 20 categorized topics. These categories were determined by researching online and examining YouTube video categories, supplemented by additional categories from various news websites to ensure comprehensive coverage. Thus, we settled on 20 categories to encompass

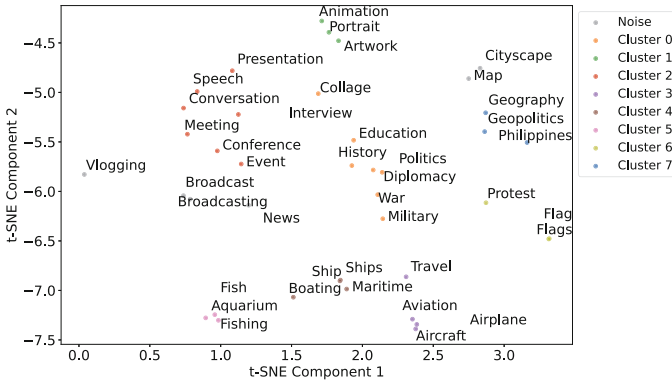


Fig. 1. General Topic Clustering for Depth 0

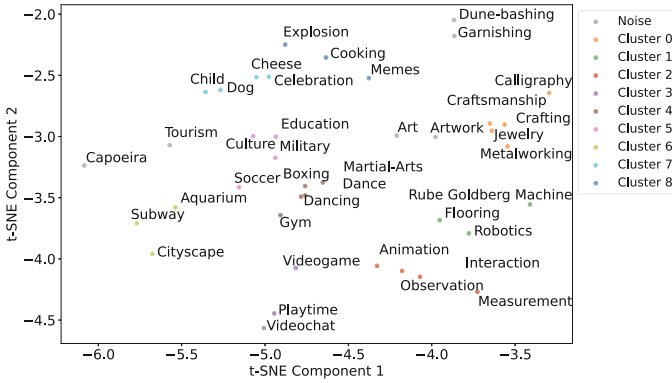


Fig. 2. General Topic Clustering for Depth 5

a broad range of topics as shown in Fig. 3. We used a lollipop chart for clarity, with the Y-axis representing the topics and the X-axis showing the topic ratios or distributions. The legend indicates depth ranges with corresponding colors. Topics for each depth range are accumulated and normalized, grouped into five classes for convenience and clarity, filtering out depths with a ratio below 0.01.

At the initial depth (depth 0), news dominates nearly 40% of the topics, followed by politics at around 15%, with other topics like history and lifestyle also present. As the recommendation algorithm suggested new videos, these new topics were neither news nor politics. News topics reduced dramatically across other depths, and political terms disappeared entirely. New topics, primarily entertainment-related, increased with each depth. Lifestyle topics remained relatively stable across depths due to their broad and encompassing nature. The graph clearly shows the topic shift happening in the recommendation algorithm.

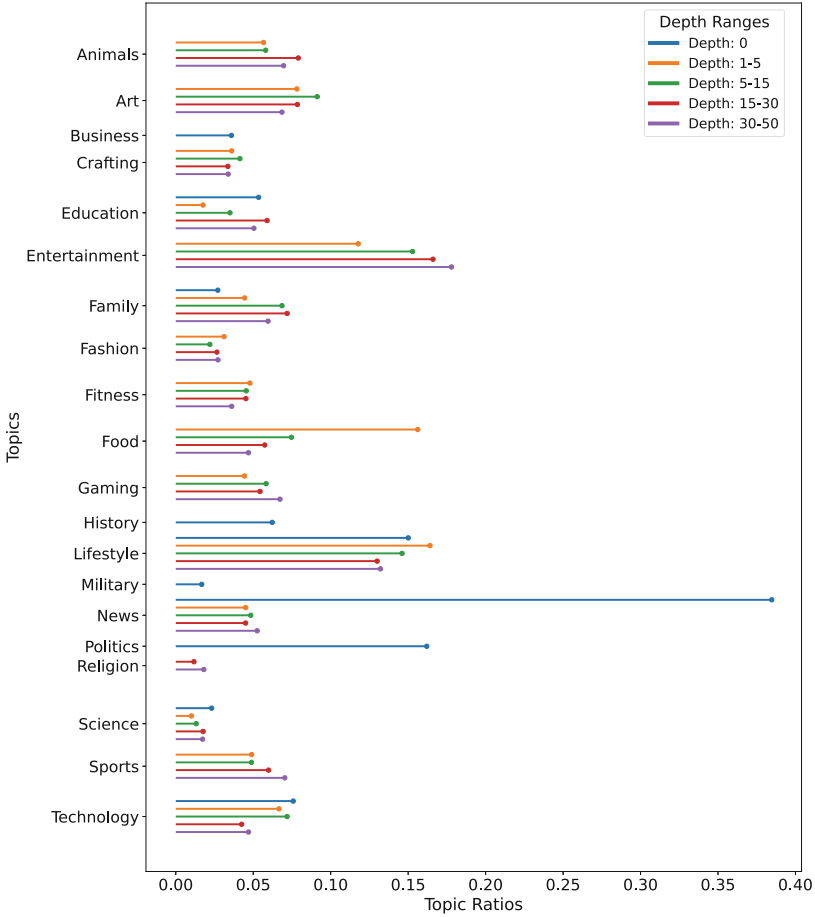


Fig. 3. The distribution of Categorized Topics Across Depths

4.3 Topic Distribution with BERTopic

In addition to generative AI, we utilized BERTopic for topic modeling, as detailed in Sect. 3.3. For illustration, we used radar charts to visualize the topics at initial, middle, and end depths as shown in Fig. 4. On the circle the topic IDs and names are represented, and the topic IDs are available on Huggingface [28] with details. We focused on the three most prevalent topics for each depth level to effectively track topic transitions.

At depth 0, the most prominent topics included flag, geography, and uniform. For example, topic 1650 (uniforms, uniformed, berets, beret) related primarily to military and soldiers, topic 935 (geography, geographic, geographical, geographer) concerned geopolitical regions around the South China Sea, and topic 111 (flags, flag, flagpole, commonwealth) referenced different nations. These topics clearly relate to the South China Sea Dispute.

As we move to deeper levels, we observed the emergence of unrelated topics, with the initial topics fading away. At later depths, we see topics like 706 (artistic, art, artwork, paintings), 5 (cuisine, cuisines, foods, culinary), and 1879 (lighting, lights, fluorescent, light). This shift indicates the recommendation algorithm’s steering away from the original subject matter towards more general and unrelated topics.

Although BERTopic may not capture topics as effectively as GPT, it vividly demonstrates the topic drift across depths. The transition from focused, relevant topics to broader, unrelated ones highlights the algorithm’s influence on content distribution, illustrating how quickly the focus can shift away from the initial subject matter.

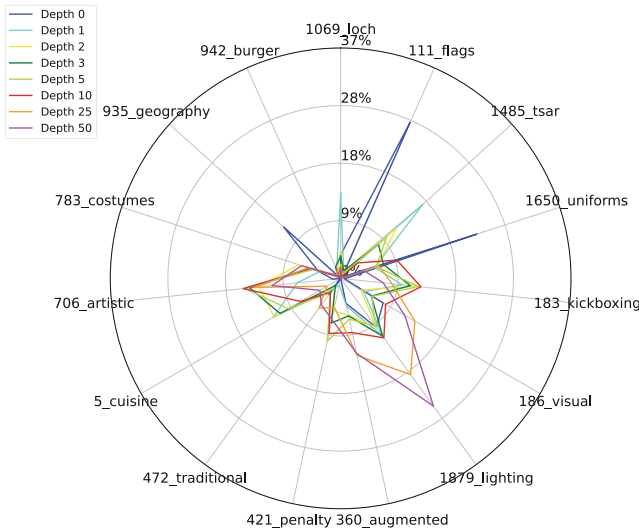


Fig. 4. Topic Distribution Using BERTopic Across Depths

5 Conclusion and Discussion

In this study, we investigated algorithmic bias in YouTube Shorts’ video recommendations, focusing on thumbnail captions within the context of the South China Sea Dispute.

Our findings indicate a clear topic shift or drift in YouTube Shorts recommendations. After the initial videos, broader and less relevant topics are suggested due to YouTube’s recommender system favoring high-engagement entertainment videos. This popularity bias results in the neglect of minority and serious issues, creating an algorithmic bias on YouTube Shorts. Consequently, these more popular but less serious videos are recommended more frequently.

For future work, we will address study limitations by comparing results with engagement scores to validate our assumptions. We will investigate various narratives, comparing well-known topics with niche subjects like the South China Sea Dispute, to understand recommendation levels. Additionally, we will incorporate interactive data collection (liking and commenting) to observe how user interactions affect recommendations and analyze text attributes such as titles, descriptions, and transcriptions.

This research highlights biases in YouTube’s algorithmic recommendations, focusing on thumbnails. Understanding these biases is essential for fair representation of diverse topics, especially serious and minority issues. Thumbnails influence user engagement, and our study shows how algorithmic preferences can skew topic visibility. By exposing these biases, we contribute to the discourse on digital media ethics and the need for transparency in recommendation systems.

Acknowledgments. This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-20-1-0262, W911NF-16-1-0189, W911NF-23-1-0011, W911NF-24-1-0078), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

1. Cakmak, M.C., Shaik, M., Agarwal, N.: Emotion assessment of YouTube videos using color theory. In: Proceedings of the 9th International Conference on Multimedia and Image Processing (ICMIP). IEEE (2024)
2. Yousefi, N., Cakmak, M.C., Agarwal, N.: Examining multimodal emotion assessment and resonance with audience on YouTube. In: Proceedings of the 9th International Conference on Multimedia and Image Processing (ICMIP). IEEE (2024)
3. Macaraig, C.E., Fenton, A.J.: Analyzing the causes and effects of the South China Sea dispute: natural resources and freedom of navigation. *J. Territ. Marit. Stud.* **8**(2), 42–58 (2021). <https://www.jstor.org/stable/48617340>
4. Fravel, M.T.: China’s strategy in the South China Sea. *Contemp. Southeast Asia* **33**(3), 292–319 (2011). <http://www.jstor.org/stable/41446232>
5. Chubb, A.: PRC assertiveness in the South China Sea: measuring continuity and change, 1970–2015. *Int. Secur.* **45**(3), 79–121 (2021). https://doi.org/10.1162/isec_a.00400
6. Violot, C., Elmas, T., Bilogrevic, I., Humbert, M.: Shorts vs. regular videos on YouTube: a comparative analysis of user engagement and content creation trends.

- In: ACM Web Science Conference (WebSci 2024). ACM (2024). <https://doi.org/10.1145/3614419.3644023>
7. Rajendran, P.T., Creusy, K., Garnes, V.: Shorts on the rise: assessing the effects of YouTube shorts on long-form video content. arXiv preprint [arXiv:2402.18208](https://arxiv.org/abs/2402.18208) (2024)
 8. Sahu, G., Gaur, L., Singh, G.: Investigating the impact of personality tendencies and gratification aspects on OTT short video consumption: a case of YouTube shorts. In: 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM), Uttar Pradesh, India, pp. 1–6 (2023). <https://doi.org/10.1109/ICIPTM57143.2023.10118122>
 9. Park, J.: The impact of YouTube’s thumbnail images and view counts on users’ selection of video clip, memory recall, and sharing intentions of thumbnail images. The Florida State University (2022)
 10. Qu, J., Hißbach, A.M., Gollub, T., Potthast, M.: Towards crowdsourcing clickbait labels for YouTube videos. In: HCOMP (WIP & Demo) (2018)
 11. Shajari, S., Alassad, M., Agarwal, N.: Characterizing suspicious commenter behaviors. In: Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, Kusadasi, Turkiye, pp. 631–635. ACM (2023). <https://doi.org/10.1145/3625007.3627309>
 12. Vitadhani, A., Ramli, K., Dewi Purnamasari, P.: Detection of clickbait thumbnails on YouTube using Tesseract-OCR, face recognition, and text alteration. In: 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), pp. 56–61 (2021). <https://doi.org/10.1109/ICAICST53116.2021.9497811>
 13. Cakmak, M.C., Okeke, O., Onyepunuka, U., Spann, B., Agarwal, N.: Analyzing bias in recommender systems: a comprehensive evaluation of YouTube’s recommendation algorithm. In: Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2023), pp. 753–760 (2024). <https://doi.org/10.1145/3625007.3627300>
 14. Alp, E., Gergin, B., Eraslan, Y.A., Çakmak, M.C., Alhajj, R.: Covid-19 and vaccine tweet analysis. In: Özyer, T. (ed.) Social Media Analysis for Event Detection. LNSN, pp. 213–229. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08242-9_9
 15. Kirdemir, B., Kready, J., Mead, E., Hussain, M.N., Agarwal, N., Adjeroh, D.: Assessing bias in YouTube’s video recommendation algorithm in a cross-lingual and cross-topical context. In: Thomson, R., Hussain, M.N., Dancy, C., Pyke, A. (eds.) SBP-BRiMS 2021. LNCS, vol. 12720, pp. 71–80. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-80387-2_7
 16. Srba, I., et al.: Auditing YouTube’s recommendation algorithm for misinformation filter bubbles. ACM Trans. Recomm. Syst. 1(1), 6 (2023). <https://doi.org/10.1145/3568392>
 17. Gurung, M.I., Bhuiyan, M.M.I., Al-Taweel, A., Agarwal, N.: Decoding YouTube’s recommendation system: a comparative study of metadata and GPT-4 extracted narratives. In: Companion Proceedings of the ACM on Web Conference 2024, pp. 1468–1472. Association for Computing Machinery (2024). <https://doi.org/10.1145/3589335.3651913>
 18. Poudel, D., Cakmak, M.C., Agarwal, N.: Beyond the click: how YouTube thumbnails shape user interaction and algorithmic recommendations. In: The 16th International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2024)

19. Okeke, O., Cakmak, M.C., Spann, B., Agarwal, N.: Examining content and emotion bias in YouTube's recommendation algorithm. In: Proceedings of the Ninth International Conference on Human and Social Analytics (HUSO 2023), Barcelona, Spain, pp. 15–20 (2023). <https://www.thinkmind.org/index.php?view=article&articleid=huso.2023.1.40.80032>
20. Cakmak, M.C., Okeke, O., Onyepunuka, U., Spann, B., Agarwal, N.: Investigating bias in YouTube recommendations: emotion, morality, and network dynamics in China-Uyghur content. In: Cherifi, H., Rocha, L.M., Cherifi, C., Donduran, M. (eds.) COMPLEX NETWORKS 2023. SCI, vol. 1141, pp. 351–362. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-53468-3_30
21. Haroon, M., Chhabra, A., Liu, X., Mohapatra, P., Shafiq, Z., Wojcieszak, M.: YouTube, the great radicalizer? Auditing and mitigating ideological biases in YouTube recommendations. arXiv preprint [arXiv:2203.10666](https://arxiv.org/abs/2203.10666) (2022)
22. Shaik, M., Cakmak, M.C., Spann, B., Agarwal, N.: Characterizing multimedia adoption and its role on mobilization in social movements. In: Bui, T.X. (ed.) 57th Hawaii International Conference on System Sciences, HICSS 2024, Hilton Hawaiian Village Waikiki Beach Resort, Hawaii, USA, 3–6 January 2024, pp. 146–155. ScholarSpace (2024). <https://hdl.handle.net/10125/106393>
23. Streamers. Youtube Scraper. APIFY (2024). <https://apify.com/streamers/youtube-scraper>. Accessed 10 Jan 2024
24. Cakmak, M.C., Okeke, O., Spann, B., Agarwal, N.: Adopting parallel processing for rapid generation of transcripts in multimedia-rich online information environment. In: 2023 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 832–837 (2023). <https://doi.org/10.1109/IPDPSW59300.2023.00139>
25. Cakmak, M.C., Agarwal, N.: High-speed transcript collection on multimedia platforms: advancing social media research through parallel processing. In: 2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). IEEE (2024)
26. OpenAI, et al.: GPT-4 technical report. [arXiv:2303.08774](https://arxiv.org/abs/2303.08774) (2024)
27. Grootendorst, M.: BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint [arXiv:2203.05794](https://arxiv.org/abs/2203.05794) (2022)
28. Grootendorst, M.: BERTopic_Wikipedia. Huggingface (2024). https://huggingface.co/MaartenGr/BERTopic_Wikipedia. Accessed 1 May 2024
29. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) (2019)
30. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008)
31. Ankerst, M., Breunig, M.M., Kriegel, H.P., Sander, J.: OPTICS: ordering points to identify the clustering structure. *ACM SIGMOD Rec.* **28**(2), 49–60 (1999)