Proceedings of the 58th Hawaii International Conference on System Sciences | 2025

Unpacking Algorithmic Bias in YouTube Shorts by Analyzing Thumbnails

Mert Can Cakmak[1] and Nitin Agarwal[1,2]
[1]COSMOS Research Center, University of Arkansas at Little Rock, AR 72204, USA
[2]International Computer Science Institute, University of California, Berkeley, CA 94720, USA
{mccakmak, nxagarwal}@ualr.edu

Abstract

As digital platforms increasingly shape our online experiences, the influence of recommendation algorithms on user behavior becomes ever more significant. This research delves into the biases inherent in YouTube Shorts' recommendation algorithms by analyzing the topical content of thumbnails through captions generated by advanced generative AI models, specifically GPT and Llama. Employing topic modeling and clustering techniques, we scrutinized a substantial dataset of YouTube Shorts to uncover patterns of bias within the recommendation process. Our findings reveal a significant drift in recommended content from serious geopolitical topics to broader, entertainment-focused themes, underscoring the impact of algorithmic preferences on user engagement. This study highlights the necessity for greater transparency and fairness in content recommendation systems, offering valuable insights into the ethical implications of algorithmic bias in digital media.

Keywords: Generative AI, YouTube Shorts, Thumbnails, Recommendation System, Topic Clustering

1. Introduction

In the contemporary digital age, humans are influenced not just by rational thought but also by emotional responses, which are significantly shaped by external stimuli. Recommendation algorithms on platforms like YouTube amplify this by curating emotionally charged or polarizing content, subtly shaping users' emotions, thoughts, and engagement patterns. This influence is especially potent in the context of sensitive geopolitical issues like the South China Sea Dispute, where these algorithms can reinforce specific narratives and biases, making it a critical topic for examining the broader implications of algorithmic influence on public discourse.

One of the most immediate and influential elements for users on YouTube is the thumbnail. Thumbnails serve as visual attractors, encapsulating the video's content in a single frame and exerting a substantial effect on a user's decision to view a video. Human brains are adept at processing visual information quickly, making thumbnails a critical component in capturing interest and influencing the user's subsequent actions. The saying "a picture is worth a thousand words" expertly describes the significant impact thumbnails have in the realm of digital content.

In addition to visual appeal, the trend towards shorter video formats is reshaping content consumption patterns. YouTube Shorts, introduced in response to the growing demand for brief, engaging videos, have gained immense popularity. These short videos, typically under 60 seconds, cater to the fast-paced lifestyles of modern viewers, who seek quick bursts of information and entertainment.

This research focuses on exploring the bias present in recommendation algorithms as it pertains to YouTube Shorts' thumbnails. By examining how these visual elements are recommended and disseminated, this study seeks to uncover patterns of bias in the recommendation process. Specifically, our research addresses the following questions:

- **RQ1:** How does the topical content of YouTube Shorts' thumbnails change over time through recommendations?
- **RQ2:** What types of topics are more frequently and less frequently recommended for YouTube

URI: https://hdl.handle.net/10125/109144
978-0-9981331-8-8

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, we utilized contemporary topic modeling and content generation techniques to conduct an in-depth analysis. Through this research, we aim to enhance the understanding of the effects of recommendation algorithms on thumbnail content and the implications for content diversity and potential user engagement.

Structured as follows, the paper aims to provide a coherent narrative: Section 2 introduces the key concepts and reviews relevant literature, setting the stage for understanding the context and importance of our study. Section 3 details our data collection processes and analytical methods, including the techniques used for topic modeling and content analysis. Section 4 presents our findings on the biases in YouTube Shorts' recommendations, supplemented by detailed graphical analyses. Finally, Section 5 summarizes our study's key insights and discusses the implications of our findings for future research and practical applications.

2. Literature Review

In this section, we explore the foundational concepts and review the relevant literature to contextualize and underscore the significance of our study.

2.1. South China Sea Dispute

The South China Sea (SCS) is a crucial geopolitical region due to its overlapping territorial claims and strategic maritime routes. This section reviews key literature on the dynamics and underlying factors of the SCS dispute.

The research by Macaraig and Fenton (2021) highlights the region's rich natural resources and vital trade routes, emphasizing their importance to claimant countries and global powers. The study by Fravel (2011) examines China's strategic approach, detailing diplomatic, economic, and military measures to consolidate its claims and the implications for regional stability and international law under UNCLOS. The work by Chubb (2020) analyzes China's assertive behavior from 1970 to 2015, noting significant policy shifts and their cumulative effects on the SCS conflict.

The SCS's significance lies in its impact on regional stability, international maritime law, and the Asia-Pacific balance of power, making it essential for understanding contemporary geopolitical dynamics.

2.2. YouTube Short Videos

YouTube Shorts signify a major shift in digital media, having become essential to the video-sharing ecosystem. Shorts, introduced in response to platforms like TikTok, allow creators to produce 60-second videos optimized for mobile viewing. Violot et al. (2024) highlight that Shorts dominate YouTube, especially in entertainment, attracting more views and likes per view than regular videos (RVs).

Shorts generate significant engagement, with creators producing more Shorts than RVs, indicating a preference for quick, digestible content. While primarily catering to entertainment, Shorts cover various other categories Sahu et al. (2023).

Understanding the impact of Shorts is crucial. Rajendran et al. (2024) note that despite high engagement, Shorts are challenging for creators to monetize because their short and fast-paced format doesn't work well with traditional revenue methods, so new strategies are needed. The adoption of Shorts reflects changes in viewer behavior, aligning with decreasing attention spans and demand for diverse, quickly consumable content.

Analyzing Shorts provides insights into algorithmic biases and content distribution, helping address biases and ensure equitable distribution across categories.

In conclusion, YouTube Shorts mark a significant evolution in digital media, essential for understanding the implications for content creators, audience engagement, and the digital ecosystem.

2.3. YouTube Thumbnails

Thumbnails on YouTube are crucial for attracting viewers and influencing video choices, significantly affecting click-through rates (CTR) and overall engagement. Park (2022) found that visually appealing thumbnails with high view counts are more likely to be selected.

The use of eye-catching, exaggerated thumbnails, or clickbait, has been studied extensively. Qu et al. (2018) note that while clickbait generates clicks, it often leads to dissatisfaction when content doesn't meet expectations, highlighting the need for balancing attractiveness with authenticity.

Thumbnails are key to a video's discoverability and engagement. Vitadhani et al. (2021) show that using OCR and face recognition can optimize thumbnails to avoid clickbait and provide a clearer content preview.

Optimizing thumbnails for higher CTR can improve algorithmic recommendations and visibility for content creators. Misleading thumbnails, however, degrade user experience and spread misinformation. Studying thumbnail impact and design strategies can help create a more transparent and user-friendly digital environment.

While most research on thumbnails focuses on full-length YouTube videos, the principles of visual appeal and engagement apply to YouTube Shorts as well. Despite the newness of Shorts and limited specific research, thumbnails still influence content visibility and user engagement, making insights from traditional video studies relevant to this format.

In conclusion, understanding and optimizing YouTube thumbnails is vital for video marketing and engagement, enhancing content creators' reach and ensuring a positive user experience.

2.4. Recommendation Bias

Recommender systems shape user content consumption, but they often embed biases and create filter bubbles. Biases can arise from user preferences, algorithmic design, and training data, necessitating analysis and mitigation to promote diverse and balanced content.

Research has shown that YouTube's recommendation algorithms can narrow the content landscape, limiting diversity and affecting information spread online Cakmak, Agarwal, and Oni (2024). Cakmak, Okeke, Onyepunuka, et al. (2024a) highlight how recommendation algorithms shape public discourse by promoting emotionally charged content.

Srba et al. (2023) audit YouTube's recommendation algorithm, focusing on misinformation and filter bubbles. Okeke et al. (2023) examine emotional bias, showing that specific emotional tones in recommendations influence user engagement. Cakmak, Okeke, Onyepunuka, et al. (2024b) further explore content drift and its implications for content diversity.

Haroon et al. (2022) reveal that YouTube's algorithm can lead to ideological bias and radicalization, especially for right-leaning users.

Addressing recommender bias and drift is crucial for fair and balanced content promotion, impacting societal discourse and user behavior. Ongoing research and intervention are needed to create more equitable recommendation algorithms.

3. Methodology

This section details our data collection and analytical methods, using generative AI models and topic clustering algorithms to identify patterns of bias in YouTube Shorts recommendations.

3.1. Data Collection

We began data collection by holding workshops with subject matter experts to generate relevant keywords for the SCS narrative. These keywords were then used to search for related YouTube Shorts videos.

Since the YouTube Data API doesn't support Shorts, we used APIFY's YouTube Scraper Streamers (2024) to gather 1,210 unique Shorts video IDs. We configured the scraper to prioritize relevance and collect only Shorts, but the number of videos collected was limited, highlighting a shortage of content on specific topics.

To address this, we expanded our keyword list using a snowball method, drawing from video titles, descriptions, and transcriptions from previous research Cakmak and Agarwal (2024) and Cakmak et al. (2023). We generated additional keywords using N-gram collocations to capture frequent or meaningful word combinations, which were then added to our list.

Table 1 illustrates some of the keywords used for data collection in the South China Sea Dispute study. These keywords cover legal rulings (e.g., UNCLOS), geopolitical tensions (e.g., China-Philippines sovereignty), and economic interests (e.g., natural resources). They are crucial for analyzing the legal, political, and economic aspects of the dispute, particularly China's territorial claims and neighboring states' responses. Ultimately, we gathered 2,094 unique video IDs.

Table 1: South China Sea Dispute Related Keywords

Keywords
Permanent Court + Arbitration + South China Sea, China + Philippines + sovereignty South China Sea + UNCLOS + 2016 ruling, Philippines + West Philippine Sea + China China + nine-dash line + historical claims, South China Sea + maritime "justice" + international law China + Philippines + economic cooperation + trade, Philippines + Scarborough Shoal + territorial dispute West Philippine Sea + environmental concerns + coral reefs, China + "Maritime Silk Road" + regional influence South China Sea + ASEAN + regional stability, Philippines + China + "mutual benefit" + diplomacy China + "artificial islands" + military presence, South China Sea + "international arbitration" + dispute resolution China + Philippines + "joint exploration" + resources

3.1.1. Scraping the Recommendations

To measure bias in the recommender system for Shorts videos, we needed to collect the recommended videos by YouTube.

Since no method was available online, we developed our own scraping method.

The previously collected unique video IDs were used as root or seed video IDs for other recommendations. We opened these videos on a fresh WebDriver instance. This ensured that the browsing history was clean, and the cookies on the device were isolated. We did not log into any account to ensure that the user profile was completely neutral and had no effect on the recommendation algorithm.

Next, this process was automated using the Selenium Python library, which allows automating interactions between a browser and websites. After opening the seed videos on WebDriver instances, we used arrow keys to scroll through the Shorts recommended by YouTube. The scrolling process continued until a predetermined depth of 50 was reached where "depth" refers to the position of a video in the recommendation chain (e.g., Depth 1 is the first recommended video, and Depth 5 is the fifth recommended video). The program then closed the browser entirely and moved on to open the next root video on a clean session.

The depth of 50 was chosen to balance thorough data coverage with practical constraints, aiming for a comprehensive analysis of recommendation patterns while keeping data sizes manageable. This depth was selected based on typical user experience.

In the end, we collected 104,700 videos. We then collected the thumbnail images of YouTube Shorts using the YouTube Data API. Some videos were not available, so we filtered those out, resulting in a final dataset of 100,300 videos.

3.2. Caption Generation

To investigate the thumbnail images more meticulously or to be able understand the context that it portrays, we generated the captions of thumbnails. The captions explains us what is inside the image and what is happening. One example illustrated is Figure 1.

We used GPT-4 Turbo to generate captions for YouTube Shorts thumbnails. This version of GPT-4 is faster and more cost-effective while maintaining similar performance. We kept the prompt simple: "What's in the image?" to ensure the captions were purely descriptive, avoiding any bias from more complex instructions.

3.3. Topic Modeling

To understand the thumbnails' captions, we aimed to classify the captions into topics. This classification helps us track how topics evolve after each recommendation.

Figure 1: Caption generated: The image shows multiple fighter jets parked on the deck of an aircraft carrier. The setting suggests a naval aviation environment with the aircraft arranged in close formation, typical for maximizing space on a carrier's flight deck. The view appears to be an overhead or aerial perspective.

3.3.1. Topic Generation with GPT-4o We used GPT-4o, which stands for GPT-4 'optimized' or 'open,' a refined and more efficient version of the original GPT-4 model. It is designed to deliver faster performance and lower operational costs while maintaining the advanced capabilities of GPT-4 in natural language understanding and generation.

The model was used to generate two types of topics. The first type, which we called general topics, was generated without any limitations, resulting in 2,314 unique topics. The second type, categorized topics, was generated with specific constraints, resulting in 20 unique topics. We have selected these 20 categories based on other platforms' categories combined with our own list, which considers many perspectives of subjects. Both types will be illustrated in the upcoming Section 4.

3.3.2. Topic Generation with Meta-Llama The Meta-Llama-3-8B-Instruct model AI@Meta (2024), part of Meta's LLaMA (Large Language Model Meta AI) series, is designed for advanced natural language processing tasks. With its 8 billion parameters, this model excels in understanding and generating complex text. Specifically fine-tuned to follow instructions, it performs exceptionally well in tasks based on user prompts. In our research project, we've utilized this model for topic classification of YouTube Shorts captions. By leveraging its advanced AI and machine learning capabilities, the Meta-Llama-3-8B-Instruct model helps in accurately categorizing the content of captions, ensuring high-quality and contextually relevant outputs tailored to our specific needs.

3.3.3. Topic Generation with BERTopic We used BERTopic, a topic modeling technique that leverages BERT embeddings to capture semantic similarities between documents, resulting in more coherent topics. This method is effective for handling large datasets, making it ideal for tasks like document classification and trend analysis.

We utilized a fine-tuned version of BERTopic Grootendorst (2024), pre-trained on around 1 million Wikipedia pages, capable of identifying 2,377 distinct topics for analyzing video content.

Recent research shows that generative AI models can produce content comparable in quality to human-generated data. Joosten et al. (2024) found AI-generated ideas often excelled in novelty and customer benefit, while Kim et al. (2024) demonstrated moderate correlations between AI and human evaluations. These findings suggest that AI-generated captions offer a credible basis for our analysis, though human verification could enhance the study.

3.4. Clustering Topics

To analyze the general topics discussed in Section 3.3.1, we clustered them due to the large number of topics.

The process began by filtering the dataset to remove uninformative topics like ‘Photograph(s)’, ‘Thumbnail(s)’, ‘Image(s)’, and ‘Video(s)’. We generated BERT embeddings, which capture the semantic meaning of text by considering both preceding and following contexts in a sentence Devlin et al. (2019).

We then reduced the embeddings to two dimensions using t-SNE (t-distributed Stochastic Neighbor Embedding) for better visualization Van der Maaten and Hinton (2008). Unlike PCA (Principal component analysis) Pearson (1901), t-SNE excels at revealing intricate local patterns, making it ideal for interpreting data clusters.

The reduced features were clustered using the OPTICS (Ordering Points To Identify the Clustering Structure) algorithm, which handles varying data densities effectively Ankerst et al. (1999). Unlike K-means Lloyd (1982), which requires a predefined number of clusters, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) Ester et al. (1996), which struggles with varying densities, OPTICS adapts to different densities without needing a set number of clusters, making it suitable for complex data structures with noise and outliers.

4. Results

This section presents our findings on biases in YouTube Shorts recommendations, supported by detailed graphical analyses.

4.1. Clustered General Topics

To visualize the topic clusters of the general topics mentioned in Section 3.3.1, we plotted them in a 2D space at various depths. The X and Y axes represent the t-SNE components, which reduce the dimensionality for visualization purposes. The legend shows noise in gray and other clusters in different colors.

4.1.1. Clusters with GPT As shown in Figure 2, which represents depth 0, the topics are clustered together. In Cluster 0, terms like politics, history, diplomacy, war, and military indicate political themes. Clusters 5, 6, and 7 contain terms like ships, aircraft, and fishing, relating to aircraft carriers and economic perspectives in the sea. Cluster 8 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 clusters, like Cluster 1, showing animated explanations of the topic. From our observations, the initial depth or seed videos were highly relevant to our investigated topic.

The topics are mostly unrelated to our original topic at depth 5, as displayed in Figure 3. For example, 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 away, and many new topics have emerged.

4.1.2. Clusters with Llama The Llama clusters exhibited similarities to those generated by GPT. As shown in Figure 4, we can observe specific terms associated with various clusters: politics in Cluster 0, war and military in Cluster 2, geopolitics in Cluster 4, news in Cluster 5, marine and aerospace in cluster 6, fishing in cluster 7, and transportation in cluster 8. These terms are highly relevant to the South China Sea Dispute, highlighting the algorithm’s ability to identify and group related content effectively.

The presence of terms like politics, war, and geopolitics directly relates to the territorial disputes

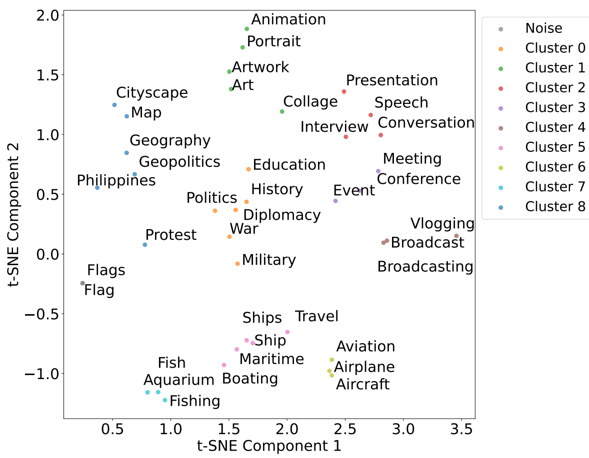

Figure 2: General Topic Clustering for Depth 0

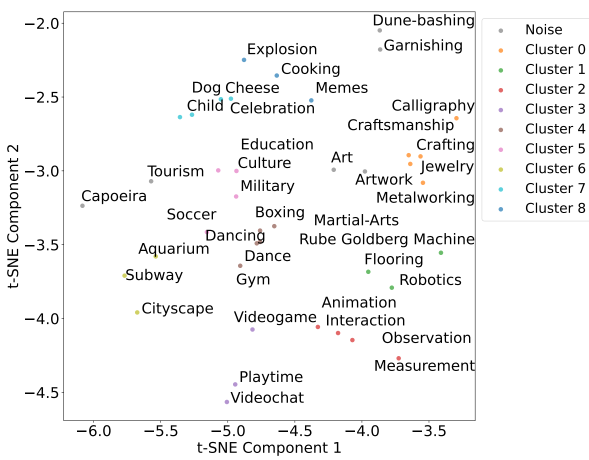

Figure 3: General Topic Clustering for Depth 5

and power dynamics in the South China Sea region. Clusters such as marine and aerospace, fishing, and transportation are pertinent due to the strategic importance of these industries in the contested waters. Additionally, the inclusion of news signifies the global attention and reporting on the ongoing conflicts and developments. This clustering underscores the multifaceted nature of the South China Sea Dispute and the algorithm's competency in capturing its various dimensions.

However, at depth 5, we observe a noticeable shift in the terms associated with the clusters. New terms such as entertainment, music, and gaming appear in Cluster 2; robotics, engineering, and technology in Cluster 4; pets and toys in Cluster 6; jewelry and crafting in Cluster 7; and memes and humorous elements in Cluster 8. Although there are a few terms still related to the original topic, the overall focus has significantly shifted to other subjects. This indicates a diversification in the content being recommended, moving away from the central theme of the South China Sea Dispute to a broader range of interests.

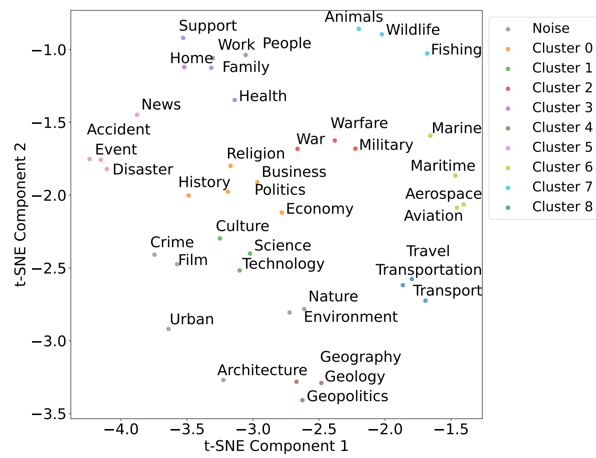

Figure 4: General Topic Clustering for Depth 0

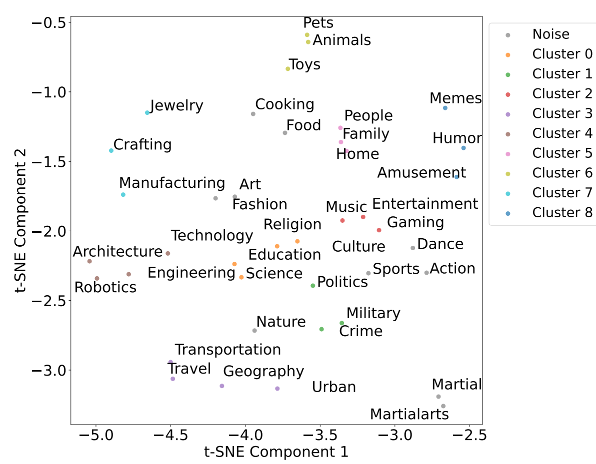

Figure 5: General Topic Clustering for Depth 5

In conclusion, both methods GPT and Llama demonstrated a consistent pattern of topic drifting at the clusters' increasing depths that we analyzed. This phenomenon was evident early on, leading to a rapid divergence from the original topic. Consequently, we chose not to include all depths in this section, as the topics had already shifted significantly in the initial stages, making further inclusion redundant. Additionally, due to space constraints, we limited our illustrations to these two depths for each method. However, in the subsequent result subsections, we will provide more detailed analyses of additional depths to further elucidate this trend.

4.2. Categorized Topics

In this section, we investigated the categorized topics mentioned earlier in Section 3.3.1. We selected 20 topics for this analysis. We chose to illustrate this data using a lollipop chart for greater clarity. The Y-axis represents the topics, while the X-axis shows the topic ratios or distributions between 0 and 1. The legend indicates the depth range and their corresponding colors in the graph. The topics for each depth range are accumulated and normalized. For convenience and clarity, the depths are grouped into five classes, highlighting the differences from the seed depth. Additionally, we filtered out depths with a ratio below 0.01 to enhance visualization.

4.2.1. Topics with GPT In Figure 6, at the initial depth, depth 0, we see that news dominates nearly 40 percent of the topics, followed by politics at around 15 percent, with a few other topics like history and lifestyle also present. After the recommendation algorithm suggested new videos or depths, it is evident that these new topics were neither news nor politics. News topics reduced dramatically across other depths, and political terms disappeared entirely. New topics emerged, primarily entertainment-related, and increased with each depth. Although the levels of lifestyle topics remained relatively stable across depths, this is likely because the lifestyle category is broad and encompasses a wide range of subjects, thereby appearing consistently throughout. Overall, we can clearly see the topic shift happening in the recommendation algorithm by looking at the graph.

4.2.2. Topics with Llama In Figure 7, the initial topics such as history, military, news, and politics are prominently featured, aligning well with our primary focus on the South China Sea Dispute. Notably, Llama provides a smoother and more balanced distribution of topics compared to GPT, which predominantly centered around news. This distribution in Llama better captures the multifaceted nature of the South China Sea Dispute. As we delve deeper into the recommendation depths, we observe a significant shift, with entertainment and other unrelated topics becoming increasingly dominant, mirroring the pattern observed in the GPT results. Overall, the graph clearly illustrates the topic shift occurring within the recommendation algorithm.

4.3. Topic Distribution with BERTopic

We also used BERTopic for topic modeling, visualizing the initial, middle, and end depth topics with radar charts (Figure 8). This method highlighted overall topic distribution, focusing on the three most prevalent topics at each depth to effectively track topic transitions.

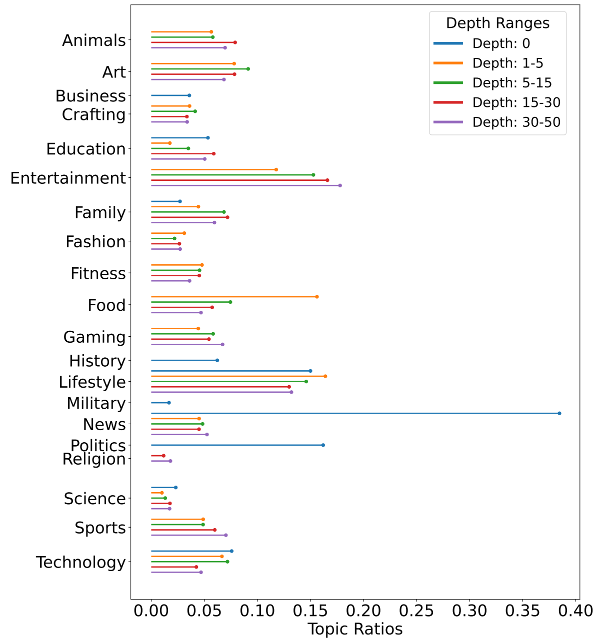

Figure 6: The distribution of Categorized Topics across Depths

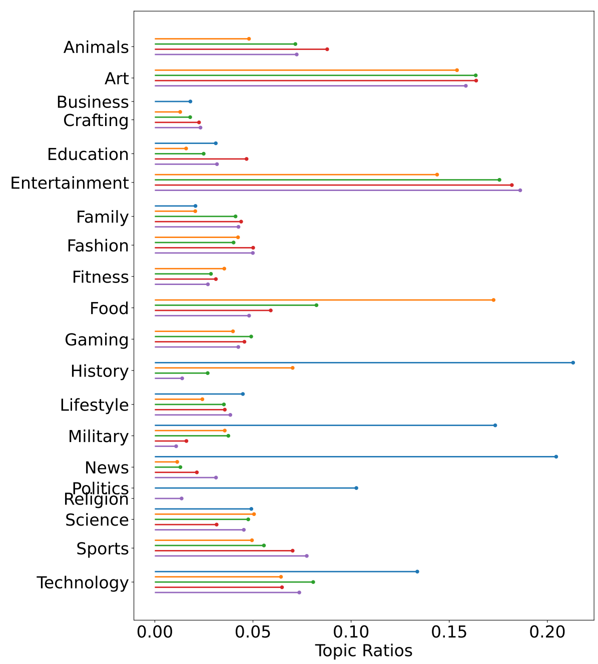

Figure 7: The distribution of Categorized Topics across Depths

Each topic ID and topic we mention is available on Huggingface Grootendorst (2024). At depth 0, the most prominent topics include flag, geography, and uniform. For example, topic ID 1650 (uniforms, uniformed, berets, beret) relates primarily to military and soldiers, while topic 935 (geography, geographic, geographical, geographer) concerns geopolitical regions around the South China Sea. Topic 111 (flags, flag, flagpole, commonwealth) references different nations. These initial topics clearly relate to the South China Sea Dispute.

As we move to deeper levels, we observe the emergence of unrelated topics, and the initial topics begin to fade away. For example, 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 that the recommendation algorithm is steering away from the original subject matter towards more general and unrelated topics.

While BERTopic may not capture topics as precisely as GPT or Llama, it effectively illustrates topic drift across depths, showing how quickly the algorithm shifts focus from relevant topics to broader, unrelated ones.

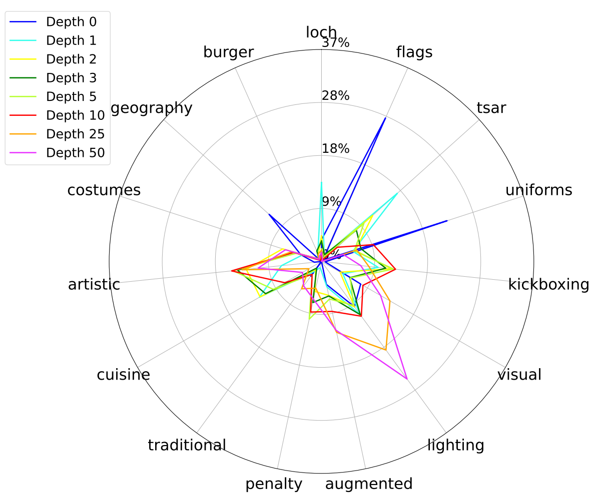

Figure 8: Topic Distribution Using BERTopic Across Depths

4.4. Engagement Statistics

We analyzed views, likes, and comments to see if YouTube's algorithm favors popular videos. By converting these metrics to logarithmic values (base 10), we made the data more comparable and visually coherent.

In Figure 9, we observe that initial engagement metrics at depth 0 are low. However, beyond this point, there is a dramatic increase in engagement metrics, with mean values reaching thousands or even tens of thousands. The algorithm appears to maintain this high level of engagement, suggesting the presence of a threshold that prevents recommendations from falling below a certain level.

This clearly indicates the presence of popularity bias on the platform, where more popular videos are recommended more frequently, leading to higher user engagement regardless of the original topic.

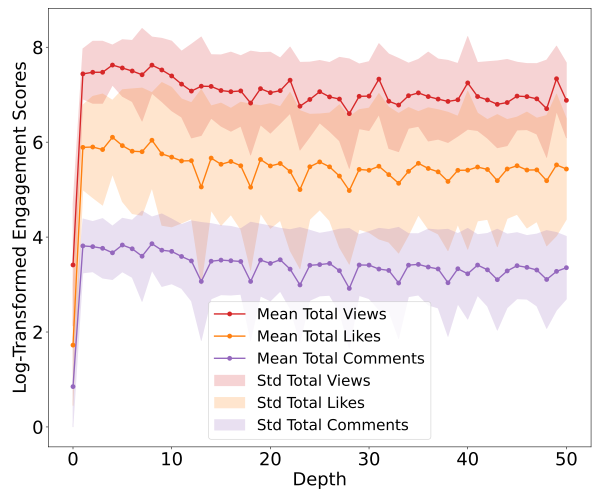

Figure 9: Engagement Values

5. Conclusion and Discussion

In conclusion, our study investigated potential algorithmic bias in YouTube Shorts' video recommendations, focusing on thumbnail captions and the South China Sea Dispute as a case study. We found a clear topic drift, where broader and less relevant videos are recommended after the initial set. This drift seems driven by YouTube's algorithm prioritizing high-engagement, entertainment-focused content, leading to the neglect of serious topics and creating algorithmic bias.

For future research, we plan to explore a wider range of topics, comparing well-known subjects with specialized ones like the South China Sea Dispute. This will help us better understand YouTube's algorithm across different content domains. Additionally, we will incorporate user interactions, such as liking and commenting, to see how engagement affects recommendation patterns and content drift. Furthermore, we will conduct human annotator experiments to verify the reliability of GPT-4-generated captions. This will be our immediate future research direction.

We also considered the potential impact of IP addresses and geolocation on recommendations. To minimize bias, we used consistent IP locations during data collection. Future studies will explore varying locations to assess their influence on the recommendation system.

While user interaction with thumbnails in YouTube Shorts may differ from full-length videos, thumbnails still play a crucial role in influencing algorithmic recommendations, particularly in contexts like recommended feeds or search results. Our focus on thumbnails offers valuable insights into the biases within these recommendations, even if user interaction with them is more subtle.

Although we collected various video attributes like titles and author information, our study specifically analyzed thumbnails to explore their unique role in YouTube Shorts recommendations. Given the limited research on this aspect, our work provides a novel perspective, with future research aiming to incorporate additional video attributes for a more comprehensive analysis.

This research is significant as it highlights the biases in algorithmic recommendations, particularly from the thumbnail perspective. Understanding these biases is crucial for ensuring a more balanced representation of diverse and serious topics on platforms like YouTube. Our study contributes to the broader conversation on digital media ethics, emphasizing the need for transparency and accountability in content recommendation systems.

Acknowledgments

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078), U.S. Office of Naval Research (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, 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. Any opinions, findings, and 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

AI@Meta. (2024). Llama 3 model card. https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). Optics: Ordering points to identify the clustering structure. *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, 49–60. https://doi.org/10.1145/304182.304187

Cakmak, M. C., & Agarwal, N. (2024). High-speed transcript collection on multimedia platforms: Advancing social media research through parallel processing. *2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), IEEE, accepted for presentation*.

Cakmak, M. C., Agarwal, N., & Oni, R. (2024). The bias beneath: Analyzing drift in YouTube's algorithmic recommendations. *Social Network Analysis and Mining*, *14*(1), 171. https://doi.org/10.1007/s13278-024-01343-5

Cakmak, M. C., Okeke, O., Onyepunuka, U., Spann, B., & Agarwal, N. (2024a). Analyzing bias in recommender systems: A comprehensive evaluation of youtube's recommendation algorithm. *Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 753–760. https://doi.org/10.1145/3625007.3627300

Cakmak, M. C., Okeke, O., Onyepunuka, U., Spann, B., & Agarwal, N. (2024b). Investigating bias in youtube recommendations: Emotion, morality, and network dynamics in china-uyghur content. In H. Cherifi, L. M. Rocha, C. Cherifi, & M. Donduran (Eds.), *Complex networks & their applications xii* (pp. 351–362). Springer Nature Switzerland.

Cakmak, M. C., Okeke, O., Spann, B., & Agarwal, N. (2023). Adopting parallel processing for rapid generation of transcripts in multimedia-rich online information environment. *2023 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 832–837. https://doi.org/10.1109/IPDPSW59300.2023.00139

Chubb, A. (2020). Prc assertiveness in the south china sea: Measuring continuity and change, 1970–2015. *International Security*, *45*(3), 79–121.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *kdd*, *96*(34), 226–231.

Fravel, M. T. (2011). China's strategy in the south china sea. *Contemporary Southeast Asia*, *33*(3), 292–319. Retrieved June 13, 2024, from http://www.jstor.org/stable/41446232

Grootendorst, M. (2024). Bertopic_wikipedia [Accessed: 2024-05-01].

Haroon, M., Chhabra, A., Liu, X., Mohapatra, P., Shafiq, Z., & Wojcieszak, M. (2022). Youtube, the great radicalizer? auditing and mitigating ideological biases in youtube recommendations.

Joosten, J., Bilgram, V., Hahn, A., & Totzek, D. (2024). Comparing the ideation quality of humans with generative artificial intelligence. *IEEE Engineering Management Review*, *52*(2), 153–164. https://doi.org/10.1109/EMR.2024.3353338

Kim, J., Lee, T. H., Bae, Y., & Kim, M. K. (2024). A comparison between ai and human evaluation with a focus on generative ai. *Proceedings of the 18th International Conference of the Learning Sciences-ICLS 2024, pp. 1722-1725*.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Macaraig, C. E., & Fenton, A. J. (2021). Analyzing the causes and effects of the south china sea dispute: Natural resources and freedom of navigation. *The Journal of Territorial and Maritime Studies*, *8*(2), 42–58. Retrieved June 13, 2024, from https://www.jstor.org/stable/48617340

Okeke, O., Cakmak, M. C., Spann, B., & Agarwal, N. (2023). Examining content and emotion bias in youtube's recommendation algorithm. *the Ninth International Conference on Human and Social Analytics, Barcelona, Spain*.

Park, J. (2022). *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.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Qu, J., Hißbach, A. M., Gollub, T., & Potthast, M. (2018). Towards crowdsourcing clickbait labels for youtube videos. *HCOMP (WIP&Demo)*.

Rajendran, P. T., Creusy, K., & Garnes, V. (2024). Shorts on the rise: Assessing the effects of youtube shorts on long-form video content.

Sahu, G., Gaur, L., & Singh, G. (2023). Investigating the impact of personality tendencies and gratification aspects on ott short video consumption: A case of youtube shorts. *2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 1–6. https://doi.org/10.1109/ICIPTM57143.2023.10118122

Srba, I., Moro, R., Tomlein, M., Pecher, B., Simko, J., Stefancova, E., Kompan, M., Hrckova, A., Podrouzek, J., Gavornik, A., & Bielikova, M. (2023). Auditing youtube's recommendation algorithm for misinformation filter bubbles. *ACM Trans. Recomm. Syst.*, *1*(1). https://doi.org/10.1145/3568392

Streamers. (2024). Youtube scraper [Accessed: 2024-01-10].

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, *9*(11).

Violot, C., Elmas, T., Bilogrevic, I., & Humbert, M. (2024). Shorts vs. regular videos on youtube: A comparative analysis of user engagement and content creation trends. *ACM Web Science Conference*. https://doi.org/10.1145/3614419.3644023

Vitadhani, A., Ramli, K., & Dewi Purnamasari, P. (2021). Detection of clickbait thumbnails on youtube using tesseract-ocr, face recognition, and text alteration. *2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST)*, 56–61. https://doi.org/10.1109/ICAICST53116.2021.9497811