Analyzing Narrative Evolution about South China Sea Dispute

Nitin Agarwal*†

*COSMOS Research Center, University of Arkansas at Little Rock, Arkansas, USA
†International Computer Science Institute, University of California, Berkeley, California, USA
e-mail: nxagarwal@ualr.edu

Abstract—YouTube has become the leading global platform for video sharing, yet most research on YouTube Narratives focuses mainly on metadata rather than video content. To address this gap, our study examines narratives embedded in 9,000 YouTube videos about the South China Sea Dispute using Generative Pre-trained Transformer (GPT-5). We also developed a visualization tool to depict connections between keywords and narratives. Our findings show that attention to the dispute surged in 2016 following China's construction and militarization of artificial islands. In 2022, misinformation-driven anti-Western narratives increased five-fold, portraying the United States as exploiting the dispute. This indicates efforts to manipulate the information environment through narrative amplification. Additionally, the dominant narrative cluster incorporated themes related to the Russia-Ukraine war and the COronaVIrus Disease of 2019 (COVID-19) pandemic.

Keywords-South China Sea; GPT-5; narrative extraction; narrative evolution; narrative visualization; YouTube; social media.

I. INTRODUCTION

Videos have become a dominant form of online content, shaping how people acquire knowledge and engage with information. Among platforms, YouTube stands out as the leading medium for video sharing and consumption, offering collaboration, diverse content, and valuable insights. Users watch 1 billion hours of video daily and upload 500 hours every minute [1], fueling exponential data growth. However, this growth also fosters divergent narratives, as some exploit YouTube to spread propaganda and misinformation. This is alarming given that nearly 70% of Americans rely on YouTube as an information source [2]. Understanding the evolution of such narratives has become increasingly complex. Most existing studies focus on metadata analysis [3], which overlooks the nuanced narratives embedded within video content. There remains a gap in computational methods that can extract and analyze these narratives for situational awareness and misinformation countermeasures.

To address this, our study employs the Generative Pre-Trained Transformer (GPT-5), an instruction-based model commonly used for text generation but adaptable for narrative extraction. A survey [4] found that zero-shot GPT-5 performs well in abstractive summarization—a task akin to narrative extraction. GPT's extensive training on large text corpora enables it to effectively process natural language [5] and grasp contextual meaning across segments [6], making it suitable for analyzing video narratives reliant on spoken language.

We apply this approach to the South China Sea dispute, a complex geopolitical conflict involving China and neighboring countries—the Philippines, Vietnam, Malaysia, Brunei, and Taiwan—over resource-rich maritime territories [7]. Covering videos from November 2014 to May 2023, our analysis examines narrative evolution and their impact on viewers' attitudes and values on YouTube.

We address the following research issues in this study:

- We explore how Generative AI, especially the Pre-Trained model (GPT-5), can be utilized for analyzing narratives in video content.
- We investigate the trend of narratives surrounding the South China Sea dispute and examine how they shift over time.

This study makes the following contributions:

- We analyze the actual video content using GPT-5 to extract narratives from YouTube videos by enhancing the understanding of embedded information.
- Our analysis uncovers narratives that intersect with other global issues, like the Russia-Ukraine conflict and COVID-19, showing how narratives within one geopolitical area can be shaped by or connected to broader global events.
- We highlight a significant five-fold rise in anti-Western, misinformation-driven narratives in 2022, that could sway public perception of the South China Sea by amplifying targeted narratives.

The paper is organized as follows. Section II presents a brief review of the literature related to narrative extraction. Section III explains the methodology. The research findings are presented in Section IV. Finally, we draw conclusions from the paper, identify its limitations, and suggest future research directions in Section V.

II. LITERATURE REVIEW

Numerous researchers have explored diverse approaches to computational narrative extraction over the years. The aim of *computational narratology* is to examine narratives from a computational and information-processing perspective, emphasizing the algorithmic processes related to narrative creation and interpretation. It involves the modeling of narrative structure using formal and computable representations, as outlined by Mani [8]. Narratives can be analyzed from various perspectives, including morality, toxicity, sentiment, emotion, plot, and story accomplishment. Initially, computational linguistics focused on shorter texts like fables and folk stories but has shifted to analyzing full-length novels and movie scripts. According to [9], the study of narratives in natural language texts can be condensed into five main stages:

pre-processing and parsing, identification and extraction of narrative components, linking components, representation of narratives, and evaluation. For example, in a survey paper, researchers combined blog posts and named entities to generate topics through network topic modeling. Parameters of the LDA model were tuned to determine the number of topics, and Natural Language Processing (NLP) techniques were used to extract sentences and identify noun and verb phrases. Grammar rules captured patterns to generate narratives, which were ranked using Term Frequency-Inverse Document Frequency (TF-IDF) scoring.

A. Narrative Extraction and Analysis

Gurung et al. [10] emphasize the narrative spread of content sharing the invocation of emotionally charged themes, triggering cognitive biases, and reinforcing existing beliefs, especially those tied to high-stakes issues like the South China Sea. Ranade et al. [11] utilize a comprehensive overview of the computational methods used to extract and analyze narratives across fragmented, multi-source datasets typical of social media. In his work, he identified the use of NLP techniques to detect central themes, tracking progressions and shifts in narrative emphasis, which are essential in understanding how South China Sea narratives evolve across different YouTube channels and over time. This, however, leads to a computational framework that reveals the extraction of narrative elements from YouTube videos and their comments that are related to territorial claims or regional power dynamics. Framing patterns and narrative arcs across different geopolitical actors were included. Kuenzler [12] applied the Narrative Policy Framework (NPF) to provide insights into how institutions and states constructed narratives manage public opinion, assert authority, and deflect criticism. The NPF examines how narratives are strategically constructed to either enhance an organization's reputation or achieve specific policy objectives. The NPF, therefore, provides a lens for analyzing how these entities utilize storytelling elements to influence public opinion on YouTube; this framework was also employed to examine how narratives evolved in response to geopolitical tensions and the role of digital media in shaping these dynamics. Sandberg [13] harnesses a methodological toolkit that could be adapted to study YouTube narratives in conflict zones. In his work, he outlines thematic, structural, performative, and dialogical approaches to narrative analysis. This thematic analysis focuses on identifying key themes such as sovereignty, security, and regional stability, while the structural analysis reveals the narratives used by different YouTube channels, showing recurring patterns in the representation of actors and adversaries in these narratives.

B. Challenges in Misinformation and Narrative Manipulation

Hussain et al. [14] explore the more troubling aspects of narrative-building on digital platforms, examining how various players manipulate social media to spread misinformation, influence public agendas, and create disinformation. They contend that YouTube's vast audience and limited content restrictions allow state-driven narratives to gain traction, which can influence public perception and contribute to a biased narrative landscape. This highlights the importance of vigilant monitoring and analytical tools capable of distinguishing between natural narratives and those crafted to influence opinions on complex geopolitical issues. Santana et al. [9] review the intricate process of extracting narratives from unstructured data, a frequent feature of digital media. Their research highlights the technical and methodological hurdles involved in parsing, annotating, and connecting narrative elements within vast amounts of text and video content, as commonly seen with YouTube data. The challenges of ensuring temporal consistency, handling multiple languages, and adapting NLP techniques for multimedia analysis emphasize the need for advanced computational tools specifically designed for the unique structure of YouTube narratives.

However, the emergence of pre-trained large language models has revolutionized these processes. According to a study by Stambach et al. [15], GPT has the ability to discern key characteristics and fulfill various roles across different domains, such as newspapers and political speeches, without the need for any additional training data beyond a prompt. Rather than relying on predefined rules, these large language models prioritize parameter tuning. In the approach proposed by Liu et al. [16], trainable continuous prompt embeddings were employed to enhance the accuracy of models like GPT and BERT by an impressive 80%. Recent studies [17] have put forward methods for comprehending figurative language, including sarcasm, metaphor, and idioms, in discriminative and generative tasks, effectively narrowing the gap between model performance and human understanding. Agarwal [18] demonstrated that GPT-4-based analysis can further reveal such biases between the text and the narratives extracted from the videos.

Narrative evolution provides insights into how stories, perspectives, and themes change over time. Understanding the evolution of narratives helps us comprehend societal shifts, cultural changes, online users' behaviors [19], and the dynamics of public opinion. For instance, authors in [20] analyzed blogs during the European migrant crisis to study narrative shifts related to refugees or migrants, using named-entity extraction. Similarly, authors in [21]–[23] identified and analyzed the influence of topics in blogs.

III. DATA AND METHODOLOGY

We collected YouTube videos focusing on the South China sea dispute. The data collection process began with the utilization of YouTube API keys, and these videos were gathered using specific keywords selected by subject matter experts. Some of these curated keywords are *Security*, *Chinese militia*, *U.S. interference*, *Philippine Coast Guard*, *Philippine Navy*, *People's Liberation Army*, and *Chinese Coast Guard/CCG*. Through this approach, we successfully amassed a dataset comprising 14,000 YouTube videos. The collected data encompassed metadata, including video ID, publication date, title, video description, and location. For transcription, we follow
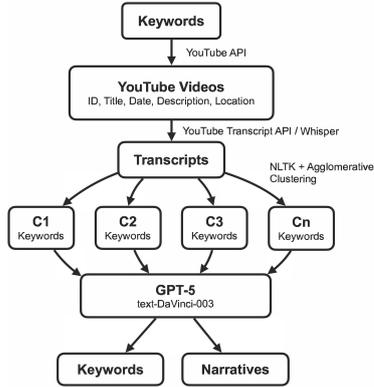
Figure 1. This figure illustrates the methodology for extracting keywords and narratives using GPT-5.

Figure 2. A drill-down style narrative visualization tool that delves into different facets associated with the South China Sea region.

the methodology suggested in [24]. There, the authors leverage parallel computing and the Python multiprocessing library to improve the speed of transcript collection from YouTube. This utilizes YouTube's Transcript API to extract YouTube-generated transcripts, and also uses OpenAI's Whisper model to generate transcriptions on videos without native YouTube transcriptions. Furthermore, the Google Translation API was employed to translate transcriptions from non-English videos. The initial number of collected videos was 14,000. However, after the transcription process, we were able to retrieve data from only 9,000 videos. This reduction in the number of videos was due to various factors, such as videos being taken down or removed during the transcription process, as well as some videos being live streams that did not have transcriptions available. After completing the transcription process, we utilized the NLTK library to eliminate punctuation and basic English stopwords. This helps remove unnecessary characters that could impact further analysis. It is important to consider that GPT-5 has a token limitation of 4096, which indicates the maximum number of tokens that can be processed in a single API call. To overcome this limitation, we implemented agglomerative clustering, as described in [25], to divide the transcripts into multiple chunks for further processing. This methodology is demonstrated in Figure 1.

A. Narrative Extraction

In our Narrative extraction process, we utilized the GPT-5 model, specifically the *text-DaVinci-003* engine. To prioritize accuracy, we set the temperature parameter to 0, ensuring precise, factual results. For the max-tokens parameter, striking the right balance was crucial for providing both a concise narrative and relevant details [26]. For keyword extraction, we set the value to 8, aiming for an average of 5 top keywords. For the macro-narrative generation, we used a value of 35 for an average of 2 to 3 sentences. One of the important parameters is prompt crafting, as a well-designed prompt is crucial when using GPT-5. In our project, we employed the prompt "Given the transcript, provide keyword" for keyword extraction. Similarly, for narrative generation, we used a prompt instructing the model to generate a narrative based on the provided keywords.
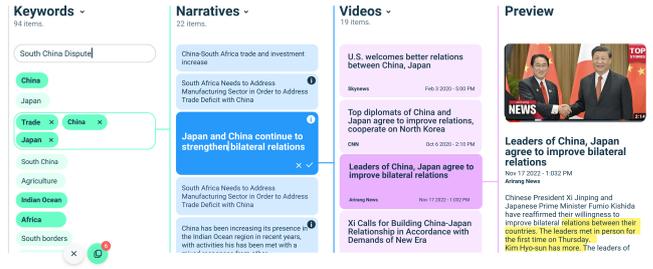
These prompts were used in [4] for news summarization. Each video's title, description, and transcript were considered; these narratives were then analyzed to gain deeper insights into their content. Furthermore, to assess the credibility of our narratives, we conducted a comparative analysis with the narratives generated by [14]. Their framework leveraged natural language processing techniques, including Part-Of-Speech (POS) tagging, chunking, and grammar rules, to extract actors and actions from the text and uncover associated narratives. We employed the Universal Sentence Encoder (USE) to compare both sets of narratives. The results indicated a significant average semantic textual similarity score of 0.7, indicating a strong alignment between the narratives.

B. Narrative Visualization

Our research introduces a custom-built narrative visualization tool as demonstrated in Figure 2, integrated into our web application called Vtracker [3], for analyzing videos. It utilizes a structured four-column format for analyzing videos. The tool displays influential entities, associated narratives, and detailed video information. It enables efficient exploration of video content in a visually appealing and user-friendly manner. Furthermore, users can edit narratives in the tool, allowing for user feedback and continuous improvement of the narrative structure. This iterative process helps refine and enhance the tool based on user preferences and feedback.

IV. RESEARCH FINDINGS

In this section, we discuss our research findings on the evolution of macro-narratives, and sentiment, emotion, morality, and toxicity analysis of the narratives.

A. Macro Narrative Evolution

From the macro-narratives displayed in Figure 3, Cluster 0 examines how China's geopolitical strategy is affected by the crisis between Russia and Ukraine. Particular micro-narratives explore how China is changing its position and tactics in reaction to the Russia-Ukraine war each month from January to December. This covers issues including China's militarization of SCS islands, its position on sanctions, and Chinese-Russian joint military drills.
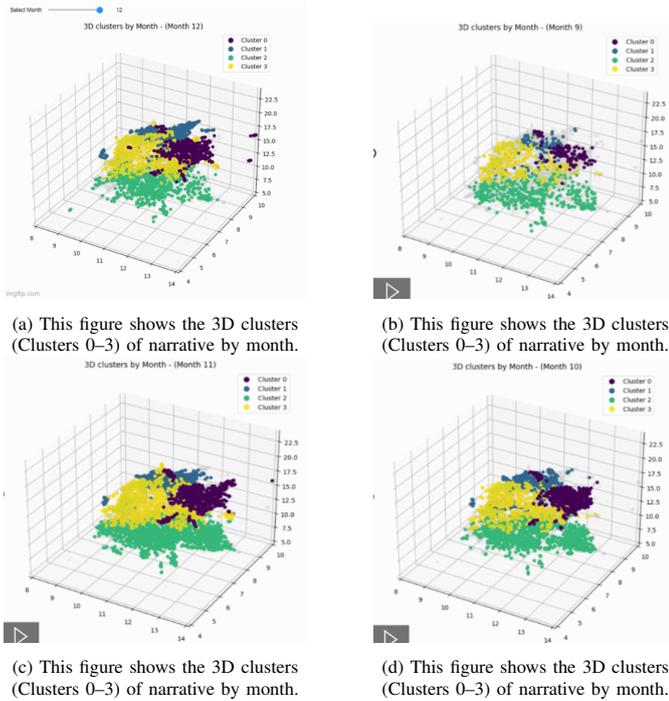
(a) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.

(b) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.

(c) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.

(d) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.

Figure 3. The figure shows the 3D clusters of narrative by month for Cluster 0, Cluster 1, Cluster 2, and Cluster 3.
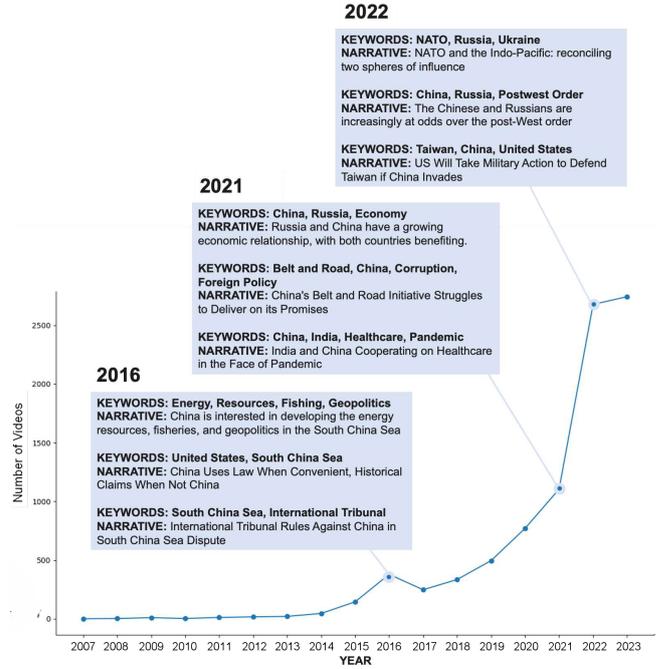

Figure 4. Posting trend of the South China Sea dispute from 2007 to 2023.

Figure 5. Narrative evolution surrounding the relationship between China and the United States.

China's hostility in the South China Sea is highlighted in Cluster 1. China's geopolitical initiatives, such as diplomatic protests, military drills, and strategic expansions, have heightened tensions with the United States and its neighbors. These developments are detailed in monthly micro-narratives. Cluster 2 highlights US security spending in the Philippines as a component of its larger Asia policy, which is entwined with a number of political and economic problems in the country. Cluster 3 macro-narratives critique President Biden's policies, about social inequality and economic difficulties in the United States, as well as unpopular diplomatic endeavors with Saudi Arabia.

Furthermore, we delve into the identified keywords and narratives associated with the dispute, thoroughly analyzing their alignment with real-life events. We draw parallels between our findings and the dynamic trends and shifts. The earliest video related to the South China Sea dispute dates back to 2007. From that period until 2014, there was limited attention given to the conflict. However, from late 2014 to 2016, there was a slight increase in discussions surrounding the matter. One prominent narrative that persisted during this time was illegal fishing, with videos highlighting fishermen's rights and shedding light on the issue. Notably, Indonesia took action by seizing 153 fishing vessels from neighboring countries, including 50 from Vietnam, 43 from the Philippines, and 1 from China [14].

From 2016 onward, discussions on "infrastructure" and "development" grew following the U.S. release of satellite images showing weapon construction on disputed South China Sea islands—a narrative that remains active. China asserted sovereignty based on historical fishing, while others argued this violated the Law of the Sea. Discourse on the dispute gradually expanded between 2016 and 2021, then surged by 108.30% from 2021 to 2022, as shown in Figure 4.
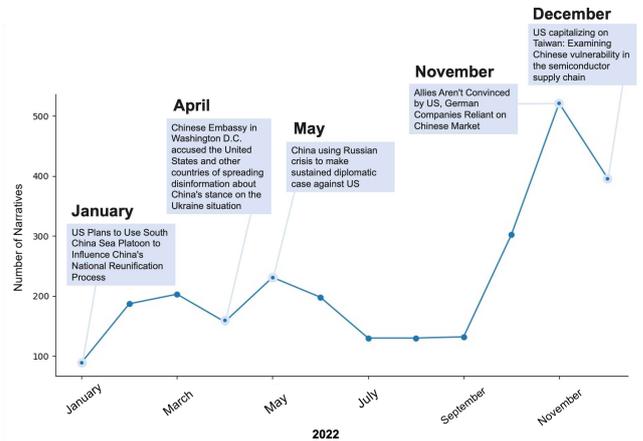
As demonstrated in Figure 5, the narratives in January and December 2022 exhibit a likely bias against the US, portraying it as attempting to exploit the South China Sea dispute. These narratives, often misleading, have been significantly magnified, being amplified fivefold. To identify various narratives, we applied K-means clustering to the 2022 narratives, resulting in the identification of 10 distinct clusters.

Among these clusters, we identified Clusters 2 and 5 as dominant narratives. These clusters are represented in Figure 6. To gain further insights, we conducted Latent Dirichlet
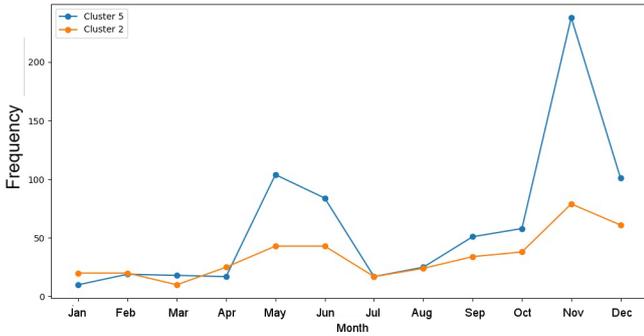
Figure 6. Trend of narratives for Clusters 2 and 5 throughout the year 2022.

Allocation (LDA) analysis on these top clusters to identify prevailing topics within the narratives.

Topics in Cluster 5:

Topic 1: ['Ukraine', 'Russia', 'Russian', 'military', 'Trump']

Topic 2: ['foreign', 'policy', 'Russia', 'climate', 'alien']

Topic 3: ['vaccine', 'global', 'war', 'COVID-19', 'Philippines']

Topic 4: ['COVID-19', 'global', 'health', 'Taiwan', 'pandemic']

Topic 5: ['foreign', 'Asia', 'human', 'alien', 'United']

In Cluster 5, May and November 2022 saw a significant number of posts focusing on the geopolitical dynamics between Ukraine and Russia. In May 2022, narratives centered around the COVID-19 pandemic and the Global COVID-19 Summit hosted by the United States, with Indonesia as a co-host. In November 2022, discussions included China's position on the Ukraine peace talks, highlighted by Chinese leader Xi Jinping during the G20 Summit in Bali. The Philippines also discussed high unemployment rates and COVID-19 deaths in comparison to other Asian countries, particularly India.

Topics in Cluster 2:

Topic 1: ['China', 'security', 'trade', 'infrastructure', 'Africa']

Topic 2: ['deterrence', 'China', 'United', 'Taiwan', 'defence']

Topic 3: ['China', 'trade', 'armed', 'politics', 'deal']

Topic 4: ['Russia', 'Ukraine', 'China', 'NATO', 'Putin']

Topic 5: ['politics', 'human', 'China', 'work', 'social']

Cluster 2 focuses on China's engagement in security, trade, and infrastructure, as well as its relationships with Africa. Topic 1 gained prominence in March 2022 and was influenced by Japanese Foreign Minister Yoshimasa Hayashi, expressing concerns about challenges faced by African nations due to the pandemic and the conflict in Ukraine [22]. Topic 2 delves into the delicate position of the United States regarding Taiwan. In this cluster, Russia and Ukraine also feature but in the context of defense and NATO's assistance to Ukraine. Disinformation surrounding Australia's nuclear submarines investment is addressed, along with concerns about China's military expansion in the South Pacific. The Quad's focus has evolved to provide

public goods, but Southeast Asian countries express mixed responses, cautious about potential tensions with China.

These findings demonstrate the use of computational toosl to detect narrative evolution and dynamics. Despite 2022 having the highest frequency of content related to the dispute, the top two narratives are focused on varied topics rather than the actual dispute itself. Socio-computational methods presented in this study enable detection of such narrative shifts that can help policy/decision makers.

V. CONCLUSIONS AND FUTURE WORK

This paper utilized GPT-5 for extracting narratives from YouTube videos discussing the South China Sea dispute. We also developed a narrative visualization tool to visualize these narratives along with their corresponding keywords. Our analysis covered the posting trend from 2007 to 2023, with a significant surge in postings related to the dispute observed between 2021 and 2022. In order to understand this surge, we conducted a thorough analysis of the year 2022. Our findings revealed a five-fold increase in anti-West narratives from January to December 2022. Additionally, we analyzed the top narrative clusters for 2022, which highlighted a noticeable divergence. This work can aid analysts in delving deeper into the narratives embedded within videos and inform strategic actions based on their findings.

This study demonstrates that combining LLMs with narrative visualization enables scalable and systematic analysis of narrative dynamics within video-based information ecosystems. By extracting and clustering narratives from YouTube content over time, the approach advances computational social science beyond text-centric analyses and provides a practical framework for detecting narrative inflection points, divergence, and escalation. The identification of a sharp rise in anti-West narratives during 2021–2022 illustrates the utility of AI-assisted methods for revealing temporal shifts that are difficult to observe through manual or episodic analysis, while visualization supports human-in-the-loop sensemaking and interpretability in high-volume, multimodal data.

Beyond methodological contributions, the findings have broader implications for international relations, media studies, and public policy. The observed narrative surge and divergence surrounding the South China Sea dispute offer empirical insight into information competition and strategic narrative formation in contested geopolitical contexts. More broadly, this work provides a practical, ethically grounded analytical framework to monitor digital narratives across domains, supporting early warning, evidence-based strategic communication, and policy evaluation. It contributes to interdisciplinary efforts to understand and respond to evolving information environments using transparent and responsible AI tools.

The current method lacks accounting for replicated videos and off-topic content, posing complexities. We manually disregarded irrelevant clusters, but we need a computational approach to remove such content accurately. In the future, we will evaluate the accuracy of our results with more existing methodologies. Furthermore, testing unexplored parameters of

GPT-5 could enhance our results. Additional content analysis, such as sentiment, emotion, toxicity, and morality analysis, could reveal deeper insights explaining which narratives resonated with the target audience and how the audience reacted. Finally, extensive testing for the narrative visualization tool is necessary to further validate its effectiveness and usability.

REFERENCES

[1] Statista Research Department, *Hours of video uploaded to youtube every minute*, Statista, statistic no. 259477, Retrieved January, 2026, 2023. [Online]. Available: https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/.

[2] B. Auxier and M. Anderson, "Social media use in 2021," Pew Research Center, Tech. Rep., 2021, Retrieved January, 2026, pp. 1–4. [Online]. Available: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

[3] T. Marcoux, O. Adeliyi, D. S. Banjo, M. I. Gurung, and N. Agarwal, "Exploring online video narratives and networks using vtracker," in *Cyber Security and Social Media Applications*, S. T. Özyer and B. Kaya, Eds., Springer, 2023. DOI: 10.1007/978-3-031-33065-0_6.

[4] T. Goyal, J. Li, and G. Durrett, "News summarization and evaluation in the era of gpt-3," *arXiv preprint arXiv:2209.12356*, vol. 1, pp. 1–4, 2022, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/2209.12356.

[5] K. Ethayarajh, "How contextual are contextualized word representations? comparing the geometry of bert, elmo, and gpt-2 embeddings," *arXiv preprint arXiv:1909.00512*, 2022, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/1909.00512.

[6] R. Dale, "Gpt-3: What's it good for?" *Natural Language Engineering*, vol. 27, no. 1, pp. 113–118, 2021. DOI: 10.1017/S1351324921000107.

[7] M. Blaine, "The geography of territorial disputes in the south china sea," *Artificial Intelligence Review*, pp. 287–303, 2022. DOI: 10.1007/s10462-021-10000-x.

[8] I. Mani, *Computational modeling of narrative*. Morgan & Claypool Publishers, 2012, vol. 5, pp. 1–142. DOI: 10.2200/S00465ED1V01Y201204HLT016.

[9] B. Santana *et al.*, "A survey on narrative extraction from textual data," *Artificial Intelligence Review*, vol. 56, pp. 8393–8435, 2023. DOI: 10.1007/s10462-022-10338-7.

[10] M. I. Gurung, N. Agarwal, and A. Al-Taweel, "Are narratives contagious? modeling narrative diffusion using epidemiological theories," in *FOSINT-SI, ASONAM 2024*, 2024.

[11] P. Ranade, S. Dey, A. Joshi, and T. Finin, "Computational understanding of narratives: A survey," *IEEE Access*, vol. 10, pp. 101575–101587, 2022. DOI: 10.1109/ACCESS.2022.3205314.

[12] J. Kuenzler, "From zero to villain: Applying narrative analysis in research on organizational reputation," *European Policy Analysis*, vol. 7, no. S2, pp. 405–424, 2021. DOI: 10.1002/epa2.1123.

[13] S. Sandberg, "Narrative analysis in criminology," *Journal of Criminal Justice Education*, vol. 33, no. 2, pp. 212–229, 2022. DOI: 10.1080/10511253.2022.2027479.

[14] M. Hussain, H. A. Rubaye, K. Bandeli, and N. Agarwal, "Stories from blogs: Computational extraction and visualization of narratives," in *Text2Story@ ECIR*, 2021, pp. 33–40.

[15] D. Stammbach, M. Antoniak, and E. Ash, "Heroes, villains, and victims, and gpt-3: Automated extraction of character roles without training data," *arXiv preprint arXiv:2205.07557*, 2022, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/2205.07557.

[16] X. Liu *et al.*, "Gpt understands, too," *arXiv preprint arXiv:2103.10385*, 2021, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/2103.10385.

[17] T. Chakrabarty, Y. Choi, and V. Shwartz, "It's not rocket science: Interpreting figurative language in narratives," *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 589–606, 2022. DOI: 10.1162/tacl_a_00485.

[18] N. Agarwal, "From metadata to meaning: Gpt-4 reveals bias trends in youtube," in *Proceedings of the First IARIA International Conference on AI-based Media Innovation (AIMEDIA 2025)*, Venice, Italy: IARIA, Jul. 2025, pp. 83–88, ISBN: 978-1-68558-330-9.

[19] S. Shajari, N. Agarwal, and M. Alassad, "Commenter behavior characterization on youtube channels," in *eKNOW 2023, Fifteenth International Conference on Information, Process, and Knowledge Management*, Apr. 2023, pp. 59–64.

[20] M. Hussain, K. Bandeli, S. Al-Khateeb, and N. Agarwal, "Analyzing shift in narratives regarding migrants in europe via blogosphere," in *Text2Story@ ECIR*, vol. 585, 2018, pp. 33–40.

[21] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, "Identifying influential bloggers in a community," in *Proceedings of the 1st ACM International Conference on Web Search and Data Mining (WSDM'08)*, Stanford, California, USA: ACM, Feb. 2008, pp. 207–217. DOI: 10.1145/1341531.1341569.

[22] M. Shaik, Z. Stine, and N. Agarwal, "Developing situational awareness from blogosphere: An australian case study," in *Proceedings of the Eleventh International Conference on Social Media Technologies, Communication, and Informatics (SOTICS 2021)*, Barcelona, Spain: IARIA, Oct. 2021.

[23] R. Strating and J. Wallis, "Maritime sovereignty and territorialisation: Comparing the pacific islands and south china sea," *Marine Policy*, p. 105110, 2022. DOI: 10.1016/j.marpol.2022.105110.

[24] M. C. Cakmak and N. Agarwal, "High-speed transcript collection on multimedia platforms," in *Proceedings of the 8th IEEE Workshop on Parallel and Distributed Processing for Computational Social Systems (ParSocial 2024)*, California, USA: IEEE, May 2024. DOI: 10.1109/IPDPSW62448.2024.00123.

[25] E. Tokuda, C. Comin, and L. Costa, "Revisiting agglomerative clustering," *Physica A: Statistical Mechanics and its Applications*, vol. 585, p. 126433, 2022. DOI: 10.1016/j.physa.2021.126433.

[26] L. Floridi and M. Chiriatti, "Gpt-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, pp. 681–694, 2020. DOI: 10.1007/s11023-020-09548-1.