Insights into Climate Change Narratives: Emotional Alignment and Engagement Analysis on TikTok

Ge Gao, Zhengyang Shan, James Crissman, Ekaterina Novozhilova, YuCheng Huang, Arti Ramanathan, Margrit Betke, Derry Wijaya
Boston University
{ggao02, shanzy, jcriss, ekaterin, ychuang2, artiram, betke, wijaya}@bu.edu

Abstract

TikTok has emerged as a key platform for discussing polarizing topics, including climate change. Despite its growing influence, there is limited research exploring how content features shape emotional alignment between video creators and audience comments, as well as their impact on user engagement. Using a combination of pretrained and fine-tuned textual and visual models, we analyzed 7,110 TikTok videos related to climate change, focusing on content features such as semantic clustering of video transcriptions, visual elements, tonal shifts, and detected emotions. (1) Our findings reveal that positive emotions and videos featuring factual content or vivid environmental visuals exhibit stronger emotional alignment. Furthermore, emotional intensity and tonal coherence in video speech are significant predictors of higher engagement levels, offering new insights into the dynamics of climate change communication on social media. (2) Our preference learning analysis reveals that comment emotions play a dominant role in predicting video shareability, with both positive and negative emotional responses acting as key drivers of content diffusion. We conclude that user engagement—particularly emotional discourse in comments—significantly shapes climate change content shareability.

1 Introduction

Over the years, social media platforms have become crucial spaces for discussing pressing issues—including climate change and sustainability—and fostering social activism, particularly among younger audiences (Hautea et al., 2021; Zulli and Zulli, 2020). TikTok has become one of the leading platforms for information gathering, with more than 120 million active users in 2024, (Statista, 2025) and one in four of them being under the age of 24 (Topics, 2024). Climate change activism on TikTok has been documented in successful movements like Just Stop Oil (2025) and Extinction Rebellion (2024), showcasing the platform's power to amplify offline activism. Given TikTok's rapid rise in popularity and crucial role in information dissemination, further research is needed to understand how audiences engage with climate change narratives and what features result in the video being more widely disseminated.

1.1 Content Features

There is limited research exploring the key content features that affect emotions and the potential for a video to be shared with others. Previous studies focused on features such as communication styles, visual imagery, and digital affordances unique to TikTok, such as stitches and other video editing techniques (Guo et al., 2024; Zulli and Zulli, 2020). However, the role of specific visual subjects of videos (e.g., environmental imagery, protests, or people) in impacting emotions or circulation is unexplored. Additionally, thematic framing has been shown to shape audience engagement and influence public discourse and action on climate change (Nabi et al., 2018). Our study fills this gap by examining how multimodal content features on TikTok influence (1) emotional alignment between videos and comments and (2) shareability.

1.2 The Role of Emotions on TikTok

The current literature presents mixed findings on how emotional valence influences engagement behavior. For instance, Ling et al. (2021) found no effect of emotional valence on TikTok virality, while Chen et al. (2021) reported that TikToks with positive titles received higher shares. Conversely, da Silva Fonseca (2023) observed that fear is not an effective emotion for driving engagement, whereas other studies found that negatively-valenced TikToks elicited more comments and stimulated discussion (Li et al., 2021; Meng et al., 2018; Cheng and Li, 2023). Studies such as Hautea et al. (2021) highlight how emotions facilitate the formation of

Proceedings of the Fourth Workshop on NLP for Positive Impact (NLP4PI), pages 128–143
July 31, 2025

"affective publics" that are more likely to participate in action both online and offline. For instance, Liu and Kuang (2024) found that fear and anxiety motivate participation in online climate actions, while Nabi et al. (2018) showed that hope positively influences offline action.

Moreover, studies on emotional contagion have shown that individuals tend to mimic the emotions they encounter (Hatfield et al., 1993; Andersson and Karlsson, 2024), suggesting that videos that resonate with viewers are more likely to elicit emotionally aligned comments. This led us to investigate what content features affect such alignment.

1.3 Engagement and Shareability

Previous research has examined social media engagement using metrics such as views, likes, comments, and shares (Perreault and Mosconi, 2018; Trunfio and Rossi, 2021; Aldous et al., 2019; Tenenboim, 2022). Since shares reflect a call to action and measure shareability, we aimed to investigate how shareability is influenced independently, without the confounding effects of other engagement metrics such as likes, views, and comments.

Prior studies examined how content features drive social media sharing (Nowak-Teter and Łódzki, 2023; Hu and Noor, 2024), but overlooked how these effects vary across different engagement levels—a crucial consideration for content creators. To bridge this gap, we classify engagement into Low, Medium, and High tiers, enabling a more nuanced understanding of how content influences sharing at different stages of audience interaction.

2 Method

2.1 Dataset Collection and Cleaning

Using TikTok's Research API, we scraped 23,878 videos tagged with "#climatechange" in the United States, along with their associated comments, posted between 1 January 2024 and 1 November 2024. For this study, we focused on videos with a narrative speech, identified based on whether the Whisper-large -v2 model detected spoken content in the audio, to analyze how spoken content influences audience (comment) emotions and engagement. Speech transcriptions were generated using OpenAI's Whisper-large-v2 model.

We observed that a significant portion of the videos tagged with "#climatechange" were unrelated to the topic and were likely included due to misuse of the hashtag. To address this, we performed a binary relevance query on each transcription using GPT-4o (the snapshot version of gpt-4o-2024-08-06)(OpenAI, 2024). Details of the query prompt and the quality of the resulting data are provided in Appendix A.1.

After applying relevance filtering, deduplication, and excluding videos without comments or those in non-English languages, we curated a dataset of 7,110 videos and 116,256 corresponding comments. We refer to this dataset as *ClimateDisc*. We compute feature vectors for 7,505,104 video pairs as described in Section 2.5. *ClimateDisc* is publicly available at `https://anonymous`, allowing academic and non-commercial use with attribution.

2.2 Emotion Detection

We analyzed emotions in both the speech and comments within *ClimateDisc* to explore emotional alignment. We deployed RoBERTa-large (Liu et al., 2019) models trained on the *GoEmotions* dataset (Demszky et al., 2020) which is composed of 58,000 curated Reddit comments labeled for 28 emotion categories. To simplify the analysis and determine the most effective combination of emotions for analysis, we grouped these 28 emotion categories that align with Plutchik's psychological study (2001) into three levels to reduce complexity by combing related emotions into 15, 8, and 5 categories (Appendix B). We trained three RoBERTa-large emotion classifiers, one for each grouping, using the respective collapsed datasets. Training was conducted for 4 epochs with a learning rate of 2e-5. Model performance metrics are provided in Appendix C, with the 5-emotions classifier achieving the best results with a F1-score of 0.660 compared to 0.596 and 0.645 for the 15 emotions and 8 emotions respectively. Consequently, our primary analysis focuses on the 5-category framework, which balances interpretability, computational efficiency, and performance. Finally, we applied these classifiers to the video speeches in *ClimateDisc*.

2.3 Feature Identification

We aim to understand what are the effects of key content features and the emotion alignment between video speeches and comments on the potential for dissemination of TikTok videos. We adopted two main approaches in our feature selection: (1) textual analysis, incorporating tone shift detection and centroid-based clustering of the semantic content in speech transcriptions, and (2) a prompt-based feature identification process on the visual elements in the video through the use the LLaVa-NeXT-Video model (Liu et al., 2024; Zhang et al., 2024).

2.3.1 Textual Feature Identification

In the *ClimateDisc* dataset, we observed that a significant portion of the speech content exhibits clear tonal shifts. These include transitions such as moving from a calm description of a phenomenon to an emotional outburst, or from a serious and analytical discussion to a humorous or lighthearted tone. To systematically identify and analyze these tonal shifts, we utilized GPT-4o, prompting it (Appendix A.2) to evaluate whether a noticeable change in tone occurred within the video speeches.

Additionally, we generated high dimensional word embeddings for each of the video speeches with the sentence transformer all-MiniLM-L6-v2 (Wang et al., 2020), mapping transcription text to a 384 dimensional dense vector space and reduced the dimension with principal component analysis (Wold et al., 1987). Our experiment shows that reduction to two-dimensional vectors yields the best result in terms of clustering performance. We then performed K-means clustering on the word embeddings of the speeches. As described in 3.1, we chose the number of clusters to be 3. After applying the clustering method to the dataset, the distribution of samples across the clusters was 29.2%, 35.9%, and 34.8% for clusters 0, 1, and 2, respectively.

2.3.2 Visual Feature Identification

In addition to analyzing the narrative speech in the video from a pure natural language processing standpoint, we wanted to also examine the visual elements in the videos to uncover more features and gain deeper insights into the videos' overall content. Through our qualitative analysis of *ClimateDisc*, we identified five broad categories of videos: (1) **hasFace**: videos of individuals, including social influencers, speaking directly to the camera and expressing their views on global warming, (2) **hasNews**: news media segments showcasing reporters and newsroom settings, (3) **hasEnvVisual**: videos featuring visual cues such as images or clips of natural environments, including melting glaciers and wildfires, (4) **hasExplanations**: explanatory or tutorial videos presenting scientific topics related to climate change, and (5) **hasProtests**: protest videos addressing climate policies.

To process the visual elements in the videos, we utilized the 7-billion-parameter version of the multimodal LLaVa-NeXT-Video model. This instruction-following model processes natural language instructions and generates corresponding responses. By incorporating temporal information through the analysis of multiple video frames, the model achieves a more comprehensive understanding of the visual content. We used custom instruction prompts for each of the five categories described above and convert the generated results into binary labels. The specific instructions used in our experiments are detailed in Appendix A.3.

To quantitatively measure reliability of the LLaVa generated labels, we conducted an inter-rater agreement study evaluated with Cohen's Kappa. The results show substantial agreement for **hasFace**, **hasNews**, **hasEnvVisual** while **hasExplanations** and **hasProtests** faced challenges due to intrinsic subjectivity. Full details of the agreement study can be found in Appendix D.

2.4 Emotion Alignment

We define a custom metric, called the Emotion Alignment Score (EAS), to quantify the degree of alignment between the emotions expressed in a video's speech and its corresponding comments. The method is demonstrated using the 5-emotions set (anger, fear, happiness, sadness, neutral) as an example, although the same process applies to the 8-emotions and 15-emotions sets.

For a video v, let the emotion detected in the video's speech be e_v, and let the video have n_c comments, with the detected emotions for the comments denoted as $e_1, e_2, ..., e_{n_c}$ for the video v.

As an illustrative example, let e_v = fear and $n_c = 3$ with e_1 = fear, e_2 = fear, e_3 = sadness. The comments for video v are encoded as $c_{enc} = [0, 2, 0, 1, 0]$, with the emotion order fixed as [anger:0, fear:2, happiness:0, sadness:1, neutral:0]. Similarly, the emotion of the speech e_v = fear, is one-hot encoded as $v_{enc} = [0, 1, 0, 0, 0]$.

The EAS is calculated using cosine similarity:

$$\text{EAS}(v) = \frac{v_{\text{enc}} \cdot c_{\text{enc}}}{||v_{\text{enc}}||\,||c_{\text{enc}}||} \tag{1}$$

The score reflects the degree to which speech and comments exhibit similar emotional patterns for a given video, with higher values indicating stronger alignment.

2.5 Engagement Analysis

Engagement metrics such as likes, views, shares, and comments influence a video's visibility, as social media algorithms prioritize highly engaged content (Gerlitz and Helmond, 2013). While these metrics are widely recognized as key drivers of content circulation, it remains unclear whether shareability is driven solely by engagement or if intrinsic content features play a significant role. Studies have primarily examined the relationship between engagement and shareability (Stappen et al., 2021). In contrast, our approach focuses exclusively on content features to determine how intrinsic video characteristics contribute to share propagation when engagement levels (#comment, #like, #views) are similar.

We used a binning approach to group videos by a single engagement metric while ignoring the other two. For instance, when binning by view count, videos within each bin had similar view counts, regardless of their comment and like counts. Using quantile-based binning, we divided videos into eight balanced bins. Within each bin, we iterated through all possible video pairs, assigning each pair a binary label based on share count. A label of 1 was assigned if the first video had a higher share count than the second, and 0 if the second video had a higher share count.

To ensure robust feature selection, we incorporate a diverse range of content-related variables. Features include **hasEnvVisual**, **hasProtests**, **hasNews**, **hasExplanations**, **hasFace**, tone change, cluster for speech text, and various emotions detected in transcriptions and comments. These features are selected based on their relevance to climate discourse and their potential to shape audience reactions and engagement (Basch et al., 2021; Nguyen, 2023; Bieniek-Tobasco, 2019).

This methodology is applied across all engagement metrics to create three datasets: *ClimateDisc-ViewCount*, where videos are binned by view count, comprising 2,847,392 video pairs; *ClimateDisc-CommentCount*, where videos are binned by comment count, comprising 2,035,664 video pairs; and *ClimateDisc-LikeCount*, where videos are binned by like count, comprising 2,622,048 video pairs. By structuring our analysis this way, we effectively disentangle content effects from engagement-driven amplification, allowing us to pinpoint which content characteristics enhance shareability independently of prior engagement.

2.6 Pairwise Preference Learning

To examine video shareability through content features, we trained a Siamese Network (Bromley et al., 1993) for pairwise preference learning, developing separate models: *Model-CommentCount*, *Model-LikeCount*, and *Model-ViewCount*. Each model was trained on its respective engagement-controlled dataset—*ClimateDisc-CommentCount*, *ClimateDisc-LikeCount*, and *ClimateDisc-ViewCount*—to predict which video in a pair was more likely to be shared while controlling for the selected engagement metric.

Each dataset was split into 70% training, 15% validation, and 15% test sets. We performed 5-fold cross-validation on the training set to select the best model configuration, then trained it on the full training set with early stopping on the validation set.

The Siamese Network (performance reported in Appendix H) consists of two identical branches, each processing one video's feature set. Given a pair of videos (v_1, v_2) , their corresponding feature representations x_1 and x_2 were processed through a shared neural network $f(\cdot)$, which maps them into a latent representation space: $h_1 = f(x_1), h_2 = f(x_2)$ where $f(\cdot)$ is a multi-layer fully connected network. The feature representations x_1 and x_2 represent the content characteristics of the videos, including visual elements (**hasEnvVisual, hasProtests, hasNewsBroadcast**), explanatory content (**hasExplanations**), facial presence (**hasFace**), tonal variation (**toneChanged**), clustering assignments (**cluster_speech**), and vectors of emotions from transcriptions and comments.

To compare the two videos, we computed the element-wise difference between their latent representations: $d = h_1 - h_2$. This difference vector was then passed through a fully connected layer with a sigmoid activation to produce a probability score: $\hat{y} = \sigma(w^T d + b)$ where w is a learned weight vector, b is a bias term, and $\sigma(\cdot)$ is the sigmoid activation function, ensuring the output falls within the range (0,1). The output $\hat{y}$ represents the predicted probability that video v_1 is more shareable than video v_2. If $\hat{y} \geq 0.5$, the model predicts that video v_1 is more shareable and assigns it a label of 1. Otherwise, it predicts that video v_2 is more shareable and assigns it a label of 0.

To assess model performance, we evaluated pairwise classification accuracy, which measures the proportion of correctly predicted video preferences. Specifically, accuracy is calculated as the ratio of correctly classified video pairs to the total number of pairs in the test set. By training separate models for each engagement-controlled dataset, we ensured that engagement metrics were neutralized, allowing for a focused analysis of the role of content-related factors in determining video shareability. This approach provides insight into whether content characteristics alone can predict shareability at comparable levels of engagement.

2.7 Feature Importance and Bin Analysis

To determine the most influential content-related features in video shareability, we performed a permutation-based feature importance analysis for *Model-CommentCount*, *Model-LikeCount*, and *Model-ViewCount* at both global and bin-specific levels. This approach quantifies each feature's impact by measuring the increase in model loss when its values are randomly permuted. A larger loss increase indicates higher importance in the model's predictions.

Since engagement metrics such as comments, likes, and views follow a power-law distribution, where a small fraction of videos receive disproportionately high engagement (Johnson et al., 2014), traditional quantile-based binning may not effectively capture meaningful engagement differences. To address this, we adopted a ranking-based binning strategy, defining engagement levels based on a video's relative position in the distribution rather than fixed thresholds.

Videos were ranked in ascending order based on a single engagement metric (e.g., comment count, like count, or view count). Each video was then assigned to a bin according to its percentile rank: low (0–33rd percentile), moderate (34–66th percentile), and high (67–100th percentile) (Appendix E). This method ensures balanced bin sizes while preserving the relative order of engagement levels, preventing extreme values from distorting the binning process.

Using our trained Siamese Network, we applied this binning strategy to analyze feature importance at different engagement levels. Within each bin, we compared video pairs with similar engagement profiles but differing share counts to isolate content-driven shareability factors. We first computed the baseline model performance on the original dataset. Then, for each feature, we randomly permuted its values across all video pairs, breaking its association with shareability. The model was re-evaluated, and the change in performance was used to quantify the feature's importance. A greater difference in performance indicated a stronger influence of the feature on shareability predictions.

3 Results

3.1 Transcription Clustering

We applied K-means clustering to speech transcriptions and used the elbow method to determine the optimal number of clusters. As shown in Figure 1, WCSS (Hartigan et al., 1979) decreases as cluster count increases, with a noticeable elbow at $k = 3$. This aligns with the highest Silhouette score (Shahapure and Nicholas, 2020), confirming well-defined clusters. Figure 8 in Appendix F further illustrates their distinct separation.

A quantitative analysis of the clusters reveals distinct thematic groupings: **Political Critique** (Cluster 0), **Sustainability and Local Knowledge** (Cluster 1), and **Personal Impacts and Cataclysmic Fears** (Cluster 2). The three data points closest to each centroid, included in Appendix G, exemplify these themes.

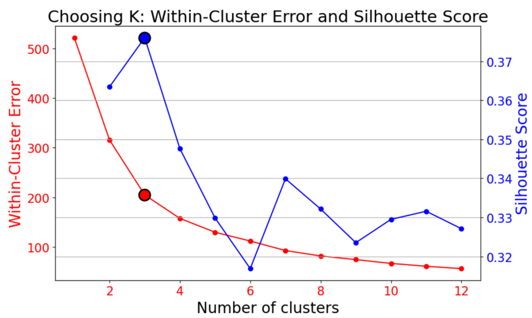

Figure 1: Within-Cluster Sum of Squares (WCSS), where lower values indicate better compactness, and Silhouette Score, where higher values reflect better-defined clusters, across varying numbers of clusters.

3.2 Emotion Alignment

3.2.1 Alignment in *ClimateDisc*

To explore the relationship between video speech and their corresponding comments, we analyzed the emotion distribution across the full *ClimateDisc* dataset. Figure 2 presents a heatmap illustrating the frequency of various emotions in the comments for each emotion detected in the video speeches. For instance, when anger is detected in a video speech, the corresponding comments exhibit 6918 instances of happiness, 2080 instances of fear, 1323 instances of sadness, and 4056 instances of anger. To assess emotion alignment, we focus on the diagonal of the heatmap, which represents instances where the emotions in the video speeches match those in the comments. A stronger intensity along the diagonal indicates greater emotional alignment

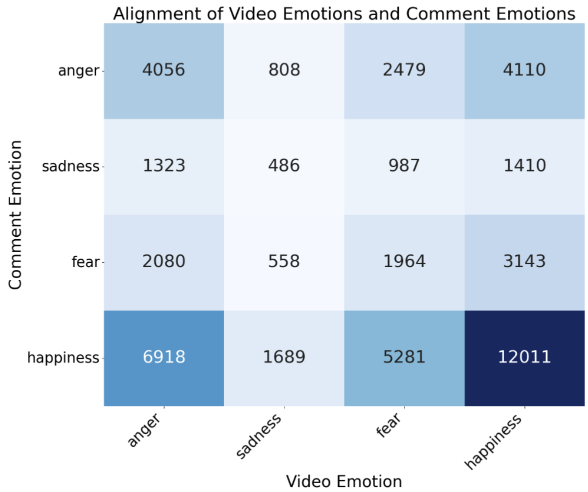

Figure 2: Frequency distribution of emotions in comments relative to the emotions expressed in videos, illustrating alignment and discrepancies across emotion categories. Neutral emotions are excluded.

between video speeches and comments within the *ClimateDisc* dataset. As shown in Figure 2, the positive emotion happiness exhibits strong alignment, with a high intensity on the diagonal. In contrast, negative emotions such as anger, sadness, and fear elicit more diverse emotional responses in the comments, with notable off-diagonal frequencies. Notably, sadness, with relatively low intensities both on and off the diagonal, appears to resonate less strongly with viewers, suggesting it may not evoke as strong or consistent reactions compared to other emotions.

The average Emotion Alignment Score (EAS) for the full *ClimateDisc* dataset is 0.532 (Eq. 1), indicating a moderate level of alignment between video speeches and their corresponding comments.

3.2.2 Alignment in Feature Groups

We grouped the videos by feature and identified the most strongly aligned emotion within each group, as presented in Table 1. Across all feature groups, happiness consistently emerged as the most strongly aligned emotion, reinforcing our earlier findings in Section 3.2.1.

Moreover, we computed each feature group's EAS, and compared it to the full *ClimateDisc* dataset's EAS. The results, summarized in Table 1, are statistically significant except for **hasProtests**. Notably, the feature group **hasEnvVisual** shows stronger alignment according to the EAS, potentially due to the emotionally compelling nature of environment images which evoke more direct and concentrated emotional responses, aligning with the video's original intent. Likewise, **hasExplanations**, which presents viewers with factual and informational content that contain less ambiguity, is successful in guiding audience emotional reactions and shows stronger alignment (EAS 0.587). In contrast, **hasFace** and **hasNews**, which exhibit lower alignment, may reflect the neutral tone of their content, leading to greater emotional discrepancies in viewer comments.

Another finding is the significantly lower alignment (0.472) when there is a **change in tone** (`toneChanged`) within the video. This tonal shift potentially confuses or alienate viewers, prompting emotional responses in the comments that deviate from the video's overarching emotional narrative.

Looking at the thematic clusters, **Cluster 1** (Political Critique) and **Cluster 2** (Sustainability and Local Knowledge) both show significantly higher alignment. As these topics typically center on factual content or reasoned argumentation (Appendix G), the emotional responses tend to remain focused on the issues at hand; thereby aligning with the video's tone. Conversely, **Cluster 3** (Personal Impacts and Cataclysmic Fears) has significantly lower alignment (0.491). The more fear-driven and subjective nature of catastrophic themes may lead viewers to respond with emotions divergent from those intended or expressed in the video.

Feature	**EAS**	**P-Val**	**Top Aligned**
`hasFace`	0.523 ↓	$< .001$	happiness
`hasNews`	0.518 ↓	$< .05$	happiness
`hasEnvVisual`	0.550 ↑	$< .05$	happiness
`hasExplanations`	0.587 ↑	$< .001$	happiness
`hasProtests`	0.553 ↑	0.480	happiness
`toneChanged`	0.472 ↓	$< .001$	happiness
`Cluster 1`	0.536 ↑	$< .001$	happiness
`Cluster 2`	0.563 ↑	$< .001$	happiness
`Cluster 3`	0.491 ↓	$< .001$	happiness

Table 1: Emotion Alignment Scores (EAS) across feature groups. Features marked with ↑ indicate greater video-comment alignment compared to the overall EAS of the full *ClimateDisc* dataset (0.532), while ↓ denotes lower alignment. P-values represent the statistical significance of the difference in EAS between each feature group and the full dataset. **Top Aligned** represents the strongest aligned emotion for each feature group.

3.3 Feature Importance in Preference Learning

Our global feature importance analysis highlights key content-related factors influencing video shareability across different engagement metrics. The global feature importance scores for *Model-*

CommentCount, *Model-LikeCount*, and *Model-ViewCount*, revealing the most significant predictors of video shareability, are presented in Fig. 3.

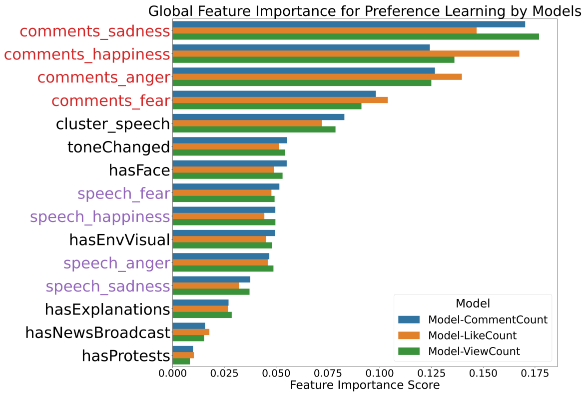

Figure 3: Global feature importance scores for *Model-CommentCount*, *Model-LikeCount*, and *Model-ViewCount*. Each feature is represented along the y-axis, with corresponding importance scores for each model.

User-generated emotional expressions in comments emerge as the strongest determinants of shareability across all models. Specifically, sadness consistently exhibits the highest feature importance, followed by happiness and anger. This trend suggests that emotionally charged discussions—whether driven by negative (sadness, anger) or positive (happiness) sentiments—play a crucial role in video dissemination. The dominance of comment-based emotions aligns with existing research on emotional contagion in social media, where emotionally engaging content is more likely to be shared and circulated (Dobele et al., 2007). Moreover, our result aligns with the finding in the study by Ziyada and Shamoi (2024) that viewer comments and reactions have a bigger impact on video popularity than raw video features.

Beyond audience sentiment, linguistic and visual attributes also contribute to shareability. The clustering of speech text (**cluster_speech**) indicates that thematic coherence enhances video distribution. Similarly, tonal variation (**toneChanged**) and facial presence (**hasFace**) play a role, suggesting that expressive speech and human faces improve engagement. These findings are consistent with research showing that facial presence enhances user engagement and perceived authenticity (Bakhshi et al., 2014).

Speech-derived emotions, including fear, happiness, and anger, rank moderately, reinforcing that emotionally expressive speech influences audience engagement. Environmental visuals (**hasEnvVisual**) and explanatory elements (**hasExplanations**) also contribute, though to a lesser extent than direct emotional expression. Notably, news and protest-related content (**hasNews**, **hasProtests**) rank among the least influential factors, suggesting that while these topics may spark discussion, they do not necessarily drive sharing behavior.

Overall, content-related features beyond engagement metrics has a significant impact on shareability. The strong influence of comment emotions highlights that audience reactions, rather than intrinsic video properties, are key to predicting virality. Additionally, the importance of thematic coherence and tonal variation underscores the role of narrative and audiovisual presentation in content dissemination.

3.4 Bin-Specific Feature Importance

To further investigate content-driven shareability at different engagement levels, we conducted a bin-specific feature importance analysis, categorizing videos into Low, Moderate, and High engagement bins. The feature importance scores across engagement levels for each model are shown in Fig. 4.

For *Model-CommentCount*, the results show a shift from content-driven to audience-driven factors. In the Low bin, speech text clusters, environmental visuals, and facial presence are the strongest predictors, suggesting that content features drive early engagement. As engagement rises to the Moderate bin, comment emotions—sadness (.10), happiness (.10), and anger (.09)—gain importance, indicating that audience responses increasingly shape shareability. In the High bin, comment sentiment dominates, with sadness (.31), anger (.17), and happiness (.13) as the top predictors, while content-based attributes lose influence. This suggests that while content features attract initial engagement, sustained virality is largely driven by audience interactions.

For *Model-LikeCount*, comment-based emotions consistently influence shareability, though their impact varies across bins. In the Low bin, happiness (.11), anger (.10), and sadness (.08) are key predictors, highlighting the early role of audience sentiment. As engagement increases, comment emotions intensify, with happiness (.14), anger (.12), and sadness (.11) dominating in the Moderate bin. In the High bin, these factors become even more pronounced, with happiness (.26), sadness (.22), and anger (.17) as the strongest predictors.

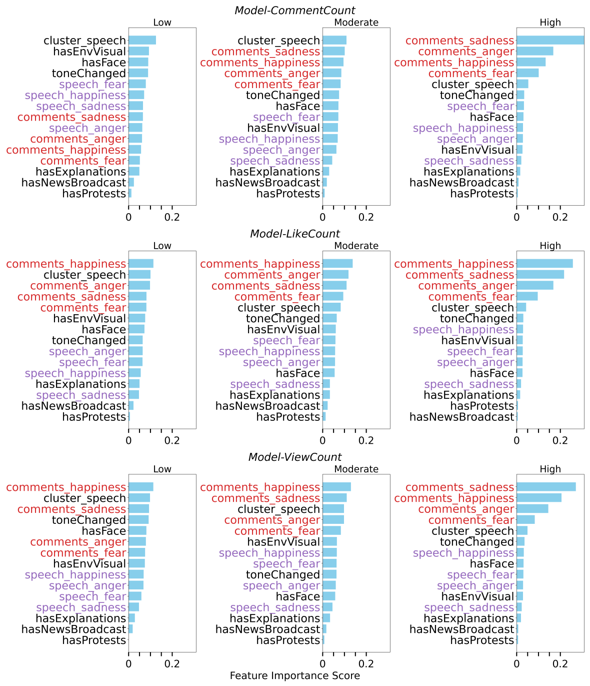

Figure 4: Bin-specific feature importance scores across Low, Moderate, and High engagement levels for *Model-CommentCount*, *Model-LikeCount*, and *Model-ViewCount*, respectively. Each feature is represented along the y-axis, with corresponding importance scores for each sub-bin.

Content-related features like speech text cluster and tonal changes remain relevant but decline in importance, reinforcing that audience emotional engagement, especially in comments, becomes the primary driver of shareability at higher engagement levels.

For *Model-ViewCount*, comment emotions consistently drive shareability, though their influence evolves with engagement. In the Low bin, happiness (.11), sadness (.09), and anger (.08) are key predictors, alongside tone changes (.09) and facial presence (.08), suggesting that both content and audience engagement contribute to early-stage shareability. As engagement grows, comment-based emotions strengthen, with happiness (.13), sadness (.11), and anger (.10) dominating in the Moderate bin. The role of speech text cluster (.10) and visual elements (.07) remains, though slightly diminished. In the High bin, comment sentiment becomes the primary predictor, with sadness (.27), happiness (.21), and anger (.15) ranking highest, while content-based features like tone changes (.04) and speech text cluster (.05) decline. As engagement increases, audience emotional responses—expressed through comments—play a larger role in shareability, while the influence of intrinsic content attributes diminishes.

4 Conclusion

This study investigates how content features influence audience responses in climate change discussions by analyzing emotional alignment and call-to-action engagement, with shareability serving as the primary measure of impact across both textual and visual feature groups.

Our analysis of emotional alignment between video speech and comments reveals that positive emotions elicit the strongest alignment in audience reactions. Furthermore, insights from visual features and semantic clustering indicate that factual and informational content, as well as visually appealing environmental elements, resonate more strongly with viewers. These findings underscore the critical role of content features and thematic focus in shaping emotional engagement. Specifically, content that reduces ambiguity fosters closer alignment between the tone of the video and audience reactions, while neutral or inconsistent content tends to invite broader emotional interpretations. Ultimately, the nature of the content—whether visually evocative, fact-based, or emotionally positive or negative, profoundly affects how effectively the intended emotional tone resonates with the audience.

From a shareability perspective, our findings reveal that comment emotions, whether positive or negative, are the most influential factors in driving call-to-action shareability. Moreover, while content-related features primarily influence initial shareability, audience emotional responses, as expressed through comments, become increasingly pivotal in shaping engagement as it grows. Unlike video content, comments are dynamic discussions where emotions can influence subsequent comments and elicit further emotions. This highlights TikTok's role as a social platform where engagement is driven by user interactions than by the informational content of the videos, despite its growing popularity as a place to seek information.

In conclusion, our study demonstrates that the nature of climate change content on TikTok—whether emotionally dynamic, visually evocative, or fact-based—significantly impacts audience engagement and action. These insights can inform content design and offer a framework for optimizing climate change communication to effectively engage and mobilize users for action.

5 Limitations

The emotion detection used in this study presents several challenges. A small portion of the video speech transcriptions exhibit shifts in emotion, such as transitioning from a serious discussion on climate change to humor. As a result, assigning a single emotion to the entire transcription may oversimplify or obscure such variations. To address this, our content feature on tone shift detection helps capture these nuances, making it a valuable addition to our analysis. Future studies could further refine this approach by segmenting videos into multiple sections to track emotional changes over time.

Additionally, assessing the quality of our visual features detected with LLaVa remains difficult. While we made efforts to evaluate inter-rater agreement between human labelers and the model-generated labels, noticeable discrepancies persisted. Moreover, even among human labelers, agreement was inconsistent, likely due to the inherent subjectivity of certain features.

Despite the effectiveness of our preference learning framework in isolating content-driven shareability factors, several limitations should be acknowledged. First, while the Siamese Network successfully models relative shareability between video pairs, it does not estimate an absolute shareability score for a given video. The pairwise classification approach captures comparative preference signals but does not provide insights into how much more shareable one video is relative to another. Future work could explore ranking-based models or regression-based approaches to quantify shareability in a more continuous manner.

Second, although the engagement binning strategy effectively controls for the confounding effects of likes, comments, and views, it inherently reduces the available dataset for each individual model. The strict binning criteria limits the number of comparable video pairs, particularly in lower or higher engagement bins, which may introduce sampling biases. Additionally, the assumption that engagement effects are sufficiently neutralized within each bin relies on the completeness of the binning process, which may not fully account for nonlinear interactions between engagement metrics.

Another limitation stems from the feature selection process. While our study incorporates a diverse set of content-related features—including visual, textual, and emotional attributes—the feature set is still constrained by observable and extracted metadata. Factors such as background music, video editing style, and implicit creator-audience relationships are not captured in the current framework. The reliance on automated emotion classifiers and multimodal embeddings, while effective, introduces potential biases due to model-specific limitations in detecting nuanced semantic and affective signals.

Furthermore, feature importance scores, whether global or bin-specific, provide an aggregate view of influence across the dataset but do not necessarily imply causal relationships. The permutation-based feature importance method captures correlations between features and model decisions but does not disentangle direct causal effects from spurious associations. Future research could integrate causal inference techniques or counterfactual analysis to validate the direct impact of content attributes on shareability.

The generalizability of our findings is constrained by the platform-specific nature of TikTok's recommendation system. The study does not account for algorithmic amplification, which dynamically adjusts content visibility based on real-time user interactions. As a result, some content features identified as significant in our models may be amplified due to algorithmic preference rather than intrinsic user interest. Extending this analysis to other social media platforms with differing recommendation dynamics would provide broader insights into content shareability mechanisms.

Lastly, although we have documented the version and snapshot of GPT-4o used, it remains a proprietary system, and we do not have access to its model weights, posing challenges to reproducibility.

6 Ethical Considerations

An important finding of our study is that emotionally charged content significantly influences shareability. However, this also raises ethical concerns regarding potential misuse. Content creators may deliberately provoke emotionally polarized discussions to artificially boost engagement, a tactic that can be further amplified through the use of generative AI. The increasing ease of generating synthetic comments or manipulating emotional discourse could distort organic user interactions, influencing public perception and the spread of climate change narratives in unintended ways.

Additionally, we take ethical considerations into account in our data collection and processing to ensure user privacy and responsible research practices. The dataset used in this study consists of publicly available TikTok videos and comments adhering to data access policies and ethical guidelines. To prevent the identification of individual users, we do not collect, store, or analyze personally identifiable information (PII). All usernames and direct user identifiers were excluded, and our analysis focuses solely on content features such as transcriptions, engagement metrics, and emotional signals. Furthermore, our codebase and data processing pipeline are designed with anonymization measures, ensuring that any shared resources do not compromise user privacy. By prioritizing ethical data handling, we mitigate risks related to content manipulation and ensure that our findings contribute to a responsible and transparent understanding of content shareability.

References

Kholoud Khalil Aldous, Jisun An, and Bernard Jim Jansen. 2019. View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In *International Conference on Web and Social Media*.

Elin Andersson and Elinor Karlsson. 2024. The role of emotional content on consumer engagement. Master's thesis, Jönköping University.

Saeideh Bakhshi, David Shamma, and Eric Gilbert. 2014. Faces engage us: photos with faces attract more likes and comments on Instagram. *Conference on Human Factors in Computing Systems - Proceedings*.

Corey Hannah Basch, Bhavya Yalamanchili, and Joseph Fera. 2021. #climate change on tiktok: A content analysis of videos. *Journal of Community Health*, 47:163 – 167.

Ashley Bieniek-Tobasco. 2019. Changing the narrative around a changing climate. *The Pursuit*.

Jane Bromley, James W. Bentz, Léon Bottou, Isabelle M Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a "siamese" time delay neural network. *Int. J. Pattern Recognit. Artif. Intell.*, 7:669–688.

Qiang Chen, Chen Min, Wei Zhang, Xiaoyue Ma, and Richard David Evans. 2021. Factors driving citizen engagement with government tiktok accounts during the covid-19 pandemic: Model development and analysis. *Journal of Medical Internet Research*, 23.

Zicheng Cheng and Yanlin Li. 2023. Like, comment, and share on tiktok: Exploring the effect of sentiment and second-person view on the user engagement with tiktok news videos. *Social Science Computer Review*, 42:201 – 223.

Davide Chicco and Giuseppe Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37 – 46.

Bruna Maria da Silva Fonseca. 2023. Do we 'like' or 'unfollow'? a social media-based analysis of public perceptions on global change. Master's thesis, Universidade do Porto, Portugal.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *Preprint*, arXiv:2005.00547.

Angela R. Dobele, Adam Lindgreen, Michael B. Beverland, Joëlle Vanhamme, and R. G. van Wijk. 2007. Why pass on viral messages? because they connect emotionally.

Extinction Rebellion. 2024. Extinction rebellion official website.

Carolin Gerlitz and Anne Helmond. 2013. The like economy: Social buttons and the data-intensive web. *New Media & Society*, 15:1348 – 1365.

Zhenhua Guo, Qi Jia, Baoyu Fan, Di Wang, Cong Xu, Yanwei Wang, Yaqian Zhao, and Rengang Li. 2024. Mvindemo: a dataset for micro video public-induced emotion prediction on social media. *Multim. Syst.*, 30:58.

John A Hartigan, Manchek A Wong, et al. 1979. A k-means clustering algorithm. *Applied statistics*, 28(1):100–108.

Elaine Hatfield, John T. Cacioppo, and Richard L. Rapson. 1993. Emotional contagion: Mechanisms of emotional contagion: I. emotional mimicry/synchrony.

Samantha Hautea, Perry Parks, Bruno Takahashi, and Jing Zeng. 2021. Showing they care (or don't): Affective publics and ambivalent climate activism on TikTok. *Social Media + Society*, 7(2):20563051211012344.

Jia Hu and Shuhaida Md Noor. 2024. Why we share: A systematic review of knowledge-sharing intentions on social media. *Behavioral Sciences*, 14.

Steven L. Johnson, Samer Faraj, and Srinivas Kudaravalli. 2014. Emergence of power laws in online communities: The role of social mechanisms and preferential attachment. *MIS Q.*, 38:795–808.

Just Stop Oil. 2025. Just stop oil official website.

J Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33 1:159–74.

Yachao Li, Mengfei Guan, Paige Hammond, and Lane E Berrey. 2021. Communicating covid-19 information on tiktok: a content analysis of tiktok videos from official accounts featured in the covid-19 information hub. *Health Education Research*.

Chen Ling, Jeremy Blackburn, Emiliano De Cristofaro, and Gianluca Stringhini. 2021. Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos. *Proceedings of the 14th ACM Web Science Conference 2022*.

Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. 2024. LLaVA-NeXT: Improved reasoning, OCR, and world knowledge.

Shujun Liu and Kai Kuang. 2024. The role of emotions in climate change communication: Examining the effects of strategy and issue framing on emotional responses and online climate action intentions. *Current Psychology*, 43:27070–27083.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *Preprint*, arXiv:1907.11692.

Jingbo Meng, Wei Peng, Pang-Ning Tan, Wuyu Liu, Ying Cheng, and Arram Bae. 2018. Diffusion size and structural virality: The effects of message and network features on spreading health information on twitter. *Computers in Human Behavior*, 89:111 – 120.

Robin Nabi, Abel Gustafson, and Risa Jensen. 2018. Framing climate change: Exploring the role of emotion in generating advocacy behavior. *Science Communication*, 40:107554701877601.

Ha Nguyen. 2023. Tiktok as learning analytics data: Framing climate change and data practices. *LAK23: 13th International Learning Analytics and Knowledge Conference*.

Ewa Nowak-Teter and Bartłomiej Łódzki. 2023. What makes news shared on facebook? social media logic and content-related factors of shareability. *Digital Journalism*, 12:451 – 475.

OpenAI. 2024. GPT-4o system card. *Preprint*, arXiv:2410.21276.

Marie-Catherine Perreault and Elaine Mosconi. 2018. Social media engagement: Content strategy and metrics research opportunities.

Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344–350. Accessed 31 Jan. 2025.

Ketan Rajshekhar Shahapure and Charles Nicholas. 2020. Cluster quality analysis using silhouette score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748.

Lukas Stappen, Alice Baird, Michelle Lienhart, Annalena Batz, and Björn Schuller. 2021. An estimation of online video user engagement from features of continuous emotions. *ArXiv*, abs/2105.01633.

Statista. 2025. Number of monthly unique TikTok users worldwide. Accessed: 2025-01-27.

Ori Tenenboim. 2022. Comments, shares, or likes: What makes news posts engaging in different ways. *Social Media + Society*, 8(4):20563051221130282.

Exploding Topics. 2024. TikTok user age, gender, demographics. Accessed: 2025-01-27.

Mariapina Trunfio and Simona Rossi. 2021. Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing*, 2021:267 – 292.

Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Preprint*, arXiv:2002.10957.

Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

Yuanhan Zhang, Bo Li, haotian Liu, Yong jae Lee, Liangke Gui, Di Fu, Jiashi Feng, Ziwei Liu, and Chunyuan Li. 2024. LLaVA-NeXT: A strong zero-shot video understanding model.

Malika Ziyada and Pakizar Shamoi. 2024. Video popularity in social media: Impact of emotions, raw features and viewer comments. *2024 Joint 13th International Conference on Soft Computing and Intelligent Systems and 25th International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, pages 1–7.

Diana Zulli and David James Zulli. 2020. Extending the internet meme: Conceptualizing technological mimesis and imitation publics on the tiktok platform. *New Media & Society*, 24:1872 – 1890.

A LLM Prompts

Here are prompts we feed to the LLMs.

A.1 Climate Change Relevance Query

After transcribing the narrative in the videos, we found that there exists a considerable amount of videos that contain the "#climatechange" tag but are not relevant to the topic of climate change, we process the transcription with the following prompt to GPT-4o and instructed the model to generate a binary response:

```
f“Content: {transcription}\n”
“Is the content related to
climate change:\n”
“1. Yes, 2. No:\n”
“Answer: ”
```

We then use the following regular expression to convert the response into binary format:

```
r“\s*(?:\d\.?\s*)?(Yes|No|1|2)”
```

To validate the quality of GPT-4o's relevance detection, we randomly sampled 100 transcriptions that the model labeled as related to climate change. Upon manual annotation, we found a 98% agreement with the model's judgments, suggesting high precision in identifying climate-relevant content.

A.2 Tone Shift Detection

The text prompt we feeded to GPT-4o to detect whether there exists a tone shift in the speech of the video:

```
f"Content: {transcription}\n"
"Does the given text exhibit a major
and sudden tone change such as
transitioning from a calm description
to an outburst, or from a serious
discussion to humor:\n"
"1. Yes, 2. No:\n"
"Answer: "
```

And then we processed it into binary format into the regular expression specified in A.1.

A.3 LLaVA instructions

The instructions we developed for extracting features by processing the visual elements were:

1. **hasProtests**:

   Yes or No: Does this video show a scene of public protest, including elements like crowds of people chanting or holding signs, or any symbolic actions (e.g., raising fists, sitting in, or blocking roads) commonly associated with demonstrations?

2. **hasEnvVisual**:

   Yes or No: Does this video show any natural elements like deserts, glaciers, forests, or oceans that are associated with the environment or climate change?

3. **hasExplanations**:

   Yes or No: Does this video feature someone explicitly explaining or demonstrating a topic, such as a scientific concept, a step-by-step tutorial, or a 'how-to' guide, with clear verbal instructions or on-screen text guiding the audience?

4. **hasNews**:

   Yes or No: Is this video a segment from a news program, containing elements like a news anchor speaking in a studio, a reporter covering an event live on-site, or official news graphics (e.g., network logos, lower-thirds, or headlines) that indicate it is part of a news broadcast?

5. **hasFace**:

   Yes or No: Does this video primarily feature a human face talking directly to the camera for most of its duration?

We then extracted the response with regular expressions and converted them into binary format:

```
r“^yes(?!,\s*no)”
```

B Emotion Collapsing

To identify and summarize relevant emotions, we utilize Plutchik's wheel of emotions that categorizes emotions by their class and intensity level. Table 2 specifies how we perform emotion collapsing. Note that "realization", "concern", "powerlessness", and "indifference" stem from Plutchik's wheel and are not included in the GoEmotions dataset.

C Emotion Classification Performances

The classification performances including accuracy, precision, recall, f1 score, and Matthews Correlation Coefficient (Chicco and Jurman, 2020) are reported in Figures 5, 6, 7. The average F1-scores for the 15 emotions, 8 emotions, and 5 emotions are 0.596, 0.645, and 0.660 respectively, displaying improved classification performances as the number of emotions decreases. The number of parameters for the 15-, 8-, and 5-emotion models are 355,375,119, 355,367,944, and 355,364,869, respectively.

D Inter-rater Agreement Interpretations

To quantitatively measure reliability of the LLaVa generated labels, we conducted an experiment using a random sample of 100 videos from *ClimateDisc* for each of the five feature groups. To maintain balance, each group comprised 50 positively

<table>
<tr><td>anger</td><td>anger</td><td>anger</td><td rowspan="5">anger</td></tr>
<tr><td>annoyance</td><td rowspan="2">irritation</td><td rowspan="4">disgust</td></tr>
<tr><td>disapproval</td></tr>
<tr><td>disgust</td><td rowspan="2">discontent</td></tr>
<tr><td>disappointment</td></tr>
<tr><td>sadness</td><td rowspan="2">sorrow</td><td rowspan="5">sadness</td><td rowspan="5">sadness</td></tr>
<tr><td>grief</td></tr>
<tr><td>remorse</td><td rowspan="3">guilt</td></tr>
<tr><td>embarrassment</td></tr>
<tr><td>realization</td></tr>
<tr><td>confusion</td><td>confusion</td><td rowspan="3">surprise</td><td rowspan="7">fear</td></tr>
<tr><td>curiosity</td><td rowspan="2">surprise</td></tr>
<tr><td>surprise</td></tr>
<tr><td>nervousness</td><td>apprehension</td><td rowspan="4">anxiety</td></tr>
<tr><td>fear</td><td>anxiety</td></tr>
<tr><td>concern</td><td rowspan="2">hopelessness</td></tr>
<tr><td>powerlessness</td></tr>
<tr><td>approval</td><td rowspan="2">empowerment</td><td rowspan="6">happiness</td><td rowspan="12">happiness</td></tr>
<tr><td>admiration</td></tr>
<tr><td>gratitude</td><td rowspan="3">happiness</td></tr>
<tr><td>pride</td></tr>
<tr><td>joy</td></tr>
<tr><td>love</td><td>calm</td></tr>
<tr><td>excitement</td><td rowspan="4">enthusiasm</td><td rowspan="6">hopeful</td></tr>
<tr><td>amusement</td></tr>
<tr><td>caring</td></tr>
<tr><td>desire</td></tr>
<tr><td>relief</td><td rowspan="2">hopefulness</td></tr>
<tr><td>optimism</td></tr>
<tr><td>indifference</td><td rowspan="2">neutral</td><td rowspan="2">neutral</td><td rowspan="2">neutral</td></tr>
<tr><td>neutral</td></tr>
</table>

Table 2: Collapsing rules defined for three collapsing mechanism: 15 emotions, 8 emotions, 5 emotions

	accuracy	precision	recall	f1	mcc	support
anger	0.963	0.498	0.556	0.525	0.507	198
anxiety	0.992	0.810	0.603	0.691	0.695	78
apprehension	0.996	0.533	0.348	0.421	0.429	23
calm	0.984	0.791	0.874	0.830	0.823	238
confusion	0.965	0.410	0.523	0.460	0.446	153
discontent	0.936	0.393	0.507	0.443	0.413	270
empowerment	0.890	0.640	0.651	0.645	0.580	834
enthusiasm	0.923	0.629	0.661	0.645	0.602	576
guilt	0.960	0.552	0.403	0.466	0.451	236
happiness	0.972	0.885	0.812	0.847	0.832	520
hopefulness	0.969	0.577	0.574	0.575	0.559	197
irritation	0.876	0.429	0.572	0.490	0.427	565
neutral	0.745	0.577	0.837	0.683	0.505	1787
sorrow	0.980	0.695	0.563	0.622	0.616	158
surprise	0.918	0.476	0.781	0.592	0.570	415

Figure 5: Classification Performances for model trained on the 15 emotions.

	accuracy	precision	recall	f1	mcc	support
anger	0.968	0.575	0.465	0.514	0.501	198
anxiety	0.989	0.691	0.684	0.687	0.681	98
disgust	0.869	0.555	0.555	0.555	0.478	800
happiness	0.889	0.788	0.815	0.802	0.725	1489
hopeful	0.917	0.701	0.688	0.695	0.646	747
neutral	0.782	0.652	0.727	0.687	0.523	1787
sadness	0.947	0.674	0.496	0.571	0.551	387
surprise	0.925	0.612	0.689	0.648	0.608	546

Figure 6: Classification Performances for model trained on the 8 emotions.

	accuracy	precision	recall	f1	mcc	support
anger	0.870	0.630	0.629	0.630	0.551	952
fear	0.910	0.609	0.680	0.642	0.592	643
happiness	0.859	0.808	0.834	0.821	0.705	2104
neutral	0.766	0.633	0.691	0.661	0.484	1787
sadness	0.939	0.579	0.519	0.548	0.516	387

Figure 7: Classification Performances for model trained on the 5 emotions.

Feature	**Cohen's Kappa**	**Size**
`hasFace`	0.69	5006
`hasNews`	0.62	532
`hasEnvVisual`	0.55	1970
`hasExplanations`	0.34	758
`hasProtests`	0.30	235

Table 3: Inter-rater agreement scores (Cohen's Kappa) between human annotations and LLaVa-generated labels for each feature group, along with the number of labeled instances (Size) for each feature.

labeled samples and 50 negatively labeled samples, thus avoiding class imbalance. A team of six researchers annotated all 500 samples manually, and the annotations were subsequently compared to the labels generated by the LLaVA model. As shown in Figure 3, we used Cohen's Kappa (Cohen, 1960) to measure the general reliability and inter-rater agreement. We can refer to the guidelines published by Landis and Koch (Landis and Koch, 1977) (in Table 4 of Appendix D) for the interpretation of the Kappa values.

Features **hasFace**, **hasNews**, **hasEnvVisual** exhibit substantial agreement, suggesting that LLaVa-generated labels for these categories can be reliably utilized. It is worth highlighting that certain features, such as **hasExplanations** and **hasProtests**, presented intrinsic challenges due to the difficulty in achieving consensus even among human annotators and they can still offer insightful values in exploratory analysis.

Cohen's Kappa	Quality
>0.8	Almost Perfect Agreement
>0.6	Substantial Agreement
>0.4	Moderate Agreement
>0.2	Fair Agreement
0-0.2	Slight Agreement
<0	Almost No Agreement

Table 4: Interpretation for different ranges of the Cohen's Kappa values.

Engagement Metric	Bin	Range	No. Videos
Comment Count	Low (0-33%)	[0 - 3]	2504
	Moderate (34-66%)	(3 - 11]	2194
	High (67-100%)	(11 - 12970]	2367
Like Count	Low (0-33%)	[0 - 34]	2366
	Moderate (34-66%)	(34 - 110]	2310
	High (67-100%)	(110 - 719256]	2389
View Count	Low (0-33%)	[0 - 453]	2334
	Moderate (34-66%)	(453 - 1365]	2329
	High (67-100%)	(1365 - 5488900]	2402

Table 5: Binning Strategy for Engagement Metrics

E Binning Strategy

To systematically analyze the role of content features in video shareability while controlling for engagement levels, we employed a percentile-based binning strategy. Videos were ranked in ascending order based on a single engagement metric—comment count, like count, or view count—while removing the influence of the other two. This approach ensures that comparisons are made within comparable engagement levels, reducing biases introduced by disparities in overall popularity.

Each engagement metric was divided into three bins: Low (0-33%), Moderate (34-66%), and High (67-100%), determined by the percentile rank of each video. For example, in the view count binning, videos with up to 453 views were classified as Low, those with 454 to 1365 views were placed in the Moderate bin, and those with more than 1365 views fell into the High bin. The same methodology was applied to likes and comments, with respective threshold ranges. The bin distributions were approximately balanced, with each bin containing about one-third of the total dataset.

Notably, the numerical ranges for comment count bins appear narrower than those for likes and views. This is due to the distributional properties of engagement metrics: comments tend to have a more compressed distribution, whereas likes and views follow a more extreme power-law pattern. Although the difference between three and eleven comments may seem small, it represents a meaningful shift in user interaction relative to the dataset distribution. This distinction ensures that even within the comment-based model, we effectively capture variations in content-driven shareability.

This binning approach allows us to examine how content features contribute to shareability at different engagement levels while ensuring that comparisons are made within a relatively homogeneous subset of videos in terms of engagement. By controlling for engagement levels, we isolate the influence of content-related factors, distinguishing between the initial attractiveness of the video (content-driven) and its amplification through engagement dynamics.

F Semantic Clustering Visualization

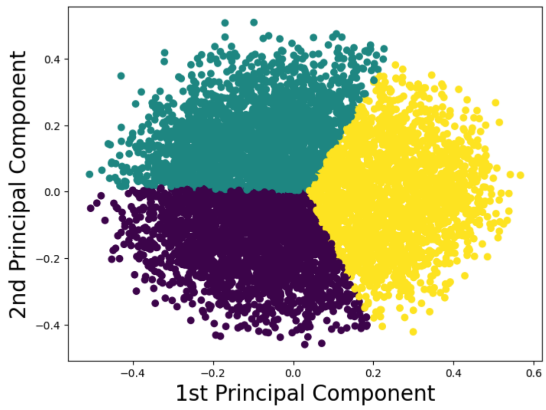

Figure 8: A two-dimensional visualization for the semantic clustering on the word embeddings of speech transcription. Each point on the scatter point corresponds to a speech transcription.

G Speech Transcriptions Closest to the Cluster Centroids

Cluster 1: (Political Critique)

1. i feel like the perfect example of trying to talk like climate change with politicians is like in game of thrones when jon snow was like frantically trying to warn everybody who were fighting wars amongst themselves that there was literally like an army of death coming for all of them and they still wanted to just bicker and fight amongst themselves and then like nobody believed him That's how scientists must feel. And then they went through all that strife to find evidence and it still didn't work. Because they just, like, Cersei just didn't care. She was like, well, we're not by the wall. They're gonna get to them first and then we can take over. Poor Jon. Poor scientists.

2. I've been trying to make this video for some time, I've been struggling how to get this idea best across. Which is this compartmentalization that I've perceived between environmental issues and climate issues on the one hand, and everything else with respect to the international legal order. Not just the situation in Gaza, of course, but multiple conflict situations all over the world. As well as the shriveling of democratic mechanisms and institutions in the US, but also in other countries. I think we have to resist compartmentalizing away issues of the environment or climate change from broader trends that we see happening in the world today. world today. From crises, from real crises that are affecting people and their rights, their human rights. These are not discrete issues. The way the world is failing to prevent genocide in Gaza will influence how the world protects or fails to protect people who are vulnerable from climate change impacts. The loss of reproductive freedom in the US and the loss of healthcare, the lack of ability to live a dignified life with a

living wage is deeply connected to who will feel the impacts most strongly from climate change. By compartmentalizing environmental issues or climate change issues away from these broader topics creates the false impression that these issues can be solved without addressing issues of power and that's just not true.

3. So from the man in high heels that brought you the don't say gay bill, it's Ron DeSantis now bringing you the don't say it's hot outside bill. As sea levels rise, Ron DeSantis signs a bill deleting climate change mentions from Florida state law. The white ranging law makes several changes to the state's energy policy, in some cases deleting entire sections of state law that talk about the importance of cutting planet warming pollution. When did pollution become political? I find it odd that there is a pro-pollution crowd. The bill would also give preferential treatment to natural gas and ban offshore wind energy even though there are no wind farms planned off Florida's coast. I have no idea why people would be against offshore wind farms, but to ban them when they are not even planned is ridiculous. He only does that because the dummies that vote for him want to see that wrote down. The bill deletes the phrase climate 8 times, often in reference to reducing the impact of global climate change through its energy policy or directing state agencies to buy climate friendly products when they are cost effective and available. Why save money? The bill also gets rid of requirements that state purchased vehicles should be fuel efficient. Again, why should the state save money? Just keep spending more taxpayer money for no d*** reason. And while standing on a milk crate shaking with anger, Ron said, Florida rejects the designs of the left to weaken our energy grid, pursue our radical climate agenda, and promote foreign adversaries. What about Wynn Farms is promoting foreign adversaries? If you're talking about the electrical grid, I assume you mean EVs. I hate to break it to everyone. I ain't rushing out getting no EV myself, I'm not a fan, but they're coming. All these big auto companies are building battery plants around this country for a reason. They're not spending tens of billions of dollars on battery plants to not build EVs.

Cluster 2 (Sustainability and Local Knowledge):

1. want to attack this commenter specifically but this is exactly my point if law if property had native species native species have a longer root systems that are able to go into the water table and absorb water which means you'd have to water them less and they'd be drought resistant like the the plants that are native to an environment are able to withstand the conditions and I think like what you're trying to say like oh it's getting hotter out yes I agree to that but it's native species longer rooted plants are able to withhold that and you could water them less often like when you're just gonna let grass die all of the biodiversity left so that's before the grass that was when the grass was green but as soon as the grass is dead it's like so now the entire system is dead and it's what so you replace your lawn like you're not just gonna water that dead dirt and it's gonna grow back like you'll have to literally replace the lawn so again it's money and when you have one you have to put fertilizer, you have to put artificial chemicals on it, you have to put weed killer, Roundup, the whole thing with Roundup. It's, of course it's all about money. Like, they don't want to have a stroke reading their, or looking at their water bill. Yeah. So it's like, they don't have the initiative, they don't have the investment to put in better systems, native plants, shaded plants, trees, like, again, uplifting the system and investing into it to see the planting trees that you'll have the shade to enjoy later. It's people want it now and people that'll have the money now. People are so fearful to do something different. I think along with that comes with money. Like I think specifically in California, they give you a tax credit to give you an incentive to rip out your lawn and like put in drought resistant plants, but again, like you have to have more money on top of that. People are already scraped thin currently. And so it's like the last thing they're gonna do is invest in their land.

2. People are going to face water shortages on the Indian subcontinent right now. You are selling out our people. You are. You're doing it. You're selling out our people. One billion people are going to have a water shortage. You're going to get a microphone if you're patient. You're going to be respectful to the people here. You're going to be respectful. You sit down and you'll be respectful to everybody else here. This is not America. You actually want to make your point, you're going to have a chance. That gentleman in the back has been waiting patiently to make a point. What's your name, sir?

3. relying on local people, relying on local knowledge. Because those people, they know a lot. They live in the forest, they live with the rivers, they live with the biodiversity, and they are there for centuries. They know a lot. So if we neglect them just because they don't speak French, they don't speak English, they don't speak those international languages, No, it's a very, very big mistake. Very big mistake.

Cluster 3 (Personal Impacts and Cataclysmic Fears):

1. Be prepared to see more and more coastal flooding just like this. Why are natural disasters popping up like crazy come 2024? Friends right now we are entering the 12,000 year cycle of cataclysms. What does that mean? Every 12,000 years we pass through a ray of cosmic ray energy. cosmic ray charges our core, charges the surrounding magma, causing this magma to rise to the surface. This leads to an extreme intensification of cataclysms and we're starting to see them just two weeks into the new year. Please research the 12,000 year cycle and please research what the Creative Society has been warning about for the last 10 years.

2. Climate change has got me all kinds of f***ed up because what do you mean I go off to check on my plants And I see there's holes in my milkweed and I count 13 caterpillar babies. We're in the second to last week of November. Why are you here? Now I buy milkweed plants in order to attract butterflies. Yes, absolutely. That is the whole point but several of my plants died several months ago and I haven't replaced them and I didn't expect to see any sort of caterpillar babies until like March or April, June at the f***ing latest, but the second to last week of November? Like, I know I'm in South Florida so we don't get freezing cold temperatures here, but you never know when a cold snap is going to happen. Like, yesterday it was 80-something degrees and today it's been in like the 70s all day. I do not know what temperature it would take to accidentally freeze these guys because I've never gotten caterpillars so late in the year before and I don't tend to bring these guys inside the house into a butterfly cage to monitor them when they're close to pupating until they are much much fatter. So now I have to worry about these caterpillars for the next two weeks as these ravenous little s***s just go about their day eating everything. I love caterpillars I I really do, but holy ****, your parents had terrible timing.

3. I honestly admit I'm scared and I'm sure other people watching this are scared of what they're seeing and why have more now disasters started occurring on earth they want to know this what is the threat and what should we be expecting next? Yeah that's a great question and to answer this question of why the number of cataclysms is increasing it's necessary to understand an important fact that all cataclysms primarily represent a release of energy in the system of our planet. Hurricanes, tornadoes, intense precipitation, all this is certainly a release of energy. Now cataclysms increasing in number and this means additional energy has appeared from somewhere in the earth system that triggers their formation. Bye!

H Preference Learning Model Performance

Model	Accuracy	Precision	Recall	F1 Score
Baselines (Accuracy Only)				
Random Guess	0.500	-	-	-
Majority (CommentCount)	0.639	-	-	-
Majority (LikeCount)	0.697	-	-	-
Majority (ViewCount)	0.678	-	-	-
Siamese Models				
Model-CommentCount	0.882	0.883	0.882	0.882
Model-LikeCount	0.825	0.824	0.826	0.825
Model-ViewCount	0.899	0.899	0.898	0.899

Table 6: Performance comparison between baseline methods and Siamese preference models. Baselines include random guessing and majority-class prediction for each engagement-based binning strategy. Siamese models are evaluated using accuracy, precision, recall, and F1 score, demonstrating significant improvements over all baselines.

The Siamese network used in our study consists of two identical branches that process paired video feature sets through a shared multi-layer feedforward architecture. The model is trained using binary cross-entropy loss and optimized with AdamW, incorporating gradient clipping and a OneCycleLR scheduler for stable and adaptive learning. Early stopping is applied based on validation loss, and the network comprises 184,769 parameters. Table 6 summarizes the performance of the three engagement-based Siamese models—*Model-CommentCount*, *Model-LikeCount*, and *Model-ViewCount*—compared to two baselines: random guessing and majority-class prediction. While baselines report accuracy only, the Siamese models are evaluated using accuracy, precision, recall, and F1 score, demonstrating substantial improvements in preference prediction.