Same Vaccine, Different Voices: A Cross-Modality Analysis of HPV Vaccine Discourse on Social Media

Mengxiao Zhu[1, 2*], **Liu He**[1*], **Han Zhao**[1], **Ruoxiao Su**[3], **Licheng Zhang**[1], **Bo Hu**[1†]

[1]University of Science and Technology of China
[2] Anhui Province Key Laboratory of Science Education and Communication
[3]University of Texas at Austin
{mxzhu, hubo}@ustc.edu.cn, {heliummn, ximang, zlczlc}@mail.ustc.edu.cn, ruoxiaosu@utexas.edu

Abstract

Despite the proven efficacy of HPV vaccines, uptake remains limited in many regions, including China. This study investigates how health beliefs and emotional responses evolve across text-, audio-, and video-based platforms by analyzing data from three representative platforms in China, including 273,357 posts from Weibo (text-based), 1,228 podcasts from Ximalaya (audio-based), and 1,225 videos from Douyin (video-based) from July 2018 to March 2023. The comparisons are conducted under four dimensions as suggested by the Health Belief Model (HBM), including susceptibility, severity, benefits, and barriers. Our findings reveal distinct modality-specific patterns. For instance, a text-based platform tends to amplify barriers and negativity, an audio-based platform enables balanced and sustained discussions, and a video-based platform highlights personal anecdotes and drives rapid sentiment shifts. By highlighting these modality-specific differences and addressing potential cross-modal incongruities at the content level, we provide actionable insights for public health communicators, policymakers, and platform designers to tailor strategies, foster informed decision-making, and ultimately enhance HPV vaccine uptake in complex social media ecosystems.

Introduction

HPV vaccination is a cornerstone of cervical cancer prevention, yet vaccination rates in many countries, including China, remain disappointingly low (Zhao et al. 2023). As governments set ambitious coverage targets, such as vaccinating 90% eligible females in China by 2025, understanding the public's health beliefs and emotional responses becomes increasingly urgent. Social media platforms now serve as primary arenas where the public encounters, debates, and disseminates health-related information (Kim 2018; Ahmed and Mao 2022). However, the landscape of social media is far from homogeneous. Text-based microblogs, audio-based podcasts, and video-focused platforms each offer distinct modes of interaction, varying narrative depth, and different degrees of emotional resonance.

Most existing research in HPV vaccine discourse on Chinese social media has centered on text-oriented services, e.g., Weibo and Zhihu, revealing problems like misinformation, cost barriers, and policy critiques (Chen et al. 2020; Li and Zheng 2020; Jiang et al. 2023). Although this body of work enhances our understanding of text-based environments, it neglects how audio and video modalities might uniquely influence user engagement, emotional reactions, and perceived risk. Indeed, Dockter et al. (Dockter et al. 2021) found that modality (e.g., videos vs. infographics) significantly impacts perceived source credibility, message effectiveness, and user engagement, especially when the topic involves varying degrees of perceived threat or risk. Such findings underscore how audio-visual formats can amplify emotional salience and heighten public perceptions of severity and susceptibility, two core dimensions of the Health Belief Model (HBM) (Bracken and Dalessandro 2017; Dockter et al. 2021). Yet comprehensive cross-modal comparisons remain scarce, leaving critical questions that require further exploration.

To fill these gaps, we draw on recent findings on conflicting affective signals conveyed by different modalities (Wang et al. 2023), and conduct a cross-modality content analysis on textual information extracted from multiple modalities. We take Weibo (text-based), Ximalaya (audio-based), and Douyin (video-based) as representative exemplars of China's broader text, audio, and video social media ecosystem, and systematically transcribe audio and video posts into text, thereby enabling a unified, content-focused comparison, rather than merely juxtaposing three specific platforms.

Furthermore, our analysis is guided by the Health Belief Model (HBM), a classic and foundational framework in health communication research (Janz and Becker 1984; Rosenstock 1974). The HBM argues that the motivations to adopt preventive health behaviors, such as screening and vaccination, are primarily due to four constructs, including susceptibility, severity, benefits, and barriers. We apply HBM to compare HPV vaccine discussions across different modalities, and investigate how the four constructs of susceptibility, severity, benefits, and barriers are framed under each modality. We further explore how emotional expressions shift in response to policy milestones, market changes, and social incidents.

To sum it up, our study aims to answer three key research questions.

•RQ1: How do text, audio, and video modalities differ in

*These authors contributed equally.
†Bo Hu is the corresponding author.

covering the susceptibility, severity, benefits, and barriers of HPV vaccination?

•RQ2: What are the representative topics in HPV vaccine-related discussions across these three modalities?

•RQ3: What emotions are expressed in HPV vaccine discourse, and how do they fluctuate under different modalities?

Related Work

Multimodality of Social Media and Its Impact on Health Communication

Social media is inherently multimodal. Users now navigate textual, auditory, and visual content across diverse platforms, including text-centric microblogs (e.g., Twitter, Weibo), audio-based platforms (e.g., Ximalaya, Spotify), and video-based networks (e.g., Douyin, TikTok, YouTube). Studies show that modality significantly influences user perception, engagement, and emotional responses to health information (Bracken and Dalessandro 2017; Dockter et al. 2021). Moreover, each modality may amplify different biases or thematic emphases, potentially affecting how users interpret and act upon vaccine-related messages.

For text-based platforms, research on HPV vaccine discourse is well-documented globally. For example, HPV vaccine-related work has been conducted on Twitter in the U.S. (Dunn et al. 2017) and Facebook in Denmark(Loft et al. 2020). In China, Weibo and Zhihu have emerged as important forums for HPV discussions, where users seeking information for HPV vaccine or related policy(Zhang, Gotsis, and Jordan-Marsh 2013; Li, Guo, and Lin 2022; Li and Zheng 2020). These studies underscore the central role of text-based platforms in shaping health knowledge and intentions (Jiang et al. 2023). Research has also documented both supportive and opposing emotional tones toward HPV vaccines of the public (Luisi 2021; Zhang et al. 2021) and identified correlations between emotional valence and network structures (Himelboim et al. 2020).

Beyond text, audio-based platforms (e.g., podcasts) are gaining prominence. Podcasts integrate social features like comments, Q&A sessions, and user-generated content, fostering an environment conducive to in-depth discussions and personal storytelling. Emerging evidence suggests podcast coverage can enhance health education, including women's health training (Dmytryshyn and Selk 2022) and HPV vaccination awareness (Dugan, McAllister, and Brennan 2022). Podcasts also convey emotional intimacy and can influence listeners' emotional orientations toward health-related subjects (Lindgren 2023; Kim, Kim, and Wang 2016). However, most podcast-related health communication research is more developed in Western contexts (Perks and Turner 2019) and systematic investigations of how podcasts in China address HPV vaccine issues remain limited, highlighting a need for further research on this auditory modality.

Video-based platforms, especially short-form video apps like Douyin (TikTok's Chinese counterpart), have also become influential arenas for health communication and shaping public perceptions and behaviors (Boatman et al. 2022; Kirkpatrick and Lawrie 2023). Studies have found that emotional tone is a critical factor for engagement, with positive emotions like hope and joy frequently emerging in vaccine-related TikTok content (Li et al. 2021). Nonetheless, misinformation and negative framing remain concerns, as seen in studies of YouTube's HPV vaccine discourse (Briones et al. 2012; Lewis and Grantham 2022), underscoring the importance of understanding how visual elements may amplify or diminish certain message facets.

In summary, text, audio, and video modalities afford distinct ways of presenting and interpreting HPV vaccine information, potentially leading to divergent emotional responses and behavioral intentions. While text-based research dominates the current literature, audio and video platforms are comparatively under-researched, as well as direct comparisons on the three modalities (Smith, Fischer, and Yongjian 2012; Thorson et al. 2013). For instance, how do the same HPV vaccine conversations about issues unfold across text-, audio-, and video-based platforms? Addressing this question through a cross-modal approach is essential for clarifying how these formats collectively influence health communication outcomes. Such insights can inform health promotion strategies, ensuring that vaccine communication leverages the affordance of each modality.

Health Belief Model as a Comparative Framework

Originally formulated in the mid-20th century, the Health Belief Model remains a widely applied theoretical framework for understanding health behaviors. The HBM theory asserts that in the health-related context, people's health behaviors are influenced by their desire to avoid diseases and their belief that a health action will prevent the disease. Specifically, the HBM consists of the following four dimensions. The first dimension is perceived susceptibility, which reflects the subjective risks of contracting a disease. The second dimension is perceived severity, which reflects the convictions concerning the seriousness of a given health problem. The third dimension is perceived benefits of health action, which reflects the beliefs of the effectiveness of the recommended health action in reducing the health threat. The fourth dimension is perceived barriers to health action, wherein people weigh the cons of the health action that it may be dangerous, expensive, inconvenient, and so on (Janz and Becker 1984; Rosenstock 1974).

Past research applying the HBM has shown that media coverage often amplifies certain dimensions over others (Quick 2010; Seiter 2021; Yang et al. 2024). Different modalities (e.g., text, audio, video) may highlight distinct aspects of HBM in health-related discourse (Briones et al. 2012; Du et al. 2020; Guidry et al. 2019; Li and Zheng 2020; Li et al. 2018), thereby shaping how users perceive risk (susceptibility, severity) or value potential gains (benefits) versus obstacles (barriers). While past research has explored these effects across modalities, the integration of HBM with multimodal and emotional analyses remains relatively underexplored, particularly in the context of HPV vaccination. By applying the HBM framework across text-, audio-, and video-based discussions of HPV vaccination, we identify how each modality's unique characteristics emphasize or downplay different HBM dimensions. Text-

Dimension	Definition	Example
Susceptibility	Assessment of the risk of getting an HPV infection.	80% of women will contract HPV in their lifetime.
Severity	Assessment of whether an HPV infection constitutes a serious health issue.	Cervical cancer ranks first in incidence and mortality among female reproductive system malignancies in China.
Benefits	Advantages of the HPV vaccine in preventing infection and related cancers.	The 9-valent HPV vaccine can prevent 92.1% of cervical cancers.
Barriers	Challenges related to HPV vaccination, such as cost or eligibility.	Anaphylaxis may occur; vaccines are becoming more expensive.

Table 1: Coding examples for each HBM dimension

based discussions often focus on providing factual information related to susceptibility and severity, while video content tends to evoke stronger emotional responses, influencing benefits and barriers. This cross-modality approach enables a deeper examination of how different communication formats might influence users' perceptions of these HBM dimensions and suggests how long-term behavioral impact could be influenced by the type of modality used. For example, video-based engagement may lead to lasting emotional shifts, while text-based engagement might foster more deliberative decision-making. By exploring the interaction between HBM constructs and emotional engagement, we aim to understand not only the immediate impact of each modality but also their potential to encourage or hinder long-term preventive behaviors.

Methodology

Dataset

To explore how text, audio, and video modalities shape discussions on HPV vaccines, we collected public data from three representative and accessible Chinese social media platforms, including Weibo, Ximalaya, and Douyin. Weibo is one of China's largest microblogging platforms, and captures text-based public opinion trends. Ximalaya, a leading platform for audio content, hosts podcasts and spoken-word discussions, emerging as the largest audio-sharing platform in China. Douyin is China's most popular short-video platform and TikTok's Chinese counterpart, which reflects video-based public engagement and storytelling.

Our dataset spans from July 2018 to March 2023, covering nearly five years of public discourse, and a detailed data collection process is described in the Appendix. The dataset comprises text, audio, and video entries, with 273,357 posts from Weibo, 1,228 audio podcasts from Ximalaya, and 1,225 videos from Douyin. It also includes related user information, enabling a unified, cross-modality exploration of HPV vaccine discussions.

Multimodal Data Preprocessing

To enable the cross-modality analysis, we adopted a text-focused approach, transcribing audio and video entry into text. This method helps address the potential incongruity that arises when different modalities carry conflicting affective or informational signals (Wang et al. 2023). The specific preprocessing steps are described as follows.

Audio data from Douyin videos were extracted using FFmpeg (Tomar 2006) and then combined with Ximalaya's raw audio dataset for voice-to-text analysis. To isolate human speech from background noise, we utilized Spleeter (Hennequin et al. 2020), a pre-trained source separation tool by Deezer Research, retaining only Mandarin Chinese content while excluding non-human speech and non-Chinese audio. The processed audio was then transcribed into text using the iFLYTEK speech transcription API[1]. To ensure the datasets focused on user-generated discourse about the HPV vaccine, additional cleaning was performed to remove irrelevant content. For the Weibo dataset, a stratified random sample of 0.5% of the whole dataset was extracted on a monthly basis to identify irrelevant entries, e.g., commercial advertisements. A dictionary of keywords and user accounts was compiled and used tofilter unrelated contents. For the Ximalaya and Douyin datasets, since they were relatively small, all contents were manually reviewed and cleaned. After preprocessing, the final dataset included 241,435 text entries from Weibo, 1,228 transcribed audio entries from Ximalaya, and 1,225 transcribed video entries from Douyin. By converting contents from different modalities into textual format, this preprocessing approach enabled cross-modal comparisons and ensured methodological consistency across modalities.

Data Coding Under Health Belief Model

We employed the HBM to systematically analyze HPV vaccine discourse by focusing on its four dimensions, including susceptibility, severity, benefits, and barriers (Janz and Becker 1984; Rosenstock 1974). Given that many entries adopted a third-person perspective and could encompass multiple dimensions, we coded each dimension separately and allowed each entry to be classified into more than one categories. To achieve this, we combined manual coding with fine-tuned deep learning models, ensuring both scalability and precision.

Development of Coding Scheme. We began by manually annotating a portion of the dataset to establish coding scheme. For Weibo, 1% of posts from each month were annotated by two independent coders, and discrepancies were resolved through discussions. Coding examples are shown in Table 1 with more details provided in Appendix A3.

Deep Learning for Weibo Classification. After manually annotating 1% of the Weibo data and building the coding scheme, we used these data to fine-tune a transformer-based model, the Chinese version of RoBERTa (Liu et al. 2019) to

[1]https://global.xfyun.cn/products/lfasr

HBM Dimension	Precision	Recall	F1-Score
Susceptibility	0.78	0.79	0.78
Severity	0.89	0.83	0.85
Benefits	0.76	0.77	0.77
Barriers	0.77	0.71	0.74

Table 2: Performance of HBM classifier on four dimensions

label the rest of Weibo data. The annotated Weibo data were split into two parts, with 80% allocated for training and the remaining 20% reserved for testing. Four binary classifiers were trained for each of the HBM dimensions with hyperparameters, such as learning rate, batch size, and dropout probability, carefully optimized to account for the unique needs in each classifier. The learning rates were tuned as follows: susceptibility (1e-5), severity (2e-5), benefits (2e-5), and barriers (1e-5), with batch sizes set at 32 and dropout rates at 0.3 for all classifiers. To address the class imbalance, Binary Focal Loss functions were adopted, with a ratio of 10:1 for susceptibility, severity, and benefits dimensions, and 1:20 for barriers dimension, to give higher weights to positive examples. The fine-tuned classifiers achieved satisfiable performance with not only relatively high precision and recall scores, but also F1 scores greater than 0.7 for all four HBM dimension, as shown in Table 2, making them suitable for large-scale classification tasks. The trained classifiers were then used to label the rest of the Weibo dataset.

Ximalaya and Douyin Classification. Manual annotation was applied to the entire of Ximalaya and Douyin datasets given their smaller samplesizes and the longer text length of individual entries. To preserve semantic integrity, each entry was coded following the finalized HBM rules established during the Weibo annotation process, with adjustments for modality-specific features. Additionally, HBM dimensions were expanded as needed to align with the unique characteristics of Ximalaya and Douyin, ensuring comprehensive and consistent coverage across all modalities.

Topic Analysis

To explore latent topics within each modality, we applied the BERTopic model (Grootendorst 2022) , which integrates BERT embeddings, dimensionality reduction, and clustering techniques to uncover hidden themes. The topic modeling was conducted on the four HBM dimensions, i.e., susceptibility, severity, benefits, and barriers, and one each of the Weibo, Ximalaya, and Douyin datasets. For topic modeling, preprocessing steps includedirrelevant elements removal (e.g., URLs, mentions, and hashtags), Chinese word segmentation using the jieba library [2], stop word removal, and conversion of all Chinese characters to simplified Chinese. To extract latent topics, short texts, as in Weibo, were processed directly, while longer transcripts, often found in Ximalaya and Douyin, were split into sentences to meet the 512-token limit of Sentence-BERT (SBERT)and preserve semantic integrity. We first used the paraphrase-multilingual-MiniLM-L12-v2 SBERT model (Reimers and Gurevych 2019) to generate sentence embeddings. This model is lightweight, supports Chinese, and is capable of efficiently handling large-scale data, such as Weibo. Next, we applied UMAP(McInnes et al. 2018) for dimension reduction, reducing the embedding space to 10 dimensions, which strikes a balance between retaining sufficient information and minimizing noise to improve efficiency. During the dimension reduction, n_neighbors was set to 15 to balance the focus on local and global structures, ensuring that the text's subtle distinctions were captured without overfitting. The reduced embeddings were then input into the HDBSCAN(McInnes, Healy, and Astels 2017) clustering algorithm, with min_cluster_size set to 200 to ensure cluster stability and effectiveness, while cluster_selection_method used the EOM (Excess of Mass) method to identify high-density regions and enhance cluster consistency. Finally, in the topic representation phase, we applied BM25 weighting and set reduce_frequent_words=true to minimize the influence of common and frequent words, ensuring that the extracted keywords were more distinctive and representative. To enhance model interpretability, we manually merged semantically similar clusters and selected the three largest and most representative clusters for each HBM dimension on each platform for further analysis. For each cluster, we also selected the representative arguments to provide insights into the thematic structure of the discussions.

[2]https://github.com/fxsjy/jieba

Emotion	Precision	Recall	F1-Score
Positive	0.90	0.93	0.90
Negative	0.90	0.87	0.90

Table 3: Performance of emotion classifier on two emotions

Emotion Analysis

To analyze emotions with high accuracy, we employed a two-step procedure for sentiment detection. Firstly, we extracted posts with neutral emotions using a six-emotion classification model trained on the SMP2020 EWECT Weibo dataset[3] (performance shown in Appendix Table A2), . Then, we classified entries with non-neutral emotions into positive or negative with the binary sentiment model trained on the Weibo_Senti_100k dataset[4](performance see Table 3). Further details on the model training are provided in the Appendix. For Ximalaya and Douyin datasets, since entries often exceeded the 512-token limit, they were split into sentences. The sentences in each entry were then classified as positive (1), negative (-1), or neutral (0), and a weighted emotion score for the full entry was generated by integrating the scores of all sentences. Entries with scores above 0 were labeled positive, below 0 as negative, and 0 as neutral. Neutral entries were further examined to identify truly neutral entries and entries with balanced emotions, i.e., positive and negative sentences canceling each other out. For Ximalaya, 38.83% entries were neutral with 0.31% balanced entries, while Douyin had 44.04% neutral, with no balanced ones. By unifying the emotional analysis pipeline across

[3]https://smp2020ewect.github.io

[4]https://dx.doi.org/10.21227/abj8-y636

short (Weibo) and long (Ximalaya, Douyin) texts, we ensured cross-modal comparability while accommodating linguistic and structural differences inherent to each modality.

Trend Analysis

To investigate how frequencies of topics in the four HBM dimensions and emotions evolve over time on each platform, we began by recording the frequency of entries in each HBM dimension across the three platforms on a monthly basis. This provided aggregatedtime-series data for further analysis. Next, we aggregated the counts of entries with each emotion type, i.e., positive, negative, and neutral, for each platform per month, creating a time-series dataset of emotional responses. These datasets allowed us to track the temporal variations in both HBM dimensions and emotional categories. Finally, we compared the time-series data for the HBM dimensions and emotions across the platforms to examine the temporal evolution of discussions. This comparison helped to reveal platform-specific trends and provided insights into how different platforms shaped the discussions of HPV vaccination over time.

Results

RQ1: How Do Text, Audio, and Video Differ in Covering the Four HBM Dimensions?

We examined how text-based (Weibo), audio-based (Ximalaya), and video-based (Douyin) modalities differ in covering the four HBM dimensions. Figure 1(A) and Figure 1(B) depict the proportion of entries and the number of entries in each dimension over time across the three modalities.

Text-based platform (Weibo): The barriers dimension overwhelmingly dominated, accounting for 87.15% of posts, while susceptibility (4.58%), severity (11.84%), and benefits (12.23%) were significantly less frequent. This disparity was statistically confirmed by a Cochran's Q test ($Q = 190,634.70$, $p < 0.01$), with post hoc McNemar tests ($p < 0.01$) revealing significant differences across all dimensions.

Audio-based platform (Ximalaya): The distribution of HBM dimensions was more balanced. Barriers appeared in 52.05% of podcasts, followed by benefits (32.27%), severity (30.43%), and susceptibility (16.80%). A Cochran's Q test ($Q = 428.90$, $p < 0.01$) indicated significant differences across dimensions.

Video-based platform (Douyin): The barriers dimension was the most frequent, appearing in 90.52% of videos, followed by severity (69.20%) and benefits (60.95%), while susceptibility was the least frequent (15.77%). Significant differences between dimensions on Douyin were validated by a Cochran's Q test ($Q = 1,247.30$, $p < 0.01$) and post hoc McNemar tests ($p < 0.01$).

These differences highlight potential modality-specific emphases. In this case, we found that text discussions were more barrier-centric, while audio and video content presented a broader coverage of all four HBM dimensions. Over time, we can also see similar patterns being consistent across the three platforms.

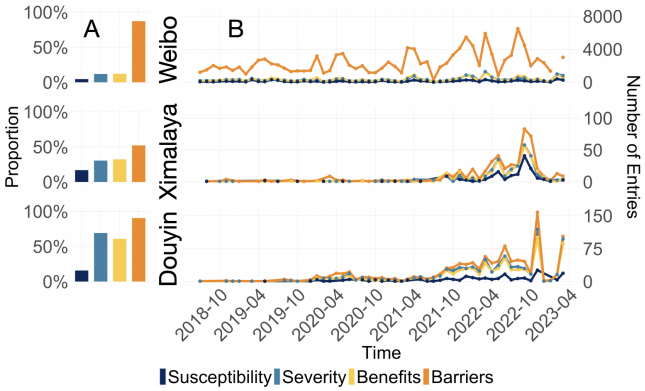

Figure 1: (A) Distribution of entries on HBM dimensions across text, audio, and video-based platforms (B) Counts of entries on HBM dimensions across text, audio, and video-based platforms over time

RQ2: What Are the Representative Topics in HPV Vaccine Discussions Across Modalities?

Using topic modeling approaches as described in the method section, we identified representative topics for each of the four dimensions of the HBM. In this section, we compare the topics across the three modalities, and the key findings on each dimension are presented below.

Barriers The representative topics in the barriers dimension are summarized in Table 4. It was found that Weibo mostly addressed public health events and policy concerns, including difficulties booking nine-valent vaccine appointments (*W0*) and reporting side effects (*W1*). The *W3* compared the prices of different vaccines. Meanwhile, the audio-based platform, Ximalaya, focused on economic and industrial themes, such as biotech companies' profits (*X2*) and age restrictions (*X0*) for vaccination. Besides, the *X1* highlighted the popularity and demand for expensive vaccines. In comparison, the video-based platform, Douyin centered on personal hurdles in obtaining the vaccine (*D0*), eligibility (*D1*), and advice on vaccination (*D2*). To sum it up, for barriers, discussions on Weibo highlights policy scheduling and side effects, Ximalaya underscores market or economic angles, and Douyin emphasizes vaccine availability and accessibility from individual standpoints.

Susceptibility The representative topics in the Susceptibility dimension are summarized in Table 5. It was found that Weibo mostly addressed HPV transmission routes (*W0*), vaccination schedules (*W1*), and target populations (*W3*). Meanwhile, the audio-based platform, Ximalaya, focused on vaccine availability and eligibility (*X0*), often noting preferences for the nine-valent vaccine; it also underscored the vaccine's safety and benefits, including access for both men and women abroad (*X1* and *X2*). In comparison, the video-based platform, Douyin, centered on everyday-life contexts (*D0*) and first-person accounts of HPV causing cervical cancer and other diseases (*D1* and *D2*). To sum it up, for susceptibility, discussions on Weibo highlight transmission methods and scheduling, Ximalaya underscores eligibility and

Plat.	Clu.	Num.	Representative Arguments
Weibo	W0	20,760	Unable to make appoint for the nine-valent vaccine, trapped by Alipay vaccine appointments.
	W1	7,429	After the vaccination, my arm started to ache, and I had bad reactions, a fever, and a bad cold.
	W2	6,585	Domestic bivalent vaccine is cheaper than imported nine-valent vaccine.
Ximalaya	X0	1,086	The HPV vaccine provides relatively little benefit to people over the age of 26.
	X1	1,070	Higher-priced HPV vaccines are more popular than lower-priced HPV vaccines and are scarcer in the market.
	X2	404	Watson Bio and Wantai BioPharm realized substantial growth in operating revenue and profit.
Douyin	D0	970	Getting the nine-valent vaccine is like winning the lottery.
	D1	732	If you are already infected by HPV, you can get the HPV vaccine.
	D2	629	It's better to get vaccinated earlier. Take any vaccines available, and don't wait for the nine-valent vaccination.

Table 4: Topics on Barriers

Plat.	Clu.	Num.	Representative Arguments
Weibo	W0	822	What to do if a woman is infected with HPV. HPV can be transmitted in a number of ways.
	W1	627	When is the best time for vaccination? As soon as possible, of course!
	W2	589	HPV vaccination could eliminate cervical cancer, and should be extended to boys and adults.
Ximalaya	X0	794	Is China using foreign vaccines when foreign countries are no longer mainlining the bivalent vaccine?
	X1	460	The HPV vaccine is quite safe, and the benefits of the vaccine far outweigh the side effects.
	X2	446	The HPV vaccine is available for both men and women.
Douyin	D0	751	The HPV causes not only cervical cancer, but also multiple other tumors and cancers, which infects both sexes.
	D1	127	We live in an environment where there are more than 100 varieties of HPV, and the quadrivalent vaccine can prevent 80% of cervical cancers and 90% of condyloma acuminatum.
	D2	85	This guy, he got HPV from playing around and infected his girlfriend without her knowing it.

Table 5: Topics on Susceptibility

safety, and Douyin emphasizes broader health impacts via personal narratives.

Severity The representative topics in the severity dimension are summarized in Table 6. It was found that Weibo mostly addressed China's low HPV vaccination rate (<1% among eligible women) and the high incidence of cervical cancer (*W1* and *W2*). The *W3* questioned the vaccine's efficacy, portraying two opposing perspectives—vaccine importance vs. limitations. Meanwhile, the audio-based platform, Ximalaya, focused on factual policy issues, such as age restrictions (*X1*) and incomplete protection (*X0*). Besides, the *X2* raised awareness about HPV-related cancers in LGBTQ+ groups. In comparison, the video-based platform, Douyin centered on personal narratives and experiences (*D0*, *D1*, and *D2*), where users shared subjective opinions in casual contexts. To sum it up, for severity, discussions on Weibo highlight policy debates on vaccination rate and cancer incidence, Ximalaya underscores policy and factual discussions, and Douyin emphasizes personal stories.

Benefits The representative topics in the benefits dimension are summarized in Table 7. It was found that Weibo mostly addressed policies, including WHO's global strategy (*W0*) and general vaccine efficacy (*W1*). The *W2* topics argued for higher vaccine valency with better protection. Meanwhile, the audio-based platform, Ximalaya, focused on safety (*X2*) and age policies (*X0*), including the extended age limit for nine-valent vaccines. Hosts often discussed early screening and treatment as complementary measures (*X1*). In comparison, the video-based platform, Douyin, centered on polarized opinions, with some videos promoted higher-valent vaccines (*D0*), others stressed affordability of bivalent options (*D1*), and early vaccination emerged as a recurring theme (*D3*). To sum it up, for benefits, discussions on Weibo highlight policy and perceived efficacy, Ximalaya underscores safety and expanded eligibility, and Douyin displays polarized views on valency and cost.

In summary, our results show that text-based platform often spotlights policy-level issues, audio-based platform delves into discussions of eligibility, safety, or industry aspects, and video-based platform emphasizes personal stories, everyday experiences, and quick recommendations. These variations align with each modality's intrinsic communication characteristics, i.e., short text provides updates, audio offers in-depth conversations, and short videos provide visual-driven content.

RQ3: What Emotions Are Expressed and How Do They Fluctuate Under Different Modalities?

Figure 2(A) illustrates the proportion of entries with neutral, negative and positive emotions across modalities. A chi-square test ($\chi^2 = 251.25$, $df = 4$, $p < 0.01$) confirmed significant differences among the three modalities. For instance, on the text-based platform, Weibo, neutral (50.10%) and emotional (positive and negative, 49.90%) posts are balanced, with negative ones (38.47%) dominate the emotional subset. We found that neutral contents often involve policy/news updates, while negative posts typically reflect user frustrations about perceived barriers, e.g., scheduling difficulties. On the audio-based platform, Ximalaya, emotions are more balanced, with 39.14% neutral, 38.63% negative, and the rest positive. We found that negative emotions, such as fear or anxiety, were triggered by susceptibility or severity contents in podcasts. The video-based platform, Douyin, shares similarities with Ximalaya (44.04% neutral, 34.33%

Plat.	Clu.	Num.	Representative Arguments
Weibo	W0	1,016	Currently, the HPV vaccination rate of school-age girls in China is less than 1 percent.
	W1	845	In China, cervical cancer is the cancer type with the second highest incidence in women.
	W2	660	Does the HPV vaccine mean you won't get cervical cancer? No!
Ximalaya	X0	1,023	HPV vaccine is a preventive vaccine that does not cure HPV infection nor prevent 100% of cervical lesions and cervical cancer.
	X1	948	HPV vaccine supply is tight and age-restricted. There are no strict restrictions in Hong Kong or Singapore.
	X2	617	HPV can also cause anal cancer, which is also present in the gay community.
Douyin	D0	691	Since I was infected with HPV, I have never had sex with a man. And I have gotten the HPV vaccine.
	D1	423	TCT test is available for cervical cancer prevention in overaged women.
	D2	419	To prevent getting infected while waiting for the 9-v vaccine, you can get the domestic 2- or 4-v vaccines.

Table 6: Topics on Severity

Plat.	Clu.	Num.	Representative Arguments
Weibo	W0	1,676	Vaccination, screening, and treatment are part of WHO's global strategy for cervical cancer elimination.
	W1	1,328	Imported and domestically produced vaccines are being inoculated in all regions to prevent cervical cancer.
	W2	661	The higher the HPV vaccine valent, the more types of HPV it protects against.
Ximalaya	X0	1,360	A new indication for Merck Sharp & Dohme's 9-valent HPV vaccine has been approved by the State Drug Administration of China. Indications extended to women of appropriate age from 9 to 45 years old.
	X1	1,281	There are lesions in between the infection and the malignant cancer, and we can detect and treat them early.
	X2	1,021	There are domestic and imported vaccines, so there may not be a significant difference in effectiveness and safety.
Douyin	D0	581	Minors can get vaccines later, people close to 45 can still get the first shot, and TCT exam is not needed.
	D1	526	The 9-valent vaccine is not cost-effective nor significantly better; the bivalent vaccine is the most effective in preventing cervical lesions and cancer, is superior to the quadrivalent vaccine, and provides cross-protection.
	D2	373	The earlier the vaccine is administered, the more protection and the lower the risk of getting infected by HPV.

Table 7: Topics on Benefits

negative, and the rest positive), emphasizing personal storytelling in amplify negativity (cost, side effects) and occasionally optimism (successful appointments). These findings highlight that the text platform leans towards being policy-driven and negative, the audio platform fosters more balanced yet strong emotions, and the video platform shares daily-life narratives and covers different emotional tones.

Emotions by HBM Dimensions Figure 2(A) also breaks down emotional compositions across the four HBM dimensions. Below are key observations and how they tie back to the topic analysis results presented in the previous section.

Barriers Chi-square test results ($\chi^2 = 153.33$, $df = 4$, $p < 0.01$) indicate significant modality differences in emotional responses to barriers. Barriers discussions remain largely negative (e.g., scheduling difficulties, high prices), though Douyin uniquely features some positive sentiment (22.56%) tied to domestic vaccine alternatives. Weibo and Ximalaya lean toward macro-level policy or economic constraints, while Douyin highlights personal experiences and occasional success stories.

Susceptibility A chi-square test ($\chi^2 = 478.51$, $df = 4$, $p < 0.01$) reveals significant differences in emotional involvement across platforms in this dimension. Ximalaya (78.26%) and Douyin (61.14%) display higher levels of emotional engagement compared to Weibo (20.27%), with the predominant sentiment being negative. Discussions often center on international comparisons, e.g., vaccine availability abroad, or personal risk, e.g., cancer transmission and gender factors, evoking fear and anxiety. The auditory depth of Ximalaya and Douyin's short-video narratives both intensify negative feelings, while Weibo's policy/news content remains less emotionally charged.

Severity Chi-square test results ($\chi^2 = 1299.70$, $df = 4$, $p < 0.01$) indicate significant modality differences in emotional responses to severity. Ximalaya leads in emotional engagement (71.52%), predominantly negative (45.08%), with a smaller positive share reflecting recognition of preventive measures. Douyin (55.61%) focuses on personal stories emphasizing HPV's consequences, whereas Weibo (17.00%) remains comparatively low on emotional responses. Discussions on severity, including topics like cancer risk or high-risk groups, are enhanced by the immersive nature of audio and video formats. These modalities amplify emotional responses, such as fear or sympathy. In contrast, the predominantly data-driven content on Weibo evokes fewer emotional reactions.

Benefits Chi-square test results ($\chi^2 = 1774.90$, $df = 4$, $p < 0.01$) shows that Douyin (60.95%) and Ximalaya (32.27%) both yield significant emotional responses to benefits, though negativity still appears when users face unmet expectations or cost barriers. Weibo exhibits the lowest emotional reaction (12.45%), aligned with its more policy-oriented style.

In summary, the audio modality, represented by Ximalaya, consistently demonstrates deeper emotional involvement, especially in the form of fear or anxiety. In contrast, the video modality, e.g., Douyin, generates mixed emotional responses, with some positivity emerging amid predominantly negative content. And, the text modality, e.g.,

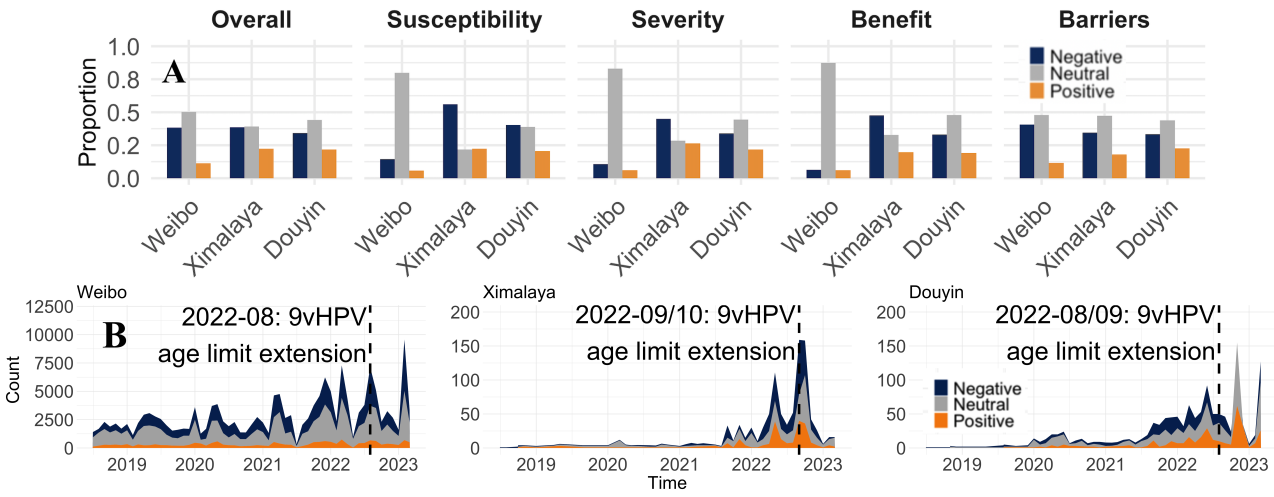

Figure 2: (A) Distribution of emotions across HBM dimensions across text, audio, and video-based platform (B) Time-series picture of emotions across text, audio, and video-based platform

Weibo, is overall the least emotionally charged but shows significant negativity when emotions are expressed.

Emotions Over Time and Across Different Modalities Figure 2(B) displays changes of entries with different emotional categories from 2018 to 2023 summarized on a monthly basis. We map these trends onto key policy milestones, market changes, and social events as follows. For the text-based platform, Weibo, rapid spikes follow major policy announcements (e.g., Mar 2022 free HPV vaccination proposals) or controversies (e.g.,Mar 2019 counterfeit vaccines), predominantly magnifying *barriers* and negative sentiment (frustration over supply, authenticity concerns). Emotional intensity subsides quickly, consistent with Weibo's fast-moving news cycle. For the audio-based platform, Ximalaya, vaccine age-extension policies and campaigns promoted by the National Health Commission (May 2022) prompt longer discussion windows, aided by the podcast format's narrative depth. Listeners and hosts explore international comparisons, scientific underpinnings, and financial aspects, sustaining diverse emotional expressions over weeks or months. And, for the video-based platform, Douyin, visual storytelling sparks swift emotional reactions, with sharp upswings in negativity during supply crises (Jun 2022 9vHPV shortage). However, positive narratives, such as those promoting domestic vaccines in November 2022, can quickly rebalance sentiment. Discussion intensity often manifests in scattered "viral" bursts rather than a single peak. At the macro level, events such as policy releases, supply fluctuations, media coverage, and scandals acted as critical triggers. However, the magnitude, duration, and nature of responses varied across platforms due to their distinct modalities and engagement patterns.

A Case Study: Nine-Valent (9v) Vaccine Coverage Expansion We chose the nine-valent HPV vaccine coverage expansion (9vExpansion) event as a case study, because it appeared as a major event on text-, audio-, and video-based platforms, yet each modality displayed a different pattern of discourse and timing. Although 9vExpansion was considered good news in general, many users still faced difficulties in making appointments, which led to mixed emotional responses. Examining how each modality responded to this event helps illuminate the distinct ways that text, audio, and video-based platforms process and circulate health-related information.

In our HBM-based coding of barriers, we distinguish "obstacles decreasing", e.g., expanded eligibility policies, and "obstacles increasing", e.g., difficulties in securing appointments for vaccines. Although the coverage expansion is positive by itself, it also leads to the surge in demand and makes it more difficult to make an appointment to get vaccinated, causing both optimism and frustration. This interplay of "barriers decreasing" and "barriers increasing" took unique forms across the three modalities.

Text-based platform (Weibo): There was an immediate spike in negative sentiment after the 9v coverage announcement, largely due to high costs and limited vaccine availability. Although some posts acknowledged the policy's promise of reducing barriers, many individuals complained about the difficulty of getting an appointment, swiftly overshadowing initial optimism. In line with Weibo's fast-paced nature, this wave of negative sentiment peaked and faded within a few weeks, reflecting a brief but strong reaction cycle.

Audio-based platform (Ximalaya): Here, the discussion cycle lasted up to two months. Early podcast episodes expressed relief over extended eligibility ("barriers decreasing"), but as more callers described ongoing supply shortages and high cost, frustration grew. This led to both positive and negative emotions coexisting for a longer period. The delayed peak and extended debate underscore how audio formats can sustain more in-depth conversations, allowing hosts and listeners to exchange personal anecdotes, analyze policy implications, and bring in expert views over time.

Video-based platform (Douyin): Discussions related to

the 9v policy were relatively fragmented, generating multiple small bursts of attention instead of one major collective wave. While some users voiced complaints about the scarcity of 9v vaccine appointments, many videos gained traction only briefly before interest shifted elsewhere. In November 2022, the ongoing shortage coincided with an emerging focus on a newly available domestic bivalent HPV vaccine, heralded as "national pride." This triggered a separate conversation spike centered on bivalent vaccines. Consequently, Douyin's short-form video structure gave rise to multiple minor peaks of discussion rather than one sustained debate.

Overall, this case study highlights how modality influences the intensity and duration of public reactions to a single policy. Text-based platforms like Weibo trigger rapid but short-lived sentiment spikes, while audio-based platforms such as Ximalaya foster more gradual and prolonged discussions, as users dissect topics over weeks. In contrast, video-based platforms like Douyin often generate brief, scattered waves of engagement, driven by shifting user attention to new topics like the domestic bivalent vaccine. These differences emphasize the critical role of modality in shaping public perceptions and debates, offering key insights for tailoring public health messages to specific media formats.

Discussion

This study examined how HPV vaccine discourse was shaped and how emotional responses evolve on the text-, audio-, and video-based platforms through the lens of the HBM. We found that the text-based platform predominantly focused on discussions around barriers to vaccination, such as side effects and costs, which aligns with previous research. On the other hand, audio- and video-based platforms provided a more balanced discourse on other HBM beliefs, including susceptibility, severity, and benefits (RQ1). Furthermore, the study found that different modalities preferred different topics and narratives. For instance, the text-based platform featured brief public discussions on public policies and prices, the audio-based platform featured longer in-depth conversations on social and economic news, and the video-based platform featured personal anecdotes in viral "hook" narratives (RQ2). The longitudinal emotional analysis revealed that emotions fluctuated over time in different patterns across modalities. For example, the text-based platforms had rapid emotional responses to policy changes with most discussions in negative emotions, the audio-based platform had slow but continuous and balanced emotional responses to policies, while the video-based platform users swung between positive and negative emotions (RQ3). These findings underscore the critical role of platform modality in shaping health communication dynamics.

To explain the observed differences in topics and emotions across three modalities, we drawing on both our analysis and the broader literature. We identified two factors that may potentially contribute to these differences, including a) content production procedures and b) platform attributes and user characteristics. Specifically, on text-based platforms, like Weibo and Twitter, the cost is minimal for users to post news snippets or comments, which foster instant discussions on negative or controversial content (Brady, Crockett, and Van Bavel 2020). On audio-based platforms, like Ximalaya and Podcast, content production often involves a longer planning stage and deeper insights, facilitating in-depth and rational user engagement (Persohn and Branson 2024). On video-based platforms, like Douyin and TikTok, users could use smartphones to easily edit visually attractive video clips. Since audiences frequently followed trending videos, videos could get viral and receive massive responses, leading to rapid emotional swings on the platform.

Implications

Firstly, our study explored the less-studied audio- and video-based platforms and found that they offered more balanced discussions, including coverage of *susceptibility*, *severity*, and *benefits*, alongside *barriers*. Video-based content, evoking stronger emotional responses, may lead to shifts in perceptions of susceptibility and severity, while text-based content encourages rational decision-making. These differences underscore the influence of modality on public beliefs about the HPV vaccine and potential behavioral changes.

Secondly, this study revealed the three-way interactions between media modalities, health beliefs, and user emotion expressions about the HPV vaccine. We found that negative emotions dominated discussions on barriers to vaccination on text-based platforms, while audio- and video-based platforms had both positive and negative emotions across all four HBM belief dimensions. These results expanded the existing finding in previous studies that examined the interactions between HBM beliefs and emotions on a single modality (e.g., (Mukhtar 2020; Cho et al. 2018; Yang et al. 2024)), and highlighted the importance of considering social media modalities in health communication.

Thirdly, by combining content and emotional analysis, this study showed how different media modalities use distinct narrative strategies and evoke dynamic emotional responses. Unlike previous studies that focused on single modalities, e.g., (Zhang et al. 2017), our findings revealed that text-based platforms triggered immediate emotional responses to news updates, mostly negative emotions about barriers to vaccination. Audio platforms encouraged in-depth conversations on the social and economic impacts of HPV policies, leading to balanced emotional engagement. Video platforms featured personal stories, with users quickly switching between positive and negative emotions. These findings offer valuable insights into the characteristics of each modality, providing practical guidance for public health communicators.

Fourth, the methodology developed in this study offers public health authorities a valuable tool for monitoring and responding to discussions and sentiment around health policies in real-time. By employing cross-modality analysis, authorities can track public perceptions and emotional responses across different platforms, enabling timely interventions to address emerging issues, correct misinformation, and support health policies. For example, the amplification of negative sentiments on text-based platforms could prompt fact-based clarifications, while the more nuanced engage-

ment on audio- and video-based platforms allows for long-term dialogue and trust-building.

Finally, our findings also emphasize the need for public health authorities to focus on addressing barriers to vaccination, such as side effects and costs, with transparent communication. This is particularly important during regulatory events, where shifts in discourse can quickly influence public sentiment. By staying agile and leveraging emotional engagement, public health agencies can better guide public understanding and foster informed decision-making.

Actionable Insights

From a public health standpoint, these findings imply that text-based platforms, e.g., Weibo, are ideal for rapid rumor control or official clarifications when new policies or crises occur. Audio-based platforms, e.g., Ximalaya, can host longer, expert-led sessions to address nuanced concerns, capitalizing on listeners' willingness for in-depth information and building trust over time. Video-based platforms, e.g., Douyin, excel at visually compelling, emotive storytelling, but appropriate regulations may be necessary to ensure accurate information transformation and to avoid viral waves of inaccurate contents.

Limitations

This study analyzed three platforms as representative examples of text, audio, and video modalities. However, platforms within the same modality may also vary significantly in user demographics, interface design, content policies, and algorithms, which could influence public discussions. Therefore, certain observed differences reported in this paper may be related to platform-specific characteristics rather than modalities alone. For example, Douyin (TikTok's Chinese counterpart) may have different discussion patterns compared to other video platforms like YouTube or Instagram Reels, despite all being video-based. Future research could explore additional platforms within each modality, e.g., Bilibili for video or Xiaoyuzhou for podcasts, to assess the generalizability of these findings and separate modality effects from platform-specific attributes.

Additionally, our study transcribed audio and video content into text, which may have resulted in the loss of important contextual features, such as tone, emphasis, or visual cues. This limitation is particularly relevant for platforms like Ximalaya and Douyin, where the multimodal nature of communication plays a significant role. Future studies could explore methods to preserve these features through multimodal analysis, integrating text, audio, and video data to more accurately capture the richness of the discussions.

The cultural and temporal scope of the dataset also warrants attention. Our analysis spanned from 2018 to 2023 without including data from earlier years. Additionally, the dataset focused on Chinese platforms, cross-cultural comparisons, longer data collection periods, or studies conducted during health crises would provide deeper insights into whether observed patterns are applicable or influenced by local norms, policies, or emergent social changes.

Future Work

While our classification models performed satisfactorily, there are opportunities for further refinement. Future work could explore the use of advanced pretrained large language models (LLMs), such as GPT-4 or T5, to enhance the detection of nuanced emotional cues, sarcasm, and platform-specific language styles. Additionally, multimodal fusion techniques could be employed to integrate features from text, audio, and video, improving the robustness and accuracy of the analysis.

Moreover, incorporating user-level data, such as interaction metrics (likes, comments, shares) and conducting network analyses could provide a more comprehensive understanding of how vaccine-related information spreads across platforms and evolves over time. This approach would allow us to identify key influencers and communities and better understand the dynamics of information dissemination.

We also recognize the potential value of integrating the HBM with other public health theories, such as the Theory of Planned Behavior (TPB), to gain a more comprehensive understanding of vaccine discussions on social media. Combining these theories with sentiment analysis (e.g., positive and negative emotions) could further enhance our understanding of the factors driving vaccination behaviors.

Lastly, while this study provides valuable insights into the differences in discussions across modalities, further work could investigate how these differences manifest across platforms or within different regions, policy contexts, or time periods. Examining these variations would help clarify whether the observed patterns are influenced more by modality or other factors such as platform-specific algorithms or user demographics. This research would be valuable in understanding the underlying dynamics of emotional engagement on different platforms. We envision future studies focusing on the interplay between modality, platform-specific factors, and external influences, such as governmental regulations or cultural contexts.

Conclusion

Our findings highlight that text-, audio-, and video-based platforms shape public discourse on HPV vaccines in distinct ways, potentially driven by modality-specific production costs, platform attributes, and user characteristics. Text-based platforms' environment amplifies short-lived controversies and negative sentiments, audio-based platforms' podcasts sustain multidimensional and reflective debates, and video-based platforms' short clips yield fragmented but emotionally charged mini-waves. Recognizing these modality-driven dynamics is crucial for designing targeted public health strategies, e.g., rapid rumor control on text-based platforms, in-depth expert engagement on audio-based platforms, and compelling visual campaigns on video-based platforms. Future research exploring additional platforms within each modality, broader data sources, advanced analytics, and cross-cultural comparisons can further refine how digital health communication strategies can adapt to each unique type of ecosystem.

Acknowledgments

This research was supported by the advanced computing resources provided by the Supercomputing Center of the USTC. Special thanks to Kaijie Jiao from USTC for assisting with data collection.

References

Ahmed, R.; and Mao, Y. 2022. Communication Research on Health Disparities and Coping Strategies in COVID-19 Related Crises. *Health Communication*, 37(12): 1455–1456.

Bailey, M.; Dittrich, D.; Kenneally, E.; and Maughan, D. 2012. The menlo report. *IEEE Security & Privacy*, 10(2): 71–75.

Boatman, D. D.; Eason, S.; Conn, M. E.; and Kennedy-Rea, S. K. 2022. Human papillomavirus vaccine messaging on TikTok: social media content analysis. *Health Promotion Practice*, 23(3): 382–387.

Bracken, C. C.; and Dalessandro, A. 2017. Modality and medium characteristics. In *The international encyclopedia of media effects*, 1–12.

Brady, W. J.; Crockett, M. J.; and Van Bavel, J. J. 2020. The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, 15(4): 978–1010.

Briones, R.; Nan, X.; Madden, K.; and Waks, L. 2012. When Vaccines Go Viral: an Analysis of HPV Vaccine Coverage on YouTube. *Health Communication*, 27(5): 478–485.

Chen, L.; Ling, Q.; Cao, T.; and Han, K. 2020. Mislabeled, fragmented, and conspiracy-driven: a content analysis of the social media discourse about the HPV vaccine in China. *Asian Journal of Communication*, 30(6): 450–469.

Cho, H.; Silver, N.; Na, K.; Adams, D.; Luong, K. T.; and Song, C. 2018. Visual cancer communication on social media: an examination of content and effects of # Melanomasucks. *Journal of medical internet research*, 20(9): e10501.

Dmytryshyn, J.; and Selk, A. 2022. Learning on the go: assessing knowledge gained from medical podcasts created for vulvovaginal disease education. *Journal of Lower Genital Tract Disease*, 26(2): 164–168.

Dockter, C. E.; Lee, S.; Boman, C. D.; Hinnant, A.; and Cameron, G. T. 2021. The impact of retransmission and modality on communicating health research findings via social media. *Health Communication*, 36(10): 1231–1241.

Du, J.; Luo, C.; Shegog, R.; Bian, J.; Cunningham, R. M.; Boom, J. A.; Poland, G. A.; Chen, Y.; and Tao, C. 2020. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine. *JAMA Network Open*, 3(11): e2022025.

Dugan, M.; McAllister, K. D.; and Brennan, E. K. W. 2022. Persuasion via performance: toward a handbook for youth to develop content that promotes HPV vaccination. *Critical Reviews in Eukaryotic Gene Expression*, 32(7): 77–91.

Dunn, A. G.; Surian, D.; Leask, J.; Dey, A.; Mandl, K. D.; and Coiera, E. 2017. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine*, 35(23): 3033–3040.

Grootendorst, M. 2022. BERTopic: neural topic modeling with a class-based TF-IDF procedure.

Guidry, J. P.; Carlyle, K. E.; LaRose, J. G.; Perrin, P.; Messner, M.; and Ryan, M. 2019. Using the Health Belief Model to Analyze Instagram Posts about Zika for Public Health Communications. *Emerging Infectious Diseases*, 25(1): 179–180.

Hennequin, R.; Khlif, A.; Voituret, F.; and Moussallam, M. 2020. Spleeter: a fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, 5(50): 2154.

Himelboim, I.; Xiao, X.; Lee, D. K. L.; Wang, M. Y.; and Borah, P. 2020. A Social Networks Approach to Understanding Vaccine Conversations on Twitter: network Clusters, Sentiment, and Certainty in HPV Social Networks. *Health Communication*, 35(5): 607–615.

Janz, N. K.; and Becker, M. H. 1984. The Health Belief Model: a decade later. *Health Education Quarterly*, 11(1): 1–47.

Jiang, S.; Wang, P.; Liu, P. L.; Ngien, A.; and Wu, X. 2023. Social media communication about HPV vaccine in China: a study using topic modeling and survey. *Health Communication*, 38(5): 935–946.

Kim, J.-N. 2018. Digital networked information society and public health: Problems and promises of networked health communication of lay publics. *Health Communication*, 33(1): 1–4.

Kim, Y.; Kim, Y.; and Wang, Y. 2016. Selective exposure to podcast and political participation: the mediating role of emotions. *International Journal of Mobile Communications*, 14(2): 133–148.

Kirkpatrick, C. E.; and Lawrie, L. L. 2023. Can videos on TikTok improve pap smear attitudes and intentions? Effects of source and autonomy support in short-form health videos. *Health Communication*, 39(10): 2066–2078.

Lewis, M.; and Grantham, S. 2022. From fork hands to microchips: An analysis of trending # CovidVaccine content on TikTok. *International Journal of Communication*, 16: 24.

Li, J.; and Zheng, H. 2020. Coverage of HPV-related information on Chinese social media: a content analysis of articles in Zhihu. *Human Vaccines & Immunotherapeutics*, 16(10): 2548–2554.

Li, W.; Nowak, G.; Jin, Y.; and Cacciatore, M. 2018. Inadequate and incomplete : chinese newspapers' coverage of the first licensed human papillomavirus (HPV) vaccine in China. *Journal of Health Communication*, 23(6): 581–590.

Li, Y.; Guan, M.; Hammond, P.; and Berrey, L. E. 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*, 36(3): 261–271.

Li, Y.; Guo, Y.; and Lin, H. 2022. Feminized virus: a content analysis of social media's representation of HPV vaccine. *Social Media + Society*, 8(3): 20563051221104232.

Lindgren, M. 2023. Intimacy and Emotions in Podcast Journalism: a Study of Award-Winning Australian and British Podcasts. *Journalism Practice*, 17(4): 704–719.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv*, abs/1907.11692.

Loft, L.; Pedersen, E.; Jacobsen, S.; Soborg, B.; and Bigaard, J. 2020. Using Facebook to increase coverage of HPV vaccination among Danish girls: an assessment of a Danish social media campaign. *Vaccine*, 38(31): 4901–4908.

Luisi, M. 2021. From bad to worse II: risk amplification of the HPV vaccine on Facebook. *vaccine*, 39(2): 303–308.

McInnes, L.; Healy, J.; and Astels, S. 2017. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11): 205.

McInnes, L.; Healy, J.; Saul, N.; and Grossberger, L. 2018. UMAP: Uniform Manifold Approximation and Projection. *The Journal of Open Source Software*, 3(29): 861.

Mukhtar, S. 2020. Mental health and emotional impact of COVID-19: applying health belief model for medical staff to general public of Pakistan. *Brain, behavior, and immunity*, 87: 28–29.

Perks, L. G.; and Turner, J. S. 2019. Podcasts and productivity: a qualitative uses and gratifications study. *Mass Communication and Society*, 22(1): 96–116.

Persohn, L.; and Branson, S. 2024. Broadening Legitimacy of Scholarly Podcasting as Knowledge Dissemination: Metrics, Opportunities and Considerations. *Publishing Research Quarterly*, 40: 269–286.

Quick, B. L. 2010. Applying the Health Belief Model to examine news coverage regarding steroids in sports by ABC, CBS, and NBC between March 1990 and May 2008. *Health Communication*, 25(3): 247–257.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Rosenstock, I. M. 1974. Historical origins of the health belief model. *Health Education Monographs*, 2(4): 328–335.

Seiter, C. R. 2021. Benefit, barrier, and self-efficacy messages in advance care planning education materials. *Health Communication*, 36(11): 1397–1404.

Smith, A. N.; Fischer, E.; and Yongjian, C. 2012. How does brand-related user-generated content differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing*, 26(2): 102–113.

Thorson, K.; Driscoll, K.; Ekdale, B.; Edgerly, S.; Thompson, L. G.; Schrock, A.; Swartz, L.; Vraga, E. K.; and Wells, C. 2013. Youtube, Twitter and the Occupy Movement: connecting content and circulation practices. *Information, Communication & Society*, 16(3): 421–451.

Tomar, S. 2006. Converting video formats with FFmpeg. *Linux J.*, 2006(146): 10.

Tricomi, P. P.; Kumar, S.; Conti, M.; and Subrahmanian, V. 2024. Climbing the Influence Tiers on TikTok: A Multimodal Study. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 1503–1516.

Wang, Y.; Li, Y.; Liang, P. P.; Morency, L.-P.; Bell, P.; and Lai, C. 2023. Cross-Attention is Not Enough: Incongruity-Aware Dynamic Hierarchical Fusion for Multimodal Affect Recognition. arXiv:2305.13583.

Yang, B.; Li, Y.; Terán, L.; Choi, E.; and Choi, Y. 2024. COVID-19 Vaccines# ForYou: Analyzing COVID-19 Vaccine Videos on TikTok During the Early Phase of the Vaccine Rollout in the US. *Health Communication*, 39(8): 1594–1605.

Zhang, C.; Gotsis, M.; and Jordan-Marsh, M. 2013. Social media microblogs as an HPV vaccination forum. *Human Vaccines & Immunotherapeutics*, 9(11): 2483–2489.

Zhang, J.; Xue, H.; Calabrese, C.; Chen, H.; and Dang, J. H. T. 2021. Understanding Human Papillomavirus Vaccine Promotions and Hesitancy in Northern California Through Examining Public Facebook Pages and Groups. *Frontiers in Digital Health*, 3: 683090.

Zhang, X.; Baker, K.; Pember, S.; and Bissell, K. 2017. Persuading me to eat healthy: a content analysis of YouTube public service announcements grounded in the health belief model. *Southern Communication Journal*, 82(1): 38–51.

Zhao, X.-L.; Hu, S.-Y.; Hu, J.-W.; Wang, H.-H.; Wen, T.-M.; Feng, Y.-S.; Qiao, Y.-L.; Zhao, F.-H.; and Zhang, Y. 2023. Tackling barriers to scale up human papillomavirus vaccination in China: progress and the way forward. *Infectious Diseases of Poverty*, 12(1): 86.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes.
   (e) Did you describe the limitations of your work? Yes.
   (f) Did you discuss any potential negative societal impacts of your work? Yes.
   (g) Did you discuss any potential misuse of your work? NA.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, in the Ethics statement.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? Yes.

(b) Have you provided justifications for all theoretical results? Yes.

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes.

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes.

(e) Did you address potential biases or limitations in your theoretical framework? Yes.

(f) Have you related your theoretical results to the existing literature in social science? Yes.

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes.

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? Yes.

(b) Did you include complete proofs of all theoretical results? NA.

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? We will make our datasets available when requested by reviewers and peer researchers.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes.

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes.

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? Yes.

(b) Did you mention the license of the assets? NA.

(c) Did you include any new assets in the supplemental material or as a URL? NA.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, in the Ethics statement

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR ? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA

Ethics Statement

Similar to previous social media studies (Tricomi et al. 2024) in ICWSM, our study uses only publicly available data and does not involve human subjects, exempting it from formal IRB review. We follow the Menlo Report guidelines (Bailey et al. 2012), reporting data in aggregated form and ensuring no risk to users. Our datasets, which follow FAIR principles, exclude raw audio and video to respect privacy settings, providing only video URLs and hydration code. Data was collected using non-invasive automated methods within platform rate limits, without causing server issues or receiving bans. The data were collected for research purposes, with personally identifiable information removed during preprocessing. All processed data is stored securely. Adherence to platform Terms of Service and ethical guidelines was prioritized throughout.

Appendix

Data Collection

We employed automated Python crawler scripts to collect comprehensive datasets encompassing all publicly accessible discussions related to HPV vaccines across three Chinese social media platforms: Weibo, Ximalaya, and Douyin.

We retrieved text data through keyword searches on Weibo's search interface[5] using three search keywords (*"HPV vaccine"*, *"Human Papilloma Vaccine"*, and *"Cervical Cancer Vaccine"*) to collect posts along with user information, restricting the dataset to Simplified Chinese posts. Initially, the dataset comprised 316,902 posts ranging from March 2011 to March 2023. For Ximalaya and Douyin, due to search limitations on these platforms, we implemented a consistent multi-step approach via their respective search interfaces[6][7]. First, we used the keywords *"HPV vaccine"*,

[5]Weibo search interface: https://s.weibo.com

[6]Ximalaya search interface: https://www.ximalaya.com/so/

[7]Douyin search interface: https://www.douyin.com/search/

"Human Papilloma Vaccine", and *"Cervical Cancer Vaccine"*, and collected one part of the dataset. Second, we used the keywords *"HPV"*, *"Human Papilloma"*, and *"Cervical Cancer"*, and then filtered the data by the keyword *"Vaccine"* in the title, description, and tag, generating the other part of the dataset. Third, we removed duplicates in the dataset collected in the previous two steps, generating the final dataset of Ximalaya and Douyin. The initial collection spanned December 2016 to March 2023 for Ximalaya and October 2018 to March 2023 for Douyin. The English-Chinese keyword mapping is provided in the Appendix Table A1.

To reduce or eliminate the influence of recommendation algorithms and ensure the collection of sufficient HPV vaccine-related discussions on Douyin, we registered multiple brand-new accounts with no prior usage history.

To align the dataset with our research questions and ensure temporal consistency across text-based Weibo, audio-based Ximalaya, and video-based Douyin, we standardized the timeframe to cover October 2018 to March 2023. The final dataset comprised 273,357 posts from Weibo, 1,228 audio podcasts from Ximalaya, and 1,225 videos from Douyin.

Additionally, for the Weibo dataset, irrelevant content such as advertisements, promotional entries, and unrelated campaigns was filtered out, refining the dataset to 241,435 posts. User contributions included 131,310 unique Weibo users, 564 Ximalaya users, and 863 Douyin users.

English Keyword	Chinese Keyword
HPV vaccine	HPV 疫苗
Human Papilloma Vaccine	人乳头瘤疫苗
Cervical Cancer Vaccine	宫颈癌疫苗
HPV	HPV
Human Papilloma	人乳头瘤
Cervical Cancer	宫颈癌
Vaccine	疫苗

Table A1: English-Chinese keyword mapping

Data Characteristic

The user and content distribution across the three platforms (Weibo, Ximalaya, and Douyin) are summarized as follows: On Weibo, there are 121,423 total users and 241,435 total posts. A small fraction of users is highly active, with the top 1% (1,214 users) contributing 24.67% of the total posts, the top 5% (6,071 users) contributing 42.69%, and the top 10% (12,142 users) contributing 50.81%. The Gini coefficient for Weibo is 0.469, indicating a relatively high concentration of posts among a small number of users. On Ximalaya, there are 565 total users and 1,227 total posts. The top 1% (5 users) contributed 9.94% of the posts, the top 5% (28 users) contributed 31.30%, and the top 10% (56 users) contributed 44.25%. The Gini coefficient for Ximalaya is 0.462, signifying a moderate level of concentration in user activity. Douyin has 863 total users and 1,224 total posts, with the top 1% (8 users) contributing 10.29%, the top 5% (43 users) contributing 23.20%, and the top 10% (86 users) contributing 31.45% of the total posts. Douyin's Gini coefficient is 0.269, indicating a more even distribution of posts compared to the other two platforms. For Weibo, our final dataset contains 241,435 posts from 131,310 unique users. A re-share analysis revealed that 17.00% of posts (41,095 posts) were re-shared, with a total of 2,128,073 re-shares and an average of 8.8 re-shares per post. The re-share distribution shows that 83.00% of posts were not re-shared, 13.30% were re-shared between 1 and 9 times, 2.90% between 10 and 99 times, 0.70% between 100 and 999 times, and only 0.10% reached over 1,000 re-shares. This indicates that while a minority of posts receive high dissemination, the overall re-sharing pattern suggests a broad participation rather than being driven solely by a handful of highly influential users.Regarding Douyin and Ximalaya, our analysis of content repetition provides further insights into user activity. On Douyin, among 1,224 videos (after filtering out advertisements), only 13 titles were repeated, resulting in a repetition rate of 2.80%, which suggests diverse content contributions. In contrast, on Ximalaya, out of 1,226 audio podcasts, 71 titles were repeated, yielding a repetition rate of 15.50%. Although this indicates a somewhat higher level of content duplication on Ximalaya, it still reflects contributions from a broader range of users.

Emotion Detection

For both emotion classification tasks, we followed the standard 8:1:1 ratio to split the data into training, validation, and test sets. This split ensures a balanced approach to model evaluation, allowing us to tune the model effectively while maintaining the integrity of the test set for final performance evaluation. Regarding the training process and hyperparameter tuning, we selected Chinese-BERT-wwm [8] as the pre-trained language model for both the six-emotion classification task and the binary sentiment classification task. For the model architecture and training parameters, we set the maximum token length for each input to 128 tokens to capture necessary context while balancing computational efficiency. A batch size of 10 was chosen to optimize memory usage and training efficiency. Additionally, we set the learning rate to 1e-5 for both tasks to ensure stable training, and used cross-entropy loss as the loss function, which is commonly used for classification tasks. Each model was trained for 10 epochs, based on empirical tuning, which resulted in a good trade-off between model performance and overfitting.

Emotion	Precision	Recall	F1-Score
Fear	0.62	0.76	0.68
Neutral	0.94	0.78	0.85
Sad	0.59	0.81	0.69
Surprise	0.72	0.53	0.61
Angry	0.86	0.80	0.82
Happy	0.80	0.82	0.81

Table A2: Performance of emotion classifier on six emotions

[8] https://huggingface.co/hfl/chinese-bert-wwm

Dimension	Definition	Primary Description	Examples
Susceptibility	The assessment of the risk of getting an HPV infection.	1. Likelihood of infection; possibility of infection.	1. In other countries, eligible males could receive HPV vaccination.
		2. Behaviors that may lead to HPV infection.	2. This disease only affects girls, so I encourage women to get vaccinated.
		3. Causes and symptoms of HPV infection.	3. 80% of women will contract HPV in their lifetime.
		4. Both males and females should/need/must receive the vaccine.	
		5. Suitability of the HPV vaccine for certain groups.	
Severity	The assessment of whether an HPV infection constitutes a serious health issue.	1. Death, mortality rate, infection rate, incidence rate.	1. Cervical cancer ranks first in incidence and mortality among female reproductive system malignancies in China.
		2. Long incubation period.	2. HPV vaccines cannot prevent all cancer-causing HPV types; screening is still necessary.
		3. Limited vaccine effectiveness; inability to prevent all HPV types.	3. The WHO recommends primary, secondary, and tertiary prevention strategies for HPV.
		4. Repeated infections.	4. Treatment is expensive and can financially devastate families.
		5. Methods of transmission.	5. The 9-valent vaccine can prevent 9 HPV subtypes, the quadrivalent vaccine 4, and the bivalent vaccine 2.
Benefits	The advantages of the HPV vaccine in preventing HPV infection and HPV-induced cancers.	1. Protective and preventive effects of the vaccine.	1. The 9-valent HPV vaccine can prevent 92.1% of cervical cancers.
		2. Necessity of vaccination.	2. HPV vaccination will benefit globally and support WHO's strategy for cervical cancer elimination.
Barriers	Challenges related to HPV vaccination, including side effects, cost, and restrictions.	1. Costs of vaccination and screening.	1. Anaphylaxis may occur; vaccines are becoming more expensive.
		2. Conflicts with other vaccines.	2. TCT screening costs 180 RMB.
		3. Age restrictions.	3. Vaccination is most effective before exposure to HPV.
		4. Eligibility conditions (e.g., sex or prior infection).	4. HPV vaccines are now available for public appointments.
		5. Concerns about reinfection or vaccine efficacy.	5. Could receiving the HPV vaccine increase the risk of infection?
		6. Changes or extensions in eligibility criteria.	6. HPV vaccination is effective after completing all doses.

Table A3: Codebook of Health Belief Model (HBM)

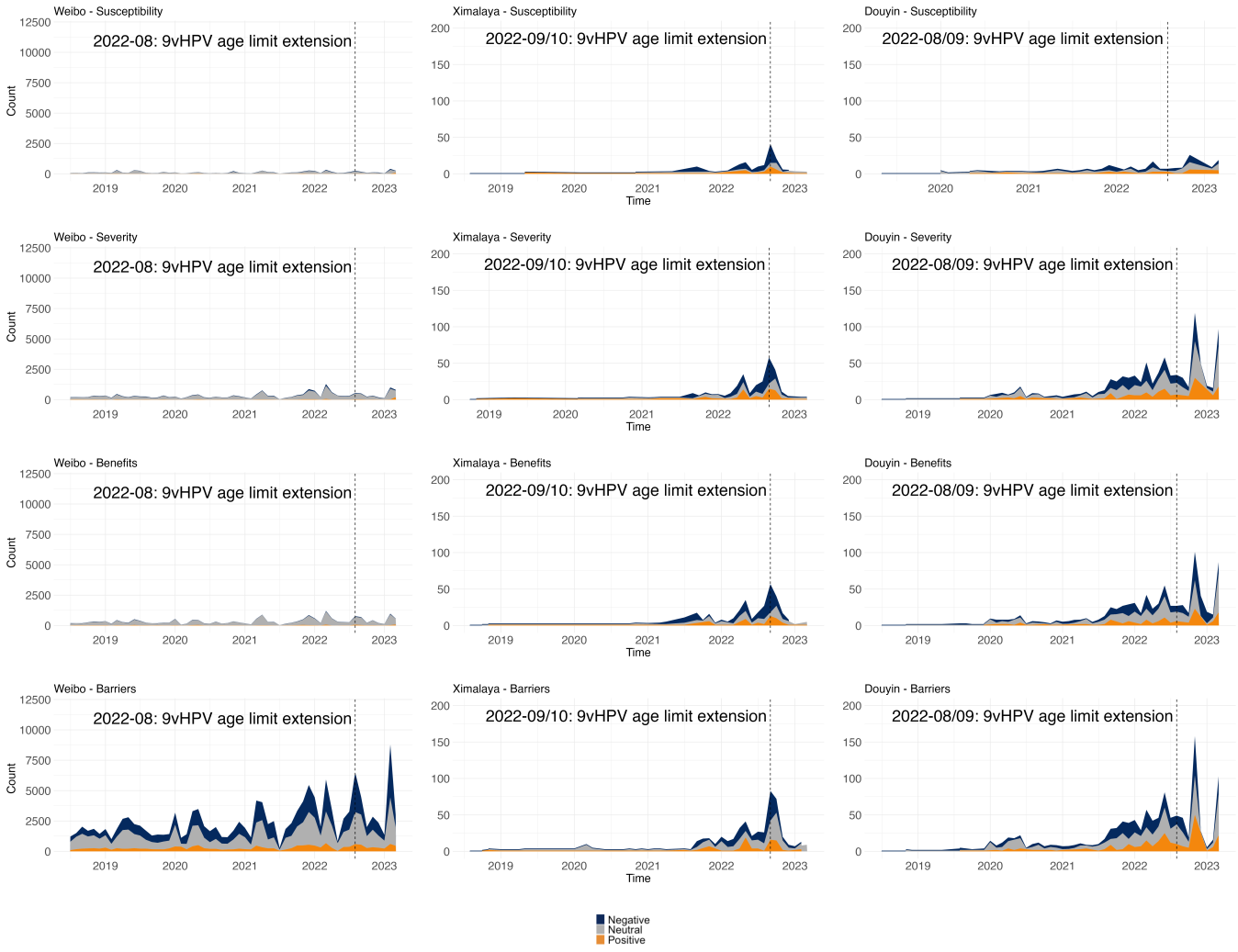

Figure A1: Time-series pictures of emotions across dimensions across text, audio, and video-based platforms

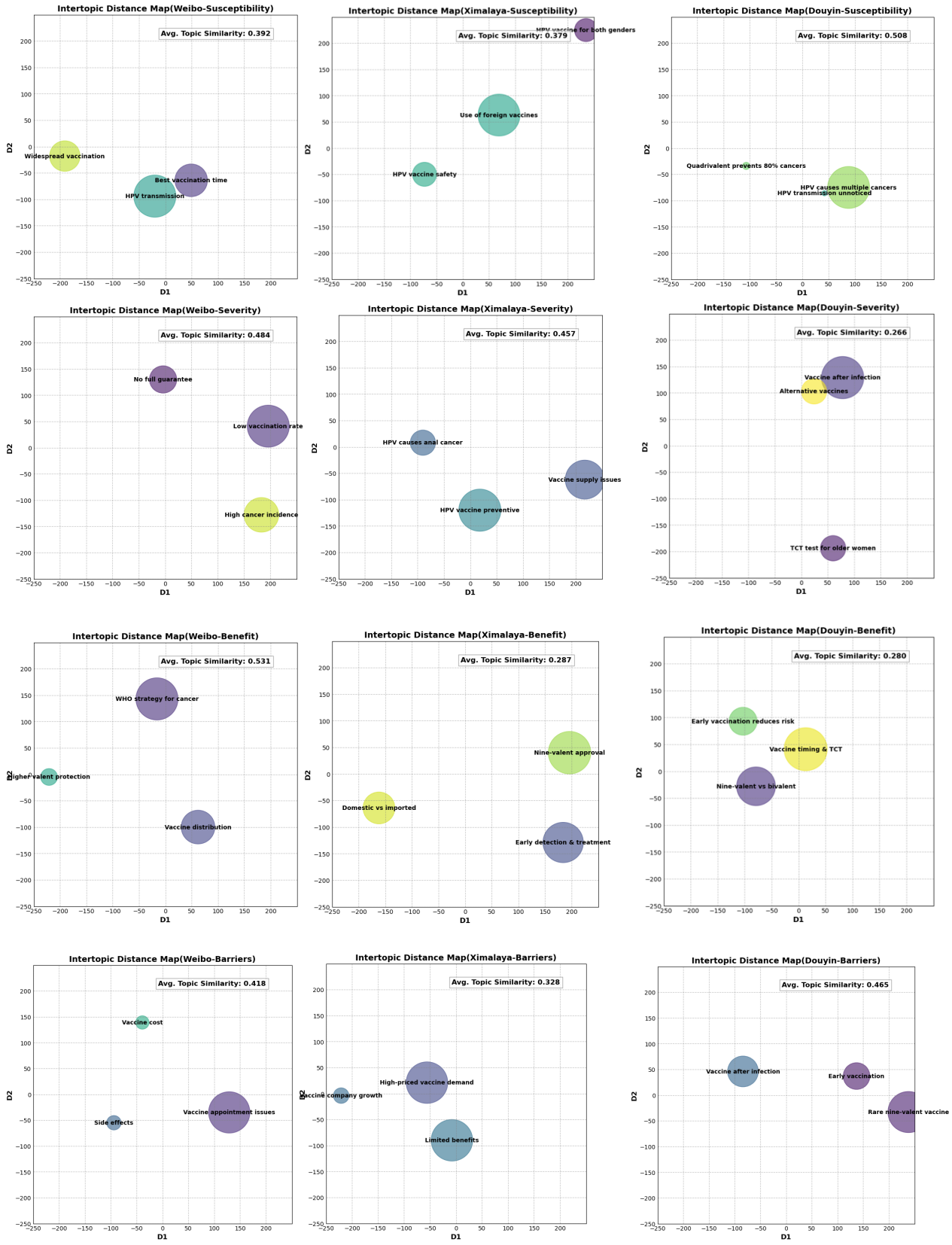

Figure A2: Topic distribution and clustering across dimensions across text, audio, and video-based platforms