Television Discourse Decoded: Comprehensive Multimodal Analytics at Scale

Anmol Agarwal*
International Institute of Information Technology
Hyderabad, India
anmol.agarwal@students.iiit.ac.in

Pratyush Priyadarshi*
International Institute of Information Technology
Hyderabad, India
pratyush.priyadarshi@students.iiit.ac.in

Shiven Sinha
International Institute of Information Technology
Hyderabad, India
shiven.sinha@research.iiit.ac.in

Shrey Gupta
International Institute of Information Technology
Hyderabad, India
shrey.gupta@students.iiit.ac.in

Hitkul Jangra
Indraprastha Institute of Information Technology
Delhi, India
hitkuli@iiitd.ac.in

Ponnurangam Kumaraguru
International Institute of Information Technology
Hyderabad, India
pk.guru@iiit.ac.in

Kiran Garimella†
Rutgers University
New Brunswick, USA
kiran.garimella@rutgers.edu

Abstract

In this paper, we tackle the complex task of analyzing televised debates, with a focus on a prime time news debate show from India. Previous methods, which often relied solely on text, fall short in capturing the multimodal essence of these debates [27]. To address this gap, we introduce a comprehensive automated toolkit that employs advanced computer vision and speech-to-text techniques for large-scale multimedia analysis. Utilizing state-of-the-art computer vision algorithms and speech-to-text methods, we transcribe, diarize, and analyze thousands of YouTube videos of a prime-time television debate show in India. These debates are a central part of Indian media but have been criticized for compromised journalistic integrity and excessive dramatization [18]. Our toolkit provides concrete metrics to assess bias and incivility, capturing a comprehensive multimedia perspective that includes text, audio utterances, and video frames. Our findings reveal significant biases in topic selection and panelist representation, along with alarming levels of incivility. This work offers a scalable, automated approach for future research in multimedia analysis, with profound implications for the quality of public discourse and democratic debate. To catalyze further research in this area, we also release the code, dataset collected and supplemental pdf[1].

*Equal contribution.
†Corresponding author
[1]https://github.com/anmolagarwal999/television-discourse-decoded

KDD '24, August 25–29, 2024, Barcelona, Spain

ACM ISBN 979-8-4007-0490-1/24/08
https://doi.org/10.1145/3637528.3671532

CCS Concepts

• **Applied computing** → *Law, social and behavioral sciences*; • **Computing methodologies** → *Natural language processing*;

Keywords

Multimodal analysis; video analysis; television; Bias detection; Incivil speech

ACM Reference Format:
Anmol Agarwal, Pratyush Priyadarshi, Shiven Sinha, Shrey Gupta, Hitkul Jangra, Ponnurangam Kumaraguru, and Kiran Garimella. 2024. Television Discourse Decoded: Comprehensive Multimodal Analytics at Scale. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3637528.3671532

1 Introduction

Television debates are a cornerstone of public discourse, serving as platforms for the exchange of ideas and viewpoints. In India, prime-time debates are viewed by millions and have a substantial impact on shaping public opinion [3]. However, these debates have recently undergone scrutiny for compromised journalistic integrity and increasing incivility [22]. Understanding the nuances at scale in these debates is important, yet a formidable task due to the multimedia nature of the content, which blends text, audio, & video.

Automated methods to analyze such content have largely been absent or inadequate, often focusing only on textual aspects [27]. These naive approaches are insufficient for two reasons: the sheer scale of televised debates available for analysis, and the intricate multimedia elements that must be considered to provide a complete picture. Previous attempts at solving this problem either employ text-based analytics that miss out on contextual cues or rely on small-scale, manual coding that lacks scalability [15, 17].

One of the most intriguing yet challenging aspects of analyzing news debates lies in their multimodal nature, which combines text,

audio, and visual elements. Each of these modalities carries crucial information that contributes to the complete understanding of a debate. While text may convey the spoken content, it misses out on the tone, pitch, and interruptions that audio captures. Similarly, video offers visual cues like facial expressions and body language that are lost in a purely textual analysis. Thus, a comprehensive analysis mandates a multifaceted approach that considers all these elements in unison.

Scale further complicates this endeavor. The vast number of televised debates—spanning thousands of episodes and millions of minutes of footage—requires a computational approach capable of scaling without loss of accuracy. Moreover, the temporal dynamics intrinsic to debates, such as topic changes and emotional fluctuations, add another layer of complexity. Capturing these dynamics over time demands sophisticated algorithms that can adapt to fast-changing contexts within a debate.

Beyond the technical aspects, subjective elements like bias and incivility pose their own challenges. Creating universally applicable metrics for these elements is particularly difficult, as perceptions of bias can differ based on individual viewpoints. Similarly, cultural and linguistic nuances like local idioms or specific styles of argumentation, especially pertinent in the Indian context, require additional considerations for accurate analysis. The presence of speech overlaps and interruptions further muddies the waters. These not only challenge the speech-to-text conversion process but also have implications for downstream analytics, potentially affecting the quality of the transcriptions and, consequently, the entire analysis. In cases where real-time analysis is required, these complexities amplify, adding an additional computational burden.

In light of these challenges, this paper introduces a novel automated toolkit designed for large-scale multimedia analysis. Our approach leverages state-of-the-art advances in computer vision algorithms and speech-to-text methods to transcribe, diarize, and analyze thousands of televised debates hosted on YouTube. We collect data spanning over 6 years from one of India's most popular prime time news debate shows (henceforth referred as which airs on one of India's most watched English news channel (henceforth referred as THE CHANNEL). The program is known for its emphasis on nationalistic discourse, its fervent critique of political adversaries, and its often intense discussions involving minority groups. It is widely perceived that the channel exhibits a preference for the incumbent government; however, this perspective has not been rigorously substantiated through quantitative research methods.

To fill this gap, we offer concrete metrics to evaluate bias in discussion topics and measure levels of incivility. Furthermore, our toolkit amalgamates textual transcriptions with video frames and audio utterances, thus capturing a comprehensive multimedia perspective. This offers a much-needed foundation for future research, making it possible to conduct studies that are both wide-ranging and deep in their analytical scope.

Our work is situated within the broader, ongoing debate about the quality of television debates in India, which have recently come under criticism for a rise in sensationalism, dramatization, and incivility. We capture these elements in our analysis to provide a comprehensive understanding of the current state of televised debates in the country.

Our analysis reveals a degree of bias in the debate show, characterized by support for the Ruling Party and a tendency to take a discrediting stance towards opposition parties and journalists. Our study also points out a notable imbalance between male and female panelists, leading to an uneven representation of social issues. It's especially concerning to see the high levels of rudeness measured by our pipeline: on average, shouting happens in about 9% of the duration of the videos. These findings have profound implications. The pronounced bias and a lack of dignified discourse cast doubt on the show's role as a fair platform for different opinions. This calls into question the show's role in fostering constructive public debate; instead, it appears to prioritize sensationalism, potentially at the cost of nuanced discussion and mutual understanding.

We make both our data analysis pipeline and the collected data publicly available. This is expected to catalyze further research in automated video analysis, extending its applicability beyond the Indian context. By doing so, we aim to unlock the untapped potential of YouTube as a tractable resource for large-scale studies.

2 Background and Related Work

2.1 Bias and Incivility in Indian media

India, the world's largest democracy, has recently experienced a decline in press freedom, currently ranking 161 out of 180 countries as per Reporters Without Borders [43]. This decline has been partly attributed to the acquisition of media outlets by influential figures who maintain close ties with political leaders. Such ownership structures have led to seemingly evident biases in media reporting, with a majority of TV channels noticeably supporting the political party in power. Given the critical role of media in a democratic setup, it becomes imperative to analyze and quantify this bias, a task that some previous work has approached qualitatively.

Since THE CHANNEL's inception, it has been the most-watched English news channel in India, commanding an average viewership of 40% [7]. Known for its sensationalist approach to news reporting, THE CHANNEL has often been criticized for displaying a pro-Hindu, pro-nationalist, and pro-government bias [39]. One of the channel's popular prime time debate show, henceforth, referred as THE SHOW epitomizes this tendency. The show attracts over five million daily viewers and is characterized by its nationalistic tone. It often targets those who appear to oppose the government's viewpoint.
Despite its status as the most-watched news TV show in India, the program has moved away from the traditional format of a balanced news debate. Instead, it now often features a heightened level of dramatization, impassioned language, and overlapping dialogue [18]. This sensational approach appears to resonate with viewers [38].

While there is a substantial body of qualitative work addressing bias, factual inaccuracies, and the dramatization of news in Indian media [3, 12, 22], our research contributes by offering quantitative evidence. Notably, some channels, including THE CHANNEL, have even acknowledged their tendencies to sensationalize news. Our study enriches this dialogue by supplying empirical data on the nature and framing of the content presented in such debate shows.

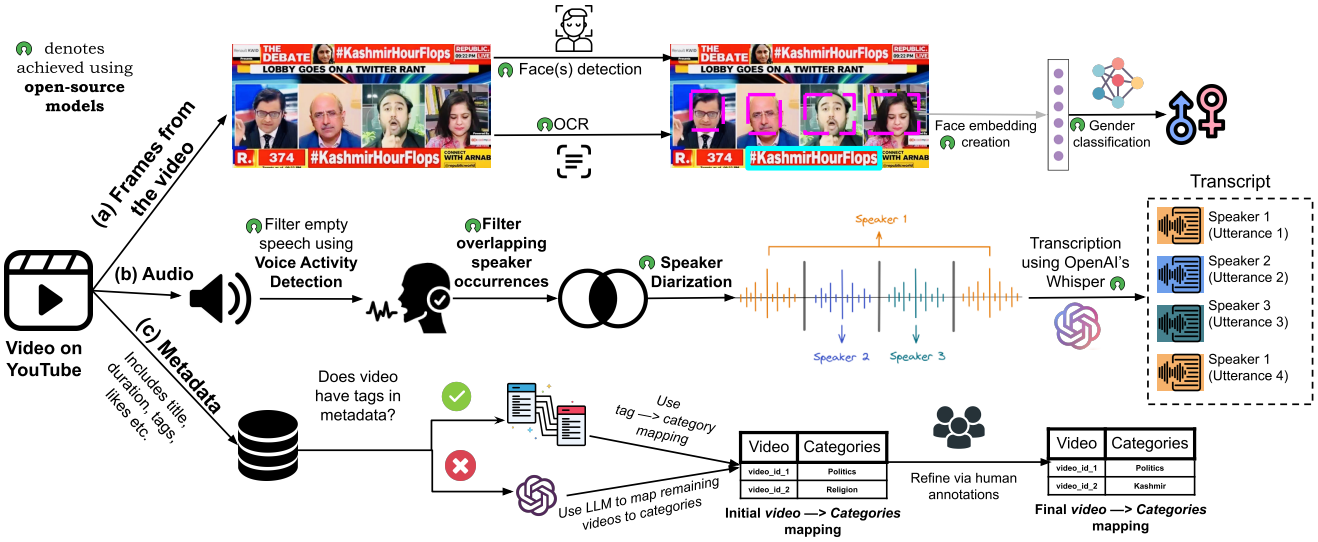


Figure 1: Pipeline overview: Branch (a) details the process for identifying gender from facial data in videos and extracting hashtags from debate screens; Branch (b) outlines the audio cleaning and speaker diarization procedures, followed by transcription of utterances into text; Branch (c) illustrates the semi-automated annotation system that leverages YouTube metadata & LLMs to streamline the categorization of videos into categories, thereby reducing human annotation workload.

2.2 Analysis of TV News and Media

In the realm of analysis of TV news and media, multiple avenues of research have emerged that address the intricate problem of media bias, the influence of media on public perception, and the role of technological platforms in shaping or amplifying these biases. One stream of work delves into detecting subtle biases in online news by examining 'gatekeeping,' coverage, and statement bias, using unsupervised methods on a geographically diverse set of news sources [32]. This line of research intersects with another that undertakes a comparative framing analysis of terrorism coverage in US and UK newspapers, revealing differing national focuses, either militaristic or diplomatic, that guide news stories' framing [24].

While these studies examine traditional media forms, a more recent shift towards social media as a news outlet is apparent in the research literature. For example, some researchers employ scalable methodologies that leverage social media's advertiser interfaces to infer the ideological slant of thousands of news outlets. This method provides granularity, capturing demographic biases that go beyond political leanings, and results in deployable systems for transparency [30]. This complements work on newspaper endorsements' influence on voting behavior, highlighting source credibility as a key factor in endorsement effectiveness.

Interestingly, research has also been conducted in the Indian context, where media bias in policy coverage has been systematically quantified. This work reveals biases in topic selection and representation of different social classes and political parties. Notably, social media platforms seem to echo rather than mitigate these biases, an insight that aligns with the earlier observations on the role of social media in amplifying traditional media biases [34]. Collectively, these studies illuminate the evolving landscape of news and media analysis, showcasing the need for comprehensive, multifaceted approaches. They underline the significance of understanding both the subtleties in traditional media framing and the influential role of social media platforms.

2.3 Multimodal Analysis Tools

Video analysis has become an increasingly significant area of research, particularly as social media platforms transition towards video-centric content. The rise of short video services like TikTok underscores the growing importance of video in the digital age. Advances in computer vision technology have reached a stage where real-world applications are not just feasible but increasingly sophisticated. Problems such as video summarization and key frame extraction have been addressed, offering novel solutions and methodologies [21, 33].

Earlier works [2] faced challenges in transcribing large volumes of audio data—284,000 hours of radio—due to the limitations in transcription models at the time. The current models for transcription have improved considerably showcasing a rapid evolution of the field. Videos present a complex interplay of multiple modalities, including visuals, text, and audio. While each can be analyzed independently, their true power lies in how they interact. Renoust et al. [29] explored this by using deep neural networks for face detection and text counting metrics to measure politicians' screen time. Their work demonstrated the capability of AI techniques in analyzing large video datasets, offering insights into complex social dynamics. The GDELT Project [13] provides web-based interfaces for analyzing caption text and other on-screen elements but lacks in-depth labelling related to voice tone or content being discussed.

Our work fills these gaps by analyzing a comparable dataset of videos and enriches it by labelling content related to what is

spoken, who is on-screen, and the tone of voice used. Overall, our research builds on recent advancements in various domains of AI. We leverage state-of-the-art models in image processing for tasks such as face and gender recognition, utilize speech processing algorithms to identify instances of shouting, and employ speech-to-text models to capture the spoken content. We aim to provide a holistic, multi-modal analysis that can serve as a robust foundation for future studies in video analytics.

3 Data Collection & Processing

Our primary dataset comprises 2,087 hours of debate footage from 3,000 videos. Initially, we used the YouTube Data API[2] to extract metadata from the official THE CHANNEL's THE SHOW playlist as of December 2022. This provided us with 3,151 unique videos dating back to May 2017. Out of these, we filtered out 67 videos because they were too short/long (i.e. their duration was less than 10 minutes or exceeded 4 hours) and filtered out an additional 84 videos because the annotators couldn't agree on their categories. We were finally left with 3,000 videos corresponding to over 2,087 hours of video content. The metadata fetched using the YouTube Data API for each video contains the title, URL, description, and a list of tags chosen by the channel[3] associated with the video.

3.1 Categorizing the Videos

To categorize the 3,000 videos in our dataset, we manually created categories. Initially, using a framework from a prior study we adopted 18 categories [8]. Each coder independently assessed a subset of videos, relying on metadata such as titles, descriptions, hashtags, and tags for initial categorization. If a video did not fit into the existing categories, a new category was proposed and discussed among coders for potential inclusion. This iterative process continued until a consensus was reached on the categories.

Recognizing that a video could span multiple topics, we implemented a two-tiered coding system comprising major and minor categories. Each video was assigned to one major category (e.g., sports, religion, international affairs) while potentially belonging to multiple minor ones, allowing for emergent sub-themes (e.g., 'Russia-Ukraine crisis', 'SSR case'). To automate video categorization, we used the "tags" present in the YouTube metadata. The tags were then mapped to categories. For example, a video with tag "budget 2019" was labelled Economy as the major category. However, after this initial categorization, we were left with 830 videos that could not be mapped to a category due to the absence of tags or the presence of generic tags. We then used GPT-4 to map these remaining videos to the categories based on the video's title.

The final step involved human refinement to correct any potential errors from the automated labelling. Two annotators independently examined and refined the labels, and their agreement was measured using the Fleiss kappa statistic, which was computed to be 0.933, indicating excellent agreement. By incorporating LLMs and tags-metadata we reduced the number of categories that can be mapped to a video to a smaller subset thereby significantly reducing the time taken in the human annotation. This hybrid approach of automated and human annotation in our pipeline allowed for an efficient and comprehensive categorization of the videos.
In a minority of the cases with disagreements (110 cases), both the annotators discussed among themselves and resolved most of the disagreements. There was no clear agreement on 84 videos which were removed from further analysis, leaving us with 3,000 videos. A complete breakdown of major and minor categories is available in Table 3. The majority of the videos fall into five dominant categories: Politics, Religion, COVID Lockdowns, International Affairs, and Crime & Justice, collectively accounting for 66% of the total dataset. A mapping from these categories to their respective tags and examples of the annotation process can be found in the supplemental pdf. Our semi-automated pipeline has been illustrated in branch (c) of Figure 1.

3.2 Transcription and Speaker Diarization

To analyze the content of the debates, it was essential to determine both what was said and who said it. Audio transcription converts speech in an audio file into written text, but debates involve multiple speakers in multi-turn interactions. Therefore, before transcription, we performed speaker diarization—a process that partitions an audio stream into segments and attributes them to specific speakers [25]. This allowed us to transcribe individual speaker segments, resulting in a conversation-format transcription for each video.

We executed two key pre-processing steps to enhance the quality of the diarization results. First, we removed segments devoid of speech, such as interstitials and speaker transitions, using the **Voice Activity Detection** feature from the Pyannote toolkit [6]. This removal improved subsequent diarization accuracy. Second, we filtered out overlapping speech segments to avoid performance degradation in speaker clustering during diarization, accomplished using the same Pyannote model [5].

After these pre-processing steps, we employed the Pyannote **diarization** module to partition the audio into homogeneous segments, each assigned to a specific speaker [5, 6]. For transcription, we leveraged OpenAI's Whisper speech-to-text model [28], notable for its robust performance on diverse accents and technical language. Whisper has demonstrated near-human-level accuracy in challenging noisy settings [20]. Combining Whisper's transcription capabilities with Pyannote's audio segmentation and speaker diarization enabled us to transcribe and accurately attribute speech (and the corresponding transcribed text) to individual speakers.

Our qualitative analysis revealed certain limitations in the Pyannote model's **overlap detection**. Specifically, the model only considered speech overlapping if all audio segments were incoherent. If one speaker's voice dominated others, the model did not recognize the speech as overlapping. This issue resulted in scenarios where multiple speakers are active, but not identified as overlapping. Additionally, the transcription quality for overlapped speech was suboptimal, likely because Whisper's training data primarily focuses on transcribing a single speaker while treating other voices as background noise.[4] Due to these overlap detection limitations, we encountered 'spurious speakers'—artifacts that appeared to be individual speakers but were actually combinations of multiple voices. Such spurious speakers also emerged when the debate

[2]https://developers.google.com/youtube/v3/docs/playlistItems/list
[3]Details: https://developers.google.com/youtube/v3/docs/videos#snippet.tags[]

[4]https://github.com/openai/whisper/discussions/434#discussioncomment-4141250

anchor played relevant footage with accompanying audio, complicating the speaker diarization process. Nevertheless, this might impact a small fraction of our video content and manual evaluations on a subset of videos showed that the overall quality of the transcripts was exceptional. The entire transcription pipeline is outlined in *Branch (b)* of Figure 1.

3.3 Face and gender detection

Gender identification from video frames, as shown in *Branch (a)* of Figure 1, entailed extracting and analyzing facial data. For facial recognition in our study, we employed the DeepFace library [35], specifically utilizing the RetinaFace detector coupled with the VGG-Face model [36]. From a given video, we sampled one frame every 3 seconds and extracted all faces from it. One challenge we encountered was the presence of spurious faces, in advertisements or images unrelated to the debate. To address this, we implemented a filtering mechanism based on the size of the face in the frame and the confidence scores provided by the model. It's important to acknowledge that our study operates within the limitation of recognizing gender in binary terms, although we recognize that gender is not a binary construct. In a small-scale experiment to validate the performance of this model, we annotated all the faces on 2,500 randomly sampled frames across our dataset and found the classifier to have a precision of 0.91 and a recall of 0.994 for males, and a precision of 0.975 and a recall of 0.81 for females.

3.4 Extracting Panelist Names from Transcripts

To study the individuals appearing in the debates, we extracted the names of panelists from the transcripts. Traditional approaches like Named-Entity Recognition (NER) on the transcripts did not perform well for three main reasons: (i) NER captured names of people mentioned in the debate but not actually panelists, (ii) multiple variations were used to refer to the same person (e.g., [General GD Bakshi, General Bakshi, Major General GD Bakshi]), and (iii) transcription errors led to inconsistent spellings of the same name (e.g., Atiqur Rahman, Atiq-ur-Rehman Sahab, Atiku Rehman). To address these issues, we adopted Meta's open-sourced LLaMA-2 13B model for this task [42].

When the transcript of an entire video exceeded the model's context length, we chunked the transcript into parts and took the union of names extracted from each chunk to identify potential panelists for the video. The prompt used for name extraction can be found in the supplemental pdf. The names returned by this approach were not completely clean, so we performed fuzzy matching and clustered similar names using a combination of Partial Token Sort Ratio and metaphone-based matching.

Using these techniques, we curated a list of 265 panelists, covering 91.7% of the videos and 50% of all appearances. We focused on frequently invited guests rather than full coverage due to the long tail distribution of debate participants. To validate our pipeline, one author manually identified panelists in 50 videos and compared them to our pipeline's results, achieving a precision of 0.901 and recall of 0.730.

Next, we identified and coded the occupation of the panelists into categories such as TV-related, academics, activist, advocate, analyst, author, civil servant, consultant, doctor, film-related, journalist, politician, religious leader, social leader, and spokesperson. We also coded their affiliations (e.g., political party support).

From the initial set of 285 people identified, 20 were removed as false positives. We only marked individuals who were part of some organization (e.g., Bombay High Court, Samajwadi Party, DMK, BJP, All India Trinamool Congress, THE CHANNEL, Congress) and marked 'None' for others.

4 What is discussed in the debates?

4.1 Bias in transcripts

Existing literature [8, 39] supports the notion that the show exhibits a pro-government stance. Our categorization, summarized in Table 3, aligns with this perspective, revealing a significant 3-to-1 ratio in favor of narratives that support the ruling party. However, unlike previous works, this paper zeroes in further on the *content* of the show to showcase a political tilt, if any. To achieve this, we work with the transcripts and adopt a methodology akin to those in [10, 23], utilizing language models to identify potentially biased attributive/contextual tokens.

Specifically, we train a classifier to determine if a sentence in the transcript pertains to the ruling party or the opposition. This classifier is based on a fine-tuned BERT-Base-Uncased model [9], equipped with a classification head.

For classifier training, we select sentences from the transcripts that explicitly reference the ruling party or the opposition, using specific keywords such as names of parties or leaders (detailed in Table 4). We exclude sentences that mention both to prevent ambiguity. To ensure the model focuses on the context rather than the keywords, we mask the specific keywords, replacing person names with <PER> and party names with <PARTY>.

Given BERT's shortcomings in handling negations [16], we exclude sentences containing negation keywords such as *not*, *won't* etc. Our final dataset comprises 16,444 sentences about the Opposition and 14,865 about the Ruling Party, divided into 80% training, 10% validation, and 10% test sets. The model is fine-tuned for 30 epochs with a batch size of 32, using the AdamW optimizer at a $2e^{-5}$ learning rate.

To make the model's decision-making process more interpretable, we use *integrated gradients* [40], a technique that effectively determines the influence of individual tokens on the model's predictions. This approach helps us pinpoint the tokens that significantly sway the model's judgment in classifying sentences as pertaining to the Ruling Party or the Opposition, in line with Ding et al. [10].

Our classifier achieved an accuracy of 85.72%. For a nuanced understanding, we sorted the words in each category by their average attribution scores across all sentences. After excluding stopwords, infrequently occurring words (less than 50 times), and generic terms to minimize noise, a qualitative analysis of these highly-attributable tokens reveals a distinct bias against the Opposition, while favouring the Ruling Party. The complete list can be found in Table 1. Below, we provide examples to illustrate this qualitatively:

Ruling Party related tokens:
(i) *Election-centric Narratives*: Tokens like 'vote' 'victory' and 'power' suggest a focus on the electoral successes of the ruling party.
(ii) *Veneration of Leadership*: Terms like 'Modi wave,' 'Modi factor,'

Table 1: Words found to be important in the context in sentences involving the Ruling Party and the Opposition. (* indicates that the word was not present in BERT vocabulary and the score is indicative of the word's subtokens. Eg: raf→rafale, par→parivar)

Ruling Party related words						
wave (*0.645*)	hate (*0.635*)	trump (*0.603*)	hatred (*0.595*)	bengal (*0.573*)	factor (*0.517*)	ji (*0.501*)
pm (*0.483*)	model (*0.443*)	cabinet (*0.4*)	voted (*0.397*)	defeat (*0.375*)	riot (*0.362*)	vote (*0.354*)
2019 (*0.354*)	uttar (*0.321*)	kashmir (*0.308*)	rallies (*0.306*)	responsible (*0.269*)	victory (*0.264*)	pakistan (*0.262*)
secular (*0.259*)	development (*0.252*)	power (*0.248*)	democracy (*0.247*)	policy (*0.232*)	poll (*0.231*)	elected (*0.198*)
economy (*0.197*)	farmers (*0.167*)	global (*0.164*)	campaign (*0.156*)	2014 (*0.154*)	security (*0.142*)	credit (*0.133*)
Opposition related words						
indira (*0.772*)	baba (*0.473*)	mother (*0.444*)	dynasty (*0.442*)	rafale * (*0.362*)	apologize (*0.348*)	vatican (*0.344*)
parivar * (*0.327*)	silent (*0.275*)	victim (*0.272*)	questioning (*0.268*)	lie (*0.262*)	age (*0.26*)	italian (*0.257*)
courage (*0.256*)	personal (*0.233*)	exposed (*0.231*)	silence (*0.23*)	concerned (*0.22*)	lobby (*0.209*)	son (*0.207*)
shame (*0.174*)	fake (*0.169*)	brother (*0.168*)	hindus (*0.165*)	secret (*0.161*)	sorry (*0.147*)	evidence (*0.122*)
president (*0.122*)	investigation (*0.121*)	corruption (*0.116*)	communal (*0.101*)	chinese (*0.092*)	xi-jinping * (*0.088*)	failed (*0.087*)

and respectful suffixes like 'ji' (as in 'Modiji') paint a picture of reverence around the party leadership. The term 'development' often co-occurs, framing the ruling party as a catalyst for progress.
(iii) *Defensive and Counter-Narratives*: Surprisingly, words like 'hatred' appear in the context of disputing the notion that animosity towards ruling party is justified. Other tokens like 'Trump' and 'Pakistan' indicate international validation or emphasize a tough stance on national security.

Opposition related tokens:

(i) *Dynastic Politics*: Usage of words like 'dynasty,' and familial references like 'mother-son-sister' aim to cast the main Opposition party in a light suggestive of nepotism.
(ii) *Name-Calling and Stereotypes*: Phrases like 'Rahul Baba,' 'Vadra Congress,' and references to 'lobby' paint the main Opposition party with connotations of naivety & questionable ethics, or disloyalty.
(iii) *Allegations and Scandals*: Terms like 'Rafale,' 'China,' and 'Jinping' are mentioned in contexts that suggest improper or unpatriotic conduct by the main Opposition party. Words like 'fake,' 'shame,' and 'lie' reinforce a narrative of dishonesty and ineptitude.

We also find similar bias in hashtags used for the show. To fetch the hashtags displayed on the screen, we sampled a frame every 30 seconds and extracted text using EasyOCR [31]. The text corresponding to the hashtags was extracted using a regular expression. We see a clear pattern in how the hashtags are chosen: while criticisms of the ruling party tend to be issue-specific and nuanced, criticisms of the Opposition are likely to be sweeping and derogatory, contributing to a narrative that could potentially influence public perception. In debates critical of the ruling party, the hashtags tend to be issue-centric rather than party-centric. For example, hashtags like #WillYogiSackMLA, and #YogiWakeUp focus on individual incidents or politicians and don't necessarily indict the ruling party as a whole. On the contrary, hashtags targeting the Opposition often portray them as either against the country or as disorganized and ineffective. Examples include #CongInsultsDemocracy and #RahulMocksForces, where the use of 'Cong' (an abbreviation for the main opposition party) implies that the entire party is undermining democratic values or the armed forces.

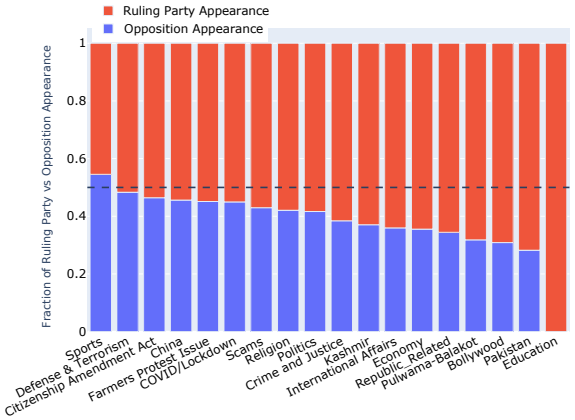


Figure 2: Fraction of panelists invited from the ruling party vs. the opposition. Pro-ruling-party panelists appear more than the opposition in almost all categories.

Further, hashtags like #MamataLosesGrip or #MayaDumpsCong indicate that the opposition parties are fractious and unreliable. The full list of hashtags used in our analysis is shown in Table 5.

By analyzing the affiliations of panelists, whose names were extracted from the transcripts, we observe a discernible bias in the selection process for the show's panelists. As illustrated in Figure 2, there is a disproportionate tendency to invite spokespeople or supporters of the ruling party across various categories.

4.2 Gender Bias

Figure 3 provides a temporal analysis of the gender distribution of faces visible during the debate videos, spanning a period of six years. The data unambiguously shows that females are consistently underrepresented when compared to their male counterparts. This trend is not isolated to specific periods but is persistent across the entire dataset's history.

We further delved into the issue by examining the representation of females in debates across various categories. Figures 4a and 4b highlight the top 5 and bottom 5 categories in terms of female

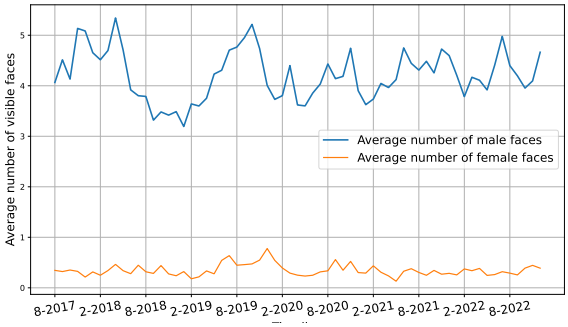


Figure 3: Average number of faces observed when a frame is randomly sampled from a video in the given month. Female guests are consistently underrepresented compared to their male counterparts.

representation, respectively. The data corroborates the presence of systemic gender bias. Notably, there are no categories where females constitute the majority. Although Bollywood-related debates are an outlier, having nearly 40% of the panelists as women, in other categories, female presence is alarmingly sparse. For instance, in critical and often polarizing topics like the Citizenship Amendment Act (CAA) or the Kashmir issue, women make up only about 20% of the panelists. This under representation becomes even more stark in debates about the Pulwama terror attack, where women occupy a mere 5% of the screentime.

In addition to presence, we assessed the screen space allocated to each gender by measuring the average size of visible faces in square pixels. Our findings show that, on average, male faces occupy 3, 798.51 sq pixels, while female faces are allotted only 2, 424.87 sq pixels. This discrepancy is not an isolated occurrence but a consistent pattern over time, as illustrated in Figure 6 in the appendix. The limited screen space for women, even when they are present, underscores the bias.

Our comprehensive dataset of 3,000 videos reveals that women account for a mere 7.5% of the total screen time, which diminishes to 7.2% in political debates. This underrepresentation is stark when compared to the presence of women in Indian politics, where females make up 14.32% of Parliament members, and around 25% of the internet population in India.

As we will discuss in Section 5, there is a correlation between categories with lower female representation and higher levels of incivility. This correlation raises concerns about the quality of discourse and suggests that the gender imbalance may contribute to a more hostile debate environment. It also challenges the inclusivity of media channels in reflecting diverse viewpoints, especially on matters of national and societal significance.

5 Incivility in the Debates

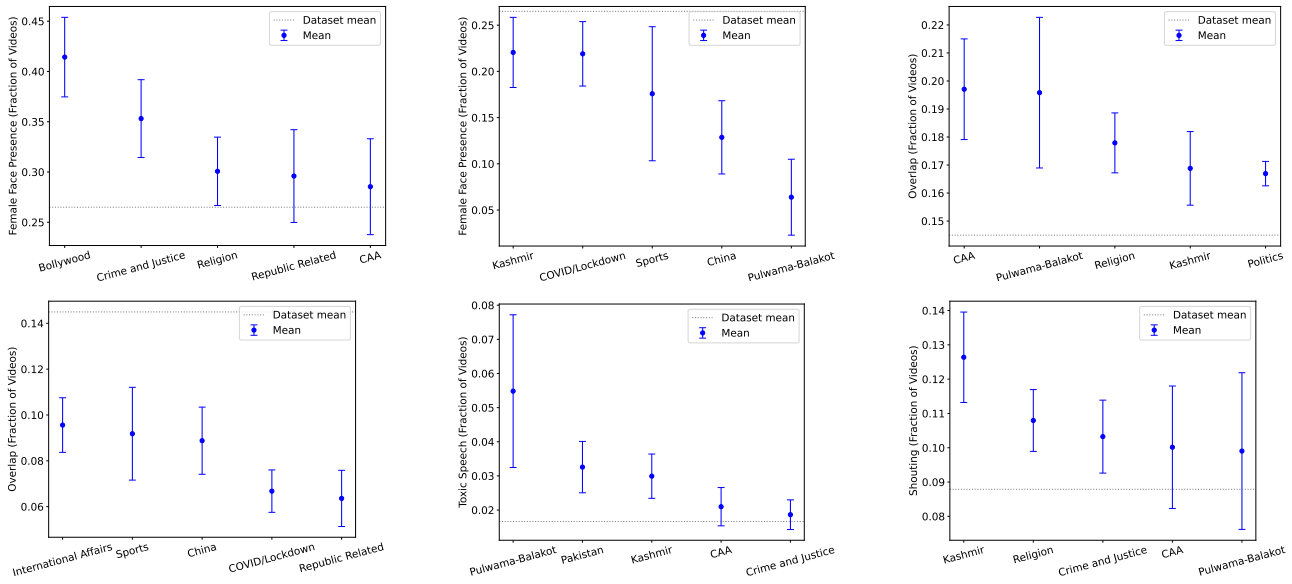
Indian television debates, particularly the one under study, are often marked by high levels of incivility and excessive dramatization, characteristics that can both entertain and polarize the audience. While these traits contribute to the show's popularity, they raise serious questions about the quality of public discourse and democratic debate in the country. In this section, we aim to quantify these elements of incivility using three carefully chosen metrics: *(1) speech overlap, (2) use of foul language, and (3) instances of shouting*. *Speech overlap* acts as a proxy for conversational decorum, with excessive overlap often indicative of a lack of respect for differing opinions. The use of *foul language*, operationalized through detecting hateful language using Google's Perspective API [19], directly reflects the tone and content of the debate, revealing any underlying animosities or prejudices. Lastly, the *frequency of shouting* by the panelists offers insights into the emotional intensity of the debate, potentially correlating with heightened levels of aggression or antagonism. Collectively, these metrics provide a comprehensive lens to quantify and understand incivility in the complex setting of Indian TV debates.

5.1 Overlapping speech and toxicity

The debates often elicit an emotional response from the panelists which either results in (1) panelists speaking over each other or (2) using foul speech to attack others' opinions [14].

To identify **overlapping speech**, we follow the procedure outlined in Section 3.2. Figures 4c and 4d show the top and bottom 5 categories which are significantly over or under the mean respectively. They indicate a pronounced pattern of overlap in specific categories of debates, with particularly elevated levels observed in discussions revolving around contentious issues like the Citizenship Amendment Act (CAA), Kashmir, Politics, and Pulwama-Balakot events [37], as well as Religion. It is striking to note that in debates on the Pulwama terror attack, the CAA, and Kashmir, over 20% of the discourse features overlapping speech. This suggests that these highly contentious issues are divisive and incite a breakdown in conversational decorum. Conversely, we find markedly lower levels of incivility in debates related to International Affairs, COVID-19, the TRP Scam related to THE CHANNEL, Sports, and Bollywood. We next turn our attention to the prevalence of toxic speech, specifically the use of **foul language**, in prime-time news debates. Contrary to what one might expect from a mainstream platform, the presence of toxic speech is not an aberration but rather an unsettling norm. To quantitatively measure toxicity, we employ the Perspective API [19], which assesses text across multiple dimensions including toxicity, identity attack, insult, profanity, severe toxicity, and threat. Our analysis, detailed in Figure 4e, shows that an average of over 1% of the duration across videos in our dataset contain some form of foul language. While this percentage may seem relatively low, it gains significance when considering the show's mass viewership, often in the millions. Most strikingly, the categories registering the highest toxicity levels are those discussing sensitive topics like Pakistan, Kashmir and terrorist attacks. These topics require the most thoughtful and nuanced discussion, yet they have been reduced to shouting matches and verbal attacks.

Elevated levels of incivility (captured through overlap speech and toxic speech) are not just isolated events but indicative of a broader trend that compromises the quality of public discourse. When panelists choose disruption over dialogue, they contribute to a media environment where aggressive and confrontational behaviour becomes the norm rather than the exception.



(d) **(e)** **(f)**

Figure 4: Confidence Intervals. (a) Top-5 categories with more females than average. (b) Bottom-5 categories with less females than average. (c) Fraction of the total duration of videos exhibiting overlapped speech for the top-5 categories, significantly exceeding the dataset's mean. The highest-ranking category has 20% of video duration overlapping speech. (d) Fraction of the total duration of videos with overlapping speech for the bottom-5 categories, significantly below the dataset's mean. (e) Fraction of the total duration of videos with toxic speech in the top-5 most toxic categories. (f) Fraction of the total duration of videos with most shouting in the top-5 categories.

Generalizability: *Though the current study focuses on Indian TV debates, our pipeline is adaptable to other content on the web, specifically to debate shows in English.* To demonstrate its generalizability and establish baselines, we applied our pipeline to four English debate/panel-based shows: The Debate Show (France 24), The Pledge Debates (Sky News, UK), Morning Joe (MSNBC, US), and US Presidential Debates (2008-2020). Our analysis (Figure 5) compared overlapping speech and toxicity in these shows and found that the shows on THE CHANNEL have a statistically significantly higher incivility ($p < 0.01$) than all these shows (refer to Tables 2a, 2b). Refer to Appendix A for details on data collection for other debates and the results.

5.2 Detecting Shouted Speech

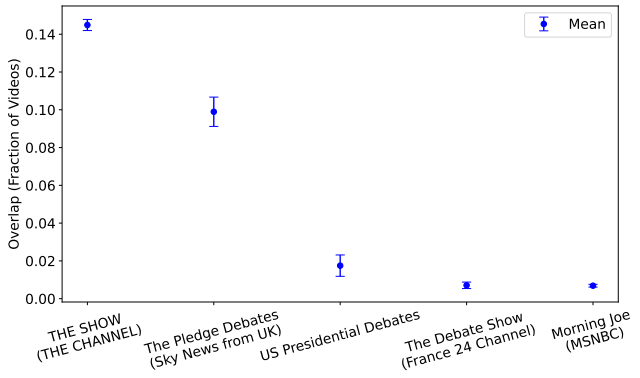
To capture incivility holistically, it is imperative to not just study what is said but how it was said. Shouting is another form of incivility used to overpower others' opinions in a debate. Shouting detection in human speech is an established area of research [26].

The Indian Broadcast News Debate (IBND) corpus [1] contains news debates from THE CHANNEL with annotations for shouted vs. normal speech. We used only the data corresponding to debates held on THE CHANNEL since all our inferences will be performed on samples from the same domain. Using the raw audio from videos in our dataset, we extract 26 MFCCs[5] per frame per audio file, with a frame size of 25ms and a gap of 10ms. On a per-audio level, we perform standard-scaling of these features and group frames into 1 second blocks. Inferences for shouting detection are performed on a per-second level.

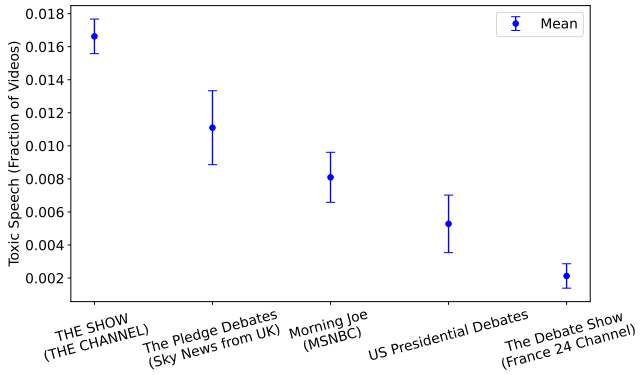
We use a Convolutional Neural Network (CNN) to perform inference on per-second samples. The CNN consists of four blocks. Each block contains a convolutional layer with a ReLU activation function, a max pooling layer for down-sampling, and a dropout layer for regularization and ends with a fully connected layer with a sigmoid activation function for binary classification. The CNN was compiled with the Adam optimization algorithm and binary cross-entropy as the loss function. We tested our approach on the IBND dataset with an 80/20 train-test split, ensuring no data leakage by dividing on a per-audio basis. The model achieved 85% accuracy and, with a high precision of 0.862, was deemed reliable for broader application. A majority voting system for continuous shouting further minimized false positives. The lower recall of 0.71 suggests that shouting instances may be underreported. Manual checks of randomly sampled shouting instances found no false positives. For validation of the classifier's performance, see Appendix B.

Figure 4f shows the average percentage of time shouting occurs in each video, focusing on the top five categories. The complete plot for all categories is included in supplemental pdf. Shouting occupies 9% of the video duration on average, suggesting a notable departure from civil discourse. Categories like Kashmir, Religion, and Crime & Justice are especially prone to high levels of shouting, corroborating the findings in Figures 4c, 4d, and 4e. This level of shouting, particularly in sensitive topics, underscores the emotionally charged nature of these debates. It raises questions about the efficacy of such discourse in fostering meaningful dialogue.

[5]Mel Frequency Cepstral Coefficients (MFCCs) of a signal are features which concisely describe the overall shape of an audio spectral wave.



(a)



(b)

Figure 5: Comparison with other TV debate channels: (a) Fraction of video duration with overlapping speech. (b) Fraction of video duration with toxic speech.

Additional analysis on panelist participation in shouting and incivility is detailed in the Appendix (Sections C.1, C.2). These results further corroborate the extensive presence of incivility and its correlation with debate dynamics.

6 Discussion

Our research employs a comprehensive toolkit, integrating state-of-the-art *open-source tools* in computer vision, speech processing, and NLP, to analyze large quantities of video content. We apply this toolkit to a case study involving one of India's most-watched prime-time television debate shows, which garners over five million daily viewers. The show has received critique for its emphasis on strong nationalistic sentiments and its approach towards minority communities. By making our code public, we aim to encourage further research and analysis in diverse contexts.

Our analysis uncovers significant bias and incivility within the debates, including a notable underrepresentation of women and a bias towards the ruling party. While there has been anecdotal evidence suggesting such biases, our research quantifies these biases. The act of delegitimizing opposition voices has far-reaching implications for the democratic discourse. Our analysis suggests that the use of sensationalism and dramatization may be a deliberate tactic rather than merely a byproduct of the show's popularity. Around 10% of the debate time involves shouting, highlighting an environment that is antithetical to civil discourse.

Television's significant influence on public opinion is concerning when coupled with the biases we've identified [4]. This becomes even more alarming considering that opposition coalitions have started boycotting certain television hosts based on similar criticisms [41], potentially furthering polarization. When millions rely on such a low-quality platform for political insights, the spread of biased information undermines democratic processes and could lead to a misinformed electorate. The high ratings of such shows despite their evident flaws introduce a complex paradox. It challenges the simplistic notion that the media merely reflects public opinion, suggesting that it may play a role in shaping/distorting it.

Overall, our findings offer more than an academic contribution; they signal an urgent call to action. They serve as a critical resource for researchers studying media ethics, democratic governance, and societal polarization. Our work raises complex questions about the ethical responsibilities of media in a democracy and the influence of media on public opinion. These issues warrant investigation and should be of concern to policymakers, civil society organizations, and the public at large.

Limitations: *(i) Scope*: Our study is limited to a single prime-time news debate show and may not apply to more informal content like TikTok videos, which have highly variable discourse quality and nature. (ii) *Manual Annotation*: The need for manual annotation in categorizing videos and identifying panelists limits scalability and could introduce bias. (iii) *Technical Constraints*: Our work is constrained by the accuracy and potential biases of the classifiers, with the risk of compounded errors throughout the pipeline stages.

Ethics Statement: While our toolkit makes large video datasets more tractable for analysis, the potential for misuse is present; for example, the ability to index and search entire video archives could pose significant privacy risks. As with any tool, the ethical implications of its application should be carefully considered according to the use case. Considering that politicians and political analysts are public figures, and taking into account the significance of research in comprehending the language employed in political debates and its consequences, we believe our work conforms to acceptable standards of privacy [11].

Future Work. This study merely scratches the surface of what can be achieved with automated, large-scale analysis of televised debates. We have not fully utilized diarization data due to clustering challenges. While speech embeddings have been tested, they need refinement for practical use. Future work could use diarization for deeper analyses like anchor bias or systemic media bias Overall, while our study has limitations, it offers a pioneering approach to multimedia content analysis, setting the stage for more comprehensive, automated methods in the future.

Acknowledgments

Hitkul is supported by TCS Research Scholar Program. Kiran Garimella's research is funded by grants from the National Science Foundation, Knight Foundation and Google.

References

[1] Shikha Baghel, Mrinmoy Bhattacharjee, SR Mahadeva Prasanna, and Prithwijit Guha. 2021. Automatic Detection of Shouted Speech Segments in Indian News Debates.. In *Interspeech*. 4179–4183.

[2] Doug Beeferman, William Brannon, and Deb Roy. 2019. Radiotalk: A large-scale corpus of talk radio transcripts. *arXiv preprint arXiv:1907.07073* (2019).

[3] Prashanth Bhat and Kalyani Chadha. 2023. Expanding public debate? Examining the impact of India's top English language political talk shows. *Media Asia* 50, 2 (2023), 244–263.

[4] Jay G Blumler. 1970. The political effects of television. *The political effects of television* (1970), 68–104.

[5] Hervé Bredin and Antoine Laurent. 2021. End-to-end speaker segmentation for overlap-aware resegmentation. In *Interspeech 2021*.

[6] Hervé Bredin, Ruiqing Yin, Juan Manuel Coria, Gregory Gelly, Pavel Korshunov, Marvin Lavechin, Diego Fustes, Hadrien Titeux, Wassim Bouaziz, and Marie-Philippe Gill. 2020. Pyannote. audio: neural building blocks for speaker diarization. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 7124–7128.

[7] Broadcast Audience Research Council. 2022. Broadcast Audience Research Council Data 2022. https://www.barc.co.in/

[8] Christophe Jafrelot and Vihang Jumle. 2020. One-Man Show: A study of 1,779 Republic TV debates reveals how the channel champions Narendra Modi. available at: https://caravanmagazine.in/media/republic-debates-study-shows-channel-promotoes-modi-ndtv.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). arXiv:1810.04805 http://arxiv.org/abs/1810.04805

[10] Xiaohan Ding, Michael Horning, and Eugenia Rho. 2023. Same Words, Different Meanings: Semantic Polarization in Broadcast Media Language Forecasts Polarity in Online Public Discourse. *Proceedings of the International AAAI Conference on Web and Social Media* 17 (06 2023), 161–172. https://doi.org/10.1609/icwsm.v17i1.22135

[11] Michael Doherty. 2007. Politicians as a Species of 'Public Figure' and the Right to Privacy. *Humanitas Journal of European Studies ()* 1, 1 (2007), 35–56.

[12] Onaiza Drabu. 2018. Who is the Muslim? Discursive representations of the Muslims and Islam in Indian prime-time news. *Religions* 9, 9 (2018), 283.

[13] GDELT. 2017. GDELT Summary: Television Explorer — api.gdeltproject.org. https://api.gdeltproject.org/api/v2/summary/summary?d=iatv. [Accessed 11-10-2023].

[14] SK Hussain. 2020. The Dirty Game Pro-Hindutva TV Channels And Their Anchors Play — old.indiatomorrow.net. https://old.indiatomorrow.net/eng/the-dirty-game-pro-hindutva-tv-channels-and-their-anchors-play. [Accessed 13-10-2023].

[15] Jungseock Joo, Erik Bucy, and Claudia Seidel. 2019. Automated Coding of Televised Leader Displays: Detecting Nonverbal Political Behavior With Computer Vision and Deep Learning. *International Journal of Communication* 13 (2019).

[16] Aditya Khandelwal and Suraj Sawant. 2020. NegBERT: A Transfer Learning Approach for Negation Detection and Scope Resolution. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 5739–5748. https://aclanthology.org/2020.lrec-1.704

[17] Lev Konstantinovskiy, Oliver Price, Mevan Babakar, and Arkaitz Zubiaga. 2021. Toward automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection. *Digital threats: research and practice* 2, 2 (2021), 1–16.

[18] Raksha Kumar. 2023. How Indian TV news became a theatre of aggression fanning the flames of populism — reutersinstitute.politics.ox.ac.uk. https://reutersinstitute.politics.ox.ac.uk/news/how-indian-tv-news-became-theatre-aggression-fanning-flames-populism.

[19] Alyssa Lees, Vinh Q Tran, Yi Tay, Jeffrey Sorensen, Jai Gupta, Donald Metzler, and Lucy Vasserman. 2022. A new generation of perspective api: Efficient multilingual character-level transformers. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3197–3207.

[20] Bo Li, Dongseong Hwang, Zhouyuan Huo, Junwen Bai, Guru Prakash, Tara N Sainath, Khe Chai Sim, Yu Zhang, Wei Han, Trevor Strohman, et al. 2023. Efficient domain adaptation for speech foundation models. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

[21] Tianming Liu, Hong-Jiang Zhang, and Feihu Qi. 2003. A novel video key-frame-extraction algorithm based on perceived motion energy model. *IEEE transactions on circuits and systems for video technology* 13, 10 (2003), 1006–1013.

[22] Naveen Mishra. 2018. Broadcast Media, Mediated Noise, and Discursive Violence-High Decibel TV Debates and the Interrupted Public Sphere. *KOME: An International Journal of Pure Communication Inquiry* 6, 1 (2018), 1–13.

[23] Shriphani Palakodety, Ashiqur R. KhudaBukhsh, and Jaime G. Carbonell. 2020. Mining Insights from Large-Scale Corpora Using Fine-Tuned Language Models. In *European Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:212412401

[24] Zizi Papacharissi and Maria de Fatima Oliveira. 2008. News frames terrorism: A comparative analysis of frames employed in terrorism coverage in US and UK newspapers. *The international journal of press/politics* 13, 1 (2008), 52–74.

[25] Tae Jin Park, Naoyuki Kanda, Dimitrios Dimitriadis, Kyu J Han, Shinji Watanabe, and Shrikanth Narayanan. 2022. A review of speaker diarization: Recent advances with deep learning. *Computer Speech & Language* 72 (2022), 101317.

[26] Jouni Pohjalainen, Tuomo Raitio, Santeri Yrttiaho, and Paavo Alku. 2013. Detection of shouted speech in noise: Human and machine. *The Journal of the Acoustical Society of America* 133, 4 (2013), 2377–2389.

[27] Sven-Oliver Proksch, Christopher Wratil, and Jens Wäckerle. 2019. Testing the validity of automatic speech recognition for political text analysis. *Political Analysis* 27, 3 (2019), 339–359.

[28] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*. PMLR, 28492–28518.

[29] Benjamin Renoust, Duy-Dinh Le, and Shin'Ichi Satoh. 2016. Visual analytics of political networks from face-tracking of news video. *IEEE Transactions on Multimedia* 18, 11 (2016), 2184–2195.

[30] Filipe Ribeiro, Lucas Henrique, Fabricio Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna Gummadi. 2018. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[31] rkcosmos. 2020. GitHub - JaidedAI/EasyOCR: Ready-to-use OCR with 80+ supported languages and all popular writing scripts including Latin, Chinese, Arabic, Devanagari, Cyrillic and etc. — github.com. https://github.com/JaidedAI/EasyOCR. [Accessed 13-10-2023].

[32] Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 1679–1684.

[33] Parul Saini, Krishan Kumar, Shamal Kashid, Ashray Saini, and Alok Negi. 2023. Video summarization using deep learning techniques: a detailed analysis and investigation. *Artif. Intell. Rev.* 56, 11 (2023), 12347–12385. https://doi.org/10.1007/s10462-023-10444-0

[34] Anirban Sen, Debanjan Ghatak, Gurjeet Khanuja, Kumari Rekha, Mehak Gupta, Sanket Dhakate, Kartikeya Sharma, and Aaditeshwar Seth. 2022. Analysis of media bias in policy discourse in india. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*. 57–77.

[35] Sefik Ilkin Serengil and Alper Ozpinar. 2020. LightFace: A Hybrid Deep Face Recognition Framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 23–27. https://doi.org/10.1109/ASYU50717.2020.9259802

[36] Sefik Ilkin Serengil and Alper Ozpinar. 2021. HyperExtended LightFace: A Facial Attribute Analysis Framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE, 1–4. https://doi.org/10.1109/ICEET53442.2021.9659697

[37] Mohammed Sinan Siyech. 2019. The Pulwama Attack. *Counter Terrorist Trends and Analyses* 11, 4 (2019), 6–10.

[38] Scroll Staff. 2016. Watch: Why Arnab Goswami's shouting worked — scroll.in. https://scroll.in/video/823774/watch-why-arnab-goswami-s-shouting-worked. [Accessed 13-10-2023].

[39] Paul Subhajit and Uttam Kr Pegu. 2021. Media Polarization and Assertion of Majoritarianism in Indian News Media. *The Journal of Communication and Media Studies* 6, 2 (2021), 1.

[40] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic Attribution for Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (Sydney, NSW, Australia) *(ICML'17)*. JMLR.org, 3319–3328.

[41] India Today. [n. d.]. INDIA bloc to boycott shows of 14 TV journalists, media panel condemns move — indiatoday.in. https://www.indiatoday.in/india/story/opposition-bloc-india-bloc-to-boycott-shows-of-14-tv-journalists-bjp-says-bullying-media-2435788-2023-09-14. [Accessed 12-10-2023].

[42] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, et al. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]

[43] Reporters without Borders. 2023. India — rsf.org. https://rsf.org/en/country/india. [Accessed 29-09-2023].

A Performance on other datasets

To demonstrate the generalizability of our pipeline, we applied our pipeline on four more debates hosted in English: The Debate Show (hosted in the France 24 Channel): 216 videos from their YouTube playlist. The Pledge Debates (hosted in Sky News from UK): used YouTube channel videos. Their channel contains both (a) entire

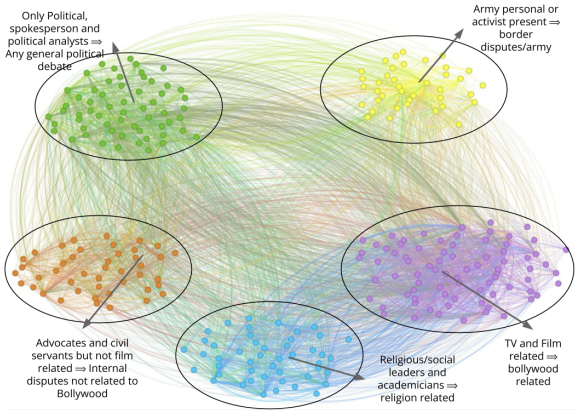


Figure 7: Found five kinds of clusters inside people affiliation coming on debate.

debate videos, (b) smaller snippets from individual debate videos. To restrict our analysis to only videos in (a), we used only those videos which had a duration of more than 20 minutes. Morning Joe (hosted on MSNBC in the US): selected videos from YouTube playlist. To ensure that our analysis focused on the main show, we included videos longer than 30 minutes. After applying this criteria, we were left with a total of 403 videos for our analysis. the US Presidential Debates (from 2008-2020): includes 38 debate videos including the main presidential and vice-presidential debates from 2008-2012 and the intra-party candidate-nomination debates. We did a two-tailed t-test with 95% confidence interval between debates of THE CHANNEL and other debates. We found that the overlap speech in debates in THE SHOW is statistically greater than all the other TV debates mentioned above. Refer to Table 2a for details. Similarly, for toxicity we find that debates in THE SHOW has statistically greater toxicity compared to France 24, US Presidential Elections and Morning show with Joe. Refer to Table 2b for details.

Quantifying Foul Speech Methodology: We processed debate video transcripts, which consist of sequential utterances by different speakers. Using the Perspective API, we assessed each utterance for categories like toxicity, severe toxicity, profanity, insult, threat, or identity attack. Utterances with a probability over 0.5 in any category were marked as foul speech. However, we observed that the API occasionally mislabels factual content as uncivil, such as news-related statements not expressing a panelist's opinion. Therefore, the reported foul speech levels might be marginally overstated.

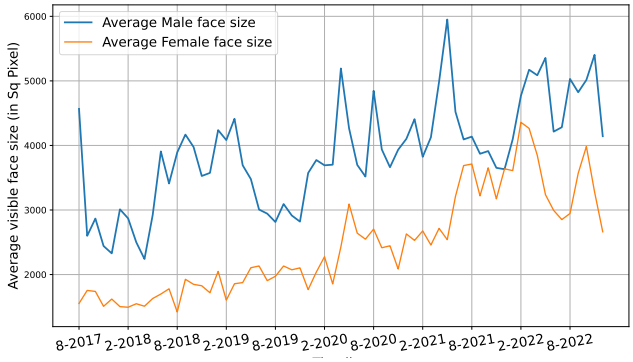


Figure 6: Average size of faces (Males: 3798 sq pixels, Females: 2424 sq pixels)

B Validation Experiments

Validation for classification of speech into shouted/non-shouted categories: We manually annotated 50 audio samples from our dataset, classifying them as shouted or non-shouted. Our classifier achieved a precision of 0.91 and a recall of 0.75. The Indian Broadcast News Debate (IBND) dataset [1], which includes debates from THE CHANNEL with shout annotations, showed our classifier had a precision of 0.86 and a recall of 0.71 on 62,375 test samples. Given the IBND dataset shares our dataset's domain, these results suggest comparable performance on our data.

C Additional Analysis

C.1 Co-attendance Networks of Panelists

Using panelists information we coded in Section 3.4, we created a co-occurrence network between the panelists. If two panelists appeared together in a debate, they were connected by an edge. We found that such a network (shown in Figure 7) was clustered along categories and occupations of the panelists, indicating that the show invites specific panelists based on topics of discussion. The five communities were automatically identified using the Louvain method for community detection. (1) Orange: Found occupation like Advocate, civil servants but not film related occupation: Not related to Bollywood internal disputes (2) Blue: All religious/social leaders and academic people: Something related to religion (3) Pink: All TV and film related people: related to Bollywood (4) Yellow: Army related personal, activists: related to border disputes/army (5) Green: Only politician, spokesperson and analyst: Any general political debate

C.2 Participants involved in shouting

We look at the number of people participating in the shouting. By matching the shouting segments with the diarized text, we identify the speakers who shouted. We wanted to understand whether the debates are being derailed by a small group of people or many panelists have to engage in such behavior to have their voices heard. Figure 8 shows the top five categories ordered by the average number of panelists engaging in shouting along with the number of speakers on average in each category. We find that, roughly half of the participants engage in shouting. It is also important to note that these categories with the highest number of shouting panelists are very different from the results we found in the rest of the figures documenting incivility (Figures 4c, 4d, 4e, and 4f).

Table 2: One-Tailed *t*-test to check difference between distribution of overlap speech and toxicity between THE CHANNEL vs other shows is statistically significant, we report *t*-stat for $\alpha = 0.05$

(a) Two-Tailed t-test for Overlap Speech

Debate	*Mean*	*t*-stat	*p*-value
THE CHANNEL	0.1448	NA	NA
France 24	0.0076	23.2468	5.98-110
Sky News UK	0.0984	5.1663	2.55-07
US Presidential Elections	0.0175	9.9175	8.21-23
Morning show with Joe	0.0069	35.2401	4.41-230

(b) Two-Tailed t-test for Toxicity

Debate	*Mean*	*t*-stat	*p*-value
THE CHANNEL	0.0166	NA	NA
France 24	0.0021	6.9221	5.43-12
Sky News UK	0.0110	1.7640	0.0778
US Presidential Elections	0.0053	2.4801	0.01
Morning show with Joe	0.0081	5.9825	2.44-09

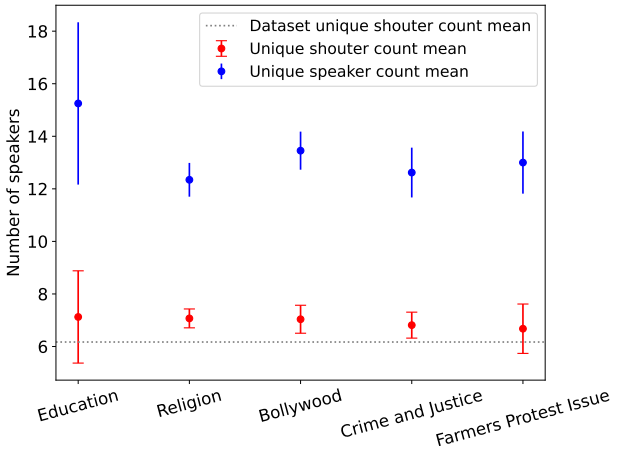


Figure 8: Average count of panelists engaged in shouting (in red) compared to the total panelist count (in blue) for top 5 categories with the highest incidence of shouting. The data indicates that 50% of panelists in these categories participate in shouting behavior.

Table 3: Frequency of various categories

Category	Videos where present as major label	Videos where present as minor label
Politics	1209	739
Religion	216	-
Crime and Justice	190	262
International Affairs	181	128
COVID/Lockdown	181	-
Pakistan	155	47
Bollywood	140	-
Kashmir	134	3
Political Scams	128	-
Citizenship Amendment Act	87	-
Republic TV related	77	-
Economy	76	3
China	58	6
Defense & Terrorism	50	288
Farmers Protest issue	42	-
Pulwama-Balakot	39	-
Sports	29	-
Education	8	-
Anti-Opposition	-	599
State level politics	-	548
Supporting-BJP	-	160
SSR_Case	-	78
Anti-BJP	-	61
Ram Mandir Babri Masjid	-	59
Russia-Ukraine	-	49
Triple Talaq	-	15
Total	3000	

Table 5: Hashtags showcasing the level of scrutiny between videos in Anti-BJP vs Anti-Opposition videos

Hashtags used in Anti-BJP videos	Hashtags used in Anti-Opposition videos
BaggaTweetArrest, YogiWakesUp, GovernorRightorWrong, WillYogiSackMLA, FightForAsifa, SadhviBackGodse, SackBJPBrat, RepublicVsBJPMLA, YogicopsStung, BJPWakeUpCall	CongRapeComment, MayaDumpsCong, CongVsCitizens, ConginsultsDemocracy, ECBansMamata, MamataLosesGrip, AAPForFreebies, KejriwalMinisterArrested, NeechPolitics, VadrasMustGo, RahulMocksForces, CongresslIsOver

Ruling party Specific Words	Opposition Specific Words
modi, narendra, shah, amit, yogi, adityanath, bjp	rahul, vadra, sonia, priyanka, gandhi, kejriwal, congress

Table 4: Keywords