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

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

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

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 multimedia 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 prime-time television debates 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. We will make our data analysis pipeline and collected data publicly available to catalyze further research in this domain.

1 INTRODUCTION

Television debates are a cornerstone of public discourse, serving as platforms for the exchange of ideas and viewpoints. Particularly in India, prime-time debates are viewed by millions and have a substantial impact on shaping public opinion [4]. However, these debates have recently come under scrutiny for compromised journalistic integrity and increasing incivility [22]. Understanding the nuances in these debates is critically 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 main 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, given that perceptions of bias can differ based on individual viewpoints. Similarly, cultural and linguistic nuances like local idioms or specific styles

*Equal contribution. Corresponding authors: anmolagarwal4453@gmail.com, kiran.garimella@rutgers.edu

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 India's most popular prime time news debate show, 'The Debate with Arnab Goswami' which airs on Republic TV[1] (the most watched English language news channel in India) [8]. The show is particularly known for its focus on hyper-nationalistic themes, aggressive attacks on political opponents, and derogatory treatment of minority communities. While there is a prevailing sentiment that the channel overtly supports the ruling party, this claim has yet to be substantiated through quantitative 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.

Furthermore, 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 seek to 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 striking degree of bias in the debate show, characterized by overt support for the BJP and a discrediting stance towards opposition parties and journalists. Furthermore, we identify a significant gender imbalance in the panelist representation, which contributes to a skewed portrayal of societal issues. What is particularly alarming are the high levels of incivility we quantified: approximately 9% of the videos, on average, feature shouting by panelists. These findings have profound implications. The pronounced bias and a lack of dignified discourse not only questions the credibility of the platform as a democratic space for diverse viewpoints but also risks perpetuating political and social divides. 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.

Upon publication, we will 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.

[1]https://www.youtube.com/@RepublicWorld

2 BACKGROUND AND RELATED WORK

2.1 Bias and incivility in 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 [45]. This decline has been partly attributed to the acquisition of media outlets by oligarchs who maintain close ties with political leaders. Such ownership structures have led to evident biases in media reporting, with a majority of television 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.

A case in point is Republic TV, an English-language news channel founded in May 2017 by journalist Arnab Goswami. Since its inception, it has been the most-watched English news channel in India, commanding an average viewership of 40%[8]. Known for its sensationalist approach to news reporting, Republic TV, and its controversial anchor Arnab Goswami, have often been criticized for displaying a pro-Hindu, pro-nationalist, and pro-government bias[40]. One of the channel's flagship programs, "The Debate with Arnab Goswami," epitomizes this tendency. The show attracts over five million daily viewers and is characterized by its hyper-nationalistic tone. It aggressively targets anyone who appears to oppose the government's viewpoint. Despite its status as the most-watched news television show in India, the program has abandoned any pretense of being a credible news debate. Instead, it has opted for a formula rife with overdramatization, shouting, foul language, and overlapping speech [18]. Disturbingly, this sensational approach appears to resonate with viewers [39].

While there is a substantial body of qualitative work addressing bias, factual inaccuracies, and the dramatization of news in Indian media [4, 12, 22], our research contributes by offering quantitative evidence. Notably, some channels, including Republic TV, have even acknowledged their tendencies to sensationalize news. This admission has had repercussions; the main opposition coalition has initiated a boycott against 14 television hosts, including Arnab Goswami, accusing them of disseminating rumors, hate, and false content aimed at opposition parties [42]. Our study enriches this dialogue by supplying empirical data on the nature and framing of the content presented in such debate shows.

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 not just in topic selection but also in the 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 [35]. 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 Multimedia 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].

Despite these advancements, earlier work like that of Beeferman et al. [3] faced challenges in transcribing large volumes of audio data—284,000 hours of radio—due to the limitations in transcription models at the time. This illustrates the speed at which the field has evolved, given that current models for transcription have improved considerably.

Videos present a complex interplay of multiple modalities, including visuals, text, and audio. While each of these 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 modern AI techniques in analyzing large video datasets, offering insights into complex social dynamics.

In comparison, the GDELT Project [13] has provided web-based interfaces for analyzing caption text and other on-screen elements, but it lacks in-depth labeling related to voice tone or specific content being discussed. Our work aims to fill these gaps. We not only analyze a comparable dataset of video but also enrich it by labeling 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 artificial intelligence. 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. In doing so, 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

We extracted the metadata about the YouTube videos corresponding to the debates using the YouTube Data API[2] from the playlist created by the official Republic TV account titled "The Debate with Arnab Goswami - Full Episodes | Republic TV"[3] at the end of December 2022, yielding 3,151 unique videos spanning the entirety of the debates, starting in 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, which corresponded to over 2,087 hours of video content.

The metadata fetched using the YouTube Data API for each video contained the title, url, description, and a list of tags chosen by the channel[4] associated with the video. Some examples of tags associated with the videos can be found in Table 10.

3.1 Categorizing the videos

To categorize the 3,000 videos in our dataset, we initially adopted 18 categories based on a prior study [9]. Utilizing an iterative, inductive coding strategy, 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 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 while potentially belonging to multiple minor ones. The minor categories were created using the same qualitative coding scheme described above, allowing for emergent sub-themes. This nuanced approach allowed us to create a more comprehensive categorization scheme. 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.

The annotation was done by two annotators who were undergraduate computer science majors familiar with the political scene in India. The Fleiss kappa was computed to be 0.933 indicating excellent agreement. In a minority of the cases with disagreements (110 cases), both the annotators discussed among themselves and were able to resolve most of the disagreements. There was no clear agreement on 84 videos which were then removed from further analysis, leaving us with 3000 videos finally.

A complete breakdown of major and minor categories is available in Table 1 . For a more granular understanding, Table 3 in the Appendix maps these categories to their respective tags. Examples of the annotation process are also included in the Appendix A.3.

[2]https://developers.google.com/youtube/v3/docs/playlistItems/list
[3]https://www.youtube.com/playlist?list=PL9iCfxJ0ri311cJAZ0gTDJWDhNlbNbg2h
[4]Details: https://developers.google.com/youtube/v3/docs/videos#snippet.tags[]

3.2 Transcription and speaker diarization

To transcribe the videos in our dataset, we leveraged OpenAI's Whisper speech-to-text model [28], which is noted for its robust performance on diverse accents and technical language. Whisper has also demonstrated near-human-level accuracy in challenging noisy settings [20]. However, transcription was just the first step in our methodology. Debates are complex, multi-speaker environments, making it critical to also perform speaker diarization, a process that partitions an audio stream into segments and attributes them to specific speakers [25].

Before undertaking diarization, we executed two key pre-processing steps to enhance the quality of the results. First, we removed segments devoid of speech, such as interstitials and speaker transitions, using the Voice Activity Detection feature from the Pyannote toolkit [7]. This removal improved the subsequent diarization accuracy. Second, we filtered out overlapping speech segments to avoid performance degradation in speaker clustering. This was accomplished using the same Pyannote model [6].

After these pre-processing steps, we employed the Pyannote diarization module to partition the audio into homogeneous segments, each assigned to a specific speaker [6, 7]. This method, which combines Whisper's transcription with Pyannote's audio segmentation and speaker diarization features, allowed us to transcribe and also accurately attribute speech to individual speakers.

Our qualitative analysis revealed certain limitations in the Pyannote model's overlap detection. Specifically, the model only considered speech as overlapping if all involved audio segments were incoherent. If one speaker's voice dominated, the model did not recognize the speech as overlapping. This issue could result in scenarios where multiple speakers were active, but not identified as such by the model. 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.[5] 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 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.

3.3 Face and gender detection

For facial recognition in our study, we employed the DeepFace library [36], specifically utilizing the RetinaFace detector coupled with the VGG-Face model [37]. From a given video, we sampled one frame every 3 seconds and extracted all the faces from it. One challenge we encountered was the presence of spurious faces, such as those 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. Details about an experiment to validate our model's performance can be found in Appendix B.1.

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

3.4 Extracting panelist names from transcripts

In order to study the people appearing in the debate, we proceeded to extract the names of the panelists from the transcript. Approaches like Named-Entity Recognition (NER) on the transcripts didn't perform well for 3 main reasons: (i) NER was also capturing names of people who are mentioned in the debate but are not panelists, (ii) there were multiple variations in the name used to refer to a person (Eg: [general gd bakshi, g.d. bakshi, mr bakshi, general bakshi, major general gd bakshi]), (iii) due to errors in the transcription, even the same name was spelled differently (Eg: atiqur rahman, atiq-ur-rehman sahab, atiku rehman, ati kaur rehman). So instead, we made use of state of the art large language models (LLMs) for this task.

We found that using Meta's Llama-2 13 billion parameter model [43] In case the transcript for an entire video did not fit in the context length of the model, we chunked the transcript into parts and considered the union of the names extracted from each chunk to be the potential panelists for the video. The prompt used for name extracted can be found in Table 5 in the Appendix. The names returned by this approach are not completely clean and we had to perform fuzzy matching and cluster similar names. The exact process of fuzzy matching and cleaning up the names is given in the Appendix (section A.3.4).

Using the above techniques, we curated a list of 265 people who span over 91.7% videos and whom we estimate to account for 50% of all appearances in the debates in our dataset. Details about an experiment to validate our approach to find panelist names can be found in Appendix B.2. Instead of ensuring full coverage, we opted for a smaller subset due to the natural distribution's long tail of guests invited to debates. Our main interest was understanding popular users who are frequently invited.

Next, we manually identified and coded the occupation of the panelist: TV related, academics, accountant, activist, advocate, analyst, author, civil servant, consultant, doctor, film related, journalist, politician, religious leader, social leader, & spokesperson and affiliation: (e.g. political party support). Since there could be many people which can have the same name we ensured we found the occupation of the panelist who were definitely part of at least one debate of Republic TV. From the initial set of 285 people identified, 20 were removed as false positives. For affiliation, we only marked people who were part of an organisation (eg: Samajwadi Party, DMK, BJP, All India Trinamool Congress, Republic TV, Congress) and marked None for others.

4 WHAT IS DISCUSSED IN THE DEBATES?

4.1 Bias in transcripts

The notion that the show is pro-government is well-supported in existing literature [9, 40]. Our categorization, summarized in Table 1, corroborates this, revealing a 3-to-1 prevalence of pro-BJP narratives. However, unlike previous works, this paper zeroes in further on the *content* of the show to scrutinize its political tilt. To achieve this, we work with the show transcripts and adopt a methodology akin to those in [1, 23], utilizing language models to identify potentially biased attributive/contextual tokens.

Initially, we filter out sentences from the transcript that are explicitly about the BJP or the Opposition based on specific keywords such as party or leader names (see Appendix Table 6). Sentences mentioning both parties are omitted to avoid ambiguity. Next, we mask these specific keywords to allow the model to focus on the surrounding context for its predictions, replacing person names with <PER> and party names with <PARTY>.

We then fine-tune a BERT-Base-Uncased model [10] with a classification head, aiming to predict whether a given sentence pertains to the BJP or the Opposition. Given BERT's shortcomings in handling negations [16], we exclude sentences containing negation keywords (Appendix Table 11). Our final dataset comprises 16,444 sentences about the Opposition and 14,865 about the BJP, divided into 80% training, 10% validation, and 10% test sets. The model is fine-tuned over 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 employ integrated gradients [41], a technique well-suited for ascertaining the influence of each token on the model's output. This serves to identify the most impactful tokens in determining whether a sentence is about the BJP or the Opposition, in line with practices from [1].

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 Congress and the Opposition, while manifestly favoring the BJP. The complete list can be found in Appendix Table 7. We provide a few examples to illustrate this qualitatively.

BJP related tokens: (i) **Election-centric Narratives**: Tokens like 'vote,' 'victory,' 'power,' and 'campaign' suggest a focus on the electoral successes of the BJP. (ii) **Veneration of Leadership**: Terms like 'Modi wave,' 'Modi factor,' 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 BJP 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 BJP is justified. Other tokens like 'Trump' and 'Pakistan' are used to indicate international validation or to emphasize a tough stance on national security.

Opposition related tokens: (i) **Dynastic Politics**: The use of words like 'dynasty,' and familial references like 'mother-son-sister' clearly aim to frame the Congress party as a nepotistic organization. (ii) **Name-Calling and Stereotypes**: Terms like 'Rahul Baba,' 'Vadra Congress,' and references to the 'lobby' contribute to an image of Congress as immature, corrupt, or even treacherous. (iii) **Allegations and Scandals**: Tokens like 'Rafale,' 'China,' and 'Jinping' are often used to impute unethical or unpatriotic behavior to the Congress. Words like 'fake,' 'shame,' and 'lie' further this narrative of deceit and incompetence.

We also find similar bias in hashtags used for the show. In order to fetch the hashtags displayed on the screen, we sampled a frame every 30 seconds and extracted text from it 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

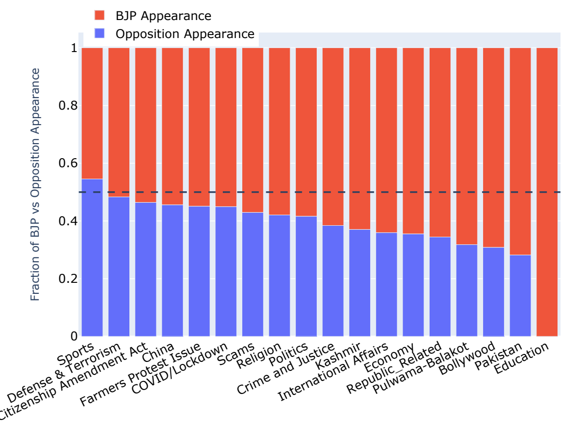

Figure 1: Fraction of panelists invited from the BJP vs. the opposition. Pro-BJP panelists appear more than the opposition in almost all categories.

chosen: while criticisms of the BJP tend to be issue-specific and nuanced, criticisms of the Opposition are more likely to be sweeping and derogatory, contributing to a broader narrative that could potentially influence public perception.

For debates that are critical of the BJP, the hashtags tend to be issue-centric rather than party-centric. For example, hashtags like #WillYogiSackMLA, and #YogiWakeUp are focused on individual incidents or politicians and don't necessarily indict the BJP 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 Congress) implies that the entire party, represented by its President Rahul Gandhi, is undermining democratic values or the armed forces. Further, hashtags like #MamataLosesGrip or #MayaDumpsCong indicate that the opposition parties are fractious and unreliable. The full list of hashtags used in our analysis are shown in Table 12 in the Appendix.

Even just looking at the number of panelists invited to the show has a significant bias, as shown in Figure 1. In most categories a larger number of BJP spokespeople or BJP supporters are invited.

4.2 Gender Bias

Figure 2 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 a persistent feature across the entire dataset's history.

We further delved into the issue by examining the representation of females in debates across various categories. Figures 3 and 4 highlight the top 5 and bottom 5 categories in terms of female 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 somewhat of an outlier, featuring women as nearly 40% of the panelists, in most other categories, female presence is alarmingly

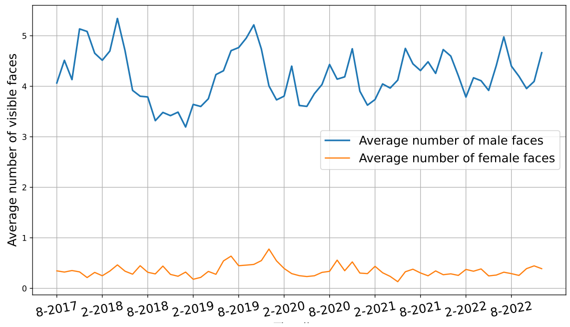

Figure 2: Average number of faces observed when a frame is randomly sampled from a video in the given month

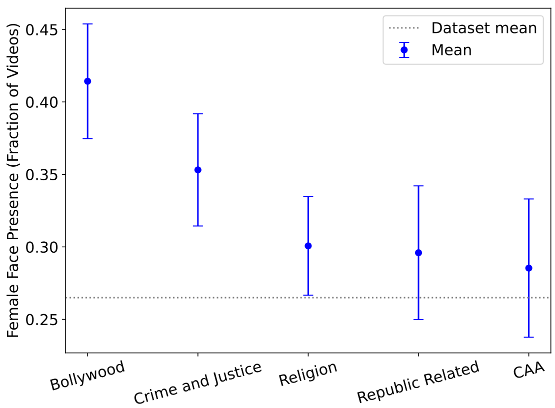

Figure 3: Top-5 categories with more females than average.

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 are present in a mere 5% of the debates.

Similarly, we also computed the screen real estate provided to men and women. We measured the average visible size in square pixels for men and women faces. On average, a male face was allocated 3798.51 sq pixels whereas a female face had only 2424.87 sq pixels. The trend also persists over time (see Figure 16 in the appendix). Even when women are present in a debate, there is a significant difference in the space given to them. This difference is consistent across time. In our dataset of 3,000 videos, women accounted for just 7.5% of the total screen-time, which drops even further to 7.2% in political debates. This suggests significant under representation of women on the show compared to their representation in Indian politics. For context, in India's Parliament 14.32% of the members are female and women represent around 25% of the internet population.

As seen in Section 5, categories with less female representation tend to exhibit higher levels of incivility. This raises questions about the quality of discourse and whether the current gender imbalance contributes to a toxic debating environment. It also calls into question the inclusivity of channels in representing diverse perspectives, particularly on issues of national and social importance.

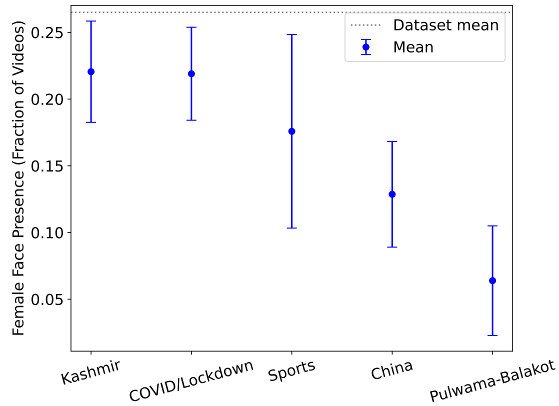

Figure 4: Bottom-5 categories with less females than average.

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: speech overlap, use of foul language, and instances of shouting. Speech overlap serves 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 through which 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 5 and 6 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 [38], 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 not only divisive but also incite a breakdown in conversational decorum. Conversely, we find markedly lower levels of incivility in debates related to International Affairs, COVID-19, the Republic TRP Scam, Sports, and Bollywood.

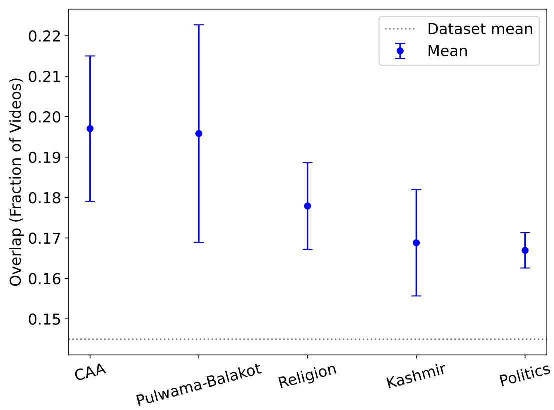

Figure 5: Fraction of videos exhibiting overlapping speech across the top-5 categories, significantly exceeding the dataset's mean. The highest-ranking category contains around 20% of videos with overlapping speech.

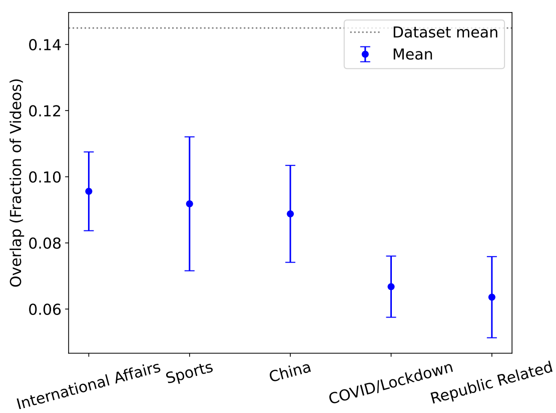

Figure 6: Fraction of videos exhibiting overlapping speech across the bottom-5 categories, significantly below the dataset's mean.

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 7, shows that an average of over 1% of the videos in our dataset contains 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 in Kashmir.

The implications of these findings are both urgent and far-reaching. The elevated levels of incivility (captured both 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

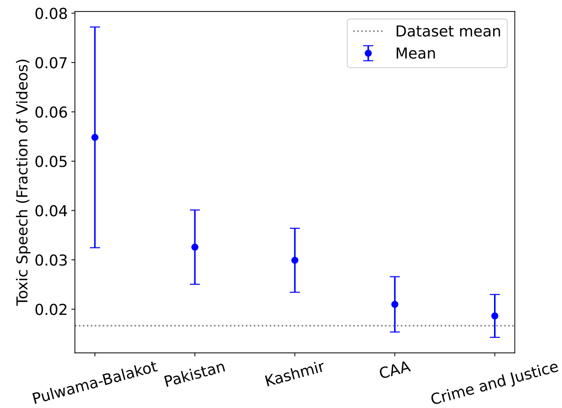

Figure 7: Fraction of videos with toxic speech in the top-5 most toxic categories: The highest-ranking category contains over 5% of videos with toxic speech.

to a media environment where aggressive and confrontational behavior becomes the norm rather than the exception. Particularly concerning is the elevated level of toxicity in discussions centered on sensitive geopolitical topics like Pakistan and Kashmir. These are the topics that require the most thoughtful and nuanced discussion, yet they are being reduced to shouting matches and verbal attacks. **Generalizability:** Though the current study focuses on Indian TV debates, our pipeline is adaptable to other multimedia 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 compared overlapping speech and toxicity in these shows and found that the shows on Republic TV have a statistically significantly higher incivility ($p < 0.01$) than all these shows. See Appendix C for data collection and results details.

5.2 Shouting detection

Finally, it is important not just to look at what was said but how it was said to capture incivility. For this, we use the detection of shouting. Shouting is another form of incivility used to overpower another opinions in a debate. Shouting detection in human speech is an established area of research in speech processing [26].

The Indian Broadcast News Debate (IBND) corpus [2] contains news debates from Republic TV along with annotations for shouted vs. normal speech. We used only the data corresponding to debates held on Republic TV, since all our inference will be performed on samples from the same domain. We obtained all the raw audio for the videos in our dataset and from each audio file we extract 26 MFCCs per frame,[6] with a frame size of 25ms and a gap of 10ms. On a per-audio level, we perform standard-scaling of these features and grouped frames in blocks of one second. Inferences for shouting detection are performed on a per-second level.

We used a Convolutional Neural Network (CNN) for performing inference on per-second samples. The CNN consists of four blocks.

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

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 evaluated this approach 80/20 train-test split on the IBND dataset. We took special care to split the train/test data on a per-audio basis instead of per-sample to prevent data leak. Our model gets an accuracy of 85%. Since the model has high precision (0.862), we felt comfortable applying it to the rest of the dataset. Additionally, the majority voting scheme used to identify continuous segments of shouting further reduces the number of false positives. Given the low recall (0.71), our results should be interpreted as under reporting the prevalence of shouting. We then applied our model on the entire dataset to identify instances of shouting. We also manually sampled a few dozen examples and checked the utterances identified as shouting. We did not find any false positives. Details about an experiment to validate this classifier's performance can be found in Appendix B.3.

Figure 8 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 the Appendix (Figure 14). Astonishingly, 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 5, 6, and 7. This prevalence 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 public dialogue and suggests that the show may be prioritizing sensationalism over substantive discussion.

We also performed additional analysis on the number of panelists participating in the shouting and the prevalence of incivility conditioned on the participants in a debate. We omit the findings due to space contstraints but refer the reader to Appendix (Sections A.8, A.9) for the results. The results strengthen our findings on the wide spread prevalence of incivility and the role of the debate setup (e.g. panelists makeup) in this process.

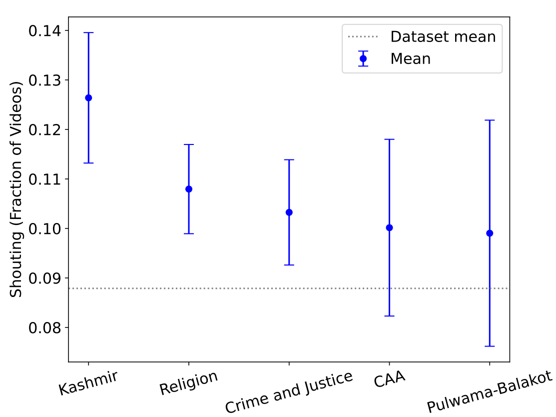

Figure 8: Fraction of videos with most shouting in the top-5 categories.

6 DISCUSSION

Our research employs a comprehensive toolkit, integrating state-of-the-art 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 been criticized for its focus on hyper-nationalism and its tendency to disparage minority communities. By making our code public, we aim to encourage further research and analysis in diverse contexts. Notably, the pipeline is designed to be language-agnostic, although the accuracy of speech-to-text components may vary for non-English languages.

Our empirical findings reveal alarming levels of bias and incivility in the analyzed debates. The data indicates a stark under representation of women and a clear skew in favor of the ruling party. While there has been anecdotal evidence suggesting such biases, our research quantifies these biases, lending empirical weight to existing criticisms. The act of marginalizing or delegitimizing opposition voices has far-reaching implications for the democratic discourse. This raises crucial ethical questions concerning the media's responsibility in a democratic society. Furthermore, our analysis unveils that sensationalism and dramatization are not just a part of the show's appeal but seem to be a calculated strategy. Astonishingly, around 10% of the debate time involves shouting, highlighting an environment that is antithetical to civil discourse. The potential impact of this dramatization on mainstreaming extreme opinions should not be underestimated.

Media, particularly television plays an important role in shaping public opinion [5]. The biases we quantify in this paper make this role particularly crucial and worrying. This becomes even more alarming considering that opposition coalitions have started boycotting certain television hosts based on similar criticisms [42], potentially furthering polarization. The low quality of a widely-watched television debate is not just concerning but potentially dangerous. When millions rely on such a 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 media merely reflects public opinion, suggesting that it may also play a role in shaping or even 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. Importantly, our work raises complex questions about the ethical responsibilities of media in democratic societies, the influence of media on public opinion, and the paradox of public endorsement of biased or uncivil media content. These issues warrant further investigation and should be of concern to policymakers, civil society organizations, and the public at large.

Limitations. (i) Scope of Analysis: Our study focuses on a single, prime-time news debate show. The approach may not generalize to less structured content, such as random TikTok videos, where the quality and nature of discourse can vary dramatically. (ii) Manual Annotation: A considerable amount of manual labor was involved in annotating video categories and identifying panelists. This makes the process less scalable and potentially introduces human bias. (iii) Technical Constraints: The study is subject to the limitations of the classifiers employed, including their accuracy and the biases they might inherently possess or propagate. These could affect the quality of the analysis, especially in cases where multiple errors accumulate across different stages of the pipeline.

Ethics Statement. While our toolkit makes large video datasets more tractable for analysis, there are ethical considerations to bear in mind. 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 the fact 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 are of the opinion that our work conforms to acceptable standards of privacy (as defined in [11]).

Future Work. This study merely scratches the surface of what can be achieved with automated, large-scale analysis of televised debates. Specifically, we have yet to fully exploit the diarization data due to technical challenges in clustering similar users. Although we experimented with speech embeddings for this purpose, the technique requires further refinement to be effective in practice. In future, the diarization data could be employed for more nuanced analyses, such as examining anchor bias or other forms of systemic bias within the media landscape. 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.

REFERENCES

[1] 2023. Same Words, Different Meanings: Semantic Polarization in Broadcast Media Language Forecasts Polarity in Online Public Discourse. 17 (Jun. 2023), 161–172. https://doi.org/10.1609/icwsm.v17i1.22135

[2] 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.

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

[4] 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.

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

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

[7] 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.

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

[9] 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.

[10] 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

[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] seatgeek. [n. d.]. thefuzz. https://github.com/seatgeek/thefuzz.

[35] 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.

[36] 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

[37] 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

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

[39] 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].

[40] 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.

[41] 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.

[42] 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].

[43] 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]

[44] James Turk. [n. d.]. GitHub - jamesturk/jellyfish: A python library for doing approximate and phonetic matching of strings. https://github.com/jamesturk/jellyfish. [Accessed 11-10-2023].

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

A APPENDIX

A.1 Tags

The tags fell into 2 types: **(Type A)** Tags which provided valuable signals about the topic being discussed in the video such as 'Sushant case CBI', 'Pulwama grenade attack', 'covishield vaccine'; and **(Type B)** Tags such as 'Republic TV exclusive', 'breaking news', 'arnab goswami projections', '8th dec with arnab' which were generic and unhelpful to comment on the topic being discussed in the video. Some examples of tags on videos are shown in Table 10.

A.2 Categorization of Videos

Finding the Categories: The debates are based on a variety of different topics such as Politics, Crime & Justice, International Affairs etc. Economy etc., as has already been documented in [9]. There are videos whose topic of discussion fell into multiple categories such as: the debate titled "Can Congress Question Centre On Fuel Prices Anymore?" is related to both the Economy but also has a political party involved in it. Moreover, we also wanted to capture whether a particular debate was "Anti-Opposition", "Anti-BJP" or "Supporting-BJP" in nature. To find the categories relevant to each debate, the following process was used:

- In order to utilize the tags present in the video metadata to map the videos to different categories, we mapped each tag T to a category C with the assumption that if a tag T present in the tags of a video, then the video is likely to have content corresponding to category C. Tags of Type-B were simply filtered out whereas tags of type A were mapped to different categories.
- To make sure that each category has sufficient set of videos for analysis, we **merged some categories under a common umbrella category**. For eg: (1) The videos corresponding to 'Sabarimala Case' and 'Triple Talaq' were grouped under Religion. Likewise, the videos corresponding to 'Regional Elections' and '2019 Elections' were grouped under Politics.
- After using the above, we were still left with 830 videos to which no category was assigned as all their tags belonged to Type-B. However, the categories for some of these videos were easily identifiable using their title. To take advantage of this, we decided to use **OpenAI's "gpt-3.5-turbo"** to assign categories for the videos based on their titles. The corresponding prompt can be found in **Table 4**.
- The final set of categories assigned to the videos was a union of the categories obtained by (1) the tag-based-mapping and (2) prompting **OpenAI's "gpt-3.5-turbo"**.

Segregating the categories into major and minor categories: In order to account for granular aspects in the content of the debate, for each video, 2 sets of categories were assigned: **Major Category** which captured the main focus of the content discussed in the debate; and **Minor Categories** which were also discussed/represented in the video but were not the most-important focus. For each video, there can only be one major label i.e. it is a singleton. For example: A video (such as "Donald Trump Praises Modi, Opposition Gets Heartburn" [7]) can be related to International Affairs as the major label but can also depict the Opposition in a bad light. When a video had multiple categories assigned to it, the major category was assigned to be the one which came higher in the preference list shown in Table 2 and the remaining categories were retained as minor category.

[7] https://www.youtube.com/watch?v=y0U5r8AuE6w

Table 1: Category Frequencies

Category	Major Label	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	

A.3 Annotation Methodology

A.3.1 Objective. We need to annotate whether the major and minor category labels per video are correct or not.

To annotate a video, your main focus would be the Title and hashtags provided in the sheet itself. If these signals aren't sufficient then you can proceed to check the video's first 5 minutes where the news anchor gives the synopsis of the debate. But watching video should be kept at a lower priority since annotation time might increase.

A.3.2 For Major Category. Below is the list of major categories that you will find during annotation. They are ordered from specific to more generic labels.

Table 2: Priority List used to select the major category when a video was mapped to several categories

Ram Mandir Babri Masjid
Farmers Protest Issue
Citizenship Amendment Act
SSR Case
Pulwama-Balakot
Kashmir
COVID/Lockdown
Republic TRP Scam
Scams
Russia-Ukraine
China
Pakistan
International Affairs
Economy
Supporting-BJP
Anti-Opposition
Anti-BJP
State level politics
Religion
Defense & Terrorism
Education
Sports
Bollywood
Crime and Justice
Politics
Miscellaneous

So for example, if the video talks about the CAA act and is political in nature, we will give preference to the CAA label since it is more specific. Politics will be part of the minor label.

If you find the major label assigned to be incorrect then mark the cell in red and in the additional comment write the suitable category you think should be present.

A.3.3 *Deciding whether a category is suitable OR not*.

A.3.4 Fuzzy matching and clustering names. While within the transcript of a video, the LLM was fairly robust to errors, we found that across different videos, there was still some variation in name for the same individual. We started with a seed set of most frequently occurring names (representing one cluster each) and iterated through all others names to see whether they were a fuzzy match for any of the names in the seed set. If a name was not a close match to any of the names in the seed set, the name was added as a new cluster to the seed set; else it was added to the cluster of the name it was fuzzy matched to. This also helped expand our seed set. We used a combination of 2 matching algorithms: (1) “Partial Token Sort Ratio” [34] (which helped match names where one name had more tokens than the other, eg: general bakshi, major general gd bakshi), (2) metaphone based matching [44] (which helped match names which had the same pronunciation but different spellings, eg: “syed asad abbas”, “sayyad asad abbas”, “sayyed asad abbas”, “syed assad abbas” all of which have the same metaphone: “SYT AST ABS”).

A.4 Bias in transcripts

Table 6 shows the list of keywords we used containing politicians and political parties.

Table 11 shows the list of negation words.

Table 7 shows the full list of keywords.

A.5 More plots related to Incivility, Overlap, Foul and gender bias

Figures 9, 12 show additional plots for all incivility.

A.6 Hashtag Bias

Table 12 shows the bias in the use of hashtags. These hashtags were extracted from the videos using OCR.

A.7 Gender Bias

Figure 16 shows the trends in the average screen real estate allocated to male and female faces over the years. We can clearly see that there is a significant gap — even if a woman is present on the screen, they are not given enough space.

Figure 17 shows the fraction videos with female faces in all categories.

A.8 Network analysis of panelists

Using the information of the panelists we manually 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. Perhaps not so surprisingly, with just this information, we found that such a network (shown in Figure 18) was clearly clustered along categories and occupations of the panelists, indicating that the show invites specific panelists based on specific topics of discussion over and over. The five communities were automatically identified using the Louvain method for community detection and correspond to topics like General politics, Religion, Bollywood, and Army related issues.

Orange: Found occupation like Advocate, civil servants but not film related occupation -> Not related to Bollywood internal disputes
Blue: All religious/social leaders and academic people -> Something related to religion
Pink: All TV and film related people -> related to Bollywood
Yellow: Army related personal, activists -> related to border disputes/army
Green: Only politician, spokesperson and analyst -> Any general political debate

A.9 Number of participants in shouting

Next, we look at the number of people participating in the shouting. By matching the shouting segments with the diarized text, we can identify the speakers who participate in the shouting. The idea

Table 3: Table showing some categories and a subset of their corresponding mapped tags

Category	Tags
China	Tiktok banned | Xi Jinping | Sonia Gandhi Xi Jinping deal | modi xi jinping | xi jinping in india | XI jinping in Mamallapuram | Bharat vs China | Boycott China | india china international news | india china border news | China clashes | China Tawang border dispute | Chinese Apps TikTok blocked | LAC dispute
Ram Mandir Babri Masjid	babri masjid | ayodhya verdict | ram mandir ayodhya fight | subramanian swamy on ram mandir | tasleem rahmani on ayodhya | CBI Babri Masjid | ayodhya land settlement | Ayodhya case settlement | ram mandir 2025 | ayodhya march | cji impeachment linked to ayodhya | Ram Mandir politics | Ayodha mediation panel
Covid	China virus | Covishield | Vaccine in India | covishield | rahul gandhi questions vaccine | Is Covishield safe | Vaccine registration | 21 day lockdown | lockdown extension | coronavirus disease | covid china | covid origin | corona in india social media | pm modi on covid vaccine | lockdown violators | Coronavirus India lockdown | lockdown 21 day | 21 days lockdown | Modi coronavirus lockdown | Coronavirus warriors attacked
Pulawama Balakot	major gaurav arya on pulwama | pulwama terror attack | surgical strike | gen gd bakshi on surgical strike | Pulwama attack | avenge pulwama | pm modi on surgical strike | Surgical Strike
Farmers Protest	farmer protest | Farmer Bill 2020 | govt farmer talks | Centre Farmer talks | Farm Bill 2020 | Farm Bills explained | Three Farm Bills | sc hearing on farm law | haryana farmers protest live | haryana farmers protest latest | haryana farmers protest study iq | haryana farmers protest lallantop
Bollywood	Deepika Padukone | Sara Ali Khan | bollywood drug party | kangana ranaut interview | ncb summons bollywood | Aryan Khan drugs case | richa chadha on indian army | me too campaign
Republic TRP Scam Case and Other Republic Related	FIR against Republic TV | Param Bir Singh case | fake trp | Param Bir Sachin Vaze secret meet | Sadhvi Pragya exposes Param Bir Singh | cbi fir trp scam case | arnab goswami trp scam | trp scam cbi | Param Bir Extortion racket | Republic Editorial Staff FIR
Economy	budget 2019 | nirmala sitharaman budget | union budget 2022 highlights | PM modi Economy push | rahul gandhi on fuel price | demonetization india | pm modi on fuel prices | arun jaitley gst | Make in India
2019 Elections	elections 2019 | 2019 lok sabha elections | who will win 2019 | election opinion poll 2019 | modi back in 2019 | modi sweep back in 2019 | manifesto 2019 | rahul gandhi 2019 manifesto | 2019 election survey | third front 2019 | bjp vs congress 2019
Sabarimala Case	Politics on Sabarimala | sabarimala updates | sabarimala temple issue | womens entry in sabarimala | sabarimala verdict supreme court | trupti desai on sabarimala | rahul easwar sabarimala debate | sabarimala protest pandalam

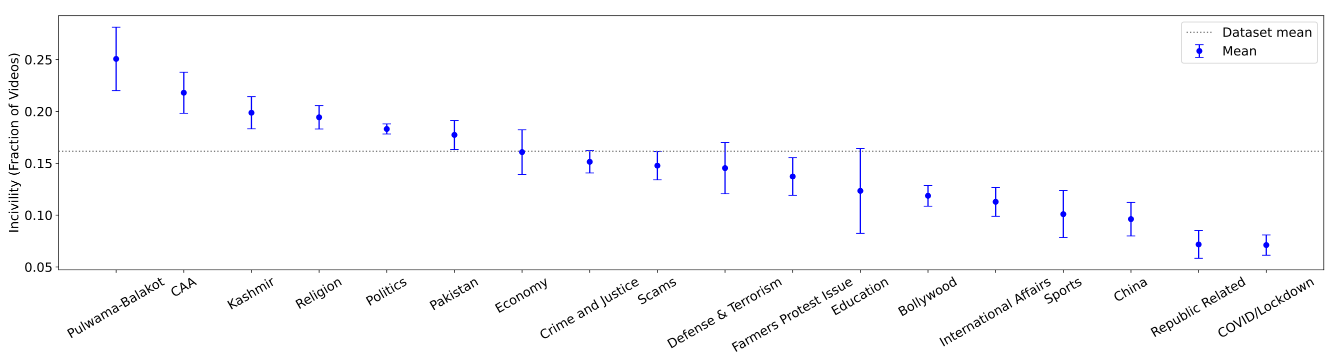

Figure 9: Incivility for all categories.

behind this line of analysis is to understand whether the debates are being derailed by a small group of people or if most of the panelists have to engage in such behavior to have their voices heard. Figure 19 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 surprisingly, most categories 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 5, 6, 7, and 8).

Table 4: Prompt fed to LLAMA-2-13b-chat to fetch names from the transcriptGPT-3.5-Turbo to fetch the categories for the videos

Please map the below debate titles for debates conducted by Arnab Goswami of Republic TV into different topics based on the title of the debates. One debate may belong to multiple topics.

List of topics:- [Politics, Supporting-BJP, Anti-BJP, Anti-Congress, State level politics, Scams, Religion, Economy, Defense & Terrorism, International Affairs, COVID/Lockdown, Bollywood, Farmers Protest Issue, Ram Mandir Babri Masjid, Citizenship Amendment Act, Kashmir, Pakistan, Pulwama-Balakot, Crime and Justice, Sports, Education]

Here is some additional context about topics:-

- Supporting-BJP: Debate titles that shows positive bias against BJP or BJP affiliated leaders. Eg: Why The Yogi Victory Is Historic 2022 Election Results.
- Anti-BJP: Debate titles that shows negative bias against BJP or BJP affiliated leaders. Eg: Shouldn't UP CM Yogi Adityanath Rise Above Namecalling, Barbaric Cruelty' Allegation Against Seema Patra - BJP Suspends But No Arrest,Is Modi Wave On The Wane.
- Anti-Congress: Debate titles that shows negative bias against Opposition such as Congress or criticisms about the Congress affiliated leaders like Vadra and Gandhis. Eg: Will Priyanka Vadra Take Responsibility For Stampede At Children's Rally, PM Modi Sets Vision For 2047, Will Opposition Rise Above Petty Politics.
- Politics: Debates including day-to-day political developments with no broader themes and political debates with no clear bias in the title. Eg: Centre Launches 'Agnipath' For India - Is It A Gamechanger Or Not, Kharge Compares PM Modi To 'Ravan'. Will The Insult Politics Backfire
- State level politics: Debate titles with any mention of an Indian state or a state leader. Eg: Eknath Shinde Stakes Claim On Shiv Sena, Is MVA Govt's Time Up,Sanjay Raut Abuses Shinde Camp; Uddhav Thackeray Extends Olive Branch To Rebels, Tamil Nadu BJP Chief K Annamalai Speaks To Arnab, Threatens 'Mass Satyagraha' Over Fuel Price.

For each debate title and id, perform the following actions:

1 - Read the debate title

2 - Iterate through all topics in above list of topics and determine whether topic is relevant to the debate

3 - Keep track of all relevant topics to the debate. Note that some of the debates may be relevant to multiple topics. In such cases, include all the topics relevant.

Separate responses for separate debates with a newline. Provide them in JSON format with the following keys:

- debate_id: <debate ID>,
- debate_title: <debate title>,
- topics_list: <list of topics mapped to>,
- reasoning: <Reasoning behind mapping>

Debate titles list (each debate on a new line with format: <debate ID>| <debate title>):

AM4jNtwHX8k | Did Rahul Gandhi Link PM Modi To "Hindu Terror"

UK4WtrqwD1U | PM Modi Sold 'Chai' & Not The Nation

a2GaKNHfLz0 | BIGGEST EXPOSE: Church Campaign Against Nationalism

lk-pPzP8K2g | Woman Misses Funeral Due To VVIP Racism

lbliEqB27nY | Rahul Gandhi To Lead The Congress Party

A.10 Incivility based on participants

In our analysis, we delved into the relationship between panelist participation and the level of incivility in the debates. Specifically, we were interested in understanding whether the presence of certain types of panelists could be linked to increased incivility. To examine this, we utilized the annotations identifying each panelist's political party affiliation. We focused on key opposition parties, such as Congress, All India Trinamool Congress (TMC), Aam Aadmi Party (AAP), DMK, and the Samajwadi Party, to create triads of participants for each video. In these triads, at least one member was affiliated with the BJP, and the remaining two were from opposition parties, or vice versa.

To ensure robust findings, we filtered out triads that occurred less than 50 times across all videos. We then calculated the average incivility for videos featuring these triads and found that the incivility was statistically higher ($p<0.05$) than the dataset's overall average incivility of 0.162. Notably, the triads comprising panelists from

Table 5: Prompt fed to LLAMA-2-13b-chat to fetch names from the transcript

<s>[INST] «SYS»

You are a helpful, respectful and honest assistant. Always answer as helpfully as possible, while being safe. Your answers should not include any harmful, unethical, racist, sexist, toxic, dangerous, or illegal content. Please ensure that your responses are socially unbiased and positive in nature.

If a question does not make any sense, or is not factually coherent, explain why instead of answering something not correct. If you don't know the answer to a question, please don't share false information.

«/SYS»

Below within backticks is the transcript from a TV debate in India. Extract the names of the speakers and panel members from the transcript. HINT: the speakers are generally introduced in the beginning of the transcript by the debate moderator. Try using this introductory statement to find the names. NOTE: Output the names as with numbering on new lines. In case the transcript has foul language, ignore the inappropriate language and personal attacks and focus only on finding the names.

Transcript:

“‘Mumbai Studios of Republic TV. It's time for Arnab Goswami on the debate.

<speaker change>

Good evening and welcome viewers. I don't believe in coincidences. The CBI director hates his number two and is fighting to stop him from taking over when he retires in a few weeks. The lawyer Prashant Bhushan steps in and challenges the appointment of the number two in the Supreme Court.

This means that the CBI director obviously loves this and would love what Prashant Bhushan is doing for him. The CBI director then meets Prashant Bhushan and two people denied a birth in the Modi government. The permanently sulking Yashwant Sinha and Arun Shourie and Rafal. Viewers I'm asking you this, is this a coincidence?

I may not name them viewers but now we all know who the lobby is. The lobby wants to engineer trouble within our country. Caste wars, they failed. Breaking down the Supreme Court, they failed. Dividing the Hindu religion, they failed. Destroying the agencies from within, they are trying. They will fail. It is our fundamental duty viewers, yours and mine, to make sure that they fail this time as well. We have to fight the lobby.

<speaker change>

....

<speaker change> And if you will permit, Arnab this is basically, Arnab is, they tried it here, now they have hatched out to the CBI, hatched out to IP, for his role in the Godhra investigation, or his role, his question, his role in Gujarat came under huge question, his role in the Godhra investigation, his role in the Amit Shah case, so and then to bring him here, Mr. Modi lost the plot, he thinks, he thought he could run everything, but he can't, he has complete control of governance, of policy, they don't know the government can't do any policy, so please do not bring in this whole thing about casting us questions, if it's to our five investigative agents, in which we have been saying all the time, we are not even talking about policy, which they don't understand a word, but this is simple administrative issue.“‘

[/INST]

Table 6: Keywords

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

opposition parties were associated with higher levels of incivility. Table 13 shows the results.

These results suggest that the composition of the panel, particularly the presence of opposition party representatives, may be a significant factor in driving incivility in these debates. This raises questions about the dynamics at play during the debates and how they may be influenced by the panelists' political affiliations. It suggests a need for further scrutiny into whether the heightened incivility is a result of the topics being discussed, the panelists' tactics, or perhaps editorial choices.

B VALIDATION EXPERIMENTS

We performed validation of different parts of our pipeline on a subset of our dataset.

Table 7: Words found to be important in the context in sentences involving the BJP 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)

BJP 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*)

Table 8: One-Tailed t-test for the hypotheses $H_{M,greater}$ and $H_{M,lesser}$, we report t-stat for $\alpha = 0.01$

(a) One-Tailed t-test for $H_{M,greater}$

Category M	$Mean(\mathbf{M})$	$Mean(Rest)$	t**-stat**	p**-value**
CAA	0.2157	0.1598	5.684	7.255e-09
Kashmir	0.1947	0.1593	5.065	2.176e-07
Politics	0.1829	0.1477	10.397	3.644e-25
Pulwama-Balakot	0.2500	0.1601	6.561	3.175e-11
Religion	0.1959	0.1584	6.146	4.551e-10

(b) One-Tailed t-test for $H_{M,lesser}$

Category M	$Mean(\mathbf{M})$	$Mean(Rest)$	t**-stat**	p**-value**
COVID/Lockdown	0.0710	0.1670	-13.804	2.804e-42
Republic TRP Scam	0.0715	0.1636	-8.362	4.796e-17
International Affairs	0.1136	0.1645	-7.252	2.655e-13
China	0.0962	0.1628	-5.581	1.311e-08
Sports	0.0940	0.1621	-3.975	3.602e-05
Bollywood	0.1185	0.1636	-5.696	6.761e-09

Table 9: One-Tailed t-test for the hypotheses that difference between distribution of overlap speech and toxicity between Republic TV vs other shows is statistically significant, we report t-stat for $\alpha = 0.05$

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

Debate	$Mean(TVDebates)$	t**-stat**	p**-value**
Republic TV	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	$Mean(TVDebates)$	t**-stat**	p**-value**
Republic TV	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

B.1 Validation for the gender-classification model

We randomly selected 50 videos from our dataset while ensuring that no more than 5 videos were taken from a single major category. For each of these 50 videos, we extracted 50 evenly spaced frames across the duration of the video. As a result, our validation dataset consists of faces from 2,500 image frames. One of the authors annotated all the frames to determine the number of male and female panelists present. The performance results show that for the female label, the precision was 0.975 and the recall was 0.81. For the male label, the precision was 0.91 and the recall was 0.994.

B.2 Validation for the extraction of names from the transcripts

We tested how well our method for finding panelist names worked by using the same 50 videos we used to check our gender classification approach. One of the authors watched each video and wrote down the names of the panelists. Then, we compared this list to

Table 10: Few examples of tags associated with the various videos

Title	Tags
CBI To Probe 'Witness Coercion' Tapes, Param Bir In Deep Trouble	CBI TRP case, Witness tapes, TRP case manipulation, TRP witness scam, TRP witness coercion tapes, india today trp scam, republic tv fake trp, india today fake trp
Justice For Sushant: Demand Grows For CBI Investigation	Sushant Singh rajput, Sushant Singh Rajput soul talks, Sushant Singh Rajput murdered by Suraj Pancholi, Rumi Jaffery, Mumbai Police Sushant Singh,Sushant Singh Rajput spirit calling, Kangana Ranaut, Kangana Ranaut Sushant Singh, Kangana Ranaut
Jyotiraditya Scindia's Exit Stings Congress	Sonia Gandhi, Rahul Gandhi, maharashtra politics, maharashtra government, maha vikas aghadi rift, NCP, Shiv sena, uddhav thackeray, sharad Pawar, Priyanka Gandhi, Scinida, jyotiraditya scindia, jyotiraditya scindia bjp join, jyotiraditya scindia quits congress
Nation backs MS Dhoni Wearing The 'Balidaan Badge'	indian premier league, Balidaan Badge, balidaan badge controversy, glove controversy, india backs ms dhoni, national pride, balidan badge indian army, ms dhoni, world cup 2019 indian team, bcci, indian cricket team, 2019 world cup indian team, bcci writes to ICC
Will Yogi Adityanath Sack Rape Accused MLA?	cm yogi adityanath, unnao rape case, kuldeep singh sengar, bjp mla rape case, mla rape case, rape victim father death, cbi investigation of unnao, uttar pradesh, unnao case, bjp mla brother arrested, up police on unnao, cm yogi on unnao case

Table 11: Negation Words

not, don't, can't, won't, shouldn't, mustn't, should not, must not, do not, cannot, will not, would not, wouldn't, isn't, is not, dare not, have not, might not, may not, need not, ought not, shall not

Table 12: 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, ItalyKeSaudagar, WillYogiSackMLA, YogiWakeUp, FightForAsifa, SadhviBackGodse, SackBJPBrat, RepublicVsBJPMLA, YogicopsStung, NetasChokeDelhi, BJPWakeUpCall	CongRapeComment, SoniaSainik, MayaDumpsCong, CongPoliticsofAbuse, CongVsCitizens, ConginsultsDemocracy, ECBansMamata, MamataLosesGrip, AAPForFreebies, CongFallsApart, KejriwalMinisterArrested, NeechPolitics, VadraCongChaos, CongRajCollapse, VadrasMustGo, RahulMocksForces, CongressllsOver, RahulGetsDumped, CongAbusesRashtrapati, RahulCheatsPoor

the names our pipeline found. The performance was found to be a precision of 0.901 and recall of 0.730.

B.3 Validation for classification of speech into shouted/non-shouted categories

Here, we manually sampled 50-audio samples from across our dataset. One of the authors then classified these samples as shouted/non-shouted speech. On cross-referencing these with the labels assigned by our classifier, we found precision to be 0.91 and recall to be 0.75. [2] contributed the Indian Broadcast News Debate (IBND) dataset, which contains news debates from Republic TV along with annotations for shouted vs. normal speech. On the IBND dataset, we found performance to be a precision of 0.86 and a recall of 0.71, based on 62,375 samples belonging to the test split. Since the domain of the IBND dataset and our own dataset is the same, i.e. "Republic TV debates", our results on the IBND dataset can also be considered a reliable indicator of similar performance on our own dataset.

C PERFORMANCE ON OTHER DATASETS

To demonstrate the generalizability of our pipeline and to establish baselines, we decided to apply our pipeline to quantify incivility on four more debate/panel-based shows hosted in English:

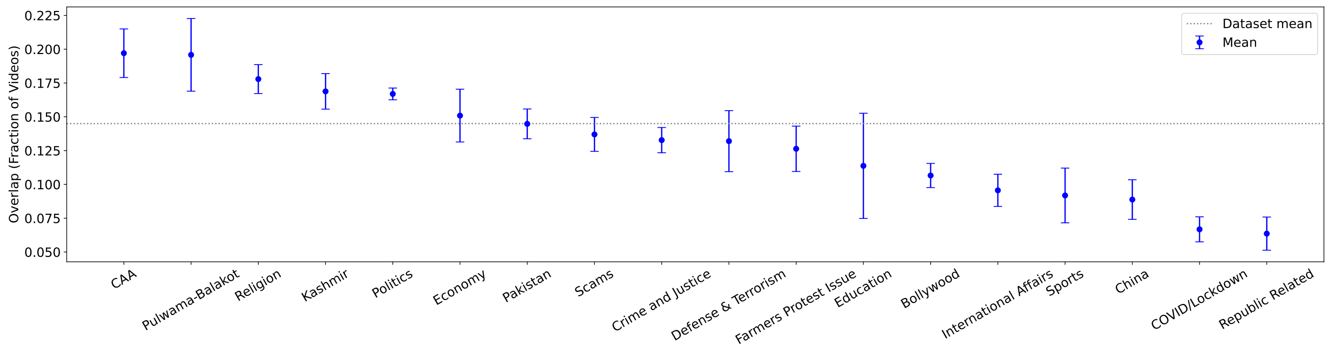

Figure 10: Overlap speech for all categories.

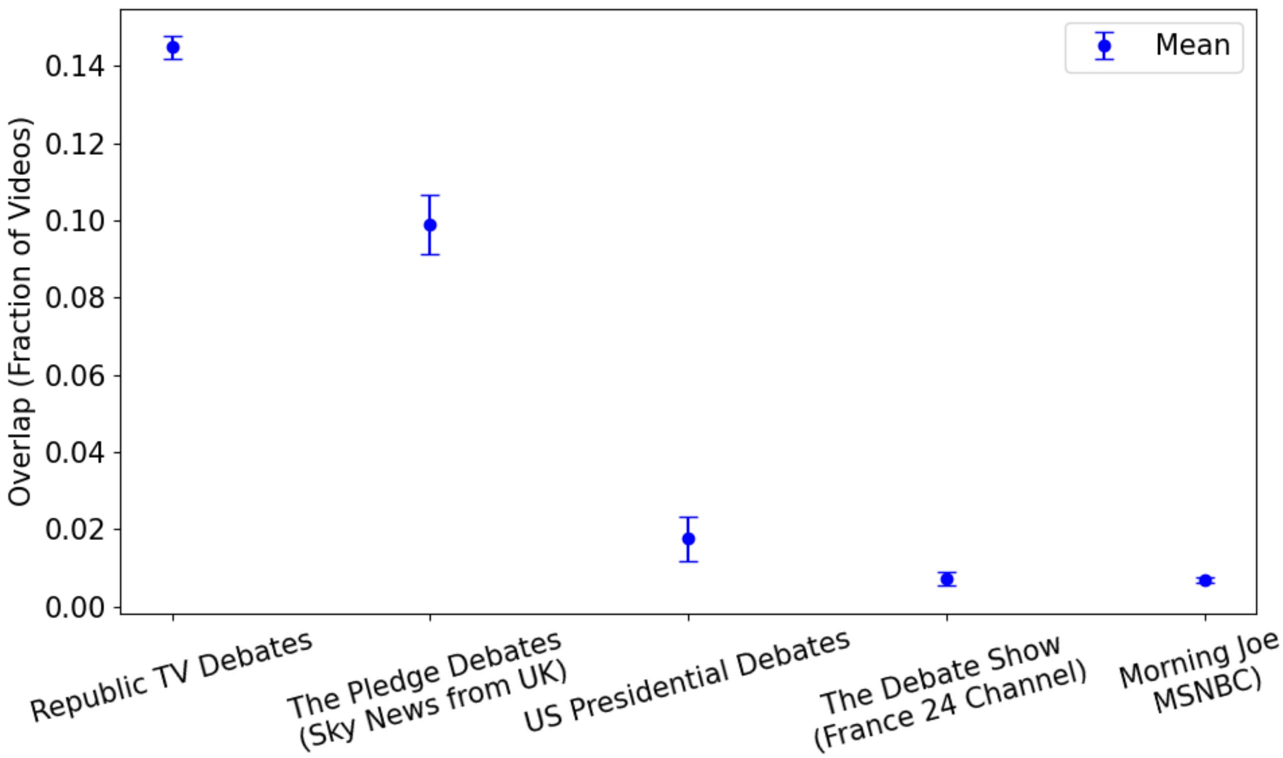

Figure 11: Comparing fraction of overlap speech in videos between different TV debate channels.

- The Debate Show (hosted in the France 24 Channel): including 216 videos from their YouTube playlist.[8]
- The Pledge Debates (hosted in Sky News from UK): we use the videos from their YouTube channel.[9] Their channel contains both (a) entire 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. After this filtering, we were left with 80 videos.
- Morning Joe (hosted on MSNBC in the US): we selected videos from the show's YouTube playlist.[10] To ensure that our analysis focused on the main show, we only included videos that were longer than 30 minutes in duration. This allowed us to exclude shorter clips from the main show that

[8] https://www.youtube.com/playlist?list=PLCUKIeZnrIUlLhXw4GoHFlUpFidIosXAT
[9] https://www.youtube.com/@thepledge/videos
[10] https://www.youtube.com/playlist?list=PLCUKIeZnrIUlLhXw4GoHFlUpFidIosXAT

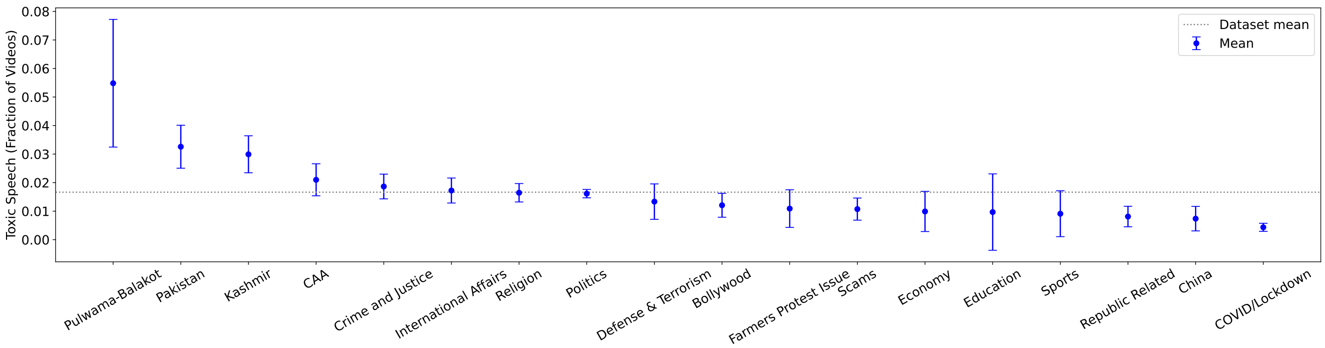

Figure 12: Fraction of videos with toxic speech in all categories

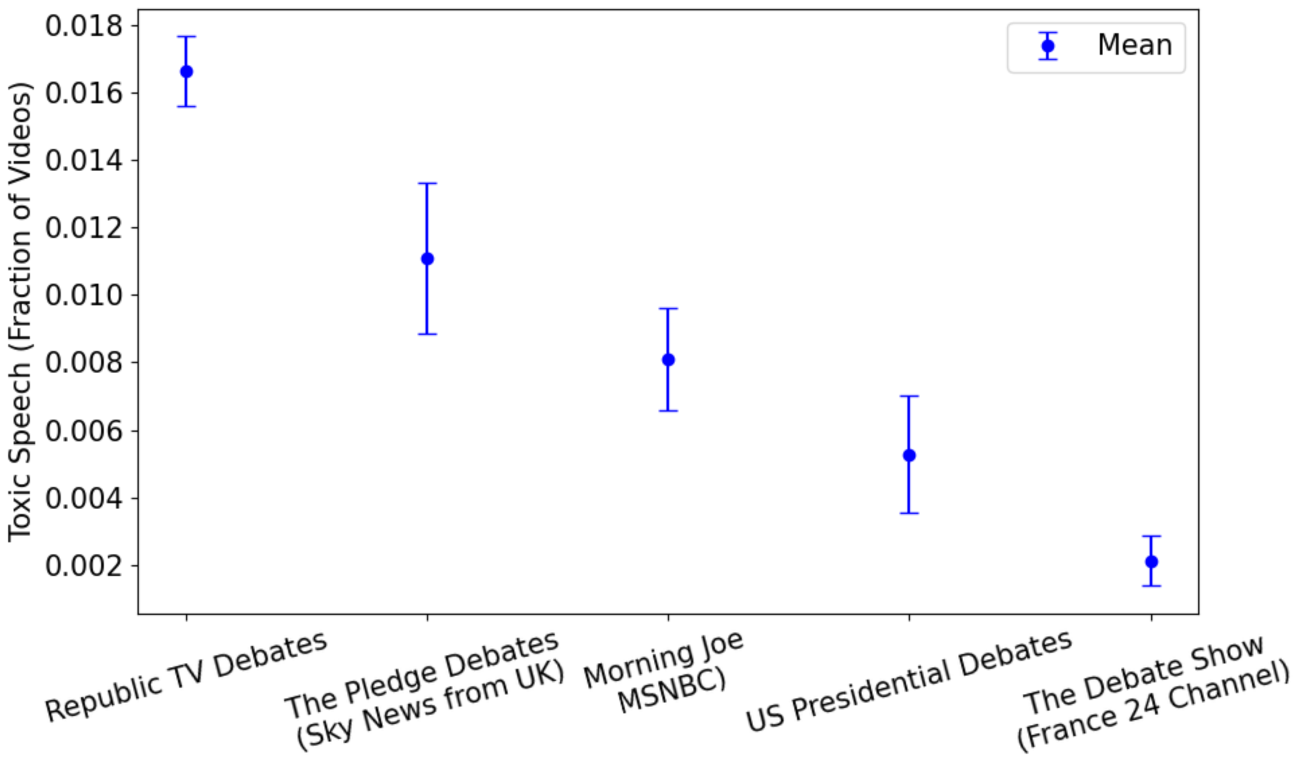

Figure 13: Comparing fraction of toxic speech in videos between different TV debate channels.

were repetitive. After applying this filtering 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 for the Democrats (for 2016 and 2020) and Republicans (for 2016).

We did a two-tailed t-test with a 95% confidence interval between Republic TV and other debates. We found that the overlap speech in Republic TV debates is statistically greater than all the other TV debates mentioned above. Refer to Table 9a for more details.

Similarly, for toxicity we find that Republic TV debates has statistically greater toxicity compared to France 24, US Presidential Elections and Morning show with Joe. Refer to Table 9b for more details.

Each debate video transcript is a list of utterances where consecutive utterances are spoken by different people. For each utterance, we use the perspective API to obtain the probability that the utterance may be interpreted as belonging to classes such as toxicity,

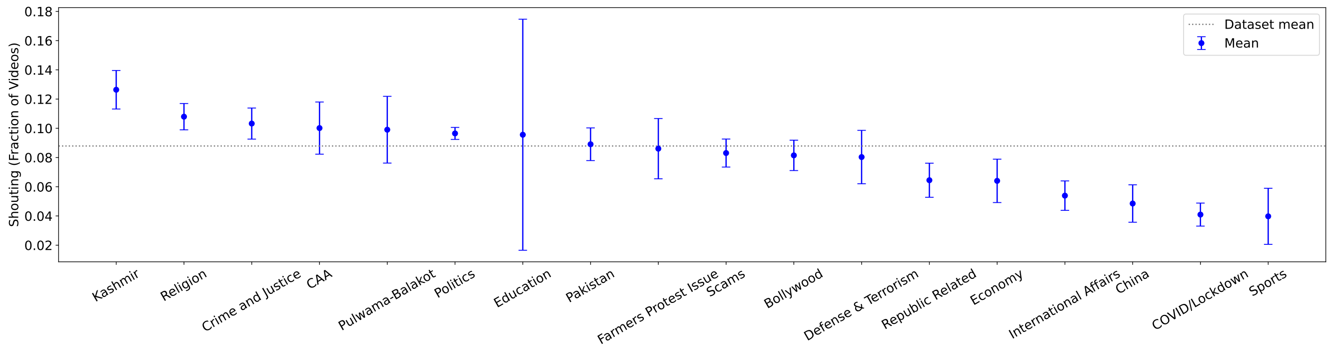

Figure 14: Fraction of videos with shouting in all categories

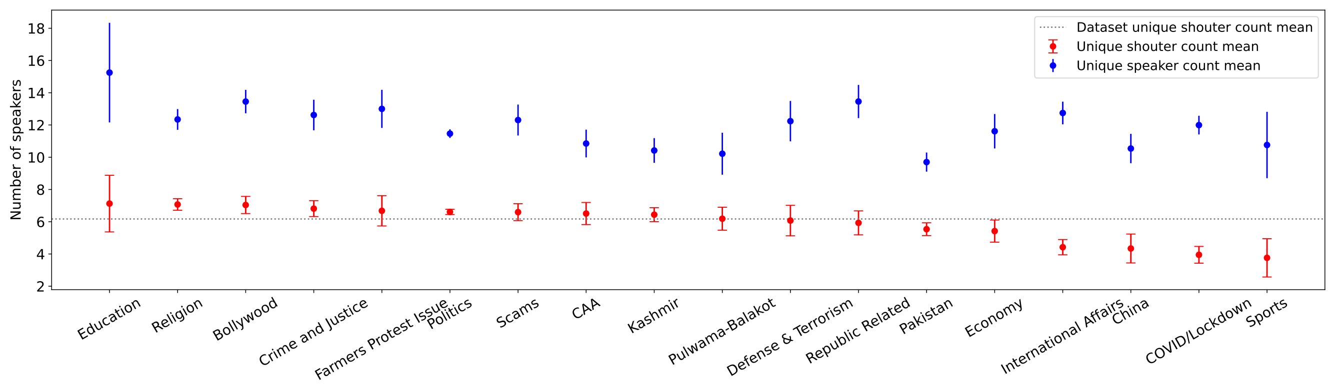

Figure 15: Number of unique shouters per video in all categories

severe toxicity, profanity, insult, threat or an identity attack. If the probability of any of these classes being present in the utterance exceeds a threshold of 0.5, we label the utterance as foul-speech. However, qualitatively, we found that there are instances where factual information related to the news item being discussed is labeled as uncivil, even though it is not the subjective opinion of any panelist. For example, the statement "He said that during his coverage of the war, he'd get anonymous calls that made him fearful for his life“ is related to the news item and not a personal viewpoint of any panelist. As a result, the fraction of foul speech shown in the plots is likely to be slightly higher than the actual amount of foul speech used by the panelists.

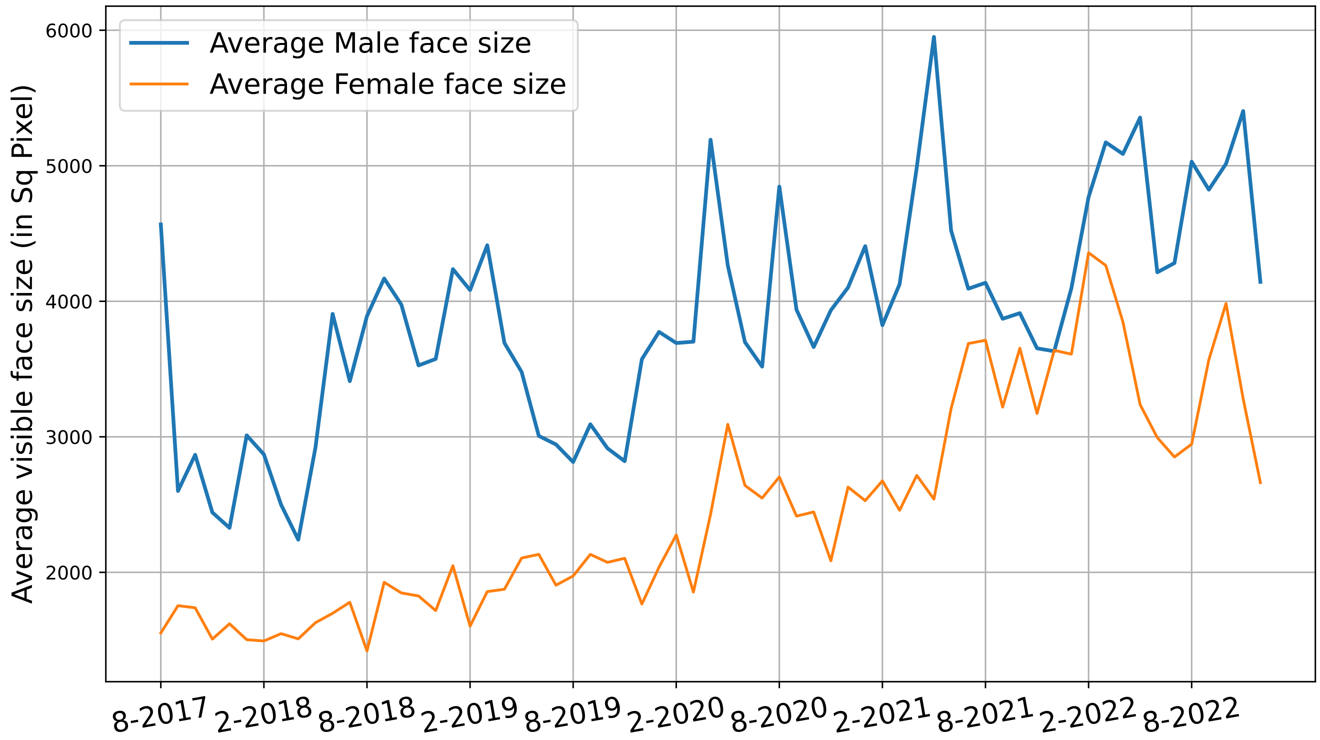

Figure 16: Average size of faces (Males: 3798 sq pixels, Females: 2424 pixels)

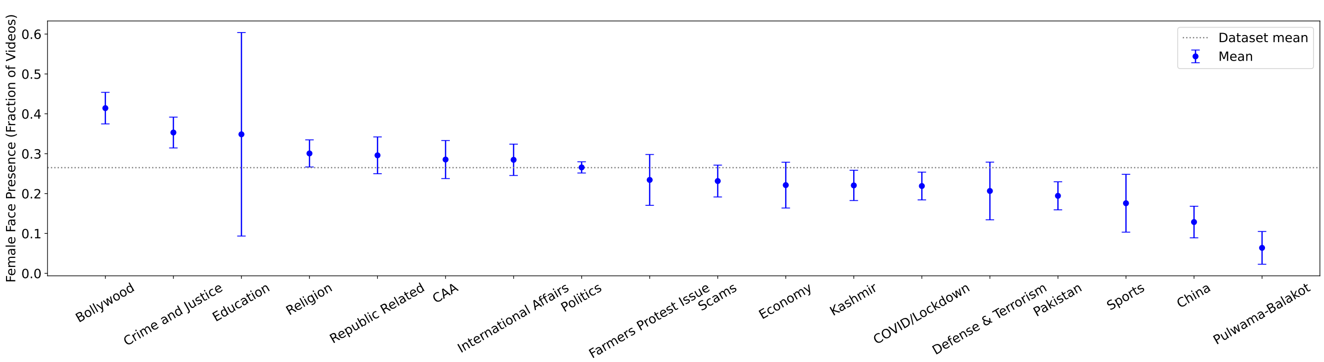

Figure 17: Fraction of videos with female faces in all categories

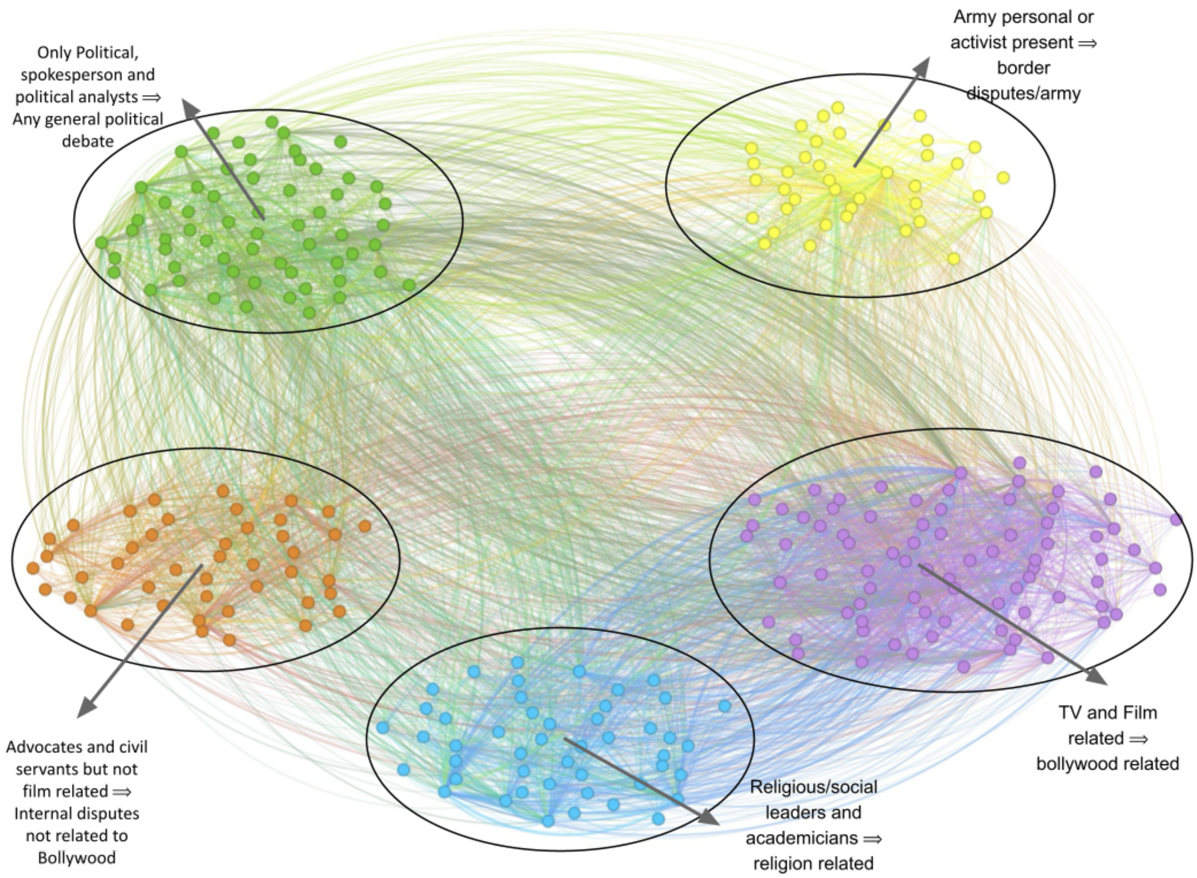

Figure 18: Found five kinds of clusters inside people affiliation coming on debate.

Table 13: Affiliation Triads with a statistically significant difference in incivility compared to overall incivility. Incivility values here are the sum of overlap and toxicity.

Triad	Average incivility fraction in videos where they occur	Frequency of them occurring in a video
DMK-BJP-BJP	0.218	72
Congress-BJP-AAP	0.209	52
BJP-AAP-BJP	0.204	88
BJP-Samajwadi party-BJP	0.189	70
TMC-BJP-BJP	0.179	148

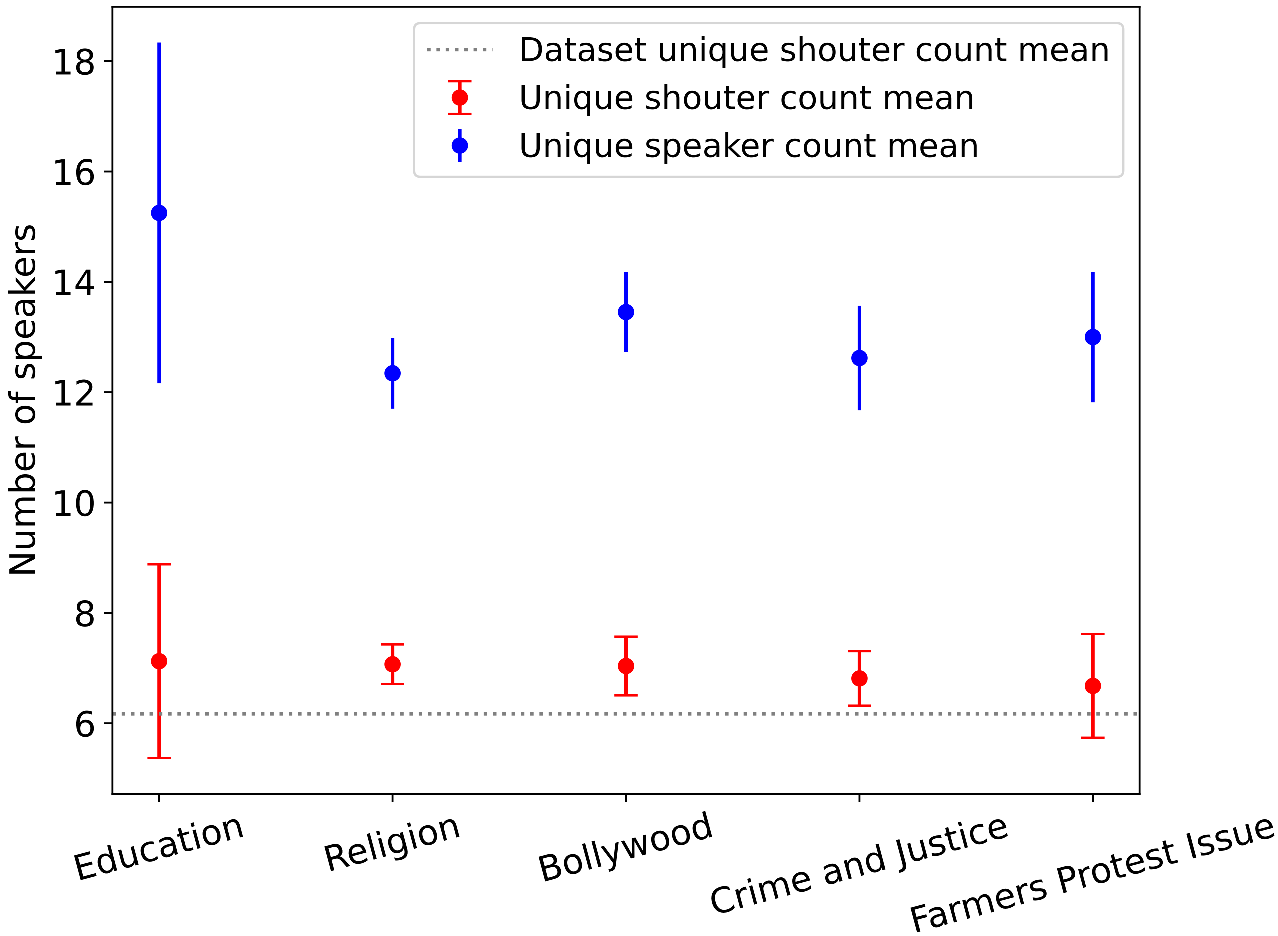

Figure 19: Average count of panelists engaged in shouting (depicted in red) compared to the total panelist count (shown in blue) for the top 5 categories with the highest incidence of shouting. The data indicates that approximately 50% of panelists in these categories participate in shouting behavior.