Supplementary Material
Toxicity Begets Toxicity: Unraveling Conversational Chains in Political Podcasts

Naquee Rizwan
Indian Institute of Technology
Kharagpur, West Bengal, India
nrizwan@kgpian.iitkgp.ac.in

Nayandeep Deb
Indian Institute of Technology
Kharagpur, West Bengal, India
nayandeepdeb125@kgpian.iitkgp.ac.in

Sarthak Roy
Indian Institute of Technology
Kharagpur, West Bengal, India
sarthak.cse22@kgpian.iitkgp.ac.in

Vishwajeet Singh Solanki
Indian Institute of Technology
Kharagpur, West Bengal, India
vsinghsolanki@kgpian.iitkgp.ac.in

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

Animesh Mukherjee
Indian Institute of Technology
Kharagpur, West Bengal, India
animeshm@cse.iitkgp.ac.in

CCS Concepts

• Computing methodologies → Speech recognition; **Discourse, dialogue and pragmatics**; **Language resources**; **Temporal reasoning**; **Information extraction**; **• Social and professional topics → Hate speech**; **Political speech**.

Keywords

Toxic conversation chains, podcasts, transcripts, change point detection, toxicity begetting toxicity

ACM Reference Format:
Naquee Rizwan, Nayandeep Deb, Sarthak Roy, Vishwajeet Singh Solanki, Kiran Garimella, and Animesh Mukherjee. 2025. Supplementary Material Toxicity Begets Toxicity: Unraveling Conversational Chains in Political Podcasts. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3746027.3754553

Warning: Contains potentially abusive/toxic contents.

1 Limitations

As future work, we plan to extend our study to incorporate audio signals as well in our analysis including a thorough investigation of the different audio features. We also plan to develop intervention strategies for this complex scenario. Further we plan to expand our dataset and scalable tooling to other political podcasts to get clear estimates on the prevalence of toxicity in podcasts. Finally, an important aspect that our paper does not answer is the impact of such toxicity on the listeners. Our findings lead us to question whether toxicity might become normalised if coming from popular podcast hosts who reach tens of millions of people, and the consequences it might have on our political discourse. Social scientists can study and answer such questions using some of the tools we developed in this work. .

2 Ethics statement

Our dataset comprises popular podcasts curated from publicly available RSS feeds, which are widely accessible and listened to by tens of millions of people. These podcasts are analyzed using automated transcription and diarization techniques. While individual speakers are segmented to enable conversational analysis, their identities are neither inferred nor used in the analysis to ensure privacy. We recognize the potential for biases in various stages of our pipeline, including transcription, diarization, toxicity detection, and change point detection. To mitigate these biases, we employed a rigorous evaluation process, comparing multiple algorithms at each stage and selecting the best-performing ones based on empirical results. While our analysis yielded promising findings, we emphasize the importance of conducting a comprehensive audit of the pipeline to further evaluate its reliability and fairness, particularly in production or high-stakes contexts. Our study adheres to ethical guidelines for research on publicly available data and prioritizes transparency, privacy, and the minimization of harm. We aim to contribute constructively to the discourse on toxicity in podcasts while acknowledging the limitations of our methods and the need for ongoing scrutiny and improvement.

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3754553

3 Employed prompts

Toxicity classifier prompt on GPT-4o: We apply the following prompt on anchor segment's diarized text:

- **"role"**: "system", **"content"**: "You are an expert in toxic speech detection. Your task is to detect whether a provided speech is toxic or not."
- **"role"**: "user", **"type"**: "text", **"content"**: "Definition of toxicity: Perspective API's primary attribute is toxicity, defined as a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion.
You are provided with the following speech: {anchor segment's diarized text}
Based on the above definition, classify the speech as either toxic or not-toxic. Output only toxic or not-toxic based on your analysis."

Change point detection prompts: An overview of the employed prompts is presented in Figure 1. The figure illustrates the combination of different prompt components gathered together to run textual & audio-based LLMs. In the case of audio, toxic chain clip (*.wav* file format) across all segments combined is also provided as input.

4 Properties of left-leaning channels

The statistics of the 7,124 toxic conversation chains from left-leaning podcast channels are presented in Figure 6. Interestingly, we observe that these properties highly align with the findings we had previously covered for right-leaning podcast channels (in the main text). Thus the dynamics of toxicity appears to be universal across the two leanings.

5 Further analysis for right-leaning channels

In this section we present some additional analysis of the right-leaning channels (the results are very similar for the left-leaning ones and hence not shown).

5.1 Additional textual properties

(i) Token count: We calculate the total number of tokens in the diarized text for each segment in a conversation chain. To break the words into tokens, we use the *word_tokenize* utility from NLTK library. Figure 3 shows that the mean token count is the largest for the anchor segment which naturally follows from the earlier observation that these segments have the longest mean duration.

(ii) Type token ratio (TTR): This metric is used to evaluate the diversity of vocabulary in a text. It is given by the formula: $\frac{\#\text{types}}{\#\text{tokens}}$, where $types$ are unique $tokens$. Higher values indicate greater lexical diversity which means that the text has less redundancy. Lower values suggest repetitive textual content. Figure 3 shows that the mean TTR for anchor segment is lower than all other segments in the chain, indicating that there is more repetition (possibly of the same hateful remark) in the anchor segment.

Table 1: Complete Dataset Statistics: We crawl a total of 52 podcast channels (31 right- and 21 left-leaning), making a total of 12,322 episodes (with 9,166 from right- and 3,156 from left-leaning channels). Statistics for the crawled dataset sorted according to the number of episodes are presented. Columns for duration (in minutes) and mean token count are shown. In these columns, numbers in parenthesis specify the standard deviation. Columns for the number of toxic episodes (those containing at least one toxic conversation chain) and the corresponding percentage distribution are also provided. First 31 channels are right-leaning followed by 21 left-leaning channels.

Podcast channel name	Number of episodes	Average episode duration (min)	Average words count	Toxic episodes	% toxic episodes
Bannon s War Room	1184	54 *(2)*	9339 *(662)*	172	15
Bill O Reilly s No Spin News and Analysis	1117	18 *(17)*	2553 *(2460)*	72	6
The Sean Hannity Show	990	37 *(5)*	6375 *(983)*	208	21
The Glenn Beck Program	725	81 *(37)*	12619 *(5703)*	275	38
The Dan Bongino Show	448	48 *(15)*	8483 *(2706)*	309	69
Mark Levin Podcast	440	103 *(18)*	14280 *(3095)*	380	86
Human Events Daily with Jack Posobiec	403	27 *(8)*	4614 *(1602)*	34	8
Conservative Review with Daniel Horowitz	398	62 *(6)*	9673 *(1210)*	95	24
The Rubin Report	372	43 *(13)*	8280 *(2513)*	161	43
The News Why It Matters	334	44 *(0)*	8313 *(423)*	181	54
The Megyn Kelly Show	332	95 *(10)*	17626 *(2117)*	180	54
Tim Pool Daily Show	321	86 *(11)*	15472 *(2295)*	114	36
The Ben Shapiro Show	229	48 *(17)*	10133 *(3664)*	62	27
The New Abnormal	220	50 *(16)*	9132 *(2919)*	198	90
Louder with Crowder	210	62 *(33)*	12588 *(6900)*	186	89
The Charlie Kirk Show	200	40 *(15)*	7127 *(2903)*	30	15
The Michael Savage Show	196	54 *(19)*	9253 *(3384)*	96	49
The Jordan B Peterson Podcast	145	99 *(21)*	16236 *(3772)*	27	19
The Michael Knowles Show	133	46 *(17)*	8112 *(3176)*	43	32
Verdict with Ted Cruz	133	43 *(11)*	7142 *(1833)*	25	19
Hold These Truths with Dan Crenshaw	115	53 *(20)*	9699 *(3478)*	13	11
Bret Weinstein DarkHorse Podcast	94	106 *(21)*	17414 *(3496)*	39	41
Pseudo Intellectual with Lauren Chen	88	11 *(2)*	2202 *(379)*	10	11
Fireside Chat with Dennis Prager	87	32 *(7)*	4309 *(1168)*	13	15
Conversations With Coleman	82	71 *(19)*	12920 *(3681)*	22	27
The One w Greg Gutfeld	71	15 *(3)*	2690 *(645)*	34	48
Candace Owens	47	29 *(14)*	5620 *(2782)*	13	28
Get Off My Lawn Podcast w Gavin McInnes	24	61 *(20)*	10012 *(3391)*	24	100
The MeidasTouch Podcast	15	52 *(40)*	9017 *(6961)*	9	60
The Matt Walsh Show	10	46 *(24)*	8081 *(4228)*	3	30
Rudy Giuliani s Common Sense	3	36 *(4)*	5414 *(750)*	2	67
The MediasTouch Podcast	431	29 *(27)*	4883 *(4748)*	115	27
Late Night with Seth Meyers Podcast	255	25 *(5)*	4616 *(840)*	73	29
Mea Culpa	231	78 *(12)*	12858 *(2126)*	225	97
Pod Save America	230	60 *(16)*	11499 *(2972)*	187	81
In the Bubble with Andy Slavitt	224	48 *(10)*	8300 *(1843)*	14	6
Fast Politics with Molly Jong-Fast	201	55 *(9)*	10036 *(1617)*	141	70
The Rachel Maddow Show	141	55 *(28)*	9496 *(4538)*	23	16
On with Kara Swisher	135	53 *(10)*	10302 *(2077)*	64	47
Political Gabfest	132	51 *(13)*	9041 *(2330)*	17	13
Pod Save the World	124	67 *(16)*	12273 *(2935)*	77	62
Lovett or Leave It	119	70 *(21)*	12423 *(3618)*	118	99
Why is This Happening with Chris Hayes	116	53 *(13)*	9933 *(2348)*	7	6
Majority 54	110	52 *(10)*	10087 *(2025)*	35	32
Krystal Kyle and Friends	108	80 *(17)*	14409 *(3101)*	104	96
Hysteria	105	71 *(15)*	12658 *(2629)*	100	95
Offline with Jon Favreau	97	54 *(13)*	10041 *(2733)*	39	40
Conversations With Coleman	96	67 *(20)*	12524 *(3805)*	23	24
Pod Save the People	92	68 *(17)*	11598 *(3039)*	12	13
Hell and High Water with John Heilemann	72	72 *(13)*	14766 *(2853)*	57	79
Intercepted with Jeremy Scahill	70	48 *(14)*	7672 *(2537)*	4	6
Lady Dont Take No	67	45 *(8)*	7828 *(1645)*	63	94

5.2 Keywords

To identify and compare the keywords present in the segments, we extract the top ten toxic key-phrases from each segment in every conversation chain using KeyBERT [1]. Each phrase is limited to a maximum of five-grams. The word embeddings for KeyBERT algorithm are obtained from the Detoxify [3] library[1]. To ensure diversity in our

[1] https://huggingface.co/unitary/toxic-bert

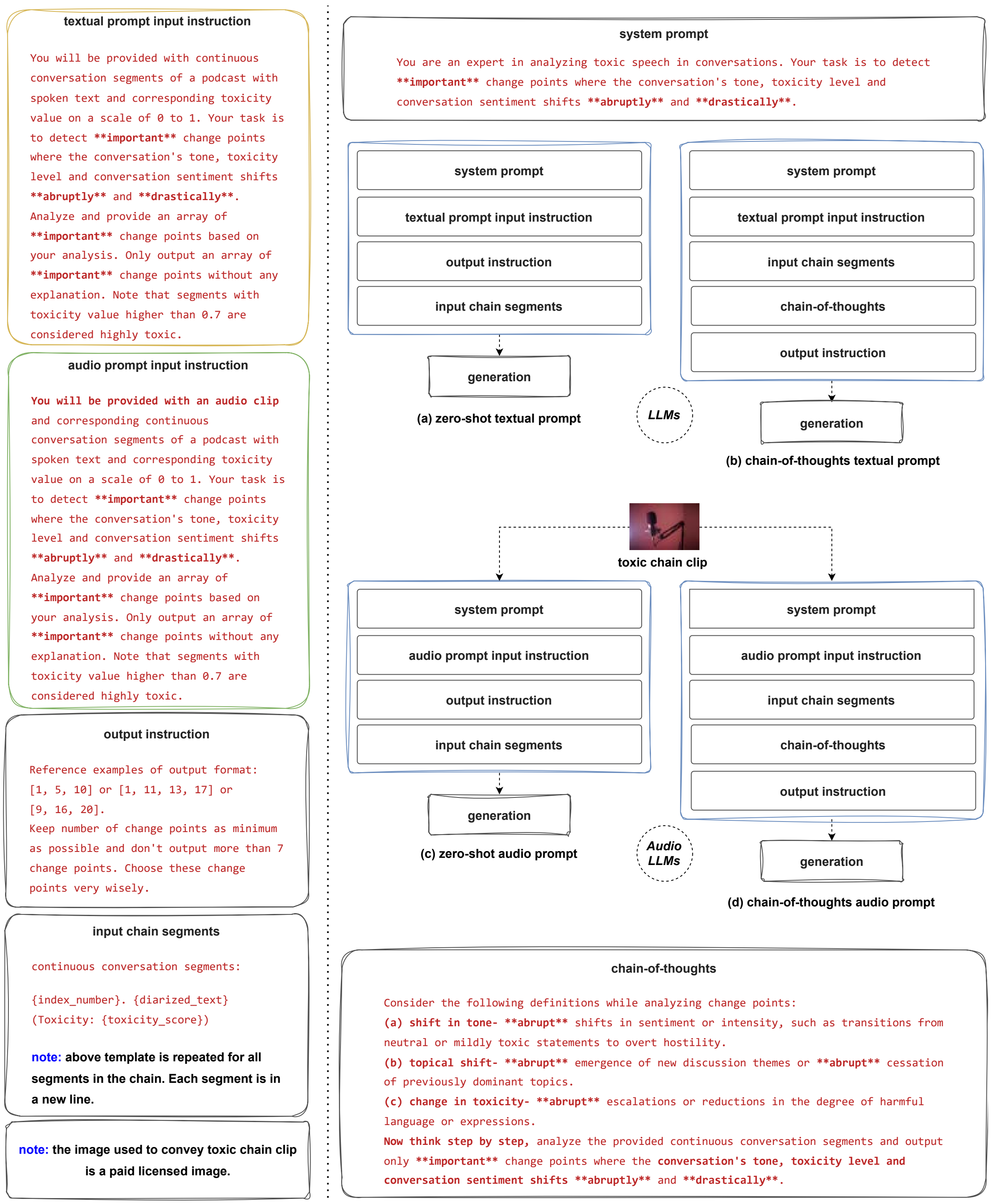

Figure 1: Employed Prompts: Textual prompts used on Qwen-2 and GPT-4o are in sub-figures (a) and (b). Similarly, sub-figures (c) and (d) are employed with Qwen-2-Audio and GPT-4o-Audio. Different components of an input prompt are separated within individual boxes. Each of the sub-figures (i.e. (a), (b), (c) & (d)) provides an overview of the organization of prompt components in different input prompt setups, i.e., in the *zero-shot* & *chain-of-thoughts* setups. Please refer to Figure 2 and a figure in main paper for examples of input chain segments.

extracted key-phrases, we use the maximal margin relevance utility of the KeyBERT algorithm, enforcing a diversity score of 0.75. In addition, we eliminate stop words & punctuation and only retain strings containing characters from the English alphabet. We combine all the keywords obtained from the preceding ten segments and represent

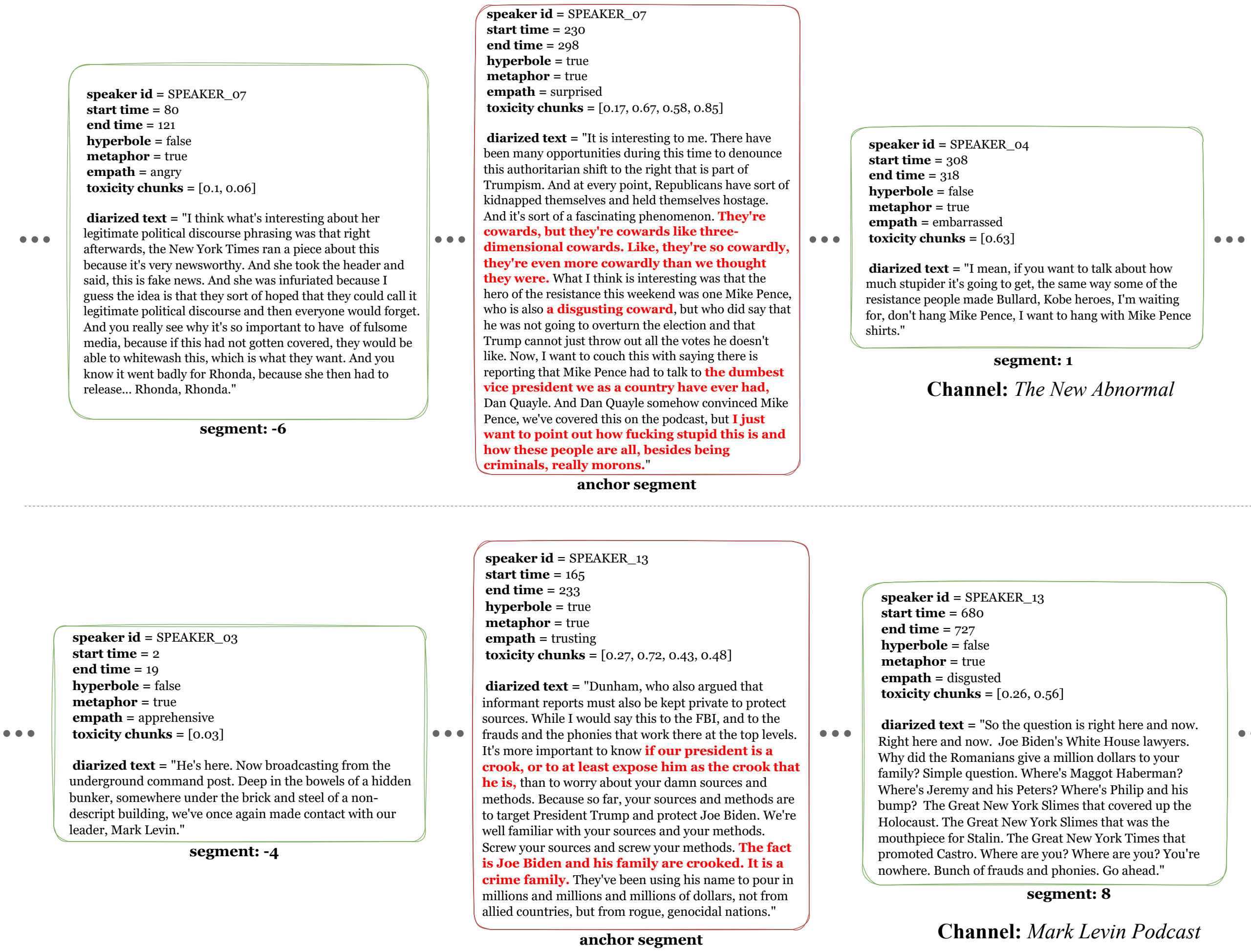

Figure 2: CONVERSATION EXAMPLES: One among previous and next segments are plotted along with anchor segment since it is not feasible to plot all segments. Toxic contents in the anchor segment are marked in red color. NOTE: start and end times are in seconds.

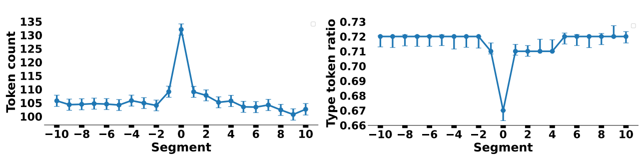

Figure 3: Token count (left) and type token ratio (right) across segments at 95% confidence intervals.

them as a word cloud; similarly, we combine all the keywords obtained from the following ten segments and represent them as another word cloud. We also obtain the word cloud for the anchor segment separately. These word clouds in series are illustrated in Figure 4. We observe that most of the words are linked to political and controversial themes like 'right', 'republican', 'democrat', 'women', 'biden', 'american' & 'liberal', which naturally follows from the choice of our dataset. The anchor segment has several toxic keywords including 'idiot', 'stupid', 'f*cking/f*ck', 'sh*t', 'a*s' & 'moron'. The high similarity between the previous and next word clouds indicates that the anchor segment introduces a disruption in the flow of the main conversation, which rewinds back to normal only at the end of the anchor segment. Finally, in both the previous and the next word clouds, words like 'want', 'know', 'people' & 'yeah' appear, which reflect an expression of demand. Thus, hostility in the speech of a particular speaker seems to be fueled by words of demand from either the anchor or other speakers/participants in podcast conversations.

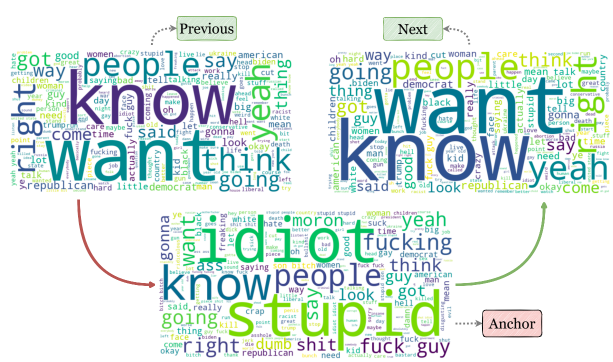

Figure 4: WORD CLOUDS: previous, anchor and next segments.

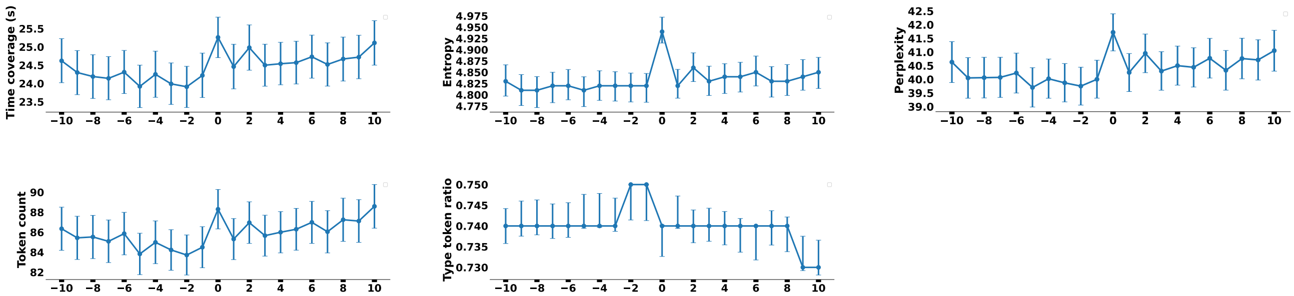

Figure 5: Control Group Statistics: Mean at 95% confidence intervals for time coverage, entropy, perplexity, token count & ttr for control group across segments.

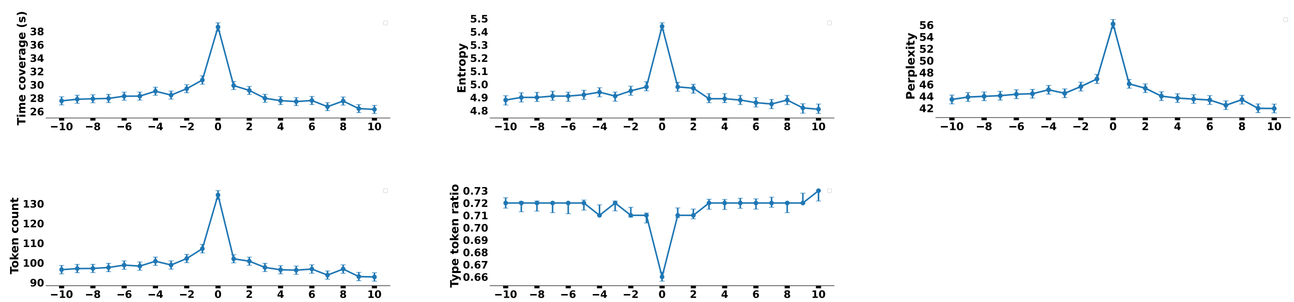

Figure 6: Left-leaning Statistics: Mean at 95% confidence intervals for time coverage, entropy, perplexity, token count & ttr for left-leaning podcast channels across segments.

Table 2: Topic transitions in the conversation chains induced by BERTopic.

Order in chain	Induced topics
Preceding	*like, know, people, go, right, yeah, think, get, say, going*
Anchor	*b**ch, stupid, son, f**k, s*it, f**king, idiot, damn, shut, guy*
Following	*like, know, people, go, right, yeah, think, get, say, going*

5.3 Topical shifts

The keyword analysis in the previous section indicates that there is a significant change in the conversation content during the transition from the previous to the anchor segment. This hints at the fact that there is a possible topical/thematic shift during such a transition. In order to establish this, we perform topic modeling using the BERTopic [2] model. We extract the topics considering the previous ten aggregated segments as one document, the next ten aggregated segments as a second document and the anchor segment as the third document. We set the number of topics to three and report the top ten most representative words for that topic, which has the highest probability of association with a document. The results are noted in Table 2, which reveals significant shifts in thematic focus and toxicity level during the transition from the previous to the anchor segment and the anchor segment to the next segment. Precisely, the anchor topic is highly toxic while the preceding and the following topics are more related to demands, thus offering insights into the progression and contextual drivers of toxic conversations.

5.4 Number of speakers in toxic chains

We present the number of speaker turns and its distribution over 8,634 toxic chains from right-leaning channels presented as (#speakers: % toxic chain)– (1: 1.26), (2: 8.78), (3: 16.89), (4: 21.35), (5: 19.21), (6: 14.94), (7: 8.92), (8: 4.91), (9: 2.11), (10 and more: 1.63). Notably, only a small percentage of chains, i.e., 1.26% have one speaker within them. Hence, we conclude that the toxic chains represent conversations rather than monologue behaviour.

5.5 Right-leaning control group

Here we cover the analysis of non-toxic chains where we choose such conversations that have the toxicity value of the anchor segment lower than 0.3 *(as per the recommendation of Perspective API)* for right-leaning channels. We also take care that the randomly selected chains have a distribution in line with the distribution of podcast channels in toxic conversation chains. Further, we exclude such chains that have the previous and next ten segments with toxicity higher than 0.3 to ensure a consistent and fair comparison.
Results are presented in Figure 5. It is evident from the figure that, unlike toxic conversation, non-toxic conversation is more organized & consistent with low randomness.

References

[1] Maarten Grootendorst. 2020. KeyBERT: Minimal keyword extraction with BERT. doi:10.5281/zenodo.4461265
[2] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794 [cs.CL] https://arxiv.org/abs/2203.05794
[3] Laura Hanu and Unitary team. 2020. Detoxify. Github. https://github.com/unitaryai/detoxify.