Emotional and Hate Speech Analysis in YouTube Comments on the 2024 U.S. Election: Harris vs. Trump
Final Report

Lena Reissinger
University of Californa, Berkeley

1 Abstract

This project investigates hate speech and emotions in the context of the 2024 U.S. presidential campaign. I focus on comments directed at Kamala Harris and Donald Trump, as well as the content of their campaign speech transcripts. By analyzing YouTube comments—a less-explored platform compared to X (formerly Twitter)—the project captures direct audience reactions tied to specific video content. Using a fine-tuned DistilBERT model, a custom classifier was developed to identify insulting comments against political opponents.

In addition to hate speech detection, the analysis explored broader emotional dynamics within comments, such as hope, frustration, and anger, while differentiating whether emotions were directed toward the candidate or their opponent. Campaign speech transcripts were further analyzed using topic modeling and emotion detection, uncovering the interplay between speech tone, topics, and hate speech prevalence in viewer comments. This dual-layered approach provides a detailed glance at the emotional landscape and hate speech dynamics within the polarized political climate of the 2024 election.

2 Introduction

A preliminary glance at every single election campaign video on YouTube confirms what my analysis aims to quantify: the 2024 presidential election was deeply polarizing, eliciting emotional responses ranging from joy to hate. My project builds upon the analysis of [7] on hate toward political opponents, extending it to the context of the upcoming election cycle. With a female candidate in the race, I hypothesize that the dynamics of political attacks have shifted significantly. With a female candidate in the race, I hypothesize that the nature of attacks against political opponents has shifted considerably. YouTube comments, which remain underexplored compared to X, offer valuable insights, particularly when analyzed alongside the video content they respond to.

When expressing a political opinion, individuals can agree, remain neutral, or disagree with positions—or with the politicians as individuals. That last category—comments directed at political opponents on a personal level—can be especially problematic. My goal was to build a classifier to identify insulting YouTube comments targeting political opponents to gauge how prevalent personal attacks are in the current political landscape.

Exploring the full range of emotions in these comments provides a deeper understanding of how the YouTube community feels about the upcoming polarizing election—their preferred candidate, the opponent, and perhaps even their own role in the discourse. By manually annotating the dataset, I aimed to create classifications that could later assess shifts in sentiment by comparing this year's comments with data from future elections. This analysis could reveal whether today's political climate reflects a lasting trend or an evolving dynamic, potentially even uncovering new patterns of behavior.

3 Related Work

Hate speech detection is a well explored topic within the NLP space and yet consits of many dimensions: As highlighted by Schmidt and Wiegand [15], various features, including sentiment analysis, lexical resources, and meta information, can be effective in hate speech detection. My project tackles hate speech in the special case of politics in the pre-election phase, It adds to existing studies on offensive speech on previous U.S. elections,

e.g. [5] [16] or [17]. Extending this analysis to a fresh dataset and amplifying it towards other dimensions, especially sexual hate speech, allows for comparison with the 2020 election and deepens the understanding of hate speech trends. My work also ties into broader NLP research on detecting toxic speech within the political discourse, which is becoming more important as social media plays a growing role in shaping political discourse. For my analysis, I'm not just focusing on comments to videos, but also replies to previous comments. Considerations on how to handle these replies to YouTube comments to capture the full context of interactions are for example described by Ashraf et al. [3] , while hate speech detection in multilingual settings was demonstrated by Aluru et al. [2]. Automated methods for hate speech annotation, such as those by Saifullah et al. [14], offer valuable tools; however, categories in this area are fluid and can evolve over time. To address this, I created a dataset specifically tailored to capture the dynamics within the communities on Trump's and Harris's YouTube channels leading up to the 2024 presidential election.

4 Methodology

4.1 Data Collection:

The data collection for this project focused on YouTube comments and replies from the official channels of Donald Trump and Kamala Harris, covering the period from August 1, 2024—the date of Harris's formal candidacy announcement—through to the election on November 5th. Using YouTube's inspection tools, I compiled a list of URLs for all relevant videos posted on these channels during this timeframe. For web scraping, I used Octoparse to capture comments and replies, with a limit of 10,000 comments per video for time reasons, to ensure good representation of audience sentiment. This resulted in a dataset of over 221,572 comments, and replies on comments.

I used the URLs to scrape transcripts of the videos by downloading the subtitles provided by YouTube. The comments were scraped in many small snippets, which I combined into one transcript for the full content of each video. In the end, my dataset included 243 fully transcribed videos.

4.2 Data Pre-Processing:

Within several preprocessing steps I wanted to prepare the dataset for my analysis. Initially, I used the `langdetect` library to identify the language of each comment, filtering out non-English comments to maintain focus on a single linguistic context. To streamline the analysis, I excluded replies and retained only top-level comments, as replies often relied on contextual information from parent comments, making them challenging to analyze in isolation.

For text cleaning, I removed emojis, special characters, and excessive whitespace to ensure textual uniformity. All comments and transcripts were converted to lowercase to reduce variation caused by capitalization. I indexed each comment and transcript with a unique identifier for easy tracking across subsequent steps in the pipeline. Additionally, I addressed missing data by removing rows with empty or invalid text fields.

4.3 Analysis of the YouTube Comments

To analyze YouTube comments, I began with exploratory steps to gain an initial understanding of the data. A standard sentiment classifier and an emotion analysis model were applied to the comments to identify general trends and potential clusters. These first insights revealed a diverse range of emotions, including anger, hope, frustration, and joy, expressed towards the candidates and their opponents. Although I initially planned to include a cross-cultural perspective, language classification using `langdetect` proved very error-prone, as many comments consisted of single words (e.g., "no," "Trump") common across multiple languages. Due to these challenges and time constraints, I decided to exclude the cross-cultural analysis.

To focus on insulting comments aimed at political opponents, I used a two-step classification process to make the annotation easier. This helped deal with the fact that hateful comments were relatively rare in the dataset.

Step 1: Negative Sentiment Classifier

In the first step, I annotated 1,100 comments to identify general negative sentiment toward either

a candidate or their opponent. Then, I created a multi-class classifier with three categories: positive, neutral, and negative comments. This data was split into 80% training and 20% validation sets and used to fine-tune a `DistilBERT-base-uncased` model. Key training parameters included a batch size of 16, the AdamW optimizer, and a maximum sequence length of 512 tokens. This lightweight yet effective model achieved a focused subset of comments with negative sentiment, which served as the foundation for the next step.

Step 2: Hate Speech Classifier

Using the negative sentiment classifier from Step 1, I sampled 1,200 comments with negative sentiment and manually annotated them for hate speech and insulting content. As expected, the proportion of hateful and insulting comments in this subset was significantly higher than in the overall dataset. The annotated data was split 80/20 into training and hold-out sets, with the latter further divided equally into validation and test sets. The final hate speech classifier, also based on `DistilBERT`, achieved an accuracy of 93% on the test dataset, demonstrating strong domain-specific performance.

This two-step process allowed for efficient annotation and analysis of hate speech, uncovering targeted negativity within political discourse on YouTube.

4.4 Analysis of Campaign Speech Transcripts

For the transcript analysis, I used Octoparse to collect subtitles from videos posted on the official YouTube accounts of Kamala Harris and Donald Trump between August 1 and November 5, 2024. These transcripts provided a clear view of the candidates' campaign messages. My analysis focused on uncovering key themes and examining the tone and emotional content of their speeches.

Topic Modeling

To identify clusters and themes within the speeches, I applied BERTopic, a state-of-the-art topic modeling tool. By generating topic embeddings, BERTopic grouped related segments of text and assigned descriptive topic labels. This approach revealed some key thematic areas, that were adressed in the videos, such as the economy, healthcare, and foreign policy. The topic modeling results served as a foundation for a deeper analysis of sentiment, emotion, and hate speech within each thematic cluster.

Sentiment Analysis

To evaluate the overall tone of the speeches, I used a pre-trained `DistilBERT-base-uncased` sentiment analysis model. Each segment of the transcripts was categorized as positive, negative, or neutral. This provided a high-level view of how the candidates conveyed their messages—whether optimistic, critical, or neutral in tone—allowing me to assess sentiment trends across different topics.

Emotion Analysis

To dive deeper into the emotional content, I employed the `distilbert-base-uncased-emotion` model, which classified transcript segments into emotions such as anger, joy, sadness, and fear. This method showed patterns of how emotions were communicated, with certain topics suggesting stronger emotional tones, such as anger during discussions of opponents' policies or fear in segments addressing national security.

Hate Speech Detection

For detecting hate-speech, I used the pre-trained `bert-base-uncased-hatexplain` model to detect potentially harmful language within the transcripts. This analysis identified sections of the speeches that included aggressive or harmful rhetoric, particularly toward opponents or specific groups. However, only in topics addressing migration, there appeared some language identifies as hate speech.

By combining these methods, the transcript analysis offered valuable insights into the tone and emotions in the candidates' campaign videos. I used this analysis to explore how language was crafted to evoke specific emotions in political discourse. Later, I compared my findings from the transcripts with the comment analysis to see how the video content influenced audience reactions.

5 Results

Insights from Comment Analysis My analysis revealed that 60.9% of insulting comments were made on videos posted on Donald Trump's YouTube channel, while 30.1% were on Kamala Harris's channel. However, this does not necessarily indicate the target of the hate speech, as insults could have been directed at either the channel owner or their political opponents. A more detailed exploration of comment content would be needed to determine the subjects of these insults.

Insights from Transcript Analysis To analyze the tone and themes of the videos, I conducted a topic modeling and emotion analysis of the transcripts. Using BERTopic, I identified major topics, including discussions on **"woman, freedom, abortion"**, **"care, price, work"**, and **"border, immigration, immigrant"**. Sentiment and emotion analysis further revealed the emotional tone of these topics. For instance:

- **Emotion Analysis**: Transcripts with the emotion **anger** were more frequently tied to topics like **"border, immigration, immigrant"** and **"vice, national security, nation"**.
- **Hate Speech in Speeches**: Certain topics, like **"woman, freedom, abortion"**, were associated with higher percentages of hateful comments in the comment section, as shown in the scatter plot.
- **Overall Sentiment**: Videos discussing economic or healthcare topics, such as **"care, price, work"**, exhibited more balanced tones, while those touching on immigration and abortion were more emotionally charged.

Insights from Combining Transcripts and Comments By linking the transcript themes and emotional tones to the comments, I investigated whether aggressive or emotionally charged content in videos corresponded to a higher prevalence of hateful comments.

1. **Topic-Level Analysis**

   As demonstrated in Figure 1, videos focusing on **"woman, freedom, abortion"** have the highest proportion of hateful comments (over 30%), followed by **"care, price, work"** and **"border, immigration, immigrant"**. These topics often evoke strong public opinions, which could explain the higher levels of negative language in the comments.

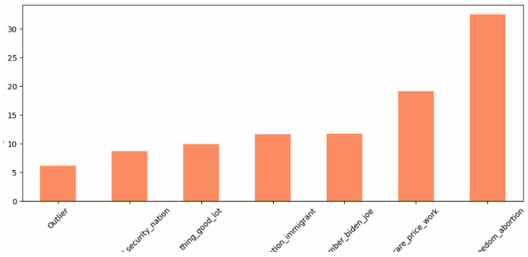

Figure 1: percentage of insulting comments towards the political opponent of the overall comments per topic

2. **Emotion and Topic Interaction**

   The scatter plot (Figure 2) demonstrates that videos with **anger** and **fear** as dominant emotions were associated with a higher percentage of hateful comments. For example:

   - Videos with **anger** discussing **"vice, national security, nation"** had one of the highest percentages of hate speech comments.
   - Emotional content tied to **"woman, freedom, abortion"** also stood out, showing significant intersections of **anger** and hate speech.

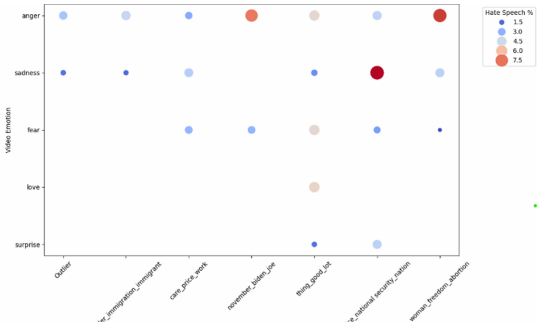

Figure 2: Percentage of insulting comments towards the political opponent per topic and per predominant emotion within the video, that were commented on

3. **General Patterns**

   The heatmap (Figure 3) further highlightes the interaction between video topics and emotions. For example, **"woman, freedom, abortion"** has the highest count of hate speech comments

linked to **sadness** and **anger**, while **"border, immigration, immigrant"** and **"vice, national security, nation"** also appear prominently in these categories.

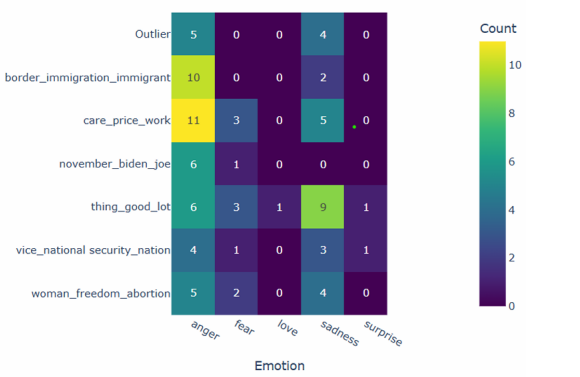

Figure 3: Predominant emotions of video per certain topics

My analysis confirms that topics and emotional tones in videos influence the prevalence of insulting comments towards the political opponent. Videos discussing socially charged issues, such as immigration and abortion, with tones of anger or fear, were most frequently associated with a higher percentage of hate speech in the comments. This demonstrates how video content shapes the emotional and linguistic patterns of audience engagement and offers a contribution on understanding the political discourse in online communities such as YouTube.

6 Discussion and Limitations

This project provides insights into the emotional dynamics and hate speech patterns surrounding the 2024 U.S. presidential election. By combining analysis of YouTube comments and campaign speech transcripts, it highlights how emotionally charged and topic-specific video content influences audience language. However, several limitations should be acknowledged.

First, the analysis was based on a total of 2,500 annotated comments, which, while substantial, is not sufficient to fully capture the breadth of political discourse on YouTube. A larger, more diverse dataset would improve the generalizability of the findings.
Second, cross-cultural analysis proved too challenging for the scope of this university project due to the limitations of language detection models, especially for short comments with single words like "Trump," "Harris," or "No," which appear in multiple languages. As a result, the study focused only on English comments, missing out on valuable insights into non-English discourse.

Another key limitation is the subjective nature of hate speech annotation. Despite efforts to maintain consistency, the definition of hate speech varies based on personal interpretation and social norms. This subjectivity may have influenced both the manual annotation process and the performance of the hate speech classifier. Please note that the annotation reflects my personal interpretation on what is considered insulting and what is not.

Future work could address these limitations by expanding the dataset to include more annotated comments, incorporating advanced multilingual language models for cross-cultural analysis, and refining the annotation framework to capture nuances in hate speech definitions. Additionally, an in-depth analysis of who is targeted by hateful language—whether the candidate hosting the video or their opponent—could provide valuable insights into the dynamics of political discourse. Exploring alternative platforms or time frames could further contextualize the findings and broaden their applicability.

References

[1] Rawan Fahad Alhujaili and Wael M.S. Yafooz. 2021. Sentiment Analysis for Youtube Videos with User Comments: Review. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pages 814–820.

[2] Sai Saketh Aluru, Binny Mathew, Punyajoy Saha, and Animesh Mukherjee. 2020. Deep Learning Models for Multilingual Hate Speech Detection. *arXiv preprint*. ArXiv:2004.06465.

[3] Noman Ashraf, Arkaitz Zubiaga, and Alexander Gelbukh. 2021. Abusive language detection in youtube comments leveraging replies as conversational context. *PeerJ Computer Science*, 7:e742. Publisher: PeerJ Inc.

[4] Bharathi Raja Chakravarthi. 2022. Hope speech detection in YouTube comments. *Social Network Analysis and Mining*, 12(1):75.

[5] Kevin Finity, Ramit Garg, and Max McGaw. 2021. A Text Analysis of the 2020 U.S. Presidential Election Campaign Speeches. In *2021 Systems and Information Engineering Design Symposium (SIEDS)*, pages 1–6.

[6] S. Forberger, L. Reisch, T. Kampfmann, and H. Zeeb. 2019. Nudging to move: a scoping review of the use of choice architecture interventions to promote physical activity in the general population. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1):77.

[7] Lara Grimminger and Roman Klinger. 2021. Hate Towards the Political Opponent: A Twitter Corpus Study of the 2020 US Elections on the Basis of Offensive Speech and Stance Detection. *arXiv preprint*. ArXiv:2103.01664 [cs].

[8] Md Saroar Jahan and Mourad Oussalah. 2023. A systematic review of hate speech automatic detection using natural language processing. *Neurocomputing*, 546:126232.

[9] Juan Pablo Latorre and Javier J. Amores. 2021. Topic modelling of racist and xenophobic YouTube comments. Analyzing hate speech against migrants and refugees spread through YouTube in Spanish. In *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)*, TEEM'21, pages 456–460, New York, NY, USA. Association for Computing Machinery.

[10] Nayeon Lee, Chani Jung, Junho Myung, Jiho Jin, Jose Camacho-Collados, Juho Kim, and Alice Oh. 2024. Exploring Cross-Cultural Differences in English Hate Speech Annotations: From Dataset Construction to Analysis. *arXiv preprint*. ArXiv:2308.16705.

[11] Jari-Mikko Meriläinen. 2024. The Role of Gender in Hate Speech Targeting Politicians: Evidence from Finnish Twitter. *International Journal of Politics, Culture, and Society*.

[12] Raphael Ottoni, Evandro Cunha, Gabriel Magno, Pedro Bernardina, Wagner Meira Jr., and Virgílio Almeida. 2018. Analyzing Right-wing YouTube Channels: Hate, Violence and Discrimination. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '18, pages 323–332, New York, NY, USA. Association for Computing Machinery.

[13] Paul Röttger, Bertie Vidgen, Dirk Hovy, and Janet B. Pierrehumbert. 2022. Two Contrasting Data Annotation Paradigms for Subjective NLP Tasks. *arXiv preprint*. ArXiv:2112.07475.

[14] Shoffan Saifullah, Nur Heri Cahyana, Yuli Fauziah, Agus Sasmito Aribowo, Felix Andika Dwiyanto, and Rafal Drezewski. 2024. Text annotation automation for hate speech detection using SVM-classifier based on feature extraction. *AIP Conference Proceedings*, 3167(1):040003.

[15] Anna Schmidt and Michael Wiegand. 2017. A Survey on Hate Speech Detection using Natural Language Processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pages 1–10, Valencia, Spain. Association for Computational Linguistics.

[16] Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis, and Sotiris Ioannidis. 2020. Analysis of Twitter and YouTube during USelections 2020. *arXiv preprint*. ArXiv:2010.08183.

[17] Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis, and Sotiris Ioannidis. 2023. What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020. *PLOS ONE*, 18(1):e0270542. Publisher: Public Library of Science.

[18] Galen Stolee and Steve Caton. 2018. Twitter, Trump, and the Base: A Shift to a New Form of Presidential Talk? *Signs and Society*, 6(1):147–165. Publisher: The University of Chicago Press.