

Analysis of Public Sentiment and Discussion Topics on the Israeli-Palestinian Conflict: An In-depth Study and Comparison of YouTube Comments Across Different News Channels

Final Project Computational Social Science course

Christian Braga, Matrikel Nr 12999840

March 18, 2025

This paper presents a comprehensive analysis of public sentiment and discussion topics regarding the Israeli-Palestinian conflict as manifested through comments on YouTube videos. Utilizing advanced natural language processing technologies, including BERT for sentiment analysis and BERTopic for topic modeling, this study explores the complex digital landscape of public opinion. The research analyzes the sentiments expressed across various news channels and over time, revealing how these factors contribute to shifts in public perception. The findings highlight a significant shift from neutral to more polarized views, underscoring the evolving nature of digital discourse in response to international events. By analyzing comments from diverse news sources, this study offers insights into the dynamic interplay between media framing and public sentiment, contributing to our understanding of the digital public sphere's role in shaping and reflecting contemporary views on longstanding geopolitical issues.

Table of contents

1	Introduction	3
2	Methodology	5
2.1	Transformers models	5
2.2	Sentiment Analysis: Bert	5
2.3	Topic Modelling: Bert-topic	6
3	Data	7
3.1	Data organization	7
3.2	Fine - Tuning Dataset	8
3.3	Inference datasets	8
3.4	Dataset Creation	9
4	Sentiment Analysis	10
5	Topic Modelling: BERT-Topic	14
6	Discussion	22
7	Conclusions	23
8	Bibliography	24

1 Introduction

The Israeli-Palestinian conflict, an enduring and complex territorial dispute, has shaped the Middle East for over a century. Originating from conflicting national aspirations and colonial legacies, this conflict involves deep-seated historical, religious, and social issues. A particularly pivotal moment in recent history was the attack on October 7 2023 [1], which dramatically escalated the conflict and captured global attention. This assault, characterized by its suddenness and the scale of violence, marked a significant turning point, bringing the long-standing conflict back into the spotlight on the world stage, particularly through social media channels.

In the current digital era, social media platforms like YouTube play a pivotal role in the dissemination of perspectives and the shaping of discourse surrounding such conflicts. These platforms provide a dynamic space where public sentiment and topical discussions are freely exchanged, offering valuable insights into global perceptions. The ability to analyze this sentiment in real-time helps not only in understanding global viewpoints but also in assisting institutions to fine-tune policies and communication strategies based on nuanced public opinions. This study focuses on extracting prevalent sentiments and discussion themes from comments across major news channels like Al Jazeera English, CNN, and Sky News.

Moreover, this study aims not only to bridge the gap between public perception and the complex dynamics of geopolitical conflicts but also to utilize recent technological advances in the field of advanced statistical models, specifically Transformers architectures [2]. Specifically designed for text processing and solving complex natural language processing (NLP) tasks, Transformers offer significant improvements over traditional methods like logistic regression and latent Dirichlet allocation (LDA) which have been widely used in similar analysis [3] for sentiment classification and topic modeling. But transformer models have proven to be superior to these classical approaches, providing more accurate and detailed results.

In the following analysis, I sought to answer various research questions designed to dissect different aspects of social media discourse. In particular i wanted to explore how digital platforms influence and shape public perceptions. The research questions are as follows:

Prevailing Sentiment and Primary Topics of Discussion: What is the prevailing sentiment among YouTube users regarding the Israeli-Palestinian conflict, and what are the primary topics of discussion for each side? This question seeks to identify the dominant emotions whether positive, negative, or neutral and the thematic concerns expressed by users in the comments sections of selected YouTube videos.

Evolution of Sentiment and Discussion: Has the sentiment and the topics of discussion evolved from the initial stages of the conflict to recent times? By examining comments from different periods, this study aims to trace any shifts in public opinion and the changing contours of the discourse over time.

Consistency Across News Channels: Is the sentiment expressed by YouTube users consistent across different news channels? Additionally, do users from different geographical or demographic backgrounds exhibit specific preferences for the news channels they use to gather information? This question addresses whether the portrayal of the conflict varies by news outlet and how that might influence viewer perceptions and preferences.

Perceptions of News Channel Bias: Is it possible to determine any relevant information about different news channels based on user sentiment and preferences, such as whether users perceive that an information network implicitly supports one faction over another? This involves analyzing comments for indications of perceived biases and evaluating how these perceptions might affect trust and credibility attributed to each channel.

By addressing these questions, the research aims to provide a nuanced understanding of how digital public spheres influence and reflect contemporary views on longstanding geopolitical issues. This inquiry not only enhances our understanding of the conflict's portrayal in media but also offers insights into the broader implications of digital media consumption on public opinion and international relations.

2 Methodology

2.1 Transformers models

The advent of transformer models in natural language processing (NLP) has marked a revolutionary shift in how text analysis is conducted, setting a new standard beyond the capabilities of traditional machine learning methods like Logistic Regression. Unlike earlier approaches that often relied on rigid, shallow learning mechanisms, transformers introduce a deep, context-aware architecture fundamentally enhancing the understanding and generation of text. Transformers, first introduced in the paper “Attention is All You Need” by Vaswani et al. (2017) [2], utilize a mechanism known as self-attention. This method allows the model to weigh the significance of different words in a sentence, irrespective of their positional distance from each other. Thus, unlike previous sequence-based models that processed data linearly (e.g., RNNs and LSTMs), transformers can interpret sentences in a non-linear fashion, capturing nuanced meanings more effectively.

This architecture has been particularly groundbreaking for its ability to handle large-scale language modeling tasks. This capability stems from two key aspects: 1. Pre-training and Fine-tuning Paradigm: Transformers are typically pre-trained on vast amounts of text in an unsupervised manner, learning a general understanding of language syntax and semantics. Once pre-trained, these models can be fine-tuned with smaller, task-specific datasets, which is highly efficient for adapting the model to specific NLP tasks such as sentiment analysis, question answering, or text summarization. 2. Text Classification Capabilities: For text classification, transformers represent a significant advancement. By understanding context across longer stretches of text and capturing subtler nuances in language, they provide superior performance in distinguishing sentiments, categorizing text, and identifying thematic elements. This deep contextual awareness allows for more accurate and nuanced classifications than was possible with traditional models, which often misinterpret the context of words.

2.2 Sentiment Analysis: Bert

To perform the sentiment analysis I used the BERT model. BERT, which stands for Bidirectional Encoder Representations from Transformers, is a groundbreaking model in NLP introduced by researchers at Google in 2018 [4]. It represents a significant advancement in the field due to its deep learning framework that fundamentally processes words in relation to all the other words in a sentence, rather than in isolation. The main characteristics and Strengths of this transformers model are:

Bidirectional Context: BERT’s major innovation is its bidirectional training of the transformer, an approach that allows the model to learn the context of a word based on all of its surroundings (left and right of the word).

Fine-Tuning Adaptability: Once pre-trained on a large corpus of text, BERT can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, including sentiment analysis, without substantial modifications to the architecture. This makes BERT incredibly versatile and efficient for adapting to any specific NLP task.

The pretrained model that can be imported from the Python library transformers has been fine-tuned on a dataset to which I manually created labels, thus enabling training in a supervised manner and creating a model capable of generalizing and recognizing the sentiment expressed in various comments about this conflict. The model was finetuned using a Google Colab GPU, specifically the NVIDIA Tesla T4, which offers substantial

computational power, accelerating the training process significantly and efficiently handling the demanding tasks of training deep learning models.

The three classes of comments that the model will attempt to classify are:

Class	Description
0	comment that support the Israeli cause
1	comment that supports the Palestinian cause
2	neutral comment

Finally, the model's accuracy will be measured using the accuracy score and a confusion matrix.

2.3 Topic Modelling: Bert-topic

In the context of this project to perform topic modelling i used another transformer model: BERTopic. BERTopic is an advanced modeling tool used for identifying and visualizing topics from large collections of textual data. Unlike traditional clustering algorithms which often rely on bag-of-words approaches, BERTopic leverages state-of-the-art language models based on BERT to enhance the understanding of textual context and semantic relationships.

BERTopic utilizes dimensionality reduction and clustering techniques to organize text data into distinct topics, making it highly effective for extracting meaningful insights from unstructured text [5]. This is particularly valuable as it allows for the automated grouping of comments into thematic categories that represent, in this case, the various viewpoints and topics of discussion related to the Israeli-Palestinian conflict.

3 Data

3.1 Data organization

As previously mentioned, this analysis aims to capture public opinion through their interactions on the on-line platform: YouTube. YouTube is an online video-sharing platform where users can upload, view, rate, share, and comment on videos. Launched in February 2005, it offers a wide range of user-generated content, including music videos, movie clips, live broadcasts, and news reports, making it a major source of entertainment and information in the digital world. Specifically, these analysis targets users of this platform who have interacted through the ‘comment’ function with videos concerning the Israeli-Palestinian conflict. The target population, therefore, consists of YouTube users, with the sample represented by users who have commented on the target videos, and the fundamental unit of analysis is the expressed comment.

However, the data provided by the platform does not allow tracking the geographical origin of the user or to have more information about them. Consequently, in order to answer the research questions, I had to adopt some measures that allowed me to organize the data in such a way that it could provide as global a view as possible of opinions regarding the conflict.

For this, I started from the assumption that YouTube users are more likely to consume content produced by news networks that have the same origin or the same vision as them. By considering videos published by networks from different parts of the world, I was able to construct a sample that was homogeneous in terms of discussion topics but heterogeneous in terms of the types of users involved.

The news networks from which I selected videos for analysis (which we will identify as ‘Inference Datasets’ in the context of this study) were:

Al Jazeera English, an international news channel in English, part of the Al Jazeera media network based in Doha, Qatar. Launched in 2006, the channel provides global news coverage, focusing attention on events in the Middle East. I chose this network under the assumption that it was visited mostly by users from the Middle East, given its origins and publishing topics.

CNN (Cable News Network) is an American cable and satellite news channel founded in 1980. CNN offers news, reporting, and analysis on politics, economics, and much more, serving a global audience and thus providing a perspective on the opinions and news disseminated by one of the main American news channels.

Sky News is a British news channel, launched in 1989 by Sky Group. It provides continuous updates on international, political, economic, and general news events. By selecting Sky News, I tried to provide a perspective from the European continent.

Furthermore, in order to deepen the analysis and also investigate the evolution of sentiment and discussion topics over time, I selected two videos for each news channel, the first dated October 2023 in conjunction with the attack by Hamas-led gunmen who breached the border fences between Gaza and Israel [1]. Thus, within the context of our analysis, this moment was identified as the starting point for user opinions regarding the conflict. The second video for each channel is dated 2024, approximately one year after that event. In this way, I was able to obtain estimates that also considered the temporal aspect and identify possible interesting information on the evolution of sentiment and discussion topics.

An additional selection that had to be made at the level of individual comments concerned the language of expression. To ensure the effective functioning of models (BERT and BERT-topic), which are designed primarily for English language processing, I had to consider only comments expressed in English. This

choice was driven by the need to maintain the accuracy and reliability of the topic modeling and sentiment analysis, as these models are optimized for English. In conclusion, within the context of the analysis, it was necessary to create two types of datasets: one dataset for fine-tuning BERT, and a second type that I called the inference dataset, composed of data taken from the previously described networks, on which I actually performed the actual analysis.

3.2 Fine - Tuning Dataset

The initial dataset created was essential for fine-tuning the BERT model. To accomplish this, it was crucial to select videos related to the Israeli-Palestinian conflict so that the model could acquire knowledge about the discussion topics and themes addressed. Specifically, videos containing highly polarizing comments rich in discussion topics were selected. This was fundamental to developing a model capable of generalizing and performing well on new comments. To this end, I manually searched for videos meeting these criteria, identifying six from which I extracted comments and mixed them randomly to ensure the creation of a uniform sample.

The videos selected for creating the fine-tuning dataset include:

- “October 7 | Al Jazeera Investigations” [6]
- “ Hamas militant’s bodycam shows how attacks on Israel began” [7]
- “Investigating war crimes in Gaza I Al Jazeera Investigations” [8]
- “ ‘People in Gaza feel abandoned by the world’: 40,000 Palestinians killed as ceasefire talks resume” [9]
- “Gaza towers collapse after explosion” [10]
- “Israel-Gaza: At least half of Gaza’s buildings damaged or destroyed, new analysis shows | BBC News” [11]

This process yielded a dataset composed of 19,121 comments, from which I manually labeled 1,486 comments, constructing the actual dataset for fine-tuning. The resulting model achieved an accuracy of 76% on the test set, a fairly good outcome considering the complexity of the classification task. In fact, I found that the comments often contain elements of irony, sarcasm, and allusions, which make the model’s task particularly challenging.

3.3 Inference datasets

For the selection of datasets on which to perform the analysis, as previously mentioned, I chose videos from the channels: Al Jazeera English, CNN, and Sky News. For each channel, I selected two videos: one published in October 2023, and another approximately one year later. Another consideration in this context was selecting videos that had at least 1000 comments, ensuring that they constituted a sufficiently large sample.

Inference Datasets:

	Title	News Channel	Pubblication Date
1	[6] October 7 Al Jazeera Investigations	Aljazeera English	20-03-2024
2	[12] Latest live Update: Over 700 killed by Is...	Aljazeera English	10-10-2023
3	[9] People in gaza feel abandoned 40k palestin...	CNN	15-08-2024

	Title	News Channel	Publication Date
4	[13] Surprise Attack on Israel October 9, 2023	CNN	9-10-2023
5	[14] Gaza before and after 7 October Israel-...	Sky News	6-10-2024
6	[15] Bombs rain down on Gaza as Hamas and Isra...	Sky News	10-10-2023

By doing this, I obtained two videos for each news channel, providing the opportunity to investigate potential correlations or to aggregate the data over time. For each dataset, the variables considered were: ‘text’ containing the text of YouTube comments and ‘sentiment’ containing the classification of the specific comment.

3.4 Dataset Creation

To create all the comment datasets, it was necessary to utilize the API provided by YouTube. The process for creating the fine-tuning dataset was as follows:

1. **API Key:** I utilized the `googleapiclient.discovery` library to access YouTube’s API, facilitating automated comment retrieval from specified videos. A crucial step in this process was the creation of an API key, a unique identifier that allowed secure and authorized access to YouTube’s video data, ensuring data retrieval was secure without exposing sensitive account details.
2. **Fetching Comments:** To systematically gather comments, first i extraced the video id from each youtube’s video, then a Python function `get_video_comments` was defined. The function iterates over all available comments for a video, extracting crucial information such as the comment text, number of likes, number of replies, and the publication date.
3. **Data Aggregation and Preprocessing:** After collecting comments from various videos, the datasets from different video sources were concatenated to form a comprehensive dataset. This dataset was then shuffled to ensure a random distribution of data, which helps in unbiased model training.
4. **Language Filtering:** To maintain consistency in language processing and enhance the accuracy of the sentiment analysis model, a filtering step was included using the `langdetect` library. This library helped identify and retain only those comments written in English, ensuring the model trained on a uniform dataset.
5. **Sentiment Column Addition:** For the purpose of model training and subsequent predictions, a ‘sentiment’ column was added to the dataset. Initially, this column was populated with default values, which were later replaced with actual sentiment labels either through manual labeling or automated predictions.

While for the inference datasets i followed the same process but without shuffled the data and without aggregating the comments of the different videos in one dataset. Before applying the models, the comments underwent a series of preprocessing steps to prepare them for analysis. This preprocessing involved converting all text to lowercase for uniformity and removing any URLs to clean the data. Additionally, unnecessary spaces were reduced to a single space, and mentions and hashtags, indicated by ‘@’ or ‘#’, were stripped from the text. To facilitate effective model processing, the text was tokenized using the BERT model’s tokenizer (`bert-base-uncased`), which breaks down the text into tokens suitable for input into the model.

4 Sentiment Analysis

Once I had trained the BERT model on the fine-tuning dataset, I used it to make predictions on the inference datasets. At this stage, one interesting aspect to investigate is the overall distribution of sentiment across all the videos I selected. This allows us to gain a global view of the opinions of YouTube users and to conduct further detailed analysis. To achieve this, I created a bar graph that combines all the datasets.

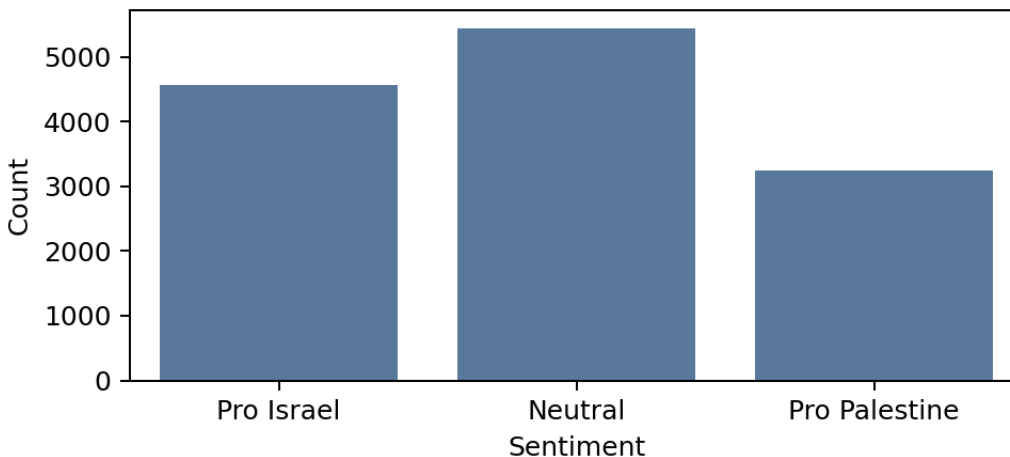
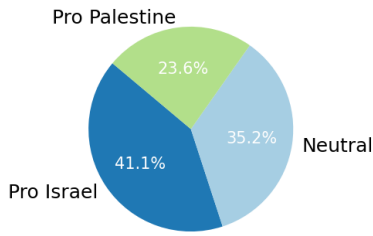


Figure 1: Distribution of Sentiment

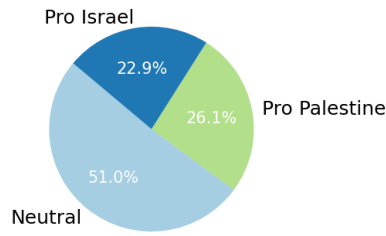
Analyzing the sentiment across all sampled datasets, it is evident that the majority of comments express a neutral opinion. This observation suggests that comments under YouTube videos serve as a platform where individuals tend to express their thoughts in a manner that may encourage dialogue rather than merely passing judgment or condemning what they see. Naturally, the resultant sentiment is influenced by the sample of videos we have selected. Additionally, it is noteworthy that the number of comments supporting Israel exceeds those in favor of Palestine. In subsequent analyses, we will delve deeper into this phenomenon to understand these data thoroughly.

The prevalence of neutral comments might stem from various factors. Firstly, viewers who comment might prefer to discuss the content of the videos in a balanced or analytical way, avoiding polarizing opinions to foster a more constructive conversation. Secondly, neutral comments could be indicative of viewers trying to understand both sides of the conflict without showing explicit bias. This neutrality might reflect a broader desire among YouTube users to engage in discussions that are informed and thoughtful rather than confrontational. To gather more detailed insights about the sentiment expressed by users, I further investigated the distribution of their opinions across different news channels.

Aljazeera Sentiment Distribution



CNN Sentiment Distribution



Sky Sentiment Distribution

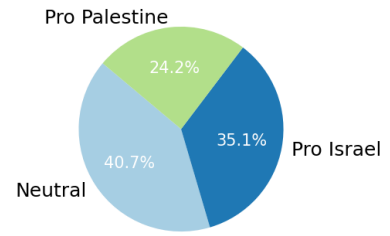


Figure 2: Distribution of Sentiment across news channels

Analyzing these charts reveals some intriguing data. We can somewhat confirm what was previously observed regarding the prevalence of neutral comments, which are proportionally the most frequent among the videos published by both Sky News and CNN. However, for Al Jazeera, the neutral comments are outnumbered by those supporting Israel. This outcome indicates that this news channel generates greater polarization of opinions; users who commented on Al Jazeera's videos tend to express their support for a particular cause more than presenting neutral arguments. The larger influx of comments supporting one faction over another can primarily be attributed to two factors: either support for the perspectives presented in the videos or strong opposition. Subsequent topic modeling analyses will be interesting to investigate and determine if we can identify patterns of preference or opposition regarding one channel over another.

Focusing on the comments that support one faction over another, it can be noted as anticipated that Al Jazeera has the highest proportion of pro-Israel comments at 41.1%, followed by Sky at 35.1% and CNN at 22.9%. Regarding pro-Palestine comments, CNN videos contain the most frequent pro-Palestine comments at 26.1%, followed by Sky at 24.2% and Al Jazeera at 23.6%, confirming our previous findings that Pro-Israel comments outnumber those supporting Palestine.

From the results obtained, it appears that there is a sort of contradiction between the channels Al Jazeera and CNN. The former indeed has the highest presence of pro-Israel comments and the lowest influx of pro-Palestine comments among the three networks. In contrast, CNN has the highest frequency of pro-Palestine comments and the lowest of pro-Israel. This aspect is important and allows us to understand that users who typically interact with these channels hold opposing views. As previously deduced, they may either show strong reluctance towards the content presented or, conversely, strong support. The topic modeling analysis will help us understand this better.

Another interesting point to analyze concerns the evolution of sentiment a year after the conflict began:

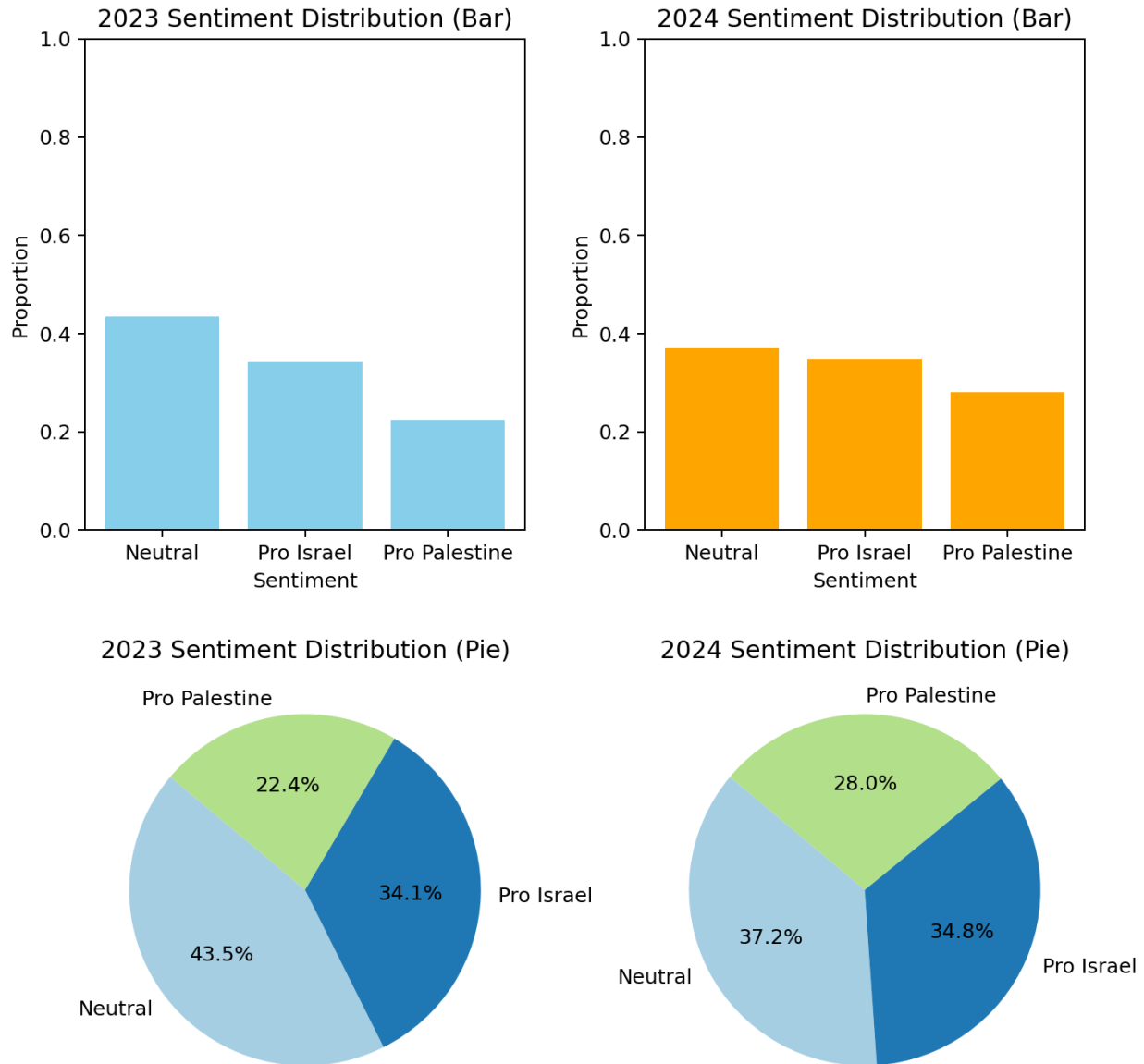


Figure 3: Evolution of the sentiment over time

As we can see from Figure 3, one year after the October 2023 attack, which we have identified as the onset of the conflict in our analysis, the sentiment expressed by users has changed. The first noticeable shift is the decrease in the proportion of neutral comments from 43.5% in 2023 to 37.2% in 2024. This could indicate that as time passes and the conflict progresses, users have formed their own opinions, resulting in increased tension. This tension is reflected in the rise of opinions supporting one faction over another. Indeed, the bar chart shows that the expressed sentiment has become more evenly distributed.

Examining the evolution of pro-Palestinian comments, there has been an increase of nearly 6%, a rise not mirrored in pro-Israel comments, which have remained almost unchanged. Given the previous decrease in neutral comments, these data suggest that after one year of conflict, the Palestinian cause appears to be more deeply felt. Based on our sample, we might infer the presence of a trend where users who initially had a neutral opinion have shifted to support the Palestinian cause. However, we cannot confirm this conclusively.

as our analysis is influenced by the fact that while the selected videos consistently deal with the Israeli-Palestinian conflict, they each present different topics, introducing bias into these considerations.

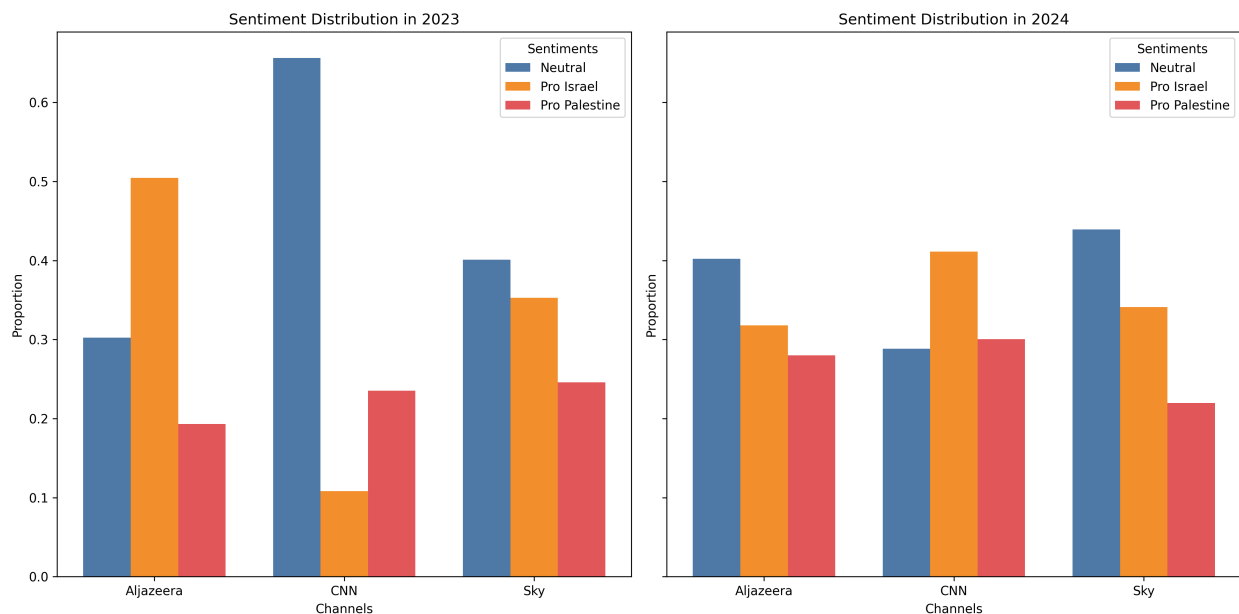


Figure 4: Sentiment distribution across channels over time

Investigating the evolution of sentiment across different news channels, we observe distinct trends. For Al Jazeera, there has been a marked decrease in pro-Israel comments and an increase in pro-Palestinian opinions, aligning with the previously identified trend. The evolution of sentiment on the CNN channel also conforms to what we observed in the general analysis. We notice a significant decrease in neutral comments on this channel, a sign of increasing polarization that manifests with a substantial increase in pro-Israel comments, making it the channel with the largest proportion under this aspect, along with a rise in pro-Palestinian comments. Analyzing Sky News, we find that the distribution of expressed sentiment has remained more or less unchanged, showing over time a higher incidence of neutral comments, followed by pro-Israel and then pro-Palestinian comments.

To conclude this temporal analysis by specific channel, we can affirm that the evolution of the conflict has led to greater polarization in the opinions expressed by YouTube users. CNN emerges as the channel where this trend is most pronounced, being the only one where comments supporting a faction outnumbered neutral ones in 2024 despite having the highest incidence of the latter in 2023. Sky News has remained the most balanced channel over time, and Al Jazeera has seen the greatest percentage increase in pro-Palestinian comments from 2023 to 2024 and the largest decrease in pro-Israel comments among the three, yet still retaining a majority.

5 Topic Modelling: BERT-Topic

The sentiment analysis has provided us with valuable insights into the distribution of sentiments. Now, using the BERTopic model, I have explored the main topics of discussion for each sentiment category. To get an initial overview of the discussion topics, I created a simple word cloud that allows us to see the most frequent words:



Figure 5: Word Cloud on the entire dataset

The word cloud clearly shows that the most frequently used words are “Israel,” “Hamas,” “people,” “Gaza,” and “War,” indicating that the discussions are focused on the geopolitical dynamics and human impacts of the conflict. Words such as “killed,” “people,” and “peace” emphasize significant concerns for human welfare and the active advocacy for action and support among viewers. To further investigate the topics of discussion across the entire dataset, I conducted a topic modeling analysis using the BERTopic model.

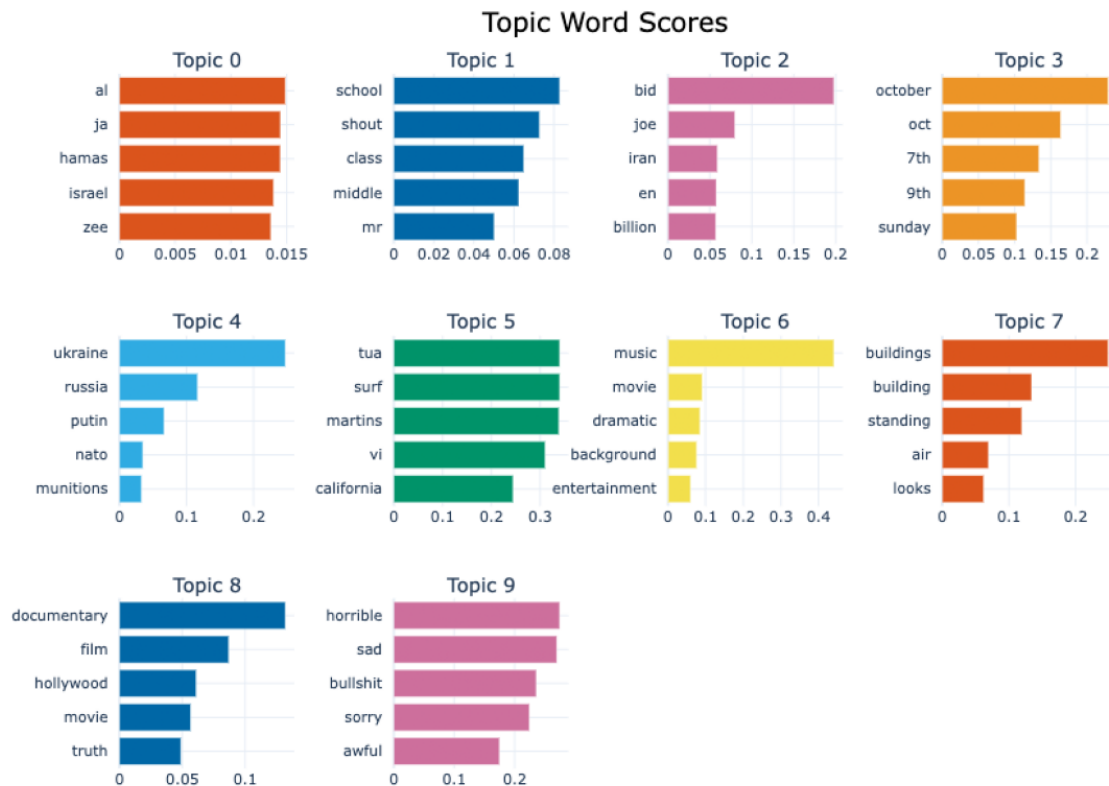


Figure 6: Top 10 discussion topics

Analyzing the ten most frequent discussion topics across the entire dataset, we can observe some intriguing insights. Topic 0, which represents the most common discussion topic, is understandably focused on the rival factions, Hamas and Israel. Subsequent topics include terms such as “school,” “shout,” “middle,” “class,” which might indicate discussions about educational or societal issues, potentially reflecting situations or events occurring in schools or involving children in conflict zones. Topic 2: Featuring words like “bid,” “joe,” “iran,” “billion,” and “en,” could be linked to international relations or economic discussions, possibly involving US policies or financial aspects concerning the Middle East. Topic 3: With words such as “October,” “oct,” “7th,” “9th,” and “Sunday,” this topic seems to capture conversations related to specific dates or events, likely referring to the October 7 event that marked the escalation of the conflict. Discussion topics also include words like “putin,” “nato,” “Russia,” “Ukraine,” likely making references or analogies to another major conflict currently ongoing in Europe.

In conclusion, we can see that emotional expressions are common, likely relating to the content of the videos or the events of the conflict, with words like “horrible,” “awful,” “sorry” demonstrating the emotional involvement of users with the reported events, or words like “movie” and “Hollywood” likely convey a widespread sense of disbelief about the events being reported. It’s interesting to note that this emotional involvement is less frequent compared to more complex discussions concerning topics such as the geopolitical landscape where the conflict unfolds and the conditions of children/schools, which were probably the subject of some videos that triggered strong reactions.

Moving forward, to determine the specific discussion topics for each faction, I conducted the same analysis specifically for the comments categorized as Pro Israel, Pro Palestine, and Neutral.



Figure 7: Pro Israel discussion topics

Analyzing the discussion topics in comments that support Israel, we can observe several key points: Topic 0 is centered around keywords such as “Hamas,” “Israel,” “Gaza,” and “people,” indicating that the discussion focuses on the principal actors of the Israeli-Palestinian conflict. The presence of geographical terms and names of involved groups suggests a direct analysis of the political and social dynamics of the conflict. Topic 1 focuses on educational themes or school-related contexts, with words like “shout,” “school,” “elementary,” “class,” and “grade.” As we have seen previously, it is a very relevant topic of discussion. Topic 2 deals with narratives of hostages and victimization, using terms like “hostages,” “don,” “victim,” “hostage,” and “free.” This indicates a significant concern among supporters of the Israeli cause about the issue of hostages and their liberation, likely referring to the events of October 7, 2023, when Hamas militants captured numerous hostages following the incident. Topic 3 is tied to specific temporal events, with terms like “October,” “8th,” “happened,” “7th,” and “Sunday.” This implies discussions on events that occurred on particular dates. Topics 5 and 6 display a mix of more casual or generic terms like “haven,” “se,” “guess,” “saw,” “won,” and words related to games or competitions such as “stupid,” “games,” “prizes,” “play,” “win.” These topics might reflect how users use games and competition as metaphors to discuss the conflict, or they might simply divert attention from the heavier themes of the conflict.

In summary, the analysis shows that pro-Israel comments cover a range of themes from political and historical to personal and metaphorical, with a strong focus on specific events and emotional reactions. This reflects the complexity of discourse surrounding the conflict, where perceptions, personal experiences, and collective history are intricately interwoven.

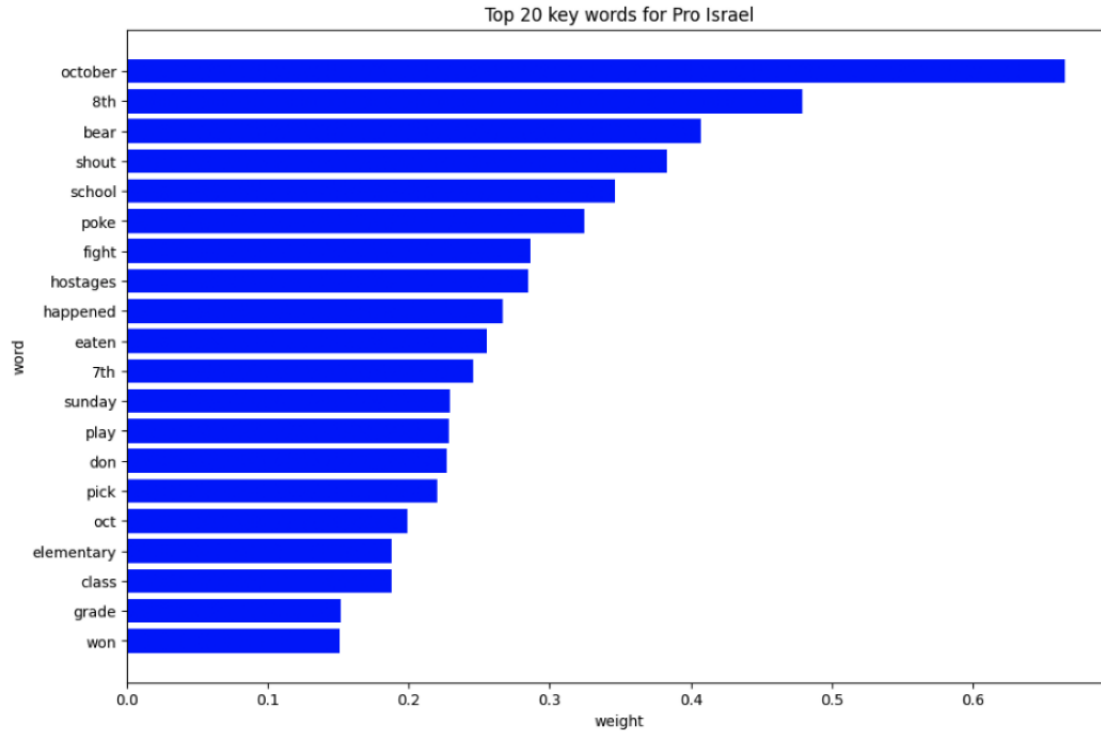


Figure 8: Top 20 words in Pro Israel comments

Analyzing the most frequent words in comments supporting the Israeli cause confirms much of what was previously stated. Heavier weighted words such as “October,” “8th,” and “bear” indicate a focused discussion on specific events, possibly referring to significant dates or occurrences during the month of October. The use of words like “shout” and “fight” suggests a tone of confrontation or direct action, while terms like “hostages” and “eaten” might reflect narratives of conflict and perceptions of aggression or victimization. Words such as “school,” “elementary,” and “class” are highlighted once again, suggesting that discussions may also touch on the impacts of the conflict on everyday life, particularly the education of young people.

In summary, the graph displays a mixture of terminology that reflects the intensity of the discussions, with a combination of direct references to the conflict, specific events, and educational themes. This variety of keywords indicates that, while predominantly focused on supporting Israel, pro-Israel commentators incorporate a range of considerations into their messages.

Moving on to comments supporting Palestine:



Figure 9: Pro Palestine discussion topics

Analyzing the comments supporting Palestine, we can see that while pro-Israel comments focus on conflict, victory, and direct discussions of events, pro-Palestinian comments touch on a more varied and complex array of themes, including historical, educational, and humanitarian aspects.

Topics 0 and 1: Similar to the pro-Israel comments, these topics include words such as “Israel,” “Gaza,” “people,” and “Palestine,” indicating a direct discussion about the conflict. In Topic 3, words like “liar,” “unk,” “real,” “tine,” and “les” suggest a debate over the truthfulness and authenticity of information, likely reflecting Palestinian supporters’ disagreement with the reported facts and their skepticism about the reliability and impartiality of the media channels disseminating this information. Subsequent discussion topics interestingly incorporate religious themes (“Jesus,” “Christ,” “God”) and concepts of love and peace (“love”), which might indicate that some users, regardless of their own religion, express feelings of closeness and love toward the Palestinian people. To conclude, in Topic 8 we observe that words like “nazis,” “hitler,” “holocaust” are used by supporters of the Palestinian cause likely to draw parallels between the actions of their adversaries and the grievous acts committed during World War II.

Regarding the most frequent words in comments supporting Palestine, they do not provide additional information beyond what has already been observed, but they do confirm the trends previously identified.

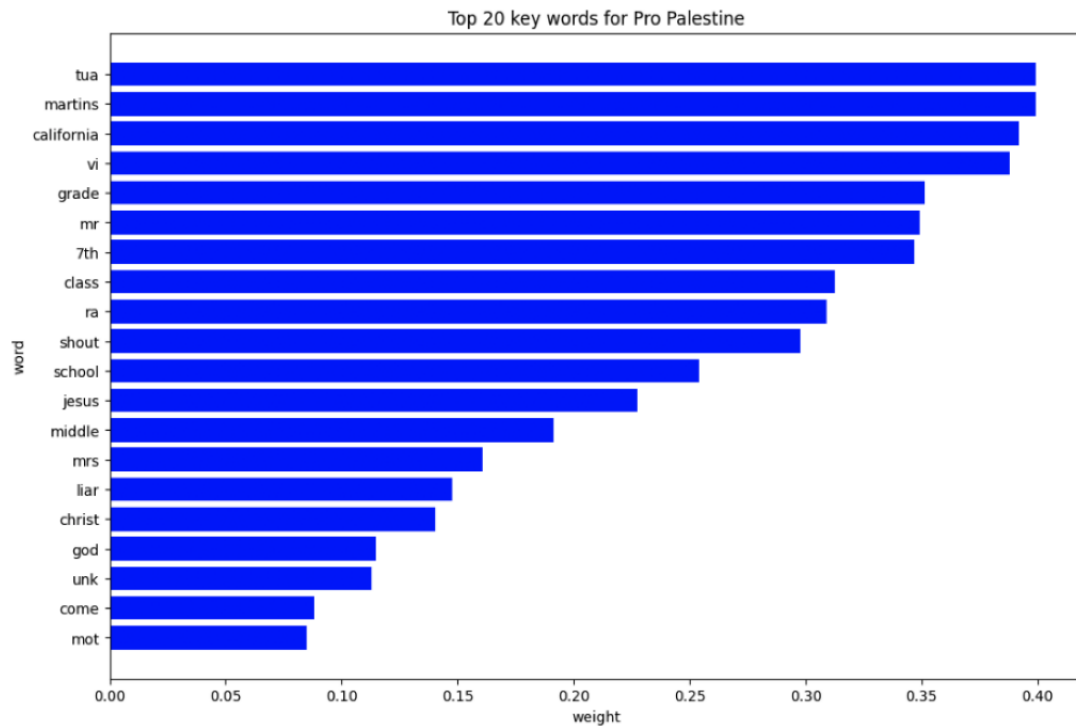


Figure 10: Top 20 words in Pro Palestine comments

Considering now the neutral comments:



Figure 11: Neutral discussion topics

In the discussion topics of comments classified as neutral, we first notice a tendency to discuss news networks. Abbreviations like ‘al’, ‘ja’, ‘zee’ seem to suggest discussions related to Al Jazeera, similarly ‘cnn’, ‘news’, ‘propaganda’, confirm the presence of contrasting opinions among users regarding the information disseminated on YouTube. Indeed, some networks are often considered biased and accused of not presenting information impartially or spreading facts deemed untruthful. The topic relative to “school”, “class” it appears to be very relevant regardless of the expressed sentiment, further confirming how deeply this theme resonates. Topics 4 and 5 focus on time-related themes and the family or victims of the conflict. The dates in Topic 4 may refer to specific events, while words like “children,” “kids,” “heart,” “killed,” and “innocent” in Topic 5 emphasize discussions on the human consequences of the conflict. Words like “sad,” “sorry,” “load,” “sham,” and “mm.” reveal an emotional response to the conflict or its portrayals, showing compassion or disappointment.

Unlike the pro-Israel and pro-Palestine comments, neutral comments tend to display a wide variety of perspectives that do not merely support a specific cause. Instead, they explore the broader implications of the conflict, the role of the media, cultural impacts, and exhibit a sensitivity toward the victims and human consequences of the conflict. This indicates a more holistic and reflective approach to the conflict, highlighting an interest in understanding its deeper causes.

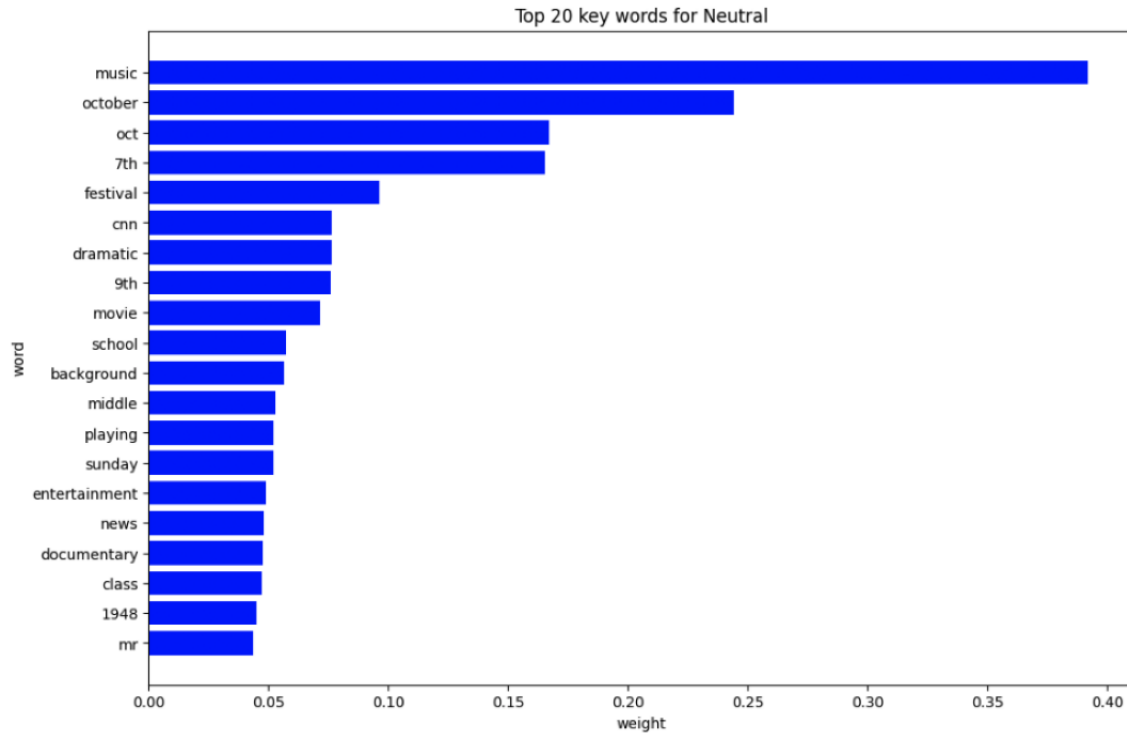


Figure 12: Top 20 words in Neutral comments

As previously highlighted, the most frequent words confirm the trend where neutral comments focus on a broader, less polarized reflection of the conflict, considering history, education, and media as central elements to understand and discuss the Israeli-Palestinian context. To conclude, this analysis reveals distinct narratives within comments categorized as pro-Israel, pro-Palestine, and neutral, reflecting the complexity and variety of user perspectives and reactions.

Pro-Israel Comments: The discussion in this category focuses on the main actors of the conflict, with frequent mentions of Hamas, Gaza, and Israel, indicating a direct discussion on the political and social dynamics. There are also themes of hostages and victimization, showing concerns for security and war actions, as well as discussions related to specific temporal events that have marked moments of tension and themes concerning children/schools in conflict zones.

Pro-Palestine Comments: In pro-Palestine comments, themes emerge that include the history and humanitarian aspects of the conflict, with a strong presence of religious components and discussions on justice and human rights. Keywords such as “Jesus,” “Nazis,” and “holocaust” indicate a tendency to link the Palestinian situation to historical contexts of oppression.

Neutral Comments: These show a wide range of themes that go beyond the direct conflict, with discussions on media, education, and culture. Words like “school,” “CNN,” and “documentary” suggest a debate on how the conflict is represented and perceived, while dates and mentions of specific events highlight sensitivity to the temporal context of events.

6 Discussion

In this study, we explored the intricate dynamics of public sentiment and discussion topics regarding the Israeli-Palestinian conflict as expressed through comments on YouTube videos. Our investigation utilized advanced NLP techniques, leveraging BERT for sentiment analysis and BERTopic for topic modeling, to dissect the varied perspectives reflected in the digital public sphere. The research addressed several key questions, providing a multi-dimensional analysis of how sentiments and discussions have evolved and varied across different viewer demographics and news channels.

The main findings that allowed me to answer the research questions are:

- **Prevailing Sentiment and Primary Topics of Discussion:** The analysis revealed a predominance of neutral sentiments, indicating a preference among YouTube users for balanced discussion or a reflective approach rather than direct confrontation. Pro-Israel and pro-Palestine sentiments were also notable but to a lesser extent. The primary topics of discussion included geopolitical dynamics, historical contexts, and human impact, with frequent mentions of key figures and locations such as Hamas, Israel, and Gaza. Discussion topics concerning the presence of children and schools in conflict areas were deeply felt, along with discussions about perceptions of the information provided by different channels.
- **Evolution of Sentiment and Discussion:** Over time, there was a noticeable shift from neutral to more polarized sentiments, reflecting an increase in user engagement and possibly growing tensions or heightened awareness of the conflict. This evolution suggests that digital platforms like YouTube not only serve as arenas for information dissemination but also for dynamic interaction that can shift public perceptions and engagement over time. More specifically, we observed an increase in support for the Palestinian cause over time.
- **Consistency Across News Channels:** Sentiments varied across different news channels. For instance, we noted a strong contrast between the channels Al Jazeera and CNN. The former indeed has the highest presence of pro-Israel comments and the lowest influx of pro-Palestine comments among the three channels. In contrast, CNN has the highest frequency of pro-Palestine comments and the lowest of pro-Israel. Based on the discussion topics analyzed, we can deduce, given the higher incidence of comments expressing concepts of opposition rather than support, that Al Jazeera presents topics that are more favorably viewed by supporters of the Palestinian cause, while the opposite is for CNN.
- **Perceptions of News Channel Bias:** The analysis revealed that users actively discuss the authenticity of the information provided by different channels, thus recognizing the presence of bias. Comments often highlighted perceived favoritism towards either Israel or Palestine, demonstrating that viewers critically engage with news content, not merely as passive recipients but as discerning critics of media reliability. Specifically, the channels most frequently discussed were Al Jazeera and CNN, while no particular criticisms were noted regarding Sky News.

7 Conclusions

The findings from this study illuminate the significant role that digital platforms play in the contemporary discourse surrounding international conflicts like the Israeli-Palestinian conflict. YouTube, as a global platform, offers a unique lens into the public's perceptions and sentiments, which are deeply influenced by historical narratives, media framing, and individual worldviews. The shift from neutral to more defined sentiments over the study period highlights the evolving nature of public engagement with such conflicts, suggesting a growing propensity among viewers to align themselves more distinctly with one side or the other as the conflict progresses.

Furthermore, the distinct topics and sentiments articulated across different viewer demographics and news channels emphasize the fragmented nature of public opinion on this issue, shaped by a complex interplay of cultural, social, and political factors. This fragmentation is also a reflection of the broader global discourse on peace, justice, and international relations.

Ultimately, this study not only advances our understanding of how the Israeli-Palestinian conflict is represented and perceived in the digital age but also sheds light on the broader implications of how modern conflicts are discussed and understood in public forums. These insights are crucial for policymakers, media professionals, and scholars interested in the intersections of media, public opinion, and international conflict.

8 Bibliography

- [1] https://en.wikipedia.org/wiki/7_October_Hamas-led_attack_on_Israel
- [2] 12 Jun 2017, Attention Is All You Need. Ashish Vaswani, Noam Shazeer et al.
- [3] Multi-Level Analysis of political sentiments using twitter data: a case study of the palestinian- israeli conflict, Iyad Al-Agha and Osama Abu-Dahrooj, December 2019.
- [4] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- [5] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure.
- [6] October 7 | Al Jazeera Investigations, https://www.youtube.com/watch?v=_0atzea-mPY&rco=1.
- [7] “ Hamas militant’s bodycam shows how attacks on Israel began”, https://www.youtube.com/watch?v=nDn10nDn_k&t=1s.
- [8] “Investigating war crimes in Gaza I Al Jazeera Investigations”,
<https://www.youtube.com/watch?v=kPE6vbKix6A&t=2837s>
- [9] “ ‘People in Gaza feel abandoned by the world’: 40,000 Palestinians killed as ceasefire talks resume”,
https://www.youtube.com/watch?v=Da_L17P5kYU
- [10] “Gaza towers collapse after explosion”, https://www.youtube.com/shorts/c5tWYj_Y60w
- [11] “Israel-Gaza: At least half of Gaza’s buildings damaged or destroyed, new analysis shows | BBC News”,
<https://www.youtube.com/watch?v=cONhigj9Po4>
- [12] “Latest live Update: Over 700 killed by Israeli forces air strike on Gaza”,
<https://www.youtube.com/watch?v=XZHXUvBcyhE>
- [13] “Surprise Attack on Israel | October 9, 2023, <https://www.youtube.com/watch?v=PuTn9g-KfR0&t=3s>
- [14] “Gaza before and after 7 October | Israel-Hamas War”, <https://www.youtube.com/watch?v=eSKq2IjmmJc&t=3s>
- [15] “Bombs rain down on Gaza as Hamas and Israel war escalates”, <https://www.youtube.com/watch?v=kBf3jm8OKyo>