Master: Social Data Science

Subject: BD and AI: Social and Evaluation Research – Prof. Mazzeo

Students: M. D'Amato, F. Guardione, V. Mascellaro, D. Montepara, N.Vuolo

Sentiment analysis of social media discourse: a case study of Cecilia Sala on BlueSky and YouTube

Introduction

Social media platforms have become powerful arenas for public discourse, shaping how individuals and society perceive events, people, and narratives. In this context, sentiment analysis offers valuable insights into the language and attitudes that emerge in digital conversations.

This study focuses on the case of Cecilia Sala, a 29 years old Italian journalist and war correspondent, who was detained in Tehran on December 19, 2024, three days after arriving on a journalist visa. She was accused of "violating the laws of the Islamic Republic," though specific allegations were not disclosed. Her arrest occurred shortly after Italian authorities detained Mohammad Abedini Najafabadi, an Iranian engineer, in Milan at the request of the United States. The Italian government, led by Prime Minister Giorgia Meloni, engaged in intensive diplomatic efforts to secure her release, including summoning the Iranian ambassador and negotiating potential diplomatic solutions.

Cecilia Sala's work in reporting from conflict zones highlights broader issues related to international negotiations and the role of journalists in disseminating critical information during crises. Her reporting brings into sharp relief the complexities of geopolitical dynamics, however, the discourse surrounding Sala on social media also unveils another layer of analysis: the intersection of gender and professional critique. Research has repeatedly demonstrated that women in public-facing roles are subject to distinct and often harsher language, characterised by biases and stereotypes that reflect persistent gender inequalities.

This paper aims to explore three key dimensions of social media discourse through the lens of sentiment analysis:

1. The framing of issues related to international negotiations
2. The specific language and sentiment employed when addressing Sala as a woman journalist
3. The broader perception of journalists operating in high-risk environments

BlueSky, a decentralized social media platform designed to prioritize user agency and reduce centralized control, presents a unique opportunity to study discourse in its early stages of community formation. Unlike larger, algorithm-driven platforms, BlueSky fosters a more conversational and participatory environment, where users engage in text-based dialogue without the same level of algorithmic amplification or monetization pressures.

This study examines the discourse on BlueSky to identify emerging sentiment patterns and evaluate whether the platform's structural and cultural differences influence public perceptions of Cecilia Sala, particularly concerning her reporting on international negotiations and her role as a journalist in high-risk areas.

YouTube, in contrast, is a well-established platform that prioritises visual and performative content. The analysis of YouTube content centres on comments and viewer interactions with Sala's videos or related discussions. As a platform optimised for video sharing and storytelling, YouTube frequently amplifies narratives through visuals and personal engagement. This study explores how the format of video content shapes the sentiment expressed towards Sala, particularly in terms of gendered language and critiques related to her professional role.

The integration of data from these two platforms is undertaken in order to reveal the nuances of how the format and culture of social platforms affect public sentiment. In doing so, an investigation is made into how societal attitudes toward gender, journalism, and international reporting are mediated by the structure and norms of both emerging and established social media spaces.

Methodology

The present study analysed a total of 1,113 comments collected from 10 **YouTube** videos selected based on a search using specific keywords: "Cecilia Sala," "Meloni," "Abedini," "journalism," "woman". The videos were chosen for their relevance to Cecilia Sala's case and related topics based on number of visualizations, ensuring a diverse representation of opinions expressed in the public discourse.

We expanded our analysis by including a dataset of 1,480 posts retrieved from the social media platform **BlueSky**, an emerging service in the social media landscape, in a period between 15/12/2024 and 15/01/2025. To collect these data, we used the official BlueSky query APIs, setting the search query to “Cecilia Sala”.

The aim of this phase was to explore public discourse surrounding Cecilia Sala on a platform distinct from YouTube, analyzing spontaneous textual content shared by users. BlueSky, being a decentralized platform with a specific demographic and technological user base, provides a unique data source that complements and enriches the analysis of public discourse.

The API queries were designed to return relevant posts filtered by keywords and sorted chronologically, enabling both qualitative and quantitative monitoring of interactions and opinions expressed within a defined temporal scope.

A text cleaning process was applied on both datasets using regular expressions (regex) to remove undesired elements such as punctuation, links, emojis and other irrelevant anomalies. Subsequently, advanced Natural Language Processing (NLP) techniques implemented via the Python library spaCy were employed to perform lemmatization, a process which reduced words to their base forms, enhancing the dataset's uniformity and the accuracy of the subsequent analysis.

After the preprocessing phase, which included text cleaning and lemmatization to standardize and prepare the textual data, we proceeded with two core analyses: **sentiment analysis**[1] and **emotion**[2]

[1] Sentiment can be taken to refer to the feeling that underlies an expressed positive or negative opinion or the feeling implied by a neutral opinion. It is therefore also called opinion mining. As summed up, a feeling is “a sensation that has been checked against previous experiences and labelled” while an emotion is “the projection/display of a feeling”.
Cit. Nip, J.Y.M.; Berthelier, B. Social Media Sentiment Analysis. *Encyclopedia* 2024, *4*, 1590–1598. https://doi.org/10.3390/ encyclopedia4040104

[2] The importance of emotion detection from text is rising expeditiously hand in hand with the Internet and online digital media. Therefore, technologists, business strategists, government agencies and political analysts accentuated and want to take advantage of this field in all aspects of the decision-making to improve businesses, reputations, etc. The Internet offers online social relationship sites such as Facebook, media sharing networks like YouTube, Instagram,

analysis. These analyses were conducted on the cleaned text field of the dataset to extract meaningful insights about the opinions and emotional tones embedded in the public discourse.

To further refine the analysis, we calculated the **relative frequency of terms**[3] within the dataset, which allowed us to quantify the prominence of specific words in relation to the total corpus. This step was essential in identifying the most recurrent linguistic patterns and understanding how certain words were associated with different sentiment and emotion categories.

Methods description

To perform these analyses, we utilized the Transformers[4] library in Python, which provides a robust and flexible interface for applying pretrained language models. Additionally, we accessed pretrained Large Language Models (LLMs) from the Hugging Face platform, a widely recognized hub for state-of-the-art NLP models. These models were selected specifically for their ability to handle text-classification tasks in the Italian language, ensuring relevance and precision in the analysis.

For the **classification of emotions**, we used the "*MilaNLProc/feel-it-italian-emotion*"[5] model. This model is optimized to detect and classify emotions into four categories:

- *Joy*: Representing positive and uplifting emotions.
- *Sadness*: Denoting feelings of grief or melancholy.
- *Anger*: Capturing expressions of hostility or frustration.
- *Fear*: Reflecting anxiety, apprehension, or worry.

By assigning one of these labels to each comment in YouTube context and to each post in BlueSky analysis, the model provided a granular view of the emotional undertones associated with the discourse surrounding the case. This level of analysis was particularly valuable for understanding

microblogging sites like Twitter, Reddit. This increase in the advent and popularity of social networks has motivated researchers to investigate online content and analyze users' online social behaviors.
Bazzaz Abkenar S, Haghi Kashani M, Mahdipour E, Mahdi Jameii S (2020) *Big data analytics meets social media: A systematic review of techniques, open issues, and future directions*, Telematics and Informatics, doi: https://doi.org/10.1016/j.tele.2020.101517.

[3] The basic methodology proposed for text corpora—a methodology successfully deployed in modern Internet search engines—reduces each document in the corpus to a vector of real numbers, each of which represents ratios of counts. A basic vocabulary of "words" or "terms" is chosen, and, for each document in the corpus, a count is formed of the number of occurrences of each word. Cit. David M. Blei, Andrew Y. Ng and Michael I. Jordan, Journal of Machine Learning Research 3 (2003) 993-1022.

[4] Transformers is an open-source library with the goal of opening up these advances to the wider machine learning community. The library consists of carefully engineered state-of-the art Transformer architectures under a unified API.
Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., … & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45.

[5] https://huggingface.co/MilaNLProc/feel-it-italian-emotion/tree/6efdabf62230414aeba764986b4ae317ce7c5c47.

the audience's emotional engagement[6] with sensitive topics like journalism, gender issues, and international events.

For **the sentiment analysis** of each comment and post, we employed the "*tabularisai/multilingual-sentiment-analysis*"[7] model. This model categorizes textual data into *five sentiment classes*[8]:

- *Very Positive*: Expressing strong approval, enthusiasm, or exceptional optimism, often accompanied by emphatic language or superlatives.
- *Positive*: Expressing approval, satisfaction, or optimism.
- *Neutral*: Denoting impartial or balanced expressions.
- *Negative*: Indicating disapproval, dissatisfaction, or pessimism.
- *Very Negative*: Expressing strong disapproval, severe criticism, or extreme pessimism, often conveyed with emotionally charged or exaggerated language.

This step allowed us to measure the general polarity of public opinion, highlighting whether the discourse was predominantly supportive, critical, or impartial regarding the case under study.

By combining these two layers of analysis, we were able to capture both the emotional depth and the attitudinal stance of the audience. Emotion analysis provided insights into the psychological impact and affective reactions, while sentiment analysis revealed the broader evaluative judgments made by commenters.

In the context of **Natural Language Processing (NLP)**[9], we computed relative frequencies using the *Counter* class from Python's collections module. This approach enabled us to systematically count word occurrences in the pre-processed text and normalize them against the total word count to obtain relative proportions. The frequency calculation followed the formula:

Relative Frequency = Count of a specific term/Total number of terms

---

[6] "The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal service which characterize the tie". Cit. Granovetter, M. S. (1973). *The strength of weak ties. American Journal of Sociology, 78(6), 1360-1380.* It was evident that groups of people, most of whom did not know each other, were cooperating in ways that had never been possible because they had mobile devices that enabled them to communicate and compute in real time, to cooperate and to adjust their behavior accordingly. Cit. Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations.* Penguin Press.

[7] https://huggingface.co/tabularisai/multilingual-sentiment-analysis

[8] Current NLP studies often define sentiments using scores on a scale, e.g., a 5-point Likert scale representing sentiments from strongly positive to strongly negative. Cit. Leimin Tian, Catherine Lai e Johanna D. Moore (2018). *"Polarity and Intensity: the Two Aspects of Sentiment Analysis"*.

[9] Natural Language Processing (NLP) combines computational linguistics, machine learning, and deep learning to analyze and generate human language. It uses syntactic analysis to define the grammatical structure of sentences and semantic analysis to understand their meaning. Techniques such as dependency and constituency parsing create structural representations useful for translations and speech recognition. Self-supervised learning optimizes model training, reducing the need for manually labeled data, thereby improving efficiency and applicability. https://www.ibm.com/it-it/topics/natural-language-processing.

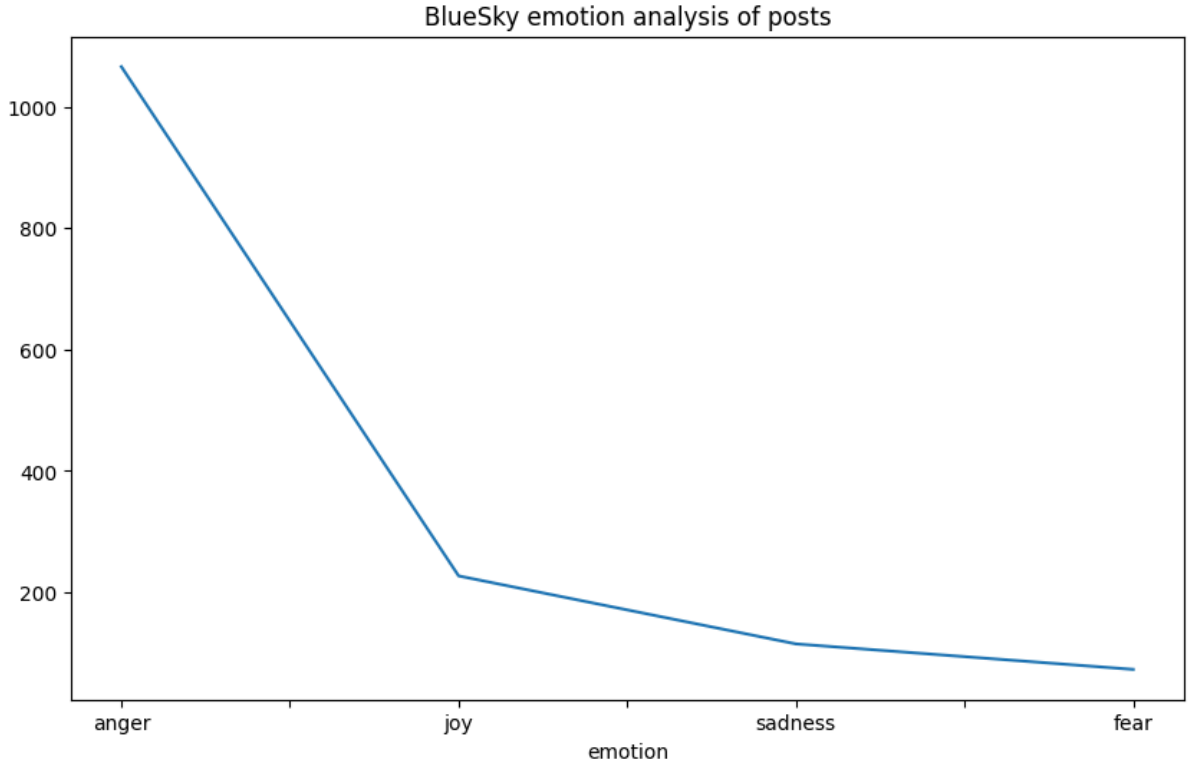
This method allowed us to account for differences in corpus size and compare the relative importance of terms across different sentiment and emotion categories. After obtaining raw frequencies, we grouped terms based on their predominant sentiment and emotion labels, enabling a structured representation of linguistic tendencies. By segmenting word frequencies according to their emotional and sentiment classifications, we were able to identify which terms were most strongly associated with specific attitudes or affective states.

This comprehensive methodology allowed us to build a nuanced understanding of how the public perceives and reacts to the themes of gender, journalism, and international incidents as they relate to the case in question.

Analysis and Results

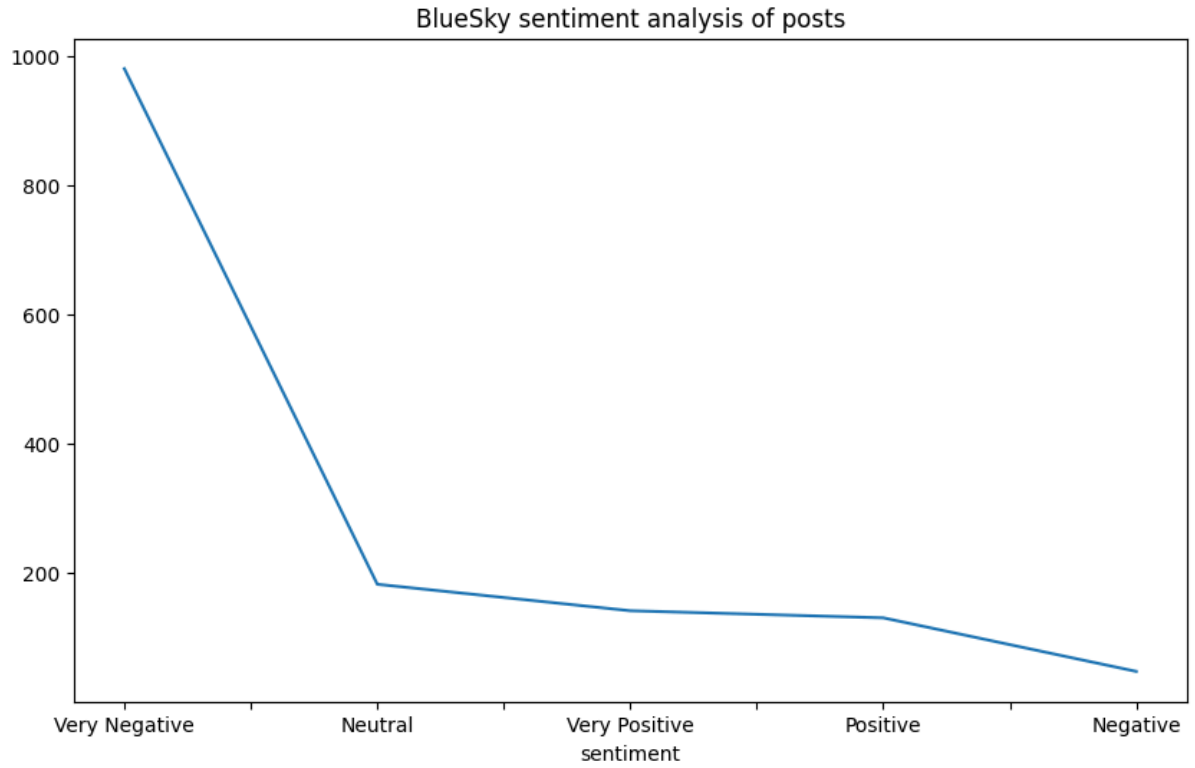
BlueSky Results

The chart illustrates the results of the emotion analysis conducted on posts from BlueSky, focusing on the public discourse surrounding Cecilia Sala. The analysis, which categorized the posts into four primary emotional responses (anger, joy, sadness, and fear) reveals a significant *predominance of anger, with over 1,000 instances recorded*.



This finding indicates that frustration or hostility is the dominant emotional tone in the discussions, potentially reflecting widespread discontent or criticism related to the case. Joy appears as the second most frequent emotion, albeit at a substantially lower rate, suggesting that a smaller segment of the audience expressed positive or uplifting sentiments. Sadness and fear, while less frequent, highlight responses of concern, grief, or apprehension, further underscoring the complexity of emotional engagement within this discourse.

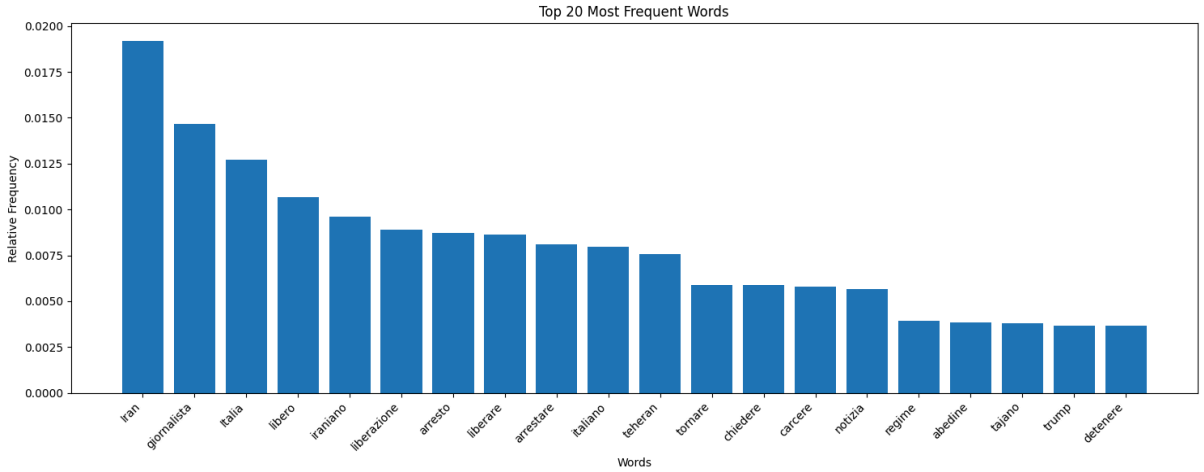
The chart provides crucial insights into the emotional dynamics of public opinion, particularly in the context of sensitive topics such as gender issues, the role of journalism in conflict zones, and international negotiations. The predominance of *anger*, juxtaposed with the presence of other emotions, suggests a multifaceted and highly charged emotional landscape, which is essential for understanding the broader implications of the case under study.



The line graph demonstrates a notably skewed distribution of sentiment classifications, with *Very Negative* sentiment dramatically dominating the dataset at approximately 950 posts, followed by a sharp decline across other sentiment categories. This distribution pattern merits careful consideration in the context[10] of social media sentiment analysis. In this instance, the *Very Negative* sentiment was not indicative of a systemic bias in the dataset but rather a genuine public reaction to a specific event the Cecilia Sala's arrest and detention in Iran. This underscores the importance of contextual analysis in sentiment studies, as certain events can naturally elicit strong negative responses[11].

[10] I. Carvalho, C .Silva and H.G. Oliveira, *The Importance of Context for Sentiment Analysis in Dialogues,* Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS, DOI 10.1109/ACCESS.2023.3304633.

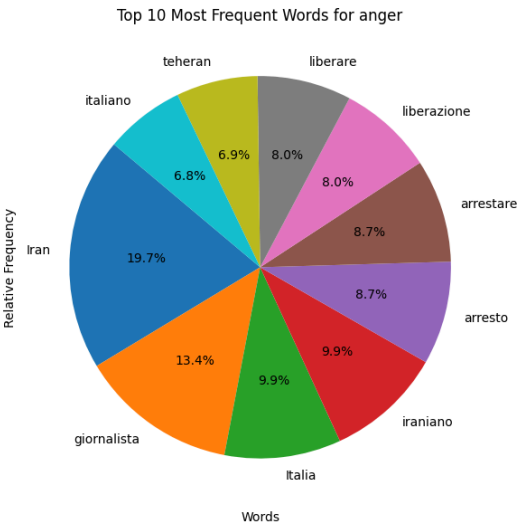
[11] A useful theoretical construct for understanding how people may react to events is the concept of affordances (Gaver, 1991; Gibson, 1977, 1986), JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY—February 2011 DOI: 10.1002/asi.

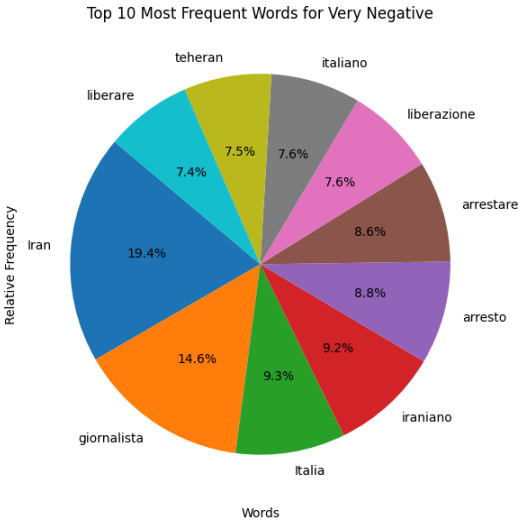


The bar chart depicts the relative frequency of the *20 most commonly occurring words* extracted from the analyzed posts on BlueSky, offering insight into the dominant topics and linguistic patterns in the public discourse surrounding Cecilia Sala. The x-axis represents the words, while the y-axis shows their normalized relative frequency within the dataset.
The bar chart depicts the relative frequency of the *20 most commonly occurring words* extracted from the analyzed posts on BlueSky, offering insight into the dominant topics and linguistic patterns in the public discourse surrounding Cecilia Sala. The x-axis represents the words, while the y-axis shows their normalized relative frequency within the dataset.

The prominence of terms such as “Iran,” “giornalista” (journalist), and “Italia” (Italy) indicates a strong thematic focus on the geopolitical context of the case and the professional role of Cecilia Sala as a journalist. Words like “liberare” (to free), “arresto” (arrest), and “carcere” (prison) suggest a significant discussion centered on her detention and calls for her release, reflecting the public’s engagement with the legal and human rights aspects of the situation. Additionally, terms such as “iraniano” (Iranian), “regime,” and “Teheran” (Tehran) emphasize the sociopolitical setting, and the perceived nature of the regime involved. The inclusion of words such as “notizia” (news) and “chiudere” (to close) points to broader discussions about journalism and media freedom.

This lexical analysis not only highlights the major themes in the discourse but also provides a quantitative basis for understanding the concerns, priorities, and framing of the case in public discussions on BlueSky. These findings contribute to the study’s overarching goal of examining how sentiment, emotion, and thematic content converge in online conversations about sensitive international and gender-related issues.





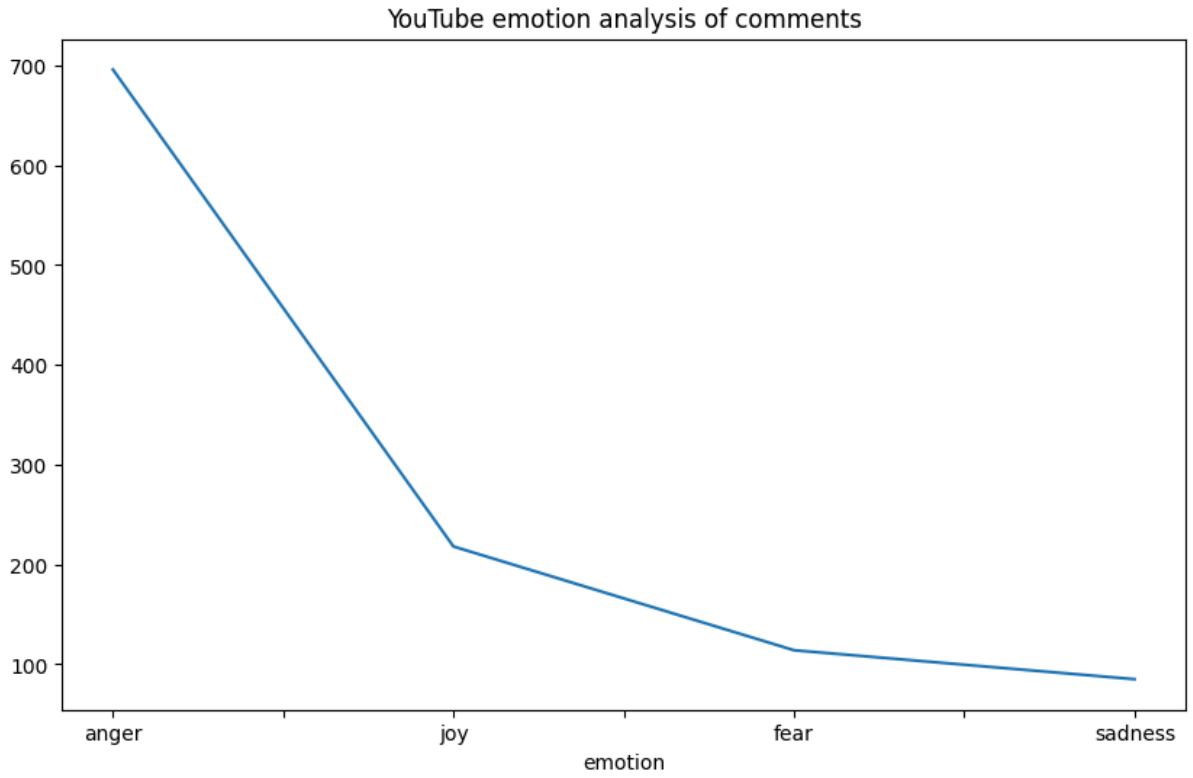
These pie charts depicting the top 10 most frequent words associated with *Anger* related content in the context of emotion analysis and *Very Negative* in the context of sentimental analysis, several interesting linguistic patterns emerge from there.

The data reveals a significant clustering around terms related to Iran and journalism, with "Iran" being the most prominent at 19.7% - 19.4% of the frequency distribution. This is followed by "giornalista" (journalist) at 13.4% - 14.6%, suggesting a strong media-related narrative context. The presence of both "italiano" (Italian, 6.8%, 7.6%) and "Italia" (Italy, 9.9% - 9.3%) indicates a cross-cultural or international focus in the discourse.
Of particular note is the prevalence of legal or enforcement-related terminology, with "arrestare" (to arrest) and "arresto" (arrest). Combined with "liberare" and "liberazione" this suggests a narrative framework centered around themes of detention and freedom.
The presence of "teheran" (6.9% - 7.5%) further reinforces the Iranian context, while "iraniano" adds to the geographical and cultural specificity of the discourse.

YouTube comments results

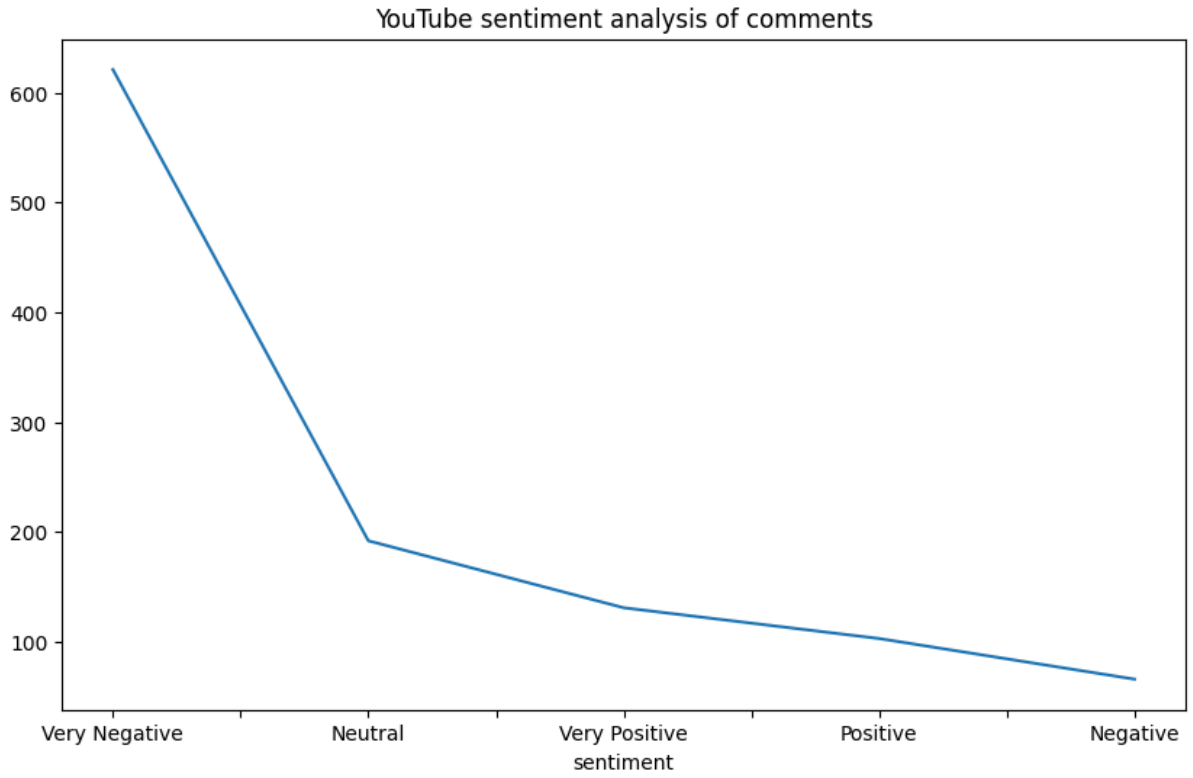
The following section presents the results derived from the analysis conducted on data extracted from YouTube comments. Specifically, it provides detailed insights into both emotion analysis and sentiment analysis, highlighting the distribution and intensity of different emotional and sentiment categories within the dataset.

Additionally, the results include the relative frequencies of key terms identified in the text, systematically grouped according to their predominant emotional and sentiment classifications. This analysis offers a comprehensive understanding of the linguistic and emotional patterns present in online discourse, allowing for a deeper exploration of the public's perception and reactions.



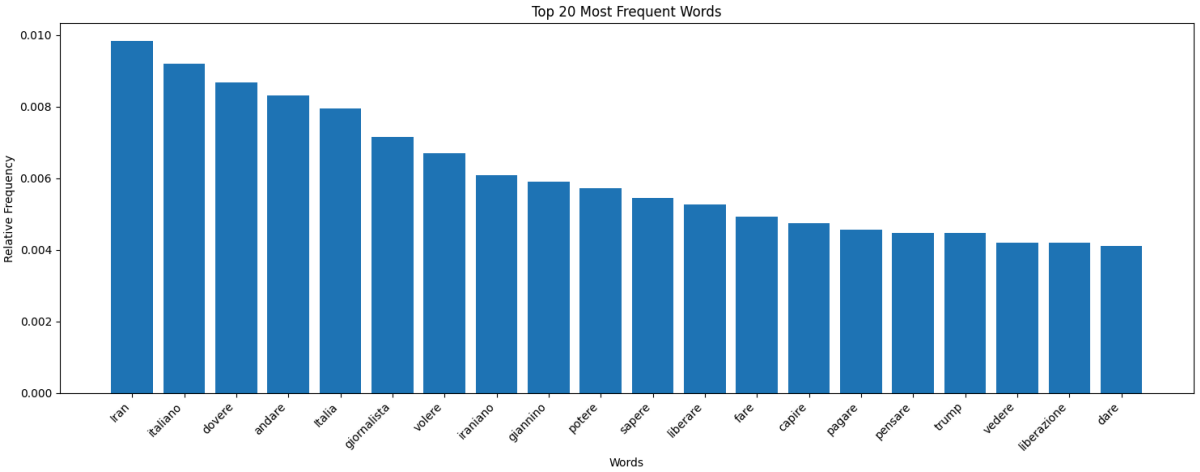
The chart presents the results of the emotion analysis performed on YouTube comments related to Cecilia Sala, illustrating the distribution of four primary emotions: anger, joy, fear, and sadness. The y-axis represents the frequency of occurrences for each emotion, while the x-axis categorises the emotions identified within the dataset.

The *data reveals that anger is the predominant emotion*, with approximately 700 instances, indicating a strong presence of negative sentiment and frustration within the discourse. Joy, on the other hand, is present but in significantly lower numbers, suggesting that while some comments express positive sentiments, they are not dominant. Fear and sadness appear with even lower frequencies, indicating that concerns and distress related to the discussed topic, while present, do not dominate the emotional landscape.



The sentiment analysis of YouTube comments regarding Cecilia Sala reveals a distinct distribution pattern, with a predominance of highly negative sentiment.The data indicate that "Very Negative" comments represent the majority, with approximately 600 instances, followed by a sharp decline in frequency as sentiment moves toward neutrality, which accounts for around 200 comments. Positive and very positive sentiments appear at significantly lower frequencies, ranging between 100 and 150, while "Negative" comments register the lowest count, below 100.

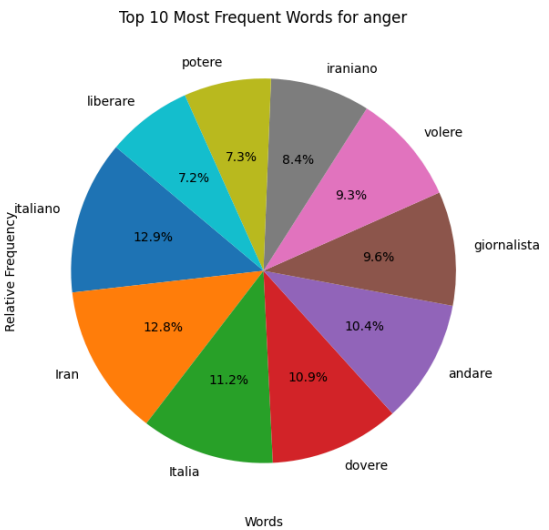
Methodologically, the analysis employs a five-point sentiment scale (Very Negative, Negative, Neutral, Positive, Very Positive), offering a granular understanding of the emotional tone within public discourse. The analysis further underscores the impact of platform-specific dynamics on discourse, reinforcing the need to examine the intersection of journalism, gender, and audience sentiment in online environments.

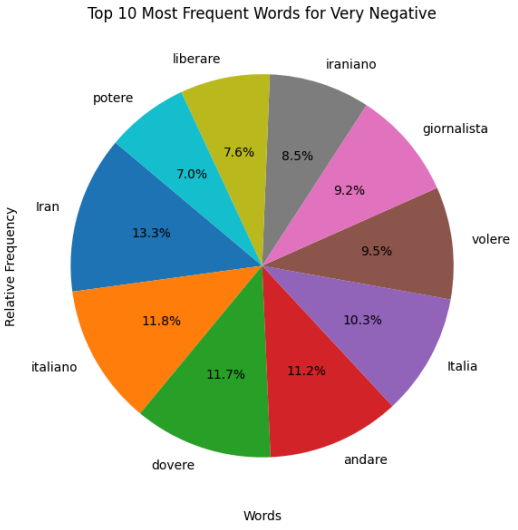


The frequency analysis presented in this chart offers a detailed examination of the lexical patterns within the discourse surrounding Cecilia Sala's journalistic work in the YouTube platform. The visualization depicts the relative frequency distributions of the twenty most frequently occurring words in the analyzed Italian-language corpus. The y-axis represents relative frequency, which normalizes word occurrences against the total corpus size, ensuring comparability across different textual datasets.

The analysis reveals that "Iran" is the most frequently occurring term (≈ 0.0095), which aligns with Sala's journalistic focus on Middle Eastern affairs. This is followed by "giornalista" and "Italia," highlighting the intersection of professional identity and national context within the discourse. The prominence of geographic and occupational terms suggests that the discussion is centered on Sala's role as an Italian journalist covering international events, particularly those related to conflict zones and foreign affairs.

Additionally, several verbs emerge with high frequency, including "liberare" (to free), "dovere" (must/should), "andare" (to go), "volere" (to want), and "pensare" (to think). These verbs indicate an action-oriented discourse, suggesting significant engagement with themes of movement, obligation, and advocacy. The recurrence of "liberare" is particularly relevant in the context of discussions concerning press freedom and journalist safety, reinforcing the presence of topics related to professional risks and geopolitical tensions.





A comparative analysis of the word frequency distributions in the discourse surrounding Cecilia Sala's journalism reveals several significant patterns that merit attention. The charts, depicting the top 10 most frequent words in each sentiment category. This method reveals striking similarities in lexical composition while highlighting subtle yet meaningful distinctions.

Both distributions prominently feature geographic markers ("Iran," "italiano," "Italia") and professional identifiers ("giornalista"), with "Iran" maintaining a notably high frequency in both contexts (12.8% and 13.3%, respectively). The presence of action-oriented verbs ("volere," "andare," "dovere") across both sentiment categories suggests a discourse centered on agency and obligation. Similarly, terms like "liberare" appear with similar frequency (7.2% and 7.0%) in both contexts.The remarkable consistency between anger and very negative sentiment lexical patterns (with only minimal percentage variations) indicates a potential correlation between these emotional valences in digital discourse surrounding female journalists covering international conflicts.

This linguistic convergence can be interpreted through the lens of affect theory and digital journalism studies, particularly when considering how professional critique intersects with gender-based digital harassment. The data also shows that geographic and professional identity markers maintain their prominence regardless of the negative emotional context, suggesting that criticism consistently focuses on the journalist's professional role and coverage area rather than diverging into unrelated topics. This finding contributes to our understanding of how professional women journalists are discussed in digital spaces, particularly when reporting on geopolitically sensitive regions.

Discussion

The findings of this study hold significant implications for social and evaluative research, particularly in the analysis of online discourse related to journalism, gender, and international affairs. By applying sentiment and emotion analysis to comments and posts from platforms such as BlueSky and YouTube, this study highlights the **public's attitudinal and emotional responses to key topics**. However, several limitations must be acknowledged, as they influence the accuracy, representativeness, and interpretability of the results.

One of the most pressing challenges relates to **API restrictions and data accessibility**[12]. The BlueSky API presents limitations in query-based search functionality and filtering by tags or metadata, restricting researchers' ability to extract targeted content. This limitation may lead to an incomplete or skewed dataset, as posts that contain relevant discussion but lack explicit keywords may be overlooked. Additionally, **YouTube comments**, while valuable for sentiment analysis, tend to be **highly polarized**, which may distort the overall perception of public sentiment, particularly in controversial discussions. The platform-specific nature of discourse means that findings from one platform may not generalize across others, further necessitating a comparative, **cross-platform approach in future research**.

Another important limitation concerns the **YouTube API**, which does **not allow filtering by date range** when searching for comments or videos. Instead, data retrieval relies on a preselected array of video IDs, meaning that the dataset must be manually curated based on available videos rather than dynamically generated from a broader temporal window. This constraint hinders the ability to perform longitudinal studies or analyze trends over time unless researchers continuously update their dataset with new video selections. As a result, the representativeness of the collected data may be influenced by the platform's algorithmic recommendations, which determine which videos receive visibility and engagement, potentially skewing the analysis towards the most popular or recently surfaced content rather than offering a balanced chronological perspective.

To overcome this limitation, future studies could explore alternative data collection approaches, such as periodic scraping combined with historical archiving of video comment sections or leveraging other platforms with more flexible API access. Additionally, integrating metadata analysis, including video publication dates and engagement metrics, could provide indirect temporal insights, helping contextualize sentiment and emotion trends within specific time frames.

---

[12] An API is a component of object-oriented programming languages that allow developers to build software for a particular application through a reference program library. The API is prescribed by a device's operating system or an application program in which a requester (another device or a client user) can make requests expecting responses from them. APIs facilitate interaction between different software programs and access to their services. It includes the specification of data structures, protocols, object classes and runtimes to communicate the consumer with the resources offered by the API. Developers can build new classes or extend existing ones to add new features or functionalities. A client API is called through an endpoint, which is a component that listens when a request is being made from the client- side of the communication to the server-side via HTTP, expecting a response to be returned. Concerning social networks and Web information sources, many of them do not offer an API to access the available information, for several reasons, such as: the data that is wanted is small or uncommon; the source does not have the infrastructure or technical ability to create an API; the data is valuable or protected and not intended to be spread widely; and even when an API does exist, there may be request volume and rate limits; also, the types and format of the data that it provides might be insufficient for the purpose. Furthermore, there are limitations on the API that include the rejection of access, if the use of the information is not enough or properly demonstrated. Data protection laws related to privacy and most of the TOS (Terms of Service) also limit their access. Dongo, I., Cardinale, Y., Aguilera, A., Martinez, F., Quintero, Y., Robayo, G., & Cabeza, D. (2020). A qualitative and quantitative comparison between Web scraping and API methods for Twitter credibility analysis. *International Journal of Web Information Systems, 17*(6), 493-516. https://doi.org/10.1108/IJWIS-05-2020-0037

Another critical issue is bias in machine learning models[13] used for sentiment and emotion classification. While pretrained NLP models such as "*MilaNLProc/feel-it-italian-emotion*" and "*tabularisai/multilingual-sentiment-analysis*" provide a structured framework for classification, they are trained on predefined datasets that may not fully capture the nuanced language used in journalistic and politically charged discussions. Bias in training data can lead to misclassification of emotions or sentiments, particularly when analyzing complex expressions such as irony, sarcasm, or implicit bias. Furthermore, sentiment models often classify text based on explicit linguistic markers, potentially overlooking contextual subtleties that are essential for an accurate interpretation of online discourse.

A specific linguistic challenge encountered during the analysis relates to the **lemmatization process**, which affected proper nouns, particularly personal names, transforming them into incorrect common words. For example, "Sala", referring to journalist Cecilia Sala, was often misinterpreted and incorrectly lemmatized as "salare" (to salt), "salo", or "sale" (salt). Similarly, "Meloni", referring to Italian Prime Minister Giorgia Meloni, was mistakenly converted into "melo" (apple tree) or "meloni" (plural of melon). This misinterpretation distorted the frequency distribution and sentiment classification, as these names, instead of being recognized as entities, were treated as generic terms and processed incorrectly by the model.

To address this issue, a text preprocessing refinement step was implemented, involving customized named entity recognition (NER) corrections and lexicon filtering. This refinement ensured that proper nouns were preserved, preventing their conversion into unrelated common words. Additionally, manual verification of frequently occurring terms helped mitigate this error, ensuring that high-profile names remained intact within the analysis. By applying these text-cleaning techniques, the study improved the accuracy of sentiment and emotion classification, avoiding distortions in lexical frequency analysis and ensuring a more faithful representation of discourse surrounding key individuals in journalistic discussions.

From an **ethical perspective**, the study raises concerns regarding user privacy and data representation. Social media users may not anticipate their comments being subjected to automated analysis, raising issues related to informed consent and data anonymization. While this study adheres to publicly available data policies, future research should consider developing more transparent methodologies and exploring ethical AI frameworks to ensure responsible data usage.

To enhance the reliability and depth of future research, several methodological improvements can be implemented. First, **expanding data collection across multiple platforms**, such as Mastodon, Reddit, or news comment sections, would provide a broader and more representative sample of public sentiment. Second, **integrating topic modeling techniques alongside sentiment and emotion** analysis could offer a more granular understanding of thematic structures in discourse, helping to uncover hidden trends that are not easily captured by direct sentiment classification. Third, **fine-tuning sentiment models with domain-specific datasets** related to journalism, digital harassment, and gender issues could reduce classification bias and improve the model's sensitivity to the specific linguistic patterns of journalistic discussions.

Finally, longitudinal studies tracking sentiment and emotion shifts over time could reveal evolving trends in public perception, particularly in response to major events or media coverage. This approach would provide deeper insights into how discourse evolves in response to social and political developments, offering a more dynamic and contextualized understanding of digital interactions.

[13] https://huggingface.co/blog/evaluating-llm-bias

By addressing these limitations and incorporating methodological refinements, future studies can strengthen the validity of sentiment and emotion analysis in social research, contributing to a more accurate and ethically responsible evaluation of digital discourse.

Conclusion

This study examined the sentiment and emotion expressed in YouTube comments and BlueSky posts related to Cecilia Sala using natural language processing (NLP) and machine learning models. The sentiment analysis revealed a predominance of negative and very negative comments, supporting existing literature on gender-based digital harassment, particularly toward female journalists operating in high-risk environments. Emotion analysis highlighted fear and anger as dominant emotional tones, reflecting public concerns over press freedom, international affairs, and political discourse. Additionally, lexical frequency analysis demonstrated a strong focus on geopolitical topics (e.g., “Iran”) and journalistic identity (e.g., “giornalista”), reinforcing the contextual relevance of the case study.

This research underscores the value of Big Data and AI-driven tools in social and evaluation research, enabling the systematic analysis of large-scale textual datasets that would otherwise be infeasible to process manually. NLP techniques, combined with machine learning models, provided a structured way to classify sentiment, extract emotions, and identify key lexical patterns, offering data-driven insights into public discourse. However, the study also highlights the inherent limitations of these technologies, including algorithmic biases, data access restrictions, and the need for continuous model validation.

Future research could expand the scope of analysis by integrating multimodal data (video and audio transcripts), cross-platform comparisons (e.g., Mastodon, Reddit), and real-time monitoring of sentiment trends. Additionally, refining AI models with domain-specific training data and improving entity recognition techniques could further enhance the accuracy and interpretability of sentiment analysis in journalistic and political contexts. Ultimately, the combination of Big Data analytics and AI-driven NLP presents a powerful framework for understanding digital discourse, informing both academic research and policy discussions on the evolving dynamics of online communication.

Bibliography

Anoop K., Manjary P. Gangan, Deepak P., Lajish V. L, *Towards an Enhanced Understanding of Bias in Pre-trained Neural Language Models: A Survey with Special Emphasis on Affective Bias,* 2022, Cornell University

Hubert Plisiecki, Paweł Lenartowicz, Maria Flakus, Artur Pokropek*, High Risk of Political Bias in Black Box Emotion Inference Models,* 2024, Cornell University

L'uso dei social per l'analisi del sentiment e dell'emotion della violenza di genere, ISTAT, 2022 **https://www.istat.it/wp-content/uploads/2024/11/Nota_metodologica_Report_sentiment_2022.pdf**

Bazzaz Abkenar S, Haghi Kashani M, Mahdipour E, Mahdi Jameii S (2020) Big data analytics meets social media: A systematic review of techniques, open issues, and future directions, Telematics and Informatics, doi: https://doi.org/10.1016/j.tele.2020.101517

Dongo, I., Cardinale, Y., Aguilera, A., Martinez, F., Quintero, Y., Robayo, G., & Cabeza, D. (2020). A qualitative and quantitative comparison between Web scraping and API methods for Twitter credibility analysis. *International Journal of Web Information Systems, 17*(6), 493-516. https://doi.org/10.1108/IJWIS-05-2020-0037

Nip, J.Y.M.; Berthelier, B. Social Media Sentiment Analysis. *Encyclopedia* 2024, *4*, 1590–1598. https://doi.org/10.3390/ encyclopedia4040104

David M. Blei, Andrew Y. Ng and Michael I. Jordan, Journal of Machine Learning Research 3 (2003) 993-1022.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., … & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45.

Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations.* Penguin Press.

Leimin Tian, Catherine Lai e Johanna D. Moore (2018).*"Polarity and Intensity: the Two Aspects of Sentiment Analysis".*

I.Carvalho, C.Silva and H.G. Oliveira, The Importance of Context for Sentiment Analysis in Dialogues*,* Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS, DOI 10.1109/ACCESS.2023.3304633.