Sentiment Analysis of Nakba Oral Histories: A Critical Study of Large Language Models

Huthaifa I. Ashqar
Arab American University, Jenin P.O. Box 240, Palestine
Huthaifa.ashqar@aaup.edu

Abstract

This study explores the use of Large Language Models (LLMs), specifically ChatGPT, for sentiment analysis of Nakba oral histories, which document the experiences of Palestinian refugees. The study compares sentiment analysis results from full testimonies (average 2500 words) and their summarized versions (300 words). The findings reveal that summarization increased positive sentiment and decreased negative sentiment, suggesting that the process may highlight more hopeful themes while oversimplifying emotional complexities. The study highlights both the potential and limitations of using LLMs for analyzing sensitive, trauma-based narratives and calls for further research to improve sentiment analysis in such contexts.

1 Introduction

The Nakba, meaning "catastrophe" in Arabic, refers to the forced displacement and dispossession of Palestinians during the 1948 occupation by the Israeli colonial power, a pivotal moment that reshaped the social, cultural, and political landscape of Palestine and the Middle East (Gluck, 2008). Narratives surrounding the Nakba are deeply embedded in the collective memory of Palestinian communities, transmitted through generations via oral histories, personal testimonies, and cultural expressions (Nur, 2008). These narratives, rich in emotional depth, political significance, and historical context, have become essential language resources for understanding the human impact of this traumatic event (Sa'di & Abu-Lughod, 2007). However, the challenge of interpreting, preserving, and analyzing such complex narratives has intensified in the age of artificial intelligence (AI) and large language models (LLMs), which have emerged as powerful tools for analyzing text at scale (Jaradat et al., 2024; Radwan et al., 2024).

Sentiment analysis, a key subfield of natural language processing (NLP), offers the potential to systematically evaluate the emotional content of texts (Assiri et al., 2024; H. Yang et al., 2024), thereby providing insights into the emotional and psychological dimensions of the Nakba narratives. By applying sentiment analysis to Nakba oral histories, we gain the ability to quantify and explore emotions like grief, loss, resistance, and resilience within these stories (Bhattacharjee et al., 2024; Q. Yang et al., 2024). However, while LLMs have demonstrated exceptional capabilities in processing and analyzing vast datasets, they are not without their limitations, particularly when applied to sensitive, culturally charged, and historically complex narratives such as those related to the Nakba (Coeckelbergh, 2023; Tian et al., 2024). These models often struggle with the subtleties of language, historical context, and the lived experiences embedded in oral histories, raising questions about their adequacy in capturing the full emotional and cultural resonance of such narratives (Shi et al., 2023; Zhang et al., 2023).

This paper critically examines the application of sentiment analysis to Nakba oral histories, exploring both the potential and the pitfalls of using LLMs to analyze such narratives. It argues that while LLMs can offer valuable insights into the emotional tone of texts, they must be used with caution, considering the unique cultural and historical context of the Nakba. The study positions Nakba narratives as crucial language resources, emphasizing their role in shaping collective memory and identity, while also critiquing the limitations of AI tools in fully capturing the intricacies of human experience and emotion. By doing so, the paper highlights the importance of interdisciplinary approaches—combining computational methods with cultural sensitivity—when applying AI-driven tools to historically significant and emotionally complex language resources such as Nakba oral histories.

Proceedings of the Proceedings of the 1st International Workshop on Nakba Narratives as Language Resources, pages 30–36
January 20, 2025.

2 Methods

2.1 Dataset

For this study, we utilized the Nakba Archive, a grassroots oral history collective established in 2002 with the aim of documenting the experiences of Palestinian refugees in Lebanon who lived through the 1948 Nakba (Allan, 2005; Hawari, 2023). This dataset is particularly valuable as it includes about 30 video interviews with first-generation Palestinian refugees from different Palestinian villages and towns that were displaced or destroyed during the creation of the Israeli colonial state. The interviews, recorded by refugees in camps in Lebanon, provide firsthand accounts of the mass displacement, dispossession, and violence experienced by Palestinian communities during the Nakba (Allan, 2005; Hawari, 2023).

The Nakba Archive offers a comprehensive and personal account of the Nakba (Fu, n.d.; Regan, 2022), which displaced approximately one million Palestinians, leaving them homeless and dispossessed of their lands. The destruction of Palestinian villages is detailed in these interviews, and the narratives offer vivid depictions of life before 1948 in Palestine. These oral histories serve as primary sources that document not only the traumatic events of the Nakba but also the emotional and psychological impacts that have shaped Palestinian collective identity across generations (Allan, 2005; Fu, n.d.; Hawari, 2023).

The dataset contains a wide range of emotional expressions and personal reflections, which are essential for the sentiment analysis conducted in this study . The interviews capture the depth of grief, loss, resilience, and resistance, reflecting the ongoing struggle for Palestinian liberation, self-determination, and the right of return. The richness and diversity of the testimonies are key to understanding the long-lasting effects of dispossession on the lives of Palestinian refugees and their descendants.

Given its depth and significance, the Nakba Archive offers a unique and invaluable resource for studying how sentiment is embedded in oral histories, especially those related to traumatic historical events (Manna', 2013; Saadah, 2021). The sentiments expressed in these interviews not only shed light on the individual and collective emotional responses to displacement but also provide insight into the broader cultural and political implications of the Nakba on Palestinian identity and memory (Allan, 2005). This dataset, therefore, serves as a critical tool in understanding the role of Nakba narratives as language resources, facilitating an initial and empirical exploration of sentiment within the context of historical trauma and its ongoing repercussions in Palestinian collective mind.

2.2 Proposed Framework

This study proposes a comprehensive framework for analyzing the sentiments embedded in the oral testimonies of Palestinian refugees from the Nakba Archive, utilizing ChatGPT as the primary tool for sentiment analysis. The framework involves a step-by-step approach that first processes the full-length testimonies and then compares the sentiment analysis results from both the original and summarized versions of the testimonies as shown in Figure 1. The framework consists of four main steps.

The first step of the framework involves using ChatGPT to perform sentiment analysis on the complete oral testimonies. For each experiment, we ran ChatGPT five times to ensure consistency, and we took the average. Each testimony, on average, contains 2,500 words, providing rich, detailed accounts of personal experiences and reflections. ChatGPT, as a language model trained on vast datasets, is employed to analyze the emotional tone, sentiment polarity (positive, negative, neutral), and emotional intensity expressed in each full testimony (Havaldar et al., 2023; Kumar et al., 2024; Patel & Fan, 2023). The sentiment analysis in this study aims to capture the emotional responses related to themes of displacement, trauma, loss, resilience, and hope that are central to the Nakba narratives (Allan, 2005; Regan, 2022). This analysis allows for an in-depth understanding of the emotions conveyed across different aspects of the refugees' experiences.

In the next step, ChatGPT is used to summarize each full testimony into approximately 300 words. This step involves distilling the key points, themes, and emotional expressions from the original testimonies while preserving their core messages. Summarization is a critical component, as it enables the analysis of more concise versions of the testimonies, making it easier to compare the sentiments without losing the essence of the original accounts (Kabadjov et al., 2009; Krugmann & Hartmann, 2024). By focusing on the most relevant parts of each narrative, the summarized version facilitates a more streamlined analysis, which can be crucial for examining large datasets of oral histories.

Following the summarization process, ChatGPT is again used to conduct sentiment analysis on the new, condensed versions of the testimonies. This second round of sentiment analysis aims to examine how the emotional tone and sentiment evolve when the testimonies are reduced to their essential elements. By comparing the sentiment expressed in these shorter summaries with the results from the full testimonies, the study can evaluate whether the sentiment is captured effectively and if any emotional nuances are lost during the summarization process (Kabadjov et al., 2009).

The final step in the framework involves comparing the sentiment analysis results from the full testimonies with those from the summarized versions. This comparison is central to understanding the impact of summarization on sentiment expression and whether important emotional nuances are preserved or altered. The study will assess key metrics such as sentiment polarity (positive, negative, neutral) and the intensity of emotional responses in both the original and summarized texts. This comparison will provide valuable insights into the relationship between text length, content condensation, and the retention of emotional depth, highlighting the potential trade-offs when working with condensed versions of oral histories.

The proposed framework allows for a structured and systematic approach to sentiment analysis, leveraging ChatGPT's natural language processing capabilities to process large amounts of qualitative data efficiently. The study investigates the ability of LLMs to understand the narrative sentiments of Nakba testimonies and to what extent it exposes bias towards the emotional nuances that are preserved in this oral history. By comparing the sentiment analysis results across both full and summarized versions of the testimonies, the study also aims to provide a deeper understanding of how different formats of narrative influence the emotional tone and content of the Nakba oral histories.

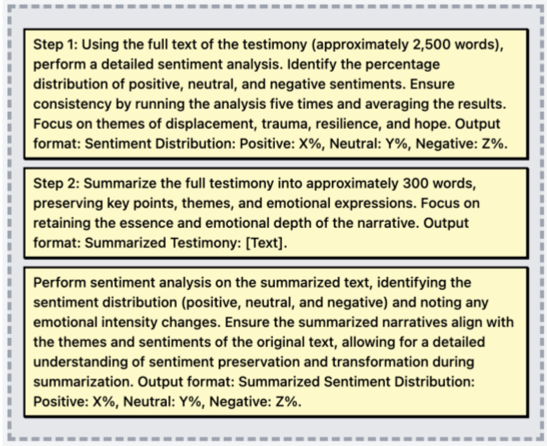

Figure 1: Prompt template for Sentiment Analysis and Summarization of Nakba Oral Histories Using ChatGPT.

3 Analysis and Results

In this study, sentiment analysis was conducted on both the full and summarized versions of ten Nakba oral testimonies, with the results broken down into positive, neutral, and negative sentiment categories for each individual testimony. The following analysis provides a comparison between the sentiment distribution of the full testimonies and their corresponding summaries, along with a critical assessment of the findings.

3.1 Sentiment Results

The sentiment analysis of the full testimonies and the summarized ones are shown in Table 2. The results suggest that the overall emotional tone of the testimonies is predominantly neutral, with a smaller portion reflecting negative sentiments. For the summarized, when compared to the full testimonies, the summarized versions show a slight increase in positive sentiment (31% vs. 26.3%) and a decrease in negative sentiment (19.5% vs. 28.7%). The neutral sentiment remains relatively consistent in both the full and summarized versions, with a small increase from 45% to 49.5%.

Sentiment	Full	Summarized
Positive	26.3%	31.0%
Neutral	45.0%	49.5%
Negative	28.7%	19.5%

Table 1: Sentiment analysis results.

3.2 Key Observations and Assessment

Increase in Positive Sentiment in Summarized Versions: One notable change between the full and summarized versions is the increase in positive sentiment. The average positive sentiment increased by 4.7 percentage points from the full testimonies (26.3%) to the summarized versions (31%). This shift suggests that the summarization process may have inadvertently emphasized more hopeful or resilient aspects of the testimonies, potentially glossing over the more negative or traumatic details in order to condense the narrative.

Decrease in Negative Sentiment in Summarized Versions: Conversely, the negative sentiment decreased significantly from 28.7% in the full testimonies to 19.5% in the summaries. This reduction may reflect the simplification of complex emotional expressions during the summarization process. It is possible that the summarized versions, by omitting certain contextual details and emotional depth, downplayed the intensity of negative sentiments such as grief, anger, and despair, which are prominent themes in the full testimonies. This reduction may not necessarily indicate a shift in the refugees' emotional experience but rather a result of truncating or neglecting the emotional complexity of the narratives.

Individual Variations: There are also notable variations across individual testimonies. For example, Testimony 10 shows a dramatic shift from a relatively balanced sentiment distribution (25% positive, 55% neutral, 20% negative) in the full version to almost completely neutral testimony (95% neutral) in the summary. This stark contrast suggests that the summarization might have completely neutralized the emotional tone of this specific testimony, possibly omitting key emotional elements or reframing the narrative in a more detached manner. In other cases, such as Testimony 3, positive sentiment increased significantly from 35% to 55%, indicating that the summarized version may have emphasized more hopeful aspects of the testimony. These variations highlight the complexity of summarization and the subjective nature of sentiment analysis using LLMs.

The comparison between the sentiment analysis of full and summarized testimonies reveals important insights into how narrative condensation affects emotional expression using LLMs. While the summarized versions exhibited an increase in positive sentiment and a decrease in negative sentiment, the neutral sentiment remained relatively stable. These changes underscore the potential risks of summarizing emotionally complex narratives using LLMs and how LLMs might not be suitable to fully understand such oral history. It seems that the summarization process may unintentionally obscure the full emotional depth of the testimonies, particularly with regard to the more negative sentiments.

4 Discussion

The results of the sentiment analysis on the full and summarized Nakba oral histories offer valuable insights into the capabilities and limitations of LLMs such as ChatGPT. Specifically, the observed shifts in sentiment—particularly the increase in positive sentiment and decrease in negative sentiment in the summarized versions—raise important questions about how LLMs process, condense, and represent emotional content in complex narratives. This discussion will explore what these shifts reveal about the behavior of LLMs in sentiment analysis and summarization tasks.

The increase in positive sentiment in the summarized testimonies, compared to the full versions, can be attributed to several factors inherent in the summarization process. In the context of the Nakba testimonies, this shift could reflect the LLM's tendency to neglect aspects such as fear, sad, or grief. Instead, LLMs emphasized positive emotions and hope, which are central to the Palestinian refugee experience but may not be as prominent in the more traumatic or sorrowful details of the longer narratives. Since summarization inherently involves reducing the complexity of emotional expression, LLMs might emphasize elements that allow for a more coherent and cohesive portrayal, possibly skewing the sentiment toward the positive end of the spectrum.

Moreover, LLMs like ChatGPT are trained on vast datasets with a significant amount of positive, optimistic language. This bias could influence the model to unintentionally highlight positive aspects of the narrative, even if those sentiments are less central in the full testimony. The summarization process could amplify this tendency, resulting in summaries that appear more positive in sentiment, even when the original narrative is emotionally complex or predominantly negative.

The reduction in negative sentiment observed in the summarized testimonies can also be understood through the lens of LLM behavior. Negative emotions, particularly those related to trauma, grief, and loss, are often more nuanced and detailed in the full testimonies. When tasked with summarizing these narratives, the LLM might omit or condense the detailed expressions of pain, anger, or sorrow to meet the constraints of brevity and focus on key events or themes.

In some cases, the model may unintentionally downplay the intensity of negative emotions by rephrasing or generalizing painful experiences. This could occur due to the model's propensity to avoid overly dramatic language or its tendency to reduce emotional complexity due to low resources of the Nakba and Palestinian narrative in the training dataset of LLMs. Furthermore, negative sentiments that are less immediately apparent or require more context might be excluded from summaries, leading to a shift in sentiment toward the neutral or positive end of the spectrum.

It is important to note that while the summarization process may reduce the explicit negativity in the text, this does not necessarily reflect a change in the underlying emotional experience of the refugees. Instead, it highlights the limitations of LLMs in capturing the full emotional depth of complex, traumatic narratives. In reducing the complexity of the text, the model may inadvertently present a version of the testimony that appears less negative, even if the full testimony conveys a much more emotionally charged story.

4.1 LLMs Behavior in Sentiment Analysis

The shifts in sentiment observed in this study provide several key insights into the behavior of LLMs, particularly in their application to sentiment analysis of sensitive, complex narratives:

Bias Toward Simplification: LLMs, when tasked with summarizing lengthy narratives, tend to simplify and condense emotional expressions. This simplification can lead to a distortion of the emotional tone of the original text. In the case of the Nakba testimonies, the reduced complexity in the summarized version may have skewed the sentiment analysis toward more positive and neutral expressions, obscuring the depth of negative emotional experiences.

Inability to Fully Capture Emotional Complexity: While LLMs are highly effective at processing and analyzing language, they often struggle with capturing the full emotional complexity of human experience, particularly in narratives shaped by trauma and historical injustice as well as experiences with low-resource data. These models are adept at identifying clear emotional signals (e.g., happiness, sadness, anger), but they may fail to fully grasp the subtleties of human emotions, especially in long, complex, and multi-faceted narratives.

Potential Bias in Summarization: In the case of Palestinian Nakba testimonies, the model may inadvertently introduce bias by favoring positive or neutral expressions that align more closely with the type of language typically found in general-purpose datasets. This could result in summaries that do not fully capture the lived realities of refugees, potentially diminishing the perceived severity of their experiences.

Context Sensitivity and Narrative Construction: LLMs may not fully understand the historical, cultural, or emotional contexts in which the Nakba testimonies were given. As a result, the summaries produced by the model may reflect a construction of the narrative that is less true to the original testimony, impacting the sentiment analysis. In other words, the emotional tone of a narrative can be shaped by the way the story is framed, and LLMs may inadvertently alter this framing during summarization.

5 Conclusion

This study aimed to evaluate the performance and behavior of LLMs in performing sentiment analysis on Nakba oral histories, with a particular focus on how summarization of complex, emotionally charged narratives influences sentiment results. The primary objective was to examine whether LLMs, specifically ChatGPT, could accurately capture and analyze the emotional tone of testimonies from Palestinian refugees, and how summarization might affect the representation of these sentiments. The contribution of this work lies in its novel application of LLMs to sensitive historical narratives, providing valuable insights into the potential and limitations of these models for sentiment analysis in emotionally complex contexts.

The study highlights several limitations. First, the sentiment analysis conducted by ChatGPT may have been influenced by biases in the model’s training data, leading to a misrepresentation of the emotional tone in the Palestinian refugee testimonies. Additionally, the summarization process itself likely oversimplified the complexity of the full narratives, which may have distorted the sentiment analysis results. These limitations point to the need for more specialized models capable of understanding the cultural and historical contexts of sensitive narratives like the Nakba. The future work will also compare different LLMs to see if the conclusions hold across them.

References

Allan, D. (2005). Mythologising al-nakba: Narratives, collective identity and cultural practice among Palestinian refugees in Lebanon. *Oral History*, 47–56.

Assiri, A., Gumaei, A., Mehmood, F., Abbas, T., & Ullah, S. (2024). DeBERTa-GRU: Sentiment Analysis for Large Language Model. *Computers, Materials & Continua*, *79*(3).

Bhattacharjee, A., Xu, S. Y., Rao, P., Zeng, Y., Meyerhoff, J., Ahmed, S. I., Mohr, D. C., Liut, M., Mariakakis, A., & Kornfield, R. (2024). " It Explains What I am Currently Going Through Perfectly to a Tee": Understanding User Perceptions on LLM-Enhanced Narrative Interventions. *ArXiv Preprint ArXiv:2409.16732*.

Coeckelbergh, M. (2023). Narrative responsibility and artificial intelligence: How AI challenges human responsibility and sense-making. *AI & SOCIETY*, *38*(6), 2437–2450.

Fu, M. (n.d.). Catastrophe and Reincarnation: Nakba and Memory in Diana Allan's Voices of the Nakba and Kamal Aljafari's Port of Memory. *Managing Editors*, 19.

Gluck, S. B. (2008). *Oral History and al-Nakbah*. Taylor & Francis.

Havaldar, S., Rai, S., Singhal, B., Liu, L., Guntuku, S. C., & Ungar, L. (2023). Multilingual language models are not multicultural: A case study in emotion. *ArXiv Preprint ArXiv:2307.01370*.

Hawari, Y. (2023). *Voices of the Nakba: A Living History of Palestine: edited by Diana Allan. London: Pluto Press, 2021. 368 pages. 23.00paper, 12.00 e-book.* Taylor & Francis.

Jaradat, S., Nayak, R., Paz, A., Ashqar, H. I., & Elhenawy, M. (2024). Multitask Learning for Crash Analysis: A Fine-Tuned LLM Framework Using Twitter Data. *Smart Cities*, *7*(5), 2422–2465. https://doi.org/10.3390/smartcities7050095

Kabadjov, M., Balahur, A., & Boldrini, E. (2009). Sentiment intensity: Is it a good summary indicator? *Language and Technology Conference*, 203–212.

Krugmann, J. O., & Hartmann, J. (2024). Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions*, *11*(1), 3.

Kumar, C. U. O., Gowtham, N., Zakariah, M., & Almazyad, A. (2024). Multimodal Emotion Recognition Using Feature Fusion: An LLM-Based Approach. *IEEE Access*.

Manna', A. (2013). The Palestinian Nakba and its continuous repercussions. *Israel Studies*, *18*(2), 86–99.

Nur, M. (2008). Remembering the Palestinian Nakba: Commemoration, oral history and narratives of memory. *Holy Land Studies*, *7*(2), 123–156.

Patel, S. C., & Fan, J. (2023). Identification and description of emotions by current large language models. *BioRxiv*, 2023–2027.

Radwan, A., Amarneh, M., Alawneh, H., Ashqar, H. I., AlSobeh, A., & Magableh, A. A. A. R. (2024). Predictive Analytics in Mental Health Leveraging LLM Embeddings and Machine Learning Models for Social Media Analysis. *International Journal of Web Services Research (IJWSR)*, *21*(1), 1–22.

Regan, B. (2022). *Diana Allan (ed.) Voices of the Nakba: A Living History of Palestine*. Edinburgh University Press The Tun-Holyrood Road, 12 (2f) Jackson's Entry ….

Saadah, M. J. (2021). The Palestinian perspective: Understanding the legacy of al-Nakba through the Palestinian narrative. *Berkeley Undergraduate Journal*, *35*(2).

Sa'di, A. H., & Abu-Lughod, L. (2007). *Nakba: Palestine, 1948, and the claims of memory*. Columbia University Press.

Shi, F., Chen, X., Misra, K., Scales, N., Dohan, D., Chi, E. H., Schärli, N., & Zhou, D. (2023). Large language models can be easily distracted by irrelevant context. *International Conference on Machine Learning*, 31210–31227.

Tian, Y., Huang, T., Liu, M., Jiang, D., Spangher, A., Chen, M., May, J., & Peng, N. (2024). Are Large Language Models Capable of Generating Human-Level Narratives? *ArXiv Preprint ArXiv:2407.13248*.

Yang, H., Zhao, Y., Wu, Y., Wang, S., Zheng, T., Zhang, H., Ma, Z., Che, W., & Qin, B. (2024). Large Language Models Meet Text-Centric Multimodal Sentiment Analysis: A Survey. *ArXiv Preprint ArXiv:2406.08068*.

Yang, Q., Ye, M., & Du, B. (2024). Emollm: Multimodal emotional understanding meets large language models. *ArXiv Preprint ArXiv:2406.16442*.

Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2023). Sentiment analysis in the era of large language models: A reality check. *ArXiv Preprint ArXiv:2305.15005*.