Brazilian Consumer Perceptions of Electric Vehicles Using LLMs on Online Comments

Kalidsa B. de Oliveira[1], **Gabriel M. Lunardi**[1], **Thiago L. T. da Silveira**[1], **Adriano Q. Oliveira**[1], **Leonardo Ramos Emmendorfer**[1]

[1]Universidade Federal de Santa Maria - UFSM, Santa Maria, RS, Brasil

kalidsa.oliveira@ecomp.ufsm.br

{gabriel.lunardi, thiago.silveira}@ufsm.br

{adriano.q.oliveira@ufsm.br, leonardo.emmendorfer}@ufsm.br

Abstract. *The transition from gasoline-powered vehicles to electric vehicles (EVs) is a major step forward, driven by both technological innovations and environmental concerns. This study analyzes the perceptions of Brazilian consumers toward EVs, using a novel dataset of YouTube and Reddit comments collected between 2015 and 2025. To handle the large volume of text, we used the LLaMA 3.2 language model with custom prompts to identify and analyze specific technical aspects of EVs. Our results show that consumers' main concerns are limited range and high cost, while the most valued aspects are energy efficiency and the technology itself.*

1. Introduction

The transition from gasoline-powered vehicles to Electric Vehicles (EVs) is a significant technological and economic advancement, driven by innovation and environmental concerns. While internal combustion engine vehicles still dominate the market, the popularity of EVs is growing rapidly [Lau 2025].

To understand this shift, we analyze how consumers perceive EVs, particularly in Brazil. Sentiment analysis is ideal for interpreting opinions expressed in text, especially with large volumes. When applied to EVs, it can reveal evaluation trends, providing strategic data for companies that produce these vehicles and organizations that formulate environmental to economic policies related to this type of motor transportation [Sharma et al. 2024, Soares et al. 2025].

In response to the challenges of analyzing large amounts of text, this study proposes a solution using Large Language Models (LLMs) to extract relevant information. The main contributions of this work are: the application of the LLaMA 3.2 model with custom prompts to identify and analyze technical aspects of electric vehicles (EVs); and the creation of a novel dataset, comprising 9,670 comments collected between 2015 and 2025 from YouTube and Reddit, providing a foundation for understanding consumer perceptions of EVs.

2. Methodology

2.1. Data Collection

The data collection phase of this work was conducted with a focus on two widely used platforms for sharing and discussion: Reddit and YouTube. Both were chosen because they allow public access to spontaneous and detailed user comments about EVs, representing a rich source of unstructured data [Oliveira et al. 2025a, Oliveira et al. 2025b]. The complete dataset is available for download in this GitHub repository[1].

For data collection on Reddit, the asyncpraw package was used, applying temporal filters to select comments from January 2015 to June 2025. This wide time window was chosen to capture the evolution of perceptions and discussions about electric vehicles. The search was conducted on relevant subreddits using specific keywords such as "electric car", "Tesla", and "BYD". The comments underwent textual pre-filtering to eliminate short entries, links, and edits, followed by sentiment analysis using the pre-trained nlptown/bert-base-multilingual-uncased-sentiment model from the transformers library. Only comments with high confidence and clear sentiments were kept, resulting in 1,039 unique comments after removing duplicates.

For the YouTube collection, we manually selected long and short videos (shorts) with high relevance and over 100 comments. The official YouTube Data API v3 was used to extract up to 5,000 comments per video, applying the same filtering and sentiment analysis criteria used for Reddit. In total, we obtained 8,676 comments from YouTube and 1,039 from Reddit, for a combined total of 9,715. However, after removing duplicates, the final dataset was reduced to 9,670. Although the code was developed to collect data over a 10-year period, the extraction was limited to publications between 2015 and 2025, with 1 comment recorded in 2015, 2 in 2017, and then 53, 40, 32, 226, 1,761, 4,058, and 3,545 occurrences for each subsequent year of the analyzed period, respectively.

2.2. Extraction and Normalization of Technical Aspects

For the qualitative analysis of the collected comments, we developed a routine for extracting technical aspects based on Large Language Models (LLMs). The LLaMA 3.2 model, deployed locally via the Ollama API, was used in conjunction with a prompt engineering technique. The specific prompts are available in the "extracao.py" file in the aforementioned repository. This approach, which has become a core strategy for using pre-trained LLMs, involves carefully crafting specific textual instructions (prompts) to guide the generation of responses without needing to reconfigure the model's internal parameters. This methodology stands out for its high adaptability and efficiency across various tasks, substantially expanding the applicability of these models in specific contexts [Sahoo et al. 2024].

In this study, the model was instructed to identify explicit mentions of technical EV features in a structured manner, such as motor performance, energy efficiency, range, and charging infrastructure, among others. It then organized these mentions into positive and negative polarity categories. Comments containing offensive language, including discriminatory terms, threats, or inappropriate content, were automatically identified and classified separately to quantify their occurrence and support complementary ethical analyses. The model's output was processed using regular expressions to identify text blocks containing the positive and negative aspects. Subsequently, a lexical normalization procedure was applied, which included the removal of Portuguese stopwords.

Each comment was classified into three distinct categories: (i) containing explicit mentions of relevant technical aspects; (ii) with no clear reference to technical features;

[1] https://github.com/Kalidsa/ERAMIA-2025-carros-eletricos

(iii) containing offensive language. The resulting records were stored in separate files to facilitate data organization and the execution of subsequent statistical analyses.

3. Results

Based on the data represented graphically, it is possible to draw important conclusions about user perceptions of EVs. Figure 1 shows that the majority of comments, specifically 7,700, do not contain relevant technical aspects, indicating a predominance of generic or uninformative evaluations. On the other hand, 696 comments exclusively highlighted positive aspects, while 977 mentioned negative ones, and 295 combined both types of perception.

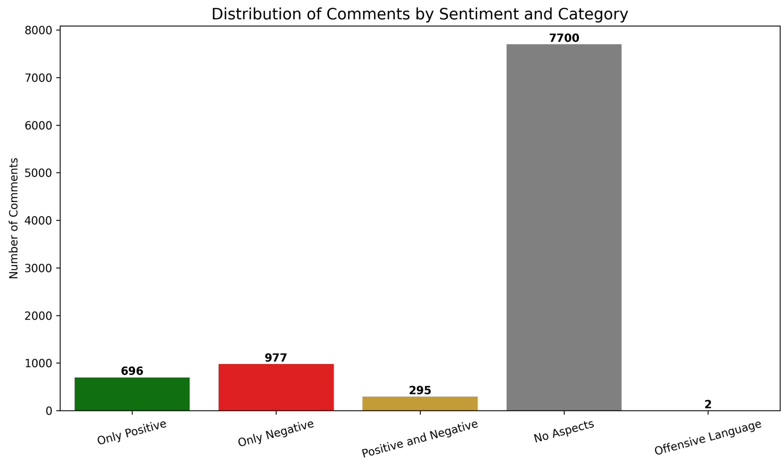

Figure 1. Total number of extracted aspects.

Furthermore, a negligible number of comments contained offensive language, demonstrating a low level of toxic discourse in the analyzed dataset. Figure 2 reveals the main technical aspects extracted from the comments, offering deeper insight into the features valued or criticized by users. Among the most frequently mentioned positive aspects are "energy efficiency", cited in 111 comments; "advanced technology", with 75 occurrences; and "extended range", mentioned in 72 records, suggesting a recognition of the technological innovations and sustainability that EVs provide.

In contrast, the most frequent negative aspects include "limited range", mentioned in 87 comments; "high cost", cited in 62 records; and the "lack of charging infrastructure", pointed out in 32 observations, reflecting recurring concerns about the practical and economic viability of these vehicles.

4. Conclusion

This study demonstrates that, although EVs are recognized for their environmental and technological benefits, they still face significant barriers in Brazil, such as the lack of charging infrastructure and high acquisition and maintenance costs. The work has contributed to an understanding of the factors that influence the adoption of this technology, providing data for researchers, public policy makers, and the automotive industry.

The study could be expanded to include data from other platforms, such as specialized forums, discussion groups, and social media. Using more advanced language models

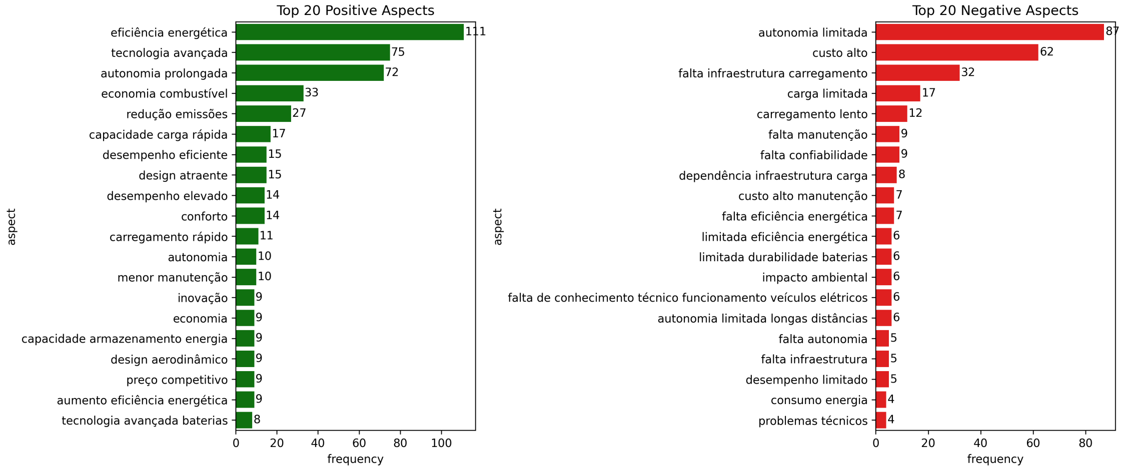

Figure 2. Top-20 aspects.

could also improve the identification of specific nuances and contexts in opinions, such as sarcasm or irony. A temporal correlation analysis could be conducted to understand how consumer perceptions evolved between 2019 and 2025. Additionally, segmenting comments by user profile (e.g., current EV owners versus potential buyers) could reveal different concerns and priorities for each group.

Acknowledgments

This section has been omitted for peer review. The authors thank CNPq for the support of the Universal Project 402086/2023-6, as well as FAPERGS for the project: ARD/ARC – Process 24/2551-0000645-1.

References

Lau, J. H. (2025). Comparative analysis of gasoline and electric vehicles: Economic, environmental, and consumer perspectives. *FE - Future Energy Studies*.

Oliveira, K., Lunardi, G., and Silva, W. (2025a). Avaliação de sentimentos de aplicativos: Uma comparação entre modelos de linguagem de grande escala. In *Anais da XX Escola Regional de Banco de Dados*.

Oliveira, K., Lunardi, G., Silva, W., Silveira, T., and Oliveira, A. (2025b). A data augmentation and validation pipeline for improving emotion classification in mobile app reviews. In *Anais do XL Simpósio Brasileiro de Bancos de Dados*.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., and Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications.

Sharma, H., Din, F. U., and Ogunleye, B. (2024). Sentiment analysis of electric vehicles using large language models. *Analytics*.

Soares, T., Costa, R., Soares, E., Calderon, I., Lunardi, G., Valle, P., Guedes, G., and Silva, W. (2025). Machine learning-assisted tools for user experience evaluation: A systematic mapping study. In *Anais do XXI Simpósio Brasileiro de Sistemas de Informação*, Porto Alegre, RS, Brasil.