Towards Improving Teacher Performance Assessment through Human-Centered AI-Powered Survey Analysis: An Approach Using Large Language Models (LLM)

Roberto E. Ramos-Rivera, Jesús García-Mancilla, Gerardo E. Cárdenas-Villa, Pedro C. Santana-Mancilla*

Abstract

Practical evaluation is crucial for improving educational quality. The blind review conducts semester surveys among students at all academic levels within the institution for a comprehensive professor performance assessment. However, processing this data manually is time-consuming and restricts in-depth analysis. With the recent advancements in Large Language Models, such as OpenAI's GPT-4, chatbots can now accurately interpret nuanced survey responses. We propose using a chatbot powered by artificial intelligence and natural language processing, specifically large language models, to streamline survey analysis. This technology aims to quickly extract important insights, reduce staff workload, and enable informed decision-making. We are utilizing user-centered design methods to create and assess a prototype, ensuring that it meets the specific needs and characteristics of the users and provides an optimal user experience in the final product. This implementation will significantly improve operational efficiency and support continuous educational enhancement at the institution.

Keywords:

Machine Learning, Large Language Models; Artificial Intelligence, Survey analysis; Text Analysis.

1 Introducción

In recent years, advancements in deep learning, particularly in the development of Large Language Models (LLMs), have significantly enhanced the implementation of systems such as chatbots in diverse domains, including healthcare [1], business [2],

Ramos-Rivera, RE., Santana-Mancilla, PC.
Universidad de Colima
Colima. México
Email: {roberto_ramos, gvilla, psantana}@ucol.mx

García-Mancilla, J.
JGMancilla.com
Houston. Texas, USA
Email: jesus@jgmancilla.com

* Corresponding author

education [3], customer service [4], and more. The evolution of LLMs has facilitated the creation of advanced chatbots like OpenAI's ChatGPT, Google's Gemini, and Microsoft's Copilot, which have demonstrated remarkable capabilities in processing and analyzing natural language.

Alongside these technological developments, periodic evaluation has emerged as a fundamental tool for ensuring continuous improvement in institutions' education quality. Through such evaluations, often conducted via surveys, students provide valuable insights into various aspects, including the quality of teaching and educational content and overall teacher performance. These insights enable educational administrators to gain critical information about students' perceptions and experiences [5]. However, analyzing open-ended responses in these surveys remains a significant challenge due to the wide variety and volume of data and the need for precise qualitative interpretation. In this context, LLMs present an opportunity, as these models excel at identifying key information embedded in survey responses [6], [7].

At the Universidad de Colima, periodic teacher performance evaluation through the Teacher Performance Assessment (*Sistema de Evaluación Docente*, or SED in the original Spanish) is a cornerstone for ensuring continuous improvement in teaching quality. These surveys are conducted every semester to collect opinions from a substantial percentage of students across all educational levels offered by the institution, providing valuable data on aspects such as teaching quality, teacher performance, and areas for improvement.

Driven by the need to enhance this process, this study proposes the development of an AI-powered chatbot leveraging Natural Language Processing (NLP), explicitly using LLMs. The chatbot aims to automate and streamline survey analysis, enabling administrative staff to efficiently extract key insights and make informed decisions to enhance the quality of education at the university.

2 Objectives and research approach

The motivation for this research stems from the need to improve the analysis process of teacher performance assessments conducted by students at the Universidad de Colima. The manual processing

amexihc

of these semesterly evaluations is a slow and error-prone task, limiting the ability to identify significant patterns and trends in open-ended responses. By implementing a chatbot based on Artificial Intelligence and Natural Language Processing (NLP) using Large Language Models (LLMs), this study aims to automate and expedite data analysis, enabling better decision-making.

This project is rooted in the field of Human-Computer Interaction. It focuses on developing an interface that is not only functional but also accessible and user-friendly for administrative staff. The design will adhere to usability and accessibility principles, ensuring that users can interact with the system intuitively without requiring advanced technical training.

Additionally, the project aligns with the principles of human-centered artificial intelligence by incorporating mechanisms to ensure ethical data handling, privacy protection for student responses, and fairness in their analysis. The system aims to accurately reflect the opinions and experiences of all students, regardless of personal characteristics, ensuring that all voices are equitably considered. These considerations are critical for creating a technologically advanced yet socially responsible system. This ethical focus is particularly relevant in survey analysis, where automated interpretations must be as impartial and accurate as possible to provide an authentic and representative view of student feedback.

This project seeks to address a technical problem by automating data analysis and integrating this solution effectively and ethically within the institution.

The research questions include:

- How can a chatbot based on Large Language Models improve the analysis of open-ended responses in teacher performance assessment compared to traditional methods?
- What patterns and common themes can be identified in open-ended survey responses, and how do these compare to findings from manual analysis?
- How can the chatbot's effectiveness and acceptance among users be evaluated, and what specific metrics should be used to assess its performance and user satisfaction?
- What design and operational strategies for the chatbot are necessary to ensure ethical data analysis, respect student privacy, and minimize bias in results?
- How can the chatbot's interaction with users be made intuitive and accessible, aligning with human-centered AI principles and providing a positive user experience?

The methodology for this research includes several stages:

1. Systematic literature review:
    - Conduct an exhaustive review of existing literature on using chatbots and Large Language Models in survey analysis, following PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*).
2. Data collection:
    - Gather data from semesterly teacher performance assessments conducted by students at the Universidad de Colima.
3. Chatbot development:
    - Design and develop an AI-powered chatbot using NLP and LLMs to analyze open-ended survey responses.
4. Evaluation and implementation:
    - Implement the chatbot within the university's assessment system and evaluate its performance using specific metrics, such as analysis accuracy and efficiency, alongside feedback from administrative staff and students.
5. User experience evaluation:
    - Conduct usability and accessibility tests with users to ensure the chatbot is intuitive and user-friendly. This phase will include iterative chatbot design improvements based on user feedback, aiming to align with human-centered AI principles.

This methodological approach will validate the hypothesis that a chatbot based on Large Language Models can significantly improve the analysis of student surveys, providing key insights more efficiently and accurately. Furthermore, the project will focus on designing an intuitive interface that facilitates natural communication, promotes effective adoption, and enhances user experience. This research emphasizes the importance of ethical and responsible interactions between humans and technology by proactively addressing critical issues such as privacy and equity.

3 Current phase of the research

The current phase of the research focuses on conducting a systematic and comprehensive literature review using chatbots for survey analysis and the interpretation of open-ended responses, following the PRISMA methodology. Key activities include collecting relevant articles and studies, selecting pertinent works, assessing the quality of the selected studies, and identifying common themes and gaps in the existing research.

- Current research on applying LLMs in higher education has been analyzed. Based on this, relevant themes and keywords were identified to carry out a systematic search across databases such as ScienceDirect, SpringerLink, IEEE Xplore, and ACM Digital Library, using the following search terms:
    - ("large language models" OR LLMs OR "GenAI" OR "Generative Artificial Intelligence" OR "Generative AI") AND ("student feedback" OR "student evaluation" OR "course feedback")
    - ("large language models" OR "LLM" OR "GenAI" OR "Generative Artificial Intelligence" OR "Generative AI") AND ("course evaluation" OR "course quality" OR "end of course")
    - ("large language models" OR "LLM" OR "GenAI" OR "Generative Artificial

Intelligence" OR "Generative AI") AND ("qualitative analysis" OR "qualitative feedback" OR "feedback analysis" OR "automated feedback")

o ("large language models" OR "LLM" OR "GenAI" OR "Generative Artificial Intelligence" OR "Generative AI") AND ("educational survey" OR "educational data" OR "higher education")

o ("large language models" OR "LLM" OR "GenAI" OR "Generative Artificial Intelligence" OR "Generative AI") AND ("teaching performance evaluation" OR "teacher evaluation" or "faculty evaluation")

After running these search queries, inclusion and exclusion criteria were established to filter the retrieved articles, as the results may include studies that match the search terms but are not relevant to this study. Subsequently, the articles were screened based on their titles and abstracts to evaluate their relevance to the research objectives.

3.1 Inclusion criteria

The inclusion criteria were as follows:

- Articles must be written in English.
- Articles must originate from peer-reviewed sources, such as books, conferences, or journals.
- Articles must be published between 2020 and 2024.
- Articles must include the search keywords in their titles or abstracts.

The initial search generated 5,908 papers to review, distributed across the four databases mentioned earlier. While the exhaustive review process is ongoing, some preliminary findings have been identified, offering an initial understanding of the subject

3.2 Preliminary findings

- Chatbots using Large Language Models have proven effective in interpreting and categorizing open-ended survey responses, improving efficiency, and overcoming the limitations of manual methods.
- Recurring themes in open-ended responses include student satisfaction, teaching quality, and the need for infrastructure improvement.
- Challenges in implementing chatbots have also been noted, such as the need for context-specific training and difficulty interpreting ambiguous or context-dependent responses.

3.3 Opportunities for improvement

- Incorporating mechanisms for continuous feedback to enhance the accuracy of chatbot analysis over time.
- Leveraging advanced Natural Language Processing techniques to interpret complex, contextually dependent responses better.

These preliminary findings provide a deep understanding of the needs and opportunities within the problem domain, guiding the development and design stages of the chatbot. They refine the hypothesis that a chatbot based on Large Language Models can significantly enhance the process of analyzing student surveys. The subsequent stages of this research involve designing and developing a prototype based on the findings from the literature review, evaluating the prototype in a controlled environment, and ensuring the chatbot's ability to deliver accurate and valuable analyses, thereby improving educational quality at the Universidad de Colima.

4 Expected contributions

This research focuses on developing an AI-powered chatbot leveraging Large Language Models (LLMs) to facilitate the automated analysis of teacher performance assessments. By implementing innovative strategies to interpret open-ended responses in these evaluations, the aim is to enhance the analysis's efficiency and accuracy. A comprehensive assessment of the chatbot's effectiveness in analyzing open-ended responses will provide detailed insights into the results. Based on the qualitative and quantitative data obtained during the implementation and evaluation phases, recommendations will be generated to ensure the continuous improvement of the process, adapting it to emerging needs.

By developing an intelligent agent that not only processes data efficiently but also interacts with users in an intuitive and accessible manner, this work fosters the creation of systems centered on user needs and experiences. This approach reduces technological barriers, promoting adoption and increasing user satisfaction.

Furthermore, by incorporating human-centered AI practices, this project addresses technical challenges and fundamental ethical issues, such as privacy, fairness, and transparency in data analysis. In doing so, it contributes to developing AI technologies that are powerful and responsible, prioritizing users and their needs.

5 Graduate consortium expectations

Participating in the MexIHC 2024 Graduate Consortium represents an invaluable opportunity for my research. I look forward to receiving critical and constructive feedback from renowned experts and other graduate students, which will significantly enrich my project. Given that my research is in its early stages, I hope to gain clear insights into how to refine and improve my methodological approach. Additionally, attending the consortium will allow me to learn firsthand about other participants' advancements and research efforts, fostering potential future collaborations. This experience will strengthen my understanding of human-centered artificial intelligence and contribute to the continuous development of my academic and professional network.

6 Acknowledgments

7 References

[1] A. Tersoo Catherine, S. K. Towfek, y A. A. Abdelhamid, "An Overview of the Evolution and Impact of Chatbots in Modern Healthcare Services", *MJAIH*, vol. 2023, pp. 71–75, dic. 2023, doi: 10.58496/MJAIH/2023/014.

[2] M. Mbaye, "Impact of Artificial Intelligence (AI) on the E-Commerce Business: Empirical Analysis for Optimal Use of the Chatbot", vol. 30, núm. 2, 2024.

[3] C. Bjelland, K. Ludvigsen, y A. Møgelvang, "Unveiling the Impact of AI Chatbots on Higher Education: Insights from Students", presentado en 18th INTED, mar. 2024, pp. 1458–1465. doi: 10.21125/inted.2024.0428.

[4] A. Rossmann, A. Zimmermann, y D. Hertweck, "The Impact of Chatbots on Customer Service Performance", en *Advances in the Human Side of Service Engineering*, vol. 1208, J. Spohrer y C. Leitner, Eds., en Advances in Intelligent Systems and Computing, vol. 1208. , Cham: Springer International Publishing, 2020, pp. 237–243. doi: 10.1007/978-3-030-51057-2_33.

[5] L. Grebennikov y M. Shah, "Student voice: using qualitative feedback from students to enhance their university experience", *Teaching in Higher Education*, vol. 18, núm. 6, pp. 606–618, ago. 2013, doi: 10.1080/13562517.2013.774353.

[6] B. J. Jansen, S. Jung, y J. Salminen, "Employing large language models in survey research", *Natural Language Processing Journal*, vol. 4, p. 100020, sep. 2023, doi: 10.1016/j.nlp.2023.100020.

[7] Z. Wang, P. Denny, J. Leinonen, y A. Luxton-Reilly, "Leveraging Large Language Models for Analysis of Student Course Feedback", en *Proceedings of the 16th Annual ACM India Compute Conference*, Hyderabad India: ACM, dic. 2023, pp. 76–79. doi: 10.1145/3627217.3627221.