



CAPABILITIES AND APPLICATION OF ARTIFICIAL INTELLIGENCE (AI) MODELS IN QUALITATIVE AND QUANTITATIVE DATA MINING, DATA PROCESSING AND DATA ANALYSIS

Moises C. Torrentira Jr.ⁱ

EnP, PhD.,

University of Southeastern Philippines,
Philippines

Abstract:

The study was conducted to determine the capabilities and application of Artificial Intelligence in the development of themes in qualitative and quantitative data analysis. Multi-stage data mining as a tool in AI data collection was employed by the researcher to process data reduction and the extraction of codes. This technique was able to generate the final themes pertaining to the capabilities and application of AI models, which include pattern recognition analysis, thematic analysis, sentiment analysis, efficient data processing, objectivity and bias reduction, and insight generation. These capabilities and applications of AI models can be used to highlight the benefits and ethical use of AI in research, especially in data mining, data processing, and data analysis.

Keywords: capabilities and application, artificial intelligence (AI) models, qualitative, quantitative, data mining, data processing, and data analysis

1. Introduction

The recent innovation and advancement of technology in terms of machine learning and data mining has brought a complete shift of paradigm in research from traditional to Artificial Intelligence or AI-driven research. According to Kelly *et al.* (2023), Artificial intelligence (AI) is the crucial technology that will shape the future of human civilization because of its capacity to upscale the capability of man without spending too much. The traditional conduct of qualitative and quantitative research is rigorous, time-consuming, laborious, and highly technical, while the AI-driven way of conducting research is lenient, fast, convenient, time-efficient, and sometimes contactless. This contrast, however, is inevitable and has even invited debates arising from its ethical consideration. Many have been alarmed about the capability of ChatGPT, Solab, Ernie Bot, Gemini,

ⁱ Correspondence: email moises.torrentira@usep.edu.ph

Claude, Grok, and Llama, Copilot (AP News, 2023), to name a few, to eventually depose human intelligence, facilitate plagiarism, and feed or misrepresent information. However, the Internet of Things has revolutionized the way researchers gather and analyze data through the Internet. In fact, the majority of the journals have already started accepting research articles that are AI-driven for possible publication. This means that Artificial Intelligence, especially ChatGPT and the like, has infiltrated the world of research (null, 2023). Mainstreaming AI in research, therefore, is not far off from reality, considering its practical acceptability.

In particular, artificial intelligence is very useful in data analysis. In terms of qualitative research, the bulk of non-numerical data gathered through interviews, focus group discussion, stakeholders' consultation, the Delphi technique, field reconnaissance, field notes and observations can be very tedious to process. It will require the right skills in data processing and analysis to engage in a meaningful play of words, codes, and themes. Although qualitative data processing can be done by one person alone, sometimes, the collaboration of two or more researchers will result in a better formulation of themes. On the other hand, under quantitative research, thematic analysis is necessary to come up with themes in exploratory factor analysis (EFA). Although the statistical software groups items, the naming of themes or dimensions will rely purely on the researcher's meaningful analysis. This is where artificial intelligence works in a powerful and efficient way. The AI may inquire about the right codes or themes based on the statistically grouped pattern of ideas using the software.

This research, therefore, identified the capabilities of artificial intelligence models, particularly in the development of themes for qualitative and quantitative data. Specifically, it answered the acceptable capabilities and applications of AI models in theme development of the gathered data sets from the AI models themselves.

2. Methodology

This study made use of artificial intelligence to generate the answers to the research questions. Considering the technical limitations, the study utilized the following AI models only: ChatGPT, Solab, Claude, YesChat, Gemini, and Copilot. Generally, these AI or language models are conceived to grasp and formulate human-like scripts based on the concepts or inquiries that they receive. They have the capacity to analyze vast amounts of text data and utilize a pattern of algorithms to process and respond to queries or commands (Asatryan, 2023).

Multi-stage data mining was used to collect the needed data or responses from the AI models. This is a kind of collecting data that is done by subjecting the AI responses to different levels of data reduction inquiries until the final themes are identified. This type of data mining is also commonly used to grasp the everchanging trends of the stock market (Huang *et al.*, 2011). However, in this study, multi-stage data mining was utilized to reduce the responses to the final themes. For the first stage of data mining, the same inquiry was loaded in each of the AI's input fields, specifically: "How helpful is Artificial

Intelligence in theme development to analyze the collected data?" The selected AI models then generated very general responses. These responses are still raw data. The second stage induced human interaction with AI by manually extracting and identifying the codes generated pertaining to the capabilities and application of Artificial Intelligence in data mining and analysis. The third stage involving the AI impelled the researcher to load manually the identified codes to ChatGPT's input field, which has the capability to deal with challenging language understanding and generation tasks in the form of conversation (Wu *et al.*, 2023). The AI then generated the themes in an instant, providing an opportunity for the researcher to gain impartial finalization of themes. Note further that without the help of AI in the production of themes, the human-generated themes are sometimes biased and limited to the mind frame of the researcher due to the dearth of information and exposure to data analysis (Beloto, 2018). The fourth stage consisted of the analysis of AI-generated data that shall be considered in the final selection of themes. Although human decision-making is highly influenced at this stage, the AI-generated themes were included in this study and were corroborated with other related literature for triangulation purposes. This is done to validate the final themes further and ensure the legitimacy and precision (Turner, 2006) of the whole process.

3. Results and Discussion

The process generated several groupings that emerged in an instant and suggested the following themes on the capabilities and application of Artificial Intelligence in the analysis of qualitative and quantitative data:

3.1 Pattern Recognition Analysis

The algorithms are designed to recognize the keywords, phrases, and concepts within the data sets. In turn, the AI efficiently processes the large volume of this data to identify patterns, trends and even correlations. Apparently, Artificial Intelligence excels at identifying subtle, recurrent themes across large data sets to represent the phenomenon being studied. Additionally, techniques like Latent Dirichlet Allocation (LDA) automatically discover abstract topics or themes present in textual data, which can form a network of modeling. Suen *et al.* (2021) emphasized how pattern recognition and artificial intelligence can be used to forecast further discoveries based on the observed sequence of eventualities.

3.2 Thematic Analysis

AI algorithms extract meaningful themes and topics from textual, numerical, and multimedia data. Social researchers can explore underlying structures and thematic clusters with qualitative data (Terry *et al.*, 2017). Further, AI generates initial thematic frameworks and presents the identified constructs representing the groups of similar trends. Natural Language Processing (NLP) techniques (Chowdhary & Chowdhary, 2020), for example, categorize information and help discover underlying themes and

concepts. This is because, with thematic analysis, the recurring ideas from the qualitative data sets can be analyzed (Riger & Sigurvinsdottir, 2016).

3.3 Sentiment Analysis

A capability that is very limited to human understanding is the capability of AI to assess the emotional tone and sentiment expressed in qualitative data. This aids in understanding the participant's attitudes and opinions in qualitative research. Aside from that, AI can recognize emotions that are expressed in texts (Taherdoost & Madanchian, 2023). By analyzing sentiment and emotion, the AI model helps uncover implicit themes and context behind words expressed as qualitative data. In fact, according to Chakriswaran *et al.* (2019), AI can uncover the exact human emotion in the context.

3.4 Efficient Data Processing

Processing large volumes of data can be burdensome and time-consuming for researchers. It can affect the productive use of time by spending so much on a bulk of data that can either be used or dumped. But the AI models can efficiently process large volumes of data, which enables the researchers to identify patterns, trends, and correlations quickly. AI assists in generalizing, categorizing, organizing, cleaning, and transforming datasets, streamlining the whole research process. Researchers can handle vast amounts of information more efficiently with the assistance of AI, which makes qualitative analysis very scalable. Especially for unstructured data, according to Baviskar, *et al.* (2021), millions of dollars are spent by businessmen annually just to extract information using traditional techniques. Nevertheless, with AI models, the bulk of data can be processed in an instant with ease and remarkable acceptability and credibility.

3.5 Objectivity and Bias Reduction

Qualitative data processing and analysis that involve human intelligence are sometimes challenged by objectivity and bias. This can usually happen when researchers select information from their own understanding and use them to label a certain phenomenon or experience (Sun, *et al.*, 2020). On the contrary, Artificial Intelligence approaches data analysis objectively, minimizing the influence of human bias. Consistency in coding and analysis is ensured by applying the same rules and algorithms across the dataset. Unlike human decision-making, which can be flawed and shaped by individual and societal biases (Silberg & Maynika, 2019), the AI then accelerates the process of coding and categorizing data, reducing the potential for human error and bias.

3.6 Insight Generation

In Philosophy, insight can be defined as the product of a deep and clear understanding of certain phenomena leading to the abstraction of significant and extraordinary ideas. The AI models can uncover these hidden insights, formulate emergent themes, and intensify the connections within data. Commonly, researchers draw more robust

conclusions and generate valuable insights from collected data. On the other hand, AI models usually excel at recognizing subtle patterns and connections, identifying latent themes not immediately apparent to human analysts. Because of their capacity to process large-scale and unstructured data sets, AI models can develop insights that cannot be predicted easily by a human mind (Ma & Sun, 2020).

4. Conclusion

The use of Artificial Intelligence in data analysis increases the capabilities of researchers to expedite research findings without much of a cost. It streamlines the research process, allowing the researchers to focus more on interpretation rather than data processing and analysis. Also, the integration of artificial intelligence in data mining, especially in the development of themes significantly enhances the researchers' capabilities to analyze the data comprehensively while extracting important insights on the research findings. However, despite the important role of AI models in data analysis, it must be noted that there is no greater knowledge than that of the human being himself. Human interpretation and contextual understanding remain crucial and should complement the power of AI.

Conflict of Interest Statement

The author declares no conflicts of interest.

About the Author

Dr. Torrentira is a licensed environmental planner and a professional teacher. He earned a Doctorate Degree in Development Administration while his Master's degrees include Master in Environmental Planning, Master in Business Administration and Master in Public Administration. His Orcid ID is <https://orcid.org/0000-0003-3812-0182>. His works are searchable in Academia and ResearchGate. He has been consistently cited by other authors in various academic fields.

References

- ["Artificial intelligence technology behind ChatGPT was built in Iowa – with a lot of water". AP News. September 9, 2023. Archived from the original on September 10, 2023. Retrieved on September 10, 2023.](#)
- Asatryan, B., Bleijendaal, H., & Wilde, A. A. (2023). Towards advanced diagnosis and management of inherited arrhythmia syndromes: Harnessing the capabilities of artificial intelligence and machine learning. *Heart Rhythm*.
- Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603-649. <https://doi.org/10.1016/j.hrthm.2023.07.001>

- Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021). Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions. *IEEE Access*, 9, 72894-72936. Retrieved from <https://ieeexplore.ieee.org/document/9402739>
- Belotto, M. J. (2018). Data analysis methods for qualitative research: Managing the challenges of coding, interrater reliability, and thematic analysis. *The qualitative report*, 23(11), 2622-2633. <http://dx.doi.org/10.46743/2160-3715/2018.3492>
- Bishop, C. M. (2006). Pattern recognition and machine learning. *Springer* 645-678. Retrieved from <https://link.springer.com/book/9780387310732>
- Chakriswaran, P., Vincent, D. R., Srinivasan, K., Sharma, V., Chang, C. Y., & Reina, D. G. (2019). Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues. *Applied Sciences*, 9(24), 5462. Retrieved from <https://www.mdpi.com/2076-3417/9/24/5462>
- Huang, C.J., Chen, P.W. & Pan, W.T. Using multi-stage data mining technique to build forecast model for Taiwan stocks. *Neural Comput & Applic* 21, 2057–2063 (2012). <https://doi.org/10.1007/s00521-011-0628-0>
- Kelly, S., Kaye, S.A., Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, Vol. 77. <https://doi.org/10.1016/j.tele.2022.101925>
- Ma, L., & Sun, B. (2020). Machine learning and AI in marketing—Connecting computing power to human insights. *International Journal of Research in Marketing*, 37(3), 481-504. <https://doi.org/10.1016/j.ijresmar.2020.04.005>
- Marshall, D. T., & Naff, D. B. (2024). The Ethics of Using Artificial Intelligence in Qualitative Research. *Journal of empirical research on human research ethics: JERHRE*, 19(3), 92–102. <https://doi.org/10.1177/15562646241262659>
- Riger, S., & Sigurvinsdottir, R. (2016). *Thematic analysis*. In L. A. Jason & D. S. Glenwick (Eds.), *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods* (pp. 33–41). Oxford University Press.
- Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). *McKinsey Global Institute*, 1(6), 1-31. Retrieved from <https://www.mckinsey.com/~media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/mgi-tackling-bias-in-ai-june-2019.pdf>
- Sun, W., Nasraoui, O., & Shafto, P. (2020). Evolution and impact of bias in human and machine learning algorithm interaction. *Plos one*, 15(8), e0235502. <https://doi.org/10.1371/journal.pone.0235502>
- Taherdoost, H., & Madanchian, M. (2023). Artificial intelligence and sentiment analysis: A review in competitive research. *Computers*, 12(2), 37. <https://doi.org/10.3390/computers12020037>
- Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. *The SAGE handbook of qualitative research in psychology*, 2(17-37), 25.

- Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods: A triangulation-based framework and roadmap. *Organizational research methods*, 20(2), 243-267. <https://doi.org/10.4135/9781526405555>
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5), 1122-1136. Retrieved from <https://ieeexplore.ieee.org/document/10113601>

Creative Commons licensing terms

Author(s) will retain the copyright of their published articles agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Education Studies shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflicts of interest, copyright violations and inappropriate or inaccurate use of any kind content related or integrated into the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).