Advances in Social Sciences Research Journal – Vol. 12, No. 08
Publication Date: August 29, 2025
DOI:10.14738/assrj.1208.19286.
Kalanda, B. F. & Cheboi, A. J. (2025). Artificial Intelligence in the Analysis of Unstructured Qualitative Data: A Literature Review. *Advances in Social Sciences Research Journal*, 12(08). 199-205.

Artificial Intelligence in the Analysis of Unstructured Qualitative Data: A Literature Review

Boniface Francis Kalanda[*†]
University of Malawi, Chancellor College,
Zomba, Malawi

Asseneth Jerotich Cheboi
University of Walden, College of Health Sciences
& Public Policy, United States of America

ABSTRACT

International development institutions, United Nations agencies, and non-governmental organizations (NGOs) generate vast quantities of unstructured qualitative data, including narrative reports, field interviews, focus group summaries, and policy documents. Despite the significant financial investments in producing this data—NGOs alone disbursing over US$23 billion annually—its potential remains underutilized due to the lack of scalable analytical methods. Traditional qualitative analysis approaches, such as content analysis, thematic analysis, narrative analysis, discourse analysis, and grounded theory, are resource-intensive, time-consuming, and often unsustainable at the scale required. Artificial Intelligence (AI) offers a promising alternative by enabling automated or semi-automated analysis of large datasets, enhancing efficiency, reliability, and the breadth of insights generated. This literature review examines the application of AI in analyzing unstructured qualitative data, focusing on its use in healthcare, social sciences, and policy research. AI techniques, including natural language processing (NLP), sentiment analysis, clustering, topic modeling, and pattern recognition, have demonstrated potential to replicate human-led content analysis outcomes in significantly reduced timeframes. For instance, in healthcare research, AI has been used to analyze chronic pain narratives and COVID-19-related qualitative datasets, often achieving results comparable to human analysis but with enhanced speed. Similarly, in education and social science contexts, AI tools such as Python's SpaCy have processed thousands of open-ended survey responses, enabling meaningful interpretation at scale. The review identifies several advantages of AI-enabled analysis. These include substantial time and cost savings, reduced analyst fatigue, improved coding accuracy, and the ability to conduct more comprehensive literature reviews and thematic explorations. AI also supports iterative and reflexive analysis, expanding the possibilities of theory development and data visualization. Such capabilities make AI particularly attractive to institutions with limited analytical resources yet large volumes of qualitative data. Nonetheless, limitations persist. AI struggles with nuanced interpretation, including detecting sarcasm, cultural references, and subtle thematic variations. Ethical concerns—ranging from data privacy to algorithmic bias—remain critical considerations.

[*] Corresponding Author: bkalanda@yahoo.com
[†] The views expressed in this paper are those of the authors and not their employers.

Reliability issues, "hallucinations" in generative AI outputs, and reduced interpretive depth compared to traditional qualitative methods underscore the need for cautious application. Studies suggest that deductive AI setups may outperform inductive approaches, but human reflexivity, contextual knowledge, and interpretive judgment remain indispensable. A recurring theme across the literature is the advocacy for hybrid analytical models where AI augments rather than replaces human expertise. In such models, AI serves as a rapid pattern-detection and preliminary coding tool, allowing human analysts to focus on higher-order interpretation and critical inquiry. This approach can maximize efficiency while preserving the depth and contextual sensitivity that define rigorous qualitative research. The paper concludes that AI holds transformative potential for the analysis of unstructured qualitative data, particularly for organizations in development, health, and policy sectors. Realizing this potential requires sustained collaboration between institutions, academia, and the private sector to refine methodologies, test across diverse datasets, and address ethical and reliability challenges. Furthermore, generational differences in AI adoption—where younger researchers are more inclined toward AI use—suggest opportunities for targeted capacity-building initiatives. By strategically integrating AI into qualitative workflows, stakeholders can unlock greater value from existing data, inform decision-making, and enhance the inclusivity and scope of research insights.

Keywords: Artificial, Intelligence, data, qualitative, unstructured.

INTRODUCTION

International development institutions such as the United Nations, Non-Government Organizations and Think Tanks generate a lot of unstructured[‡] qualitative data. This data is in the form of reports with text, field interviews, focus group discussions summaries, policy and advocacy reports amongst others and it is in large quantities (Pope, 2000). It is difficult to precisely estimate the quantity of such data. Lewis & Kanji (2009) estimates that there are 35,000 NGOs producing such data. In addition, there are over twenty (20) UN agencies operating in over 190 countries that generate such data through their implementing partners. All these entities generate large amounts of qualitative data that is of varied quality. Mostly such data is not generated to answer specific structured research questions.

Unlike structured quantitative and qualitative data, there are limited methods of analyzing unstructured qualitative data. This lack of methods of analyzing unstructured qualitative data presents a missed opportunity to get value from the investments made in collecting this data. Older estimates show that only NGOs received and disbursed over US$ 23 billion (Riddell 2007: 53). It has been shown elsewhere that such data can be used to study hard to reach and marginalized groups (Masefield et al., 2020). In terms of both financial investments and uses of such data, this data is of enormous value.

Content analysis has been the main method of analyzing unstructured qualitative data. Other methods include thematic analysis (Braun et al, 2006), narrative analysis, discourse analysis and grounded theory. It is however difficult and expensive to analyze such data using these methods due to the large volumes of unstructured qualitative data being generated. This

[‡] Unstructured data refers to data that does not follow a clear, predefined format, does not have an organizing model and is not stored in spread sheets or relational databases.

requires dedicated staff, constantly pouring over reports leading to fatigue and abandonment of the exercise. It is not sustainable. Bijker et al (2024) posits that analysis of large qualitative data sets is labour intensive and acknowledges that AI can be a useful tool in assisting with qualitative analysis taking into consideration reliability of outcomes.

Artificial intelligence could be an option of analyzing this data so that value is generated from it (Yang & Ma, 2005), Hitch (2023). AI offers the potential to automate and enhance the qualitative analysis process. AI could support pattern recognition, classification, clustering, sentiment analysis, topic modeling, and even theory generation (Khan et al, 2020). Others have proposed that AI could augment content analysis (Kuckartz & Rädiker, 2024) and all the above functions.

This paper is a literature review of use of AI in analyzing unstructured qualitative data. Use of AI has limitations that originate from its own design (OpenAI, 2023) including bias (Panteli, et al, 2025). This paper however does not address the inherent limitations of AI. It assumes AI works as expected. This paper does also not define what AI is. Others such as Russel & Norgiv (2016) and Cormen et. al (2009) have comprehensively defined AI elsewhere.

Definition of Unstructured Qualitative Data

Unstructured data refers to data that does not follow a clear, predefined format, does not have an organizing model and is not stored in spread sheets or relational databases**.** Unstructured qualitative data is "context specific," adding to its complexity in analysis. Examples include narratives, conversation transcripts, text from social media, audio-visual recordings, and images. This type of data is rich in meaning and context but difficult to analyze using traditional statistical methods.

APPLICATION OF AI IN ANALYSIS OF UNSTRUCTURED QUALITATIVE DATA

AI in Analysis of Health Care Data

There are examples of using AI to analyze qualitative data of patients with chronic pain. While limitations have been identified, Cheligeer et al (2022) found that AI enabled analysis is more efficient and accurate. Generative AI tools, such as ChatGPT and BardAI, summarize data into text for expedited information-gathering and content creation, and are gaining use in clinical and operations settings to help nursing staff improve productivity and decrease waste of resource inputs by eliminating menial tasks and enabling more informed and efficient clinical decisions (Carroll, 2023). Towler et al (2023), in analyzing COVID related qualitative data found that both AI and human analysis arrived at the same results. AI enabled analysis was much shorter. However, Danielle (2023) found that use of natural language processing-based AI using ChatGPT platform in health research has some limitations. For example, it is not capable of carrying out contextual interpretation, especially for thematic analysis where it requires interpretations of social dynamics and cultural meanings. Triangulation found higher similarities all be that differences from human nuances were observed (Towler et al, 2023).

AI in Analysis of Social Science Data

During the COVID pandemic, AI was used to analyze teacher and students perceptions and feelings arising from school closures. In Canada, Rocha (2021), used AI (Python, SpaCY) to analyze responses from 50,000 teachers. Using AI to analyze unstructured qualitative data from 9,400 respondents, it was possible to conduct meaningful analysis in a reduced time (Rocha,

URL: http://dx.doi.org/10.14738/assrj.1208.19286

2021). AI has also been used in programmes on income recipients (Hamilton, 2023), generating similarities between AI and human led analysis. AI has been used in general in social sciences among other tasks to track complex theories through word occurrence in large qualitative data sets and identification of social agents (Macanovic, 2022).

Potential for AI in Analysis of Unstructured Data

There are many reasons why investments in AI to analyze structure data should continue. Analysis of structure data using AI takes less time (Iliev, 2025), saving researchers time and money in analysis (Lenon, 2022) and leads to improved accuracy unlike human led content analysis (Haouam, 2025). AI has also shown reliability in analysis of qualitative data (Lennon, 2021).

AI has also been used in conducting literature reviews in large areas of research (Christou, 2023). Because of the computing power of AI, literature reviews can be more comprehensive. AI has also been recommended for exploration of thematic areas of analyses from large qualitative data sets (Dengel, et al, 2023). Unlike simple categorization, AI supports iterative analysis and reflexive interpretation (Friese, 2025).

Others have also documented the accuracy (Haouma, 2025) and efficiency (Nguyen-Trung, 2025) of AI in coding, identification and theme recognition and visualization of qualitative data.

Limitations for AI in Analysis of Unstructured Data

The limitations of AI such as ethics (Gal, 2023), "hallucinations" (Lackshmanan, 2022, Friese, 2025) and interpretivism (Williams, 2024) have been addressed by various others including (OpenAI, 2023). This section is limited to highlighting the limitations of AI in its use of analysis of unstructured qualitative data.

AI can be employed in analysis of unstructured descriptive data. It is easy to use, can save staff time and it can analyze a lot of data in a short period of time. This supports efficiency in data analysis (Arora, 2024). With AI models that can analyze data as expected, then large volumes of data can be analyzed in a short time, with reduced human error, better reliability, and reduced costs of employing large numbers of dedicated staff focusing on analysis of unstructured qualitative data.

While AI can easily reproduce major themes, its limitations include inability to interpret subtle themes (Morgan, 2023). These may include nuanced language, slang, sarcasm, and culturally specific issues, leading to oversimplification of complex experiences.

Akramul Kabir et al (2025) have also documented limitations in reliability of AI tools such as AvidNote, arguing further that AI cannot yet replace "human instincts". Other areas that need more research including using AI for deductive versus inductive analysis, with deductive set up of AI performing better (Siiman et al, 2025). Morgan (2023) has also shown that other AI models such as ChatGpt are more successful with descriptions other than interpretative aspects of analysis.

There is increasing advocacy for a hybrid model where AI augments human judgment rather than replaces it (Chen, 2025). Researchers emphasize the importance of maintaining

reflexivity, contextual interpretation, and critical inquiry. Morgan (2023) has suggested that AI should be better used as an assistant to find what is in the data than using it as a full-fledged analysis tool. This is also a view by Hitch (2023) and Kuckartz and Rädiker (2024) who has argued that AI can enhance human led analysis of qualitative data using various methodologies. While AI has been shown to reduce time of analysis and come to the same conclusions as human led content analysis, traditional analysis still provides better nuance (Guetterman, et all, 2018).

CONCLUSIONS

Artificial intelligence (AI) promises to transform analysis of large volumes of unstructured qualitative data. While there are limitations, the use of AI in this area is very promising. Institutions which generate a lot of data need to partner with academia and private sector to continue evaluating, using varied data sets, use of AI in analysis of unstructured qualitative data. This partnership will improve knowledge, application and ultimate use of AI for analysis of unstructured qualitative data to generate valuable insights. There is also a generational gap in adoption of use of AI, with younger researchers embracing AI more (Chatzichristos, 2025). This gap calls for intentional targeting of younger researchers to champion use of AI in analysis of qualitative data.

References

Pope, C. (2000). Analysing qualitative data. British Medical Journal: 2000; 320:114.

Bijker R, Merkouris S, Dowling N, Rodda S (2024). ChatGPT for Automated Qualitative Research: Content Analysis J Med Internet Res 2024;26:e59050 URL: https://www.jmir.org/2024/1/e59050 DOI: 10.2196/59050

Chatzichristos G (2025). Qualitative Research in the Era of AI: A Return to Positivism or a New Paradigm. International Journal of Qualitative Methods. Volume 24: 1–12. DOI: 10.1177/16094069251337583

Cheligeer C., Yang L., Nandi T., Doktorchik C., Quan H., Zeng Y., Singh S. (2022). Natural language processing (NLP) aided qualitative method in health research. Journal of Integrated Design and Process Science, Preprint, 1–18.

Hitch D (2023). Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future? Qualitative Health Research. 2023;34(7):595-606. doi:10.1177/10497323231217392

Lewis D, Kanji N. Non-government organizations and development. London: Routledge; 2009.

Riddell, R. (2007) Does Foreign Aid Really Work? Oxford: Oxford University Press.

Mende, J (2022). Extended Qualitative Content Analysis: Researching the United Nations and other International Institutions, in: Qualitative Research Journal.

Masefield SC, Megaw A, Barlow M, White PCL, Altink H, Grugel J (2020). Repurposing NGO data for better research outcomes: a scoping review of the use and secondary analysis of NGO data in health policy and systems research. Health Res Policy Syst. 2020 Jun 8;18(1):63. doi: 10.1186/s12961-020-00577-x. PMID: 32513183; PMCID: PMC7278191.

Kuckartz, U. & Rädiker, S. (2024). Integrating artificial intelligence (AI) in qualitative content analysis. https://qca-method.net/documents/kuckartz-raediker-2024-integrating-ai-inqualitative-content-analysis.pdf

Hitch D. (2023). Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future?. Qualitative Health Research. Vol. 34(7) 595–60.

Yang, Y., Ma, L. (2005). Artificial intelligence in qualitative analysis: a practical guide and reflections based on results from using GPT to analyze interview data in a substance use program. Qual Quant (2025). https://doi.org/10.1007/s11135-025-02066-1.

Morgan, D. L. (2023). Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. International Journal of Qualitative Methods, 22. December 2023.

Lakshmanan L. (2022). Why large language models like ChatGPT are bullshit artists, and how to use them effectively anyway. LinkedIn. https://www.linkedin.com/pulse/

Gal U. (2023, August 4th). ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned. The Conversation. https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283

Russel S., Norvig P. (2016). Artificial intelligence: A modern approach. Pearson.

Cormen T., Leiserson C., Rivest R., Stein C. (2009). Introduction to algorithms (3rd ed.). MIT Press.

Arora, K (2024). Ultimate Guide to AI Qualitative Data Analysis [Tools Included]

Explore top tools and techniques to effectively analyze qualitative data using AI. Qualitative Data Analysis: Methods, Tools, and FAQs Included

Khan, R; Shrivastava, P; Kapoor, A; Tiwari, A; Mittal, A (2020). Social media analysis with AI: Sentiment analysis techniques for the analysis of twitter COVID-19 data. Journal of Critical Reviews. ISSN- 2394-5125 Vol, 7, Issue 9.

Ronaghi, S.L., Emmilie Aveling, E., Sara Singer, S. (2025). Integrating AI into Qualitative Analysis. AcadamyHealth (March 12, 20205). Integrating AI into Qualitative Analysis | AcademyHealth.

Akramul Kabir, S.M; Ali, F; Lotfy Ahmed, R; Ruqayya Sulaiman-Hill, R (2025). Exploring the Use of AI in Qualitative Data Analysis: Comparing Manual Processing with Avidnote for Theme Generation. International Journal of Qualitative Methods Volume 24, April 2025.

Panteli, D., Adib,K., Buttigieg,S., Goiana-da-Silva, F., Ladewig, K., Natasha Azzopardi-Muscat, N., Figueras, J., Novillo-Ortiz, D., McKe, M (2025). Artificial intelligence in public health: promises, challenges, and an agenda for policy makers and public health institutions. Lancet Public Health 2025; 10: e428–32 Published Online February 28, 2025 https://doi.org/10.1016/ S2468-2667(25)00036-2 .

Siiman, L.A., Rannastu-Avalos, M., Pöysä-Tarhonen, J., Häkkinen, P., Pedaste, M. (2023). Opportunities and Challenges for AI-Assisted Qualitative Data Analysis: An Example from Collaborative Problem-Solving Discourse Data. In: Huang, YM., Rocha, T. (eds) Innovative Technologies and Learning. ICITL 2023. Lecture Notes in Computer Science, vol 14099. Springer, Cham. https://doi.org/10.1007/978-3-031-40113-8_9

Chen, W. (2025). Exploring the Application of AI to Qualitative Data Analysis: A Comparative Study in the Field of Industrial Design Education. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2025. Lecture Notes in Computer Science, vol 15820. Springer, Cham. https://doi.org/10.1007/978-3-031-93415-5_8

Carroll W. M. (2023). Generative AI in clinical practice and operations. Nursing management, 54(10), 56. https://doi.org/10.1097/nmg.0000000000000056

Towler L; Bondaronek P; Papakonstantinou T; Amlôt R; Chadborn T; Ainsworth B; Yardley L (2023). Applying machine-learning to rapidly analyze large qualitative text datasets to inform the COVID-19 pandemic response: comparing human and machine-assisted topic analysis techniques. Front Public Health. 2023 Oct 31; 11:1268223. doi: 10.3389/fpubh.2023.1268223. PMID: 38026376; PMCID: PMC10644111.

Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VGV (2018). Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study. J Med Internet Res. 2018 Jun 29;20(6):e231. doi: 10.2196/jmir.9702. PMID: 29959110; PMCID: PMC6045788.

Rocha, R (2021). Using NLP to analyze open-ended responses in surveys. Using NLP to analyze open-ended responses in surveys – Roberto Rocha

Iliev R, Dehghani M, Sagie E (2015). Automated text analysis in psychology: methods, applications, and future developments. Language and Cognition. 2015;7(2):265-290. doi:10.1017/langcog.2014.30.

Christou P (2023). How to Use Artificial Intelligence (AI) as a Resource, Methodological and Analysis Tool in Qualitative Research?. The Qualitative Report 2023 Volume 28, Number 7, 1968-1980. https://doi.org/10.46743/2160-3715/2023.6406

Haouam, I (2025). Leveraging AI for Advancements in Qualitative Research Methodology. Journal on Artificial Intelligence. V(7). DO - 10.32604/jai.2025.064145.

Williams R.T (2024). Paradigm shifts: exploring AI's influence on qualitative inquiry and analysis. Front Res Metr Anal. 2024 Dec 5;9:1331589. doi: 10.3389/frma.2024.1331589. PMID: 39703400; PMCID: PMC11656929.

Hamilton L; Elliott, D; Quick, A; Smith, S; Choplin V (2023). Exploring the Use of AI in Qualitative Analysis: A Comparative Study of Guaranteed Income Data. International Journal of Qualitative Methods. V(22). DO - 10.1177/16094069231201504

Dengel, A; Gehrlein, R; Fernes, D; Görlich, S; Maurer, J; Pham, H; Großmann, G; Eisermann, N (2023). Qualitative Research Methods for Large Language Models: Conducting Semi-Structured Interviews with ChatGPT and BARD on Computer Science Education. Informatics (10).DO - 10.3390/informatics10040078

Macanovic, A (2022). Text mining for social science – The state and the future of computational text analysis in sociology. Social Science Research 108 (2022) 102784.

Friese, S. (2025). From coding to conversation: A new methodology for AI-assisted qualitative analysis [Manuscript submitted for publication].

Nguyen-Trung, K (2025). ChatGPT in thematic analysis: Can AI become a research assistant in qualitative research?. Quality & Quantity https://doi.org/10.1007/s11135-025-02165-z

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3 (2). pp. 77-101. ISSN 1478-0887 Available from: http://eprints.uwe.ac.uk/11735

Lennon R, Calo W, Miller E, Zgierska A, Van Scoy L, Fraleigh R (2022). Using artificial intelligence to support rapid, mixed-methods analysis: Developing an automated qualitative assistant (AQUA). Ann Fam Med. 2022 Apr 1;20(20 Suppl 1):2893. doi: 10.1370/afm.20.s1.2893. PMID: 36944057; PMCID: PMC10549007.

Lennon RP, Fraleigh R, Van Scoy LJ, Keshaviah A, Hu XC, Snyder BL, Miller EL, Calo WA, Zgierska AE, Griffin C. Developing and testing an automated qualitative assistant (AQUA) to support qualitative analysis. Fam Med Community Health. 2021 Nov;9(Suppl 1):e001287. doi: 10.1136/fmch-2021-001287. PMID: 34824135; PMCID: PMC8627418.

OpenAI. (2023). How should AI systems behave, and who should decide? OpenAI. https://openai.com/blog/how-should-ai-systems-behave