DOI 10.18471/rbe.v38.61024

TRANSFORMATIVE TECHNOLOGIES: ARTIFICIAL INTELLIGENCE AND LARGE LANGUAGE MODELS IN QUALITATIVE RESEARCH

TECNOLOGIAS TRANSFORMATIVAS: INTELIGÊNCIA ARTIFICIAL E GRANDES MODELOS DE LINGUAGEM NA PESQUISA QUALITATIVA

TECNOLOGÍAS TRANSFORMADORAS: INTELIGENCIA ARTIFICIAL Y GRANDES MODELOS LINGÜÍSTICOS EN LA INVESTIGACIÓN CUALITATIVA

Grzegorz Bryda[1]
António Pedro Costa[2]

How to cite this article: Bryda G, Costa AP. Transformative technologies: artificial intelligence and large language models in qualitative research. Rev baiana enferm. 2024;38:e61024

Objective: To discuss the integration of artificial intelligence and advanced visualization techniques into computer-assisted qualitative data analysis software tools. Method: analytical and critical reflection that explores the possibilities, consequences and ethical aspects of this integration. Results: This integration has brought immense benefits and improved coding and analysis processes, the ability of artificial intelligence to deal with large sets of textual data, through innovative methods of visualizing qualitative data. Researchers can interpret and communicate differentiated findings more effectively by transforming data into visual representations with techniques such as diagrams, thematic maps and 2D/3D maps. Conclusion: The incorporation of artificial intelligence, machine learning and natural language processing technologies into computer-assisted qualitative data analysis software tools has automated the identification of themes and relationships in the data, speeding up the research process.

Descriptors: CAQDAS. Digital Humanities. Artificial Intelligence. Natural Language Processing. Large Language Models. ChatGPT. Automation.

Objetivo: discutir sobre a integração da inteligência artificial e as técnicas avançadas de visualização em ferramentas de software de análise de dados qualitativos assistidas por computador. Método: reflexão analítica e crítica que explora as possibilidades, consequências e aspectos éticos dessa integração. Resultados: esta integração trouxe imensos benefícios e aperfeiçoou os processos de codificação e análise, da capacidade da inteligência artificial para lidar com grandes conjuntos de dados textuais, mediante métodos inovadores de visualização dos dados qualitativos. Os pesquisadores podem interpretar e comunicar descobertas diferenciadas de forma mais eficaz, transformando dados em representações visuais com técnicas como diagramas, mapas temáticos e mapas 2D/3D. Conclusão: a incorporação de tecnologias da inteligência artificial, aprendizado de máquina e processamento de linguagem natural em ferramentas de software de análise de dados qualitativos assistidas por computador, automatizou a identificação de temas e relacionamentos dos dados, acelerando o processo de pesquisa.

Corresponding author: Grzegorz Bryda, grzegorz.bryda@uj.edu.pl

[1] CAQDAS TM, Institute of Sociology, Lab Jagiellonian University, Kraków, Polan. https://orcid.org/0000-0002-8892-099X
[2] University of Aveiro, Aveiro, Portugal. https://orcid.org/0000-0002-4644-5879

Descritores: CAQDAS. Humanidades Digitais. Inteligência Artificial. Processamento de Linguagem Natural. Grandes Modelos de Linguagem. ChatGPT. Automatização.

Objetivo: Discutir la integración de la inteligencia artificial y las técnicas avanzadas de visualización en las herramientas informáticas de análisis cualitativo de datos asistido por ordenador. Método: Reflexión analítica y crítica que explora las posibilidades, consecuencias y aspectos éticos de esta integración. Resultados: Esta integración ha aportado inmensos beneficios y ha mejorado los procesos de codificación y análisis, la capacidad de la inteligencia artificial para tratar grandes conjuntos de datos textuales, mediante métodos innovadores de visualización de datos cualitativos. Los investigadores pueden interpretar y comunicar conclusiones diferenciadas con mayor eficacia transformando los datos en representaciones visuales con técnicas como diagramas, mapas temáticos y mapas 2D/3D. Conclusiones: La incorporación de tecnologías de inteligencia artificial, aprendizaje automático y procesamiento del lenguaje natural a las herramientas informáticas de análisis de datos cualitativos asistido por ordenador ha automatizado la identificación de temas y relaciones en los datos, acelerando el proceso de investigación.

Descriptores: CAQDAS. Humanidades digitales. Inteligencia Artificial. Procesamiento del lenguaje natural. Grandes modelos lingüísticos. ChatGPT. Automatización.

Introduction

In the ever-changing landscape of contemporary qualitative research, the integration of technology, mainly through Computer-Assisted Qualitative Data Analysis Software (CAQDAS), has brought about significant changes in methodologies and practices. Debates surrounding the adoption of qualitative research software emerged during the late 1980s and early 1990s, inciting concerns about reliability, the decontextualization of data and the potential dilution of qualitative depth with technology.

Scholars such as Raymond Lee and Nigel Fielding (1–4) and Woods, Macklin and Lewis[5] have examined the impact of technology on qualitative research and highlighted issues of reflexivity, critically questioning the role of software in defining research findings. Although critics have emphasized the need for a balanced approach to research methodologies, the efficiency and advances offered by CAQDAS tools cannot be overlooked.

The integration of artificial intelligence (AI), machine learning (ML) and natural language processing (NLP) technologies into CAQDAS has introduced transformative capabilities in data analysis. AI and ML can automate coding processes, identify patterns and themes more efficiently and provide deeper insights from vast and complex data sets. NLP enables more sophisticated analysis of textual data, including sentiment analysis and extraction of differentiated meanings. The integration of large language models (LLMs) further enhances these capabilities, allowing researchers to process and understand large volumes of textual data in ways that mimic human comprehension.

These advances in CAQDAS tools have been shaped by qualitative research paradigms such as grounded theory and content analysis, emphasizing the importance of maintaining reliable research methodologies. The introduction of CAQDAS provided a structured workflow for data analysis, ensuring efficiency and organization. While these tools have brought enormous benefits, such as ease of collaboration and data management, researchers must remain aware of the potential drawbacks, including financial costs and the need for a careful balance between technology and traditional qualitative analysis.

This reflection considers the efficiency and depth of analysis that these technologies provide, ethical considerations, the potential introduction of bias, the impact of these technologies on researcher-participant relationships, data privacy and the authenticity of qualitative research will also be examined. Through this analysis, the article seeks to provide a balanced perspective, recognizing the transformative potential of technology while advocating for its conscientious and ethical use to preserve the integrity and depth of qualitative research.

The discussion will revolve around pertinent questions such as: How do AI and ML technologies influence traditional qualitative research practices while preserving integrity and depth? How has NLP revolutionized textual data analysis and how can researchers mitigate the risk of losing differentiated understanding? How do LLMs compare to traditional processing and understanding of large volumes of textual data? How can researchers navigate the ethical implications and biases introduced by technology? How can the development of CAQDAS be influenced by the challenges and opportunities faced in using technology for qualitative analysis? These questions aim to deepen understanding of the intersection between technology and qualitative analysis, promoting ongoing dialog and critical examination of these tools for qualitative research.

This article therefore critically analyzes and explores the possibilities, consequences and ethical aspects of integrating AI, ML, NLP and LLMs into CAQDAS.

AI-driven coding and data analysis

Qualitative data coding is an iterative process that requires researchers to constantly move between the data and the codes generated, refining the coding scheme to capture the complexities of the data. A systematic and consistent approach to coding qualitative data is required, through the use of a codebook or framework that details the coding process and defines criteria for codes, categories and concepts. The crucial step involves developing a structure of codings derived from theory or data[6–8].

Researchers usually employ two strategies for analyzing qualitative data: top-down and bottom-up analysis, directly related to coding schemes and research questions. Top-down analysis is often associated with a macro perspective on the data, covering a wide range of macro attributes or variables. Deductive coding stems from an established theoretical or analytical framework, indexing segments of data that align with predefined codes or categories. It is based on research questions, existing knowledge, protocols or literature.

On the other hand, the bottom-up approach involves a more focused examination of specific characteristics and micro attributes. In the inductive coding approach, the researcher generates codes or categories directly from qualitative data. This approach is based on the researcher's intuition, knowledge and experience, as well as the inherent content of the data. Codes emerge naturally from raw texts, rooted in the language of the data itself. Grounded theory exemplifies an inductive approach, employing three levels of coding: open, axial and selective[9].

Open coding divides, compares and categorizes data; axial coding connects data by establishing relationships between codes or categories after open coding, and selective coding relates the main category to other categories. Overall, coding simplifies the meaning of data into broad codes or categories, making it easier to retrieve and link data for similar purposes[1] .

Advances in CAQDAS methodology and computer *software* have increasingly blurred the boundaries between *top-down* and *bottom-up*, as well as quantitative and qualitative approaches[9-10]. In practice, researchers use both deductive (top-down) and inductive (bottom-up) strategies, depending on the direct approach to the research problem and the object of investigation.

Coding is a vital component of qualitative research methodology, which relies on the association between codes, their definitions and the meanings within text segments[2] . However, computer-assisted (manual) coding can be time-consuming and mentally demanding, especially when dealing with substantial volumes of unstructured, textual or linguistic data. After collecting the data, researchers must interact with it to understand and extract meaningful structures.

The coding process requires researchers to manage, interact with and interpret data. Coding facilitates data organization, retrieval and text interpretation by connecting data elements with shared properties. Researchers must systematically engage with the data, condensing extensive data sets into smaller, analyzable units, generating codes, categories and concepts derived from the data. Although the computer-assisted coding

process can be similar for quantitative and qualitative data, their coding objectives differ.

Qualitative coding allows data to be identified, reordered and visualized differently by assigning text codes. As well as aiding data analysis, coding also serves as a data management technique. Organizing data into codes and categories allows specific data to be located and retrieved quickly.

Researchers should take breaks during the coding process and occasionally step away from coding the data to avoid fatigue and maintain objectivity. They should be open-minded and avoid preconceived notions or prejudices, ensuring that the data, and not the researcher's assumptions or prejudices, drive the coding process. Consequently, CAQDAS users are looking for ways to automate repetitive and routine analytical procedures.

In recent decades, CAQDAS tools have been designed to support qualitative coding and analysis processes. The basis of CAQDAS is coding and text search. However, text retrieval procedures often need to be used more in the early stages of the computational analysis of qualitative data. Retrieval procedures can help researchers with diverse and unstructured textual datasets by supporting coding and the development of codebooks[3] .

Retrieval of text, keywords/key phrases in context, coded segments and pattern searching are usually not employed prior to data summarization or coding. Although search procedures are standard in large qualitative datasets such as linguistic corpuses, retrieval methods can also help with coding and codebook development during the early stages of data analysis using various unstructured textual databases. However, these retrieval techniques are less commonly used than data coding.

Two strategies can thus be distinguished which combine coding and retrieval procedures. Researchers can code segments manually, as well as search for similar uncoded segments and choose to code them manually or automatically; this method helps with pattern matching. On the other hand, researchers can explore the content to identify similar text segments and then automatically encode these segments. This strategy focuses on finding patterns before coding. This latter approach parallels the methodologies used in *Big Data* and knowledge discovery databases[11-12].

Manual or automated coding with text retrieval techniques such as search and pattern matching is a process in which the structure of the codebook is constantly evolving. Researchers investigate the semantic context and structure of qualitative data.

Coding and retrieval are undoubtedly the basis of computer-based qualitative data analysis, which produces satisfactory, data-driven results. Although this activity requires long hours in front of a computer, the emergence of new technologies and changing expectations among qualitative researchers are transforming computer-based qualitative data analysis methodologies.

CAQDAS tools have fundamentally transformed the qualitative research landscape, significantly simplifying coding and research procedures. These advances make data management easier, more efficient and improve the processing and analysis of complex data sets, and are a milestone in the way researchers engage with qualitative data.

Researchers use CAQDAS in conjunction with various methodological approaches, including Discourse Analysis[14] Media Analysis[15] and Conversation Analysis[16]. However, the core of qualitative research - data collection, analysis and interpretation - is being revolutionized by AI and big language models.

The new qualitative data handling and processing methods are based on AI, corpus linguistics, text mining and NLP procedures (tagging and fragmentation of grammatical classes with NLP; text or sentence extraction; named entity recognition - NER or linguistic annotation) which are implemented in CAQDAS.[16-23 apud5]. The integration of AI, text mining and NLP technologies with CAQDAS offers the potential for more sophisticated, efficient and accurate qualitative data analysis. As these technologies continue to advance, researchers can expect improvements in the automation and efficiency of coding and search procedures within CAQDAS and generally in QDA.

AI-driven tools and LLMs based on pre-trained general transformer models can automate the transcription of interviews, offer preliminary thematic or interpretative analysis, and help identify sentiment in large data sets. The methodological integration of AI with qualitative research primarily aims to increase the efficiency, accuracy and accessibility of automatic data coding and analysis.

Methods include the use of AI platforms and tools, such as AQUA, IBM Watson, QualiGPT and LLMs, to automate coding processes, reduce manual labor and improve analytical accuracy. The scoping literature analysis approach takes advantage of AI's capacity for thematic analysis, content analysis and deductive coding. It compares its performance with human coders and employs AI to support traditional research methodologies, offering significant advances in coding time reduction, economy and methodological innovation.

For example, AQUA was presented, an AI platform designed to streamline the coding of qualitative data using a graph-theoretic approach to topic extraction and clustering.[24-25]. These addresses reduce coding time and costs, while overcoming the challenges of maintaining accuracy across multiple linguistic expressions. Lee et al.[26] evaluated the effectiveness of IBM Watson in content analysis, highlighting AI's ability to reduce manual labor while significantly improving analytical accuracy by interacting directly with data analysis processes.

Zhang et al.[27] introduced QualiGPT, a tool that leverages LLMs for qualitative data analysis, increasing efficiency, transparency and user accessibility. This addresses the challenges of automatic coding, usability and economy, contributing directly to advances in coding and data analysis. A methodology that uses LLMs to aid deductive coding for qualitative research[28]. They demonstrated an experiment where LLMs analyze texts with a codebook and compare their performance with human coders. This shows the potential of LLMs to increase research accuracy and efficiency in data coding, i.e. the potential of ChatGPT to improve thematic analysis in qualitative research by demonstrating the development of cue structures to improve thematic analysis.[27].

This is a direct application of AI to facilitate and refine the data analysis process. De Pauli[29] explores the use of GPT-3.5-Turbo to emulate aspects of inductive Thematic Analysis in qualitative research, examining its support for data analysis in comparison to human analysis. It is a direct engagement with coding and data analysis using AI. Chew et al.[30] use LLMs such as GPT-3.5 to streamline the deductive coding process in qualitative research, showing effective achievement of accuracy comparable to human coders and offering strategies for integrating AI into content analysis methodologies.

Xiao et al.[31] demonstrated the application of LLMs for deductive coding in qualitative analysis, aiming to come to terms with expert coding without training in specific tasks, which addresses the challenges and opportunities of using AI for coding and data analysis, specifically ChatGPT for qualitative data analysis, re-analyzing datasets to assess the tool's effectiveness in identifying themes compared to manual coding[32] . Their article[32] evaluates the support of AI and the potential improvement of traditional research methods.

Ciechanowski et al.[33] highlighted how AI tools can be used to collect data, analyze sentiment and social networks without coding, making data science more accessible to qualitative researchers. Leeson et al.[34] provided proof of concept for the use of NLP to support traditional qualitative analysis in Public Health research, comparing NLP techniques with open coding in interview transcripts. This illustrates the direct use of NLP in coding and data analysis. Janasik et al.[35] explored the application of a self-organizing map (SAM) for text mining in qualitative research to improve the quality of inference in data analysis. This is a direct application of AI techniques to coding and data analysis. Cheligeer et al.[36] explored the use of AI in the analysis of qualitative data from health research, presenting a method that combines AI with manual analysis for greater efficiency and depth. Bryda and Sadowski[37] proposed two coding approaches for qualitative data analysis, incorporating AI

to develop automated theme extraction and advanced coding methods: Generative Semantic Coding based on ChatGPT-4 and Lexical Pattern Coding based on NLP and ChatGPT-4.

AI algorithms analyze text data and propose inductive coding structures based on codebooks and dictionaries for more accurate analysis. The focus is on ensuring the reliability of the analysis of coding models, addressing methodological and research challenges, and emphasizing the need for interdisciplinary collaboration and the development of qualitative research skills.

Furthermore, the implementation of the ChatGPT-4 model in *software* such as Atlas TI, MAXQDA or NVivo represents an innovative advance in contemporary methodologies in qualitative research. By integrating ChatGPT's advanced language processing features, these platforms can support researchers in summarizing, coding, analyzing and interpreting large amounts of textual data with greater precision and speed. This integration allows for a more nuanced and context-rich analysis of qualitative data, helping to deepen the *insights* derived from interviews, surveys and other textual sources.

Researchers can take advantage of ChatGPTs to automatically generate codes, identify themes and suggest new areas of investigation based on the patterns and relationships inherent in the data, enriching the qualitative research process. This fusion of qualitative analysis *software* with cutting-edge AI technology such as ChatGPT paves the way for more sophisticated, efficient and comprehensive research results.

On the other hand, the implications of the identified applications of non-coding and data analysis suggest a more in-depth examination of the role of AI in qualitative research, beyond mere data processing. Van Manen[38] focuses on the impact of AI technologies, such as ChatGPT, on qualitative health research methodologies and highlights a reflexive discourse on the integration of AI, questioning how it can refine or disrupt established research practices. Gibson and Beattie[39] question the evaluation of AI's role as a participant in research and highlight a critical examination of the authenticity and reliability of AI-generated data, challenging researchers to discern between human and machine contributions. Finally, Stegenga et al.[40] make ethical considerations that emphasize the need for a conscious approach to incorporating AI into qualitative research, advocating transparency and the ethical treatment of data. Collectively, these discussions point to a differentiated understanding of the potential and limitations of AI, through a balanced and ethical integration of technology into research methodologies.

Automating coding and analysis with AI substantially reduces the time needed to process and analyze large volumes of data, making research projects more efficient and allowing researchers to focus on interpretation and theory development instead of manual coding. By reducing reliance on manual labor for data analysis, AI automation can substantially reduce the costs associated with qualitative research, making it more accessible to researchers and institutions.

AI tools can provide consistent and replicable results across data sets, increasing the reliability of the analysis. However, the accuracy of AI in understanding and interpreting differentiated human language and complex topics depends on the sophistication of the algorithms and the quality of the training data. These consequences point to a transformative change in qualitative research methodologies, where AI can augment and, in some cases, partially replace traditional manual processes. However, this transition also requires careful consideration of the ethical implications, the potential loss of depth in understanding complex human experiences and the need for rigorous validation of AI-generated results against established research standards.

The influence of AI on CAQDAS visualizations

The integration of artificial intelligence and language models into CAQDAS tools has increased interest in visualizing qualitative data. This interest stems from the advanced capabilities of these technologies to process and analyze large sets of textual data. As a result, there is

a need for innovative visualization techniques to effectively interpret and communicate the complex patterns and *insights* gained through AI-based analysis.

The use of AI and language models has revolutionized qualitative data analysis methodologies and driven the evolution and use of data visualization strategies. Visualization has therefore become a crucial tool for researchers looking to deepen their understanding of data and improve the communication of their findings.

CAQDAS visualization has revolutionized the way researchers approach and interpret qualitative data, allowing them to better understand, interpret and communicate their findings, leading to more robust and impactful research results. Thus, researchers are increasingly using visualization techniques to better understand their data and communicate their findings more specifically and clearly[41–43].

Various visualization techniques, such as detailed diagrams and thematic maps, are crucial for effectively analyzing, exploring and interpreting complex data sets, allowing for better interpretation and disclosure of subtle concepts and patterns, transforming qualitative data into visual representations, which facilitates a more intuitive and comprehensive examination of the qualitative domain.

Data visualization is a rapidly evolving topic in contemporary computer qualitative analysis and can take many forms, but its main objective is to facilitate the analysis, exploration and interpretation of qualitative data. This goal underscores the importance of visualization in making complex data more accessible and understandable.

Visual aids offer many benefits, especially when it comes to understanding complex concepts or interpretations[44]. They transform abstract data into more concrete and interpretable forms, enhancing the analytical process. In order to communicate an idea effectively, it must first be fully understood, so using visuals to aid interpretation and thinking is more beneficial for communication.[45]. Visuals act as a bridge between raw data and human cognition, simplifying the transmission of complex ideas.

Exploratory visuals of qualitative data (as opposed to explanatory visuals) focus on the researcher's interaction with the data. This focus encourages a more engaged and insightful analysis, with a deeper understanding of the data. This interaction is a vital step that must precede a qualitative researcher's attempt to share the results of their analysis and interpretation of knowledge with others.

Interaction ensures that the *insights* shared are based on comprehensive data exploration. As already mentioned, numerous tools are available for data visualization in popular commercial CAQDAS, such as webQDA, Maxqda, AtlasTI, NVivo and QDAMiner. These tools offer various visualization options, catering to different analytical needs and preferences. They include multiple techniques, such as document portraits, code browsers, 2D/3D maps, word/tag clouds, word trees, diagrams, tables, graphs, networks and analysis of cluster and grouping techniques.[46].

The aforementioned techniques allow researchers to present their data in a visual, engaging and informative way. With the advance of new information technologies, the possibilities for visualizing qualitative data continue to grow. This growth is exemplified by the development of sophisticated *software* and algorithms that enhance visualization capabilities. This is particularly relevant for open source CAQDAS programs and those that enable publicly available *Internet* data visualization libraries, such as *Gephi, R and Python*.

Such developments democratize access to powerful visualization tools, expanding the potential for innovative data analysis and presentation. Visualization is a distinctive artistic technique for showing and explaining analytical findings. These emerging trends have revolutionized the way researchers handle and understand qualitative data, providing a greater understanding of intricate details. Visualization techniques help researchers identify patterns, trends and relationships that can be difficult to discern in textual or numerical formats alone.

According to a review of CAQDAS tools, these trends can be classified into four categories:

interactive visualization, (semantic) network analysis, geospatial mapping and mixed-methods visualization. I) Interactive visualization allows users to explore qualitative data by interactively manipulating visual elements such as nodes, *links* and *clusters* in real time, which enables faster and more intuitive *insights* and promotes immersive analysis. II) Network analysis makes it easier to examine relationships and connections within data. By representing data as networks of interconnected nodes, it allows for more efficient identification of patterns and associations, deepening the understanding of underlying themes and dynamics. III) Geospatial mapping in CAQDAS tools allows data to be visualized in geographical contexts. By exploring the spatial dimensions of qualitative data on maps, patterns, trends and relationships can be observed that would take some time to highlight through traditional text-based analysis. Changes in qualitative data can be displayed over time; timelines and sequence diagrams help track the development of themes and associations in the data, making it possible to perceive the dynamics and evolution of the research object. IV) With the mixed methods approach, CAQDAS tools have begun to incorporate visualizations that integrate qualitative and quantitative data. These hybrid visualizations make it possible to examine complex data, revealing relationships and patterns that might otherwise remain hidden.

The modern CAQDAS tools have significantly advanced data visualization capabilities, offering researchers high customization and adaptability[42]. *Online* (cloud, server) and *offline* tools allow researchers to create visualizations tailored to their specific needs and preferences. In addition, the integration of AI, NLP and text mining technologies into the CAQDAS tools has expanded the data visualization potential.[47].

Visualization techniques efficiently examine, analyze and convey complex qualitative data and by using advanced techniques, researchers can automate the identification of themes, patterns and relationships within their data, resulting in more sophisticated and differentiated visualizations.[48]. As technology advances, the potential for creative and interactive visualization techniques further enriches qualitative research.

The CAQDAS visualization allows researchers to thoroughly explore and analyze their data, promotes reflexivity and facilitates the effective communication of research results.[49]. Data storytelling through CAQDAS visualization promotes greater engagement with the data, allowing researchers new perspectives, conclusions and more informed interpretations. By presenting data in visually appealing formats, researchers can convey their ideas more easily, as clear and engaging visualizations make complex ideas more accessible, promoting a better understanding of research findings and their implications.

However, the use of AI in qualitative research also raises questions about the role of human interpretation, the potential for algorithmic bias and the need to validate and evaluate automated coding procedures. The integration of AI algorithms into qualitative research and data analysis requires attention to ethical and methodological issues to ensure that AI contributes to enhancing, rather than replacing, the fundamental principles of qualitative methodology. Researchers need to be aware of the potential ethical and methodological risks, such as algorithmic bias and the need for them to be more critical in their data interpretations.

Further studies could discuss the possible advances of CAQDAS tools, including refining data analysis, improving interoperability, enhancing visualizations and increasing accessibility through mobile and cross-platform compatibility. Emphasize the importance of personalization, collaboration in open source communities and the integration of AI for a more nuanced and multifaceted research environment.

Conclusion

As for the mention of qualitative research methodologies and the integration of AI, ML, NLP and LLMs, it is important to highlight the benefits and challenges that arise from this integration. The incorporation of artificial intelligence, machine learning and natural language processing

technologies into computer-assisted qualitative data analysis *software* tools has automated the identification of themes and relationships in the data, speeding up the research process.

However, it is also crucial to mention concerns around algorithmic bias, human oversight and the need for ethical considerations. A critical, human-centered approach to the use of AI in qualitative research is needed, as well as emphasizing the role of researchers in ensuring validity, reliability and ethical integrity. Collaboration between researchers and AI developers can lead to more transparent and context-aware tools for better quality and greater impact of qualitative studies.

Collaborations:

1 – project conception and planning: Grzegorz Bryda and António Pedro Costa.

2 – data analysis and interpretation: Grzegorz Bryda and António Pedro Costa.

3 – writing and/or critical review: Grzegorz Bryda and António Pedro Costa.

4 – approval of the final version: Grzegorz Bryda and António Pedro Costa.

Conflicts of interest

There are no conflicts of interest.

Financing

National funds financed the work of the second author through FCT - Fundação para a Ciência e a Tecnologia, IP, within the scope of the Stimulus to Scientific Employment - Institutional Competition - [CDL-CTTRI-248-SGRH/2022] and CIDTFF (projects UIDB/00194/2020 and UIDP/00194/2020).

Notes

[1]Coffey and Atkinson[50]point out that coding simplifies or complicates processes.

[2]Coding qualitative data differs from tagging in Information Science or annotation in Digital Humanities.

[3] Code-and-retrieve is the main function of most free and paid qualitative data analysis software.

[4]The survey is a popular procedure in content analysis or mixed-methods oriented CAQDAS (e.g. QDAMiner and WordStat, Maxqda, NVivo, webQDA).

[5]Using these methods, we must differentiate between traditional text indexing (coding and retrieval) and the text extraction procedure. Indexing and extracting information are two approaches to the same problem: combining data and 'noise' in the original text[51]. Indexing keeps the text as it was and adds information about where the raw data belonging to the various categories is located (indexing themes) or which category states or values are located in a specific position (indexing content). This defines all information that has not been indexed as noise. Extracting content means separating the relevant information from the text, subsuming it into categories and storing it separately for later processing. The noise remains with the text that is no longer analyzed.

[6]The researcher is responsible for the method of data analysis and the interpretation of the results. Therefore, the selection of an appropriate research methodology is crucial for the effective use of CAQDAS programs in data analysis. The inherent qualities of the data should guide and inform this choice of methodology[53-54].

[7]ChatGPT is an advanced language model developed by OpenAI based on the GPT (Generative Pre-trained Transformer) architecture. It is designed to generate human-like responses from text input, making it useful in a variety of applications such as chatbots, virtual assistants and text-based conversational agents. Deep learning algorithms power ChatGPT has been trained on vast sets of text data, allowing it to produce contextually relevant and coherent responses. The latest version is ChatGPT-4, released in March 2023.

References

1. Lee RM, Fielding N. Qualitative data analysis: representations of a technology: a commentary on Coffey, Holbrook and Atkinson. Sociol Res Online [Internet]. December 1996 [cited January 30, 2021];1(4):15-20. Available from: http://journals.sagepub.com/doi/10.5153/sro.1326

2. Fielding N. The diverse worlds and practices of qualitative software research. Forum Qual Sozialforsch / Forum Qual Soc Res [Internet]. 2012 Sep 26; Vol 13: No-2 (2012): Qualitative Computing: Diverse Worlds. Available at: http://www.qualitative-research.net/index.php/fqs/article/view/1845

3. Fielding N, editor. Using computers in qualitative research (reprinted with updated resources section). Wise; 1993

4. Fielding N. The role of computer-assisted qualitative data analysis: Impact on emerging methods in qualitative research. In: Handbook of emerging methods. 2008

5. Madeiras M, Macklin R, Lewis GK. Researcher reflexivity: exploring the impacts of using CAQDAS. Method Int J Soc Res. 2016;19(4):385-403

6. Miles MB, Huberman AM, Saldaña J. Qualitative data analysis: a methods reference book. 3rd edition. SAGE Publications, Inc; 2014

7. Miles MB, Huberman AM. Qualitative data analysis: an expanded reference book. 2nd edition. SAGE Publications Inc.; 1994

8. Saldaña J. The coding manual for qualitative researchers. 2nd edition. Los Angeles: SAGE; 2013

9. Strauss A, Corbin JM. Basics of qualitative research: procedures and techniques of grounded theory. Basics of qualitative research: procedures and techniques of grounded theory. Thousand Oaks, CA, USA: Sage Publications, Inc; 1990. 270 pgs

10. Fielding N, Fielding NG, Lee RM. Computational analysis and qualitative research. 1. public. London: SAGE; 1998. (New technologies for social research)

11. Kelle U, Prein G, Bird K. Computer-assisted qualitative data analysis: theory, methods and practice. London; Thousand Oaks, California: Sage Publications; 1995

12. Fan J, Li D. An overview of data mining and knowledge discovery. J Comput Sci Technol [Internet]. 1998 Jul [cited April 12, 2023];13(4):348-68. Available from: https://doi.org/10.1007/BF02946624

13. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. IA Mag. 1996;17(3)

14. MacMillan K. More than just coding? Evaluating CAQDAS in the discourse analysis of news texts. Forum Qual Sozialforsch [Internet]. 2005;6(3). Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-27544449032&partnerID=40&md5=6de5a6361870f5f2cbe3eb064857b585

15. Melgar-Estrada L, Koolen M. Audiovisual media annotation using qualitative data analysis software: A comparative analysis. Qual Rep. 2018;23(13):40-60

16. Paulus TM, Lester JN. ATLAS.ti for conversation studies and discourse analysis. Int J Soc Res Method. 2016;19(4)

17. Brosz M, Bryda G, Siuda P. Big Data i CAQDAS a procedury badawcze w polu socjologii jakościowej [Big Data, CAQDAS and Research Procedures in the Field of Qualitative Sociology]. Przegląd Socjol Jakościowej [Qualitative Sociol Rev [Internet]. 2017;13(2). Available at: https://ruj.uj.edu.pl/xmlui/handle/item/50481

18. Bryda G. From CAQDAS to text mining. A Domain Ontology as a Model for Representing Knowledge about Qualitative Research Practices. In: Costa AP, Reis LP, Moreira A, editors. Advances i. Springer International Publishing AG; 2020. p. 72-88

19. Bryda G, Tomanek K. Od Caqdas do Text Miningu. Nowe techniki w analizie danych jakościowych [From Caqdas to Text Mining. New techniques in qualitative data analysis]. In: Metody i techniki odkrywania wiedzy Narzędzia CAQDAS w procesie analizy danych jakościowych [Methods and techniques of knowledge discovery CAQDAS tools in the process of qualitative data analysis]. Wydawnictwo Uniwersytetu Łódzkiego; 2014

20. Bryda G. Digital Qualitative Sociology Uncovering the Linguistic Worldview of Local Community Life. Wydawnictwo Uniwersytetu Jagiellońskiego; 2024

21. Bryda G. CAQDAS a badania jakościowe w praktyce [CAQDAS and Qualitative Research in Practice]. Przegląd Socjol Jakościowej [Qualitative Sociol Rev [Internet]. 2014;10(2). Available at: http://www.qualitativesociologyreview.org/PL/Volume26/PSJ_10_2_Bryda.pdf

22. Wiedemann G. Openness to Big Data: Computer-Assisted Analysis of Textual Data in the Social Sciences. Forum Qual Sozialforsch / Forum Qual Soc Res. 2013;14(2)

23. Wiedemann G. Text Mining for Qualitative Data Analysis in the Social Sciences: A Study of Democratic Discourse in Germany. 1st edition. Springer Fachmedien Wiesbaden: Imprint: Springer VS; 2016

24. Yu C, Jannasch-Pennell A, DiGangi S. Compatibility between text mining and qualitative research from the perspectives of grounded theory, content analysis and reliability. Qualified Representative [Internet]. October 19, 2014 [accessed April 12, 2023]; Available from: https://nsuworks.nova.edu/tqr/vol16/iss3/6/

25. Lennon R, others. Developing and testing an automated qualitative assistant (AQUA) to support qualitative analysis. Cura Comunitaria Fam Med. 2021;9

26. Lee L, Dabirian A, McCarthy I, Kietzmann J. Making sense of text: content analysis enabled by artificial intelligence. Eur J Mark. 2020;54:615-44

27. Zhang H, Wu C, Xie J, Kim C, Carroll JM. QualiGPT: GPT as an easy-to-use tool for qualitative coding. ArXiv [Internet]. 2023;abs/2310.0. Available from: https://api.semanticscholar.org/CorpusID:263835317

28. Tai RH, others. Using large language models to assist in textual data analysis. bioRxiv. 2023; De Paoli S. Performing an inductive thematic analysis of semi-structured interviews with a broad language model: an exploration and provocation on the limits of the approach. Soc Sci Comput Rev [Internet]. December 7, 2023; Available from: http://journals.sagepub.com/doi/10.1177/08944393231220483

29. Chew RF, Bollenbacher J, Wenger M, Speer J, Kim A. LLM-assisted content analysis: using large language models to support deductive coding. 2023.

30. Xiao Z, Yuan X, Liao QV, Abdelghani R, Oudeyer PY. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In: 28th International Conference on Intelligent User Interfaces (IUI '23). Sydney NSW Australia: ACM; pp. 75-8.

31. Hitch D. Augmented qualitative analysis of artificial intelligence (AI): the way of the future? SSRN Electron J [Internet]. 2023; Available from: https://www.ssrn.com/abstract=4451740

32. Ciechanowski L, Jemielniak D, Gloor PA. TUTORIAL: AI research without coding: The art of fighting without fighting: Data science for qualitative researchers. J Bus Res [Internet]. September 2020;117:322-30. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0148296320303854

33. Leeson W, Resnick A, Alexander D, Rovers J. Natural language processing (NLP) in qualitative public health research: a proof-of-concept study. Int J Qual Methods. 2019;18.

34. Janasik N, Honkela T, Bruun H. Text mining in qualitative research Application of an unsupervised learning method. Organ resolution methods. 2009;12(3):436-60.

35. Cheligeer C, Yang L, Nandi T, Doktorchik C, Quan H, Zeng Y, et al. Natural language processing (NLP) aided qualitative method in health research. Gurupur VP, Wan TTH, Rudraraju RR, Kulkarni SA, editors. J Integr Des Process Sci [Internet]. 2023, October 17;27(1):41-58. Available at: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JID-220013

36. Bryda G, Sadowski D. From words to themes. Coding and analyzing qualitative data with AI technology. In: JR, CB, MN, JK, AP C, editors. Computer-supported qualitative research WCQR 2024 Advances in intelligent systems and computing. 2024

37. van Manen M. What does ChatGPT mean for qualitative health research? Qual Saúde Res [Internet]. November 28, 2023;33(13):1135-9. Disponível em: http://journals.sagepub.com/doi/10.1177/10497323231210816

38. Gibson AF, Beattie A. More or less than human? Assessing the role of AI as participant in online qualitative research. Qual Res Psychol [Internet]. February 5, 2024; 1-25. Disponível em: https://www.tandfonline.com/doi/full/10.1080/14780887.2024.2311427

39. Stegenga SM, Steltenpohl CN, Lustick H, Meyer MS, Renbarger R, Standiford Reyes L, et al. Qualitative research at the crossroads of open science and big data: ethical considerations. Soc Personal Psychol Compass [Internet]. 2024 January 27;18(1). Available at: https://compass.onlinelibrary.wiley.com/doi/10.1111/spc3.12912

40. Charmaz K. Building grounded theory. 2nd edition. Sage; 2014

41. Jackson K, Bazeley P. Qualitative data analysis with NVivo. 3rd edition. SÁBIO; 2019

42. Maxwell JA. Qualitative research design: an interactive approach. SAGE Publications; 2012

43. Silver C, Lewins A. Using software in qualitative research: a step-by-step guide. 2nd edition. SÁBIO; 2014

44. Lieberman MD, Cunningham WA. Concerns about type I and type II errors in fMRI research: rebalancing the scale. Soc Cogn Affect Neurosci [Internet]. 2009 December [cited 2023 May 3];4(4):423-8. Available from: https://doi.org/10.1093/scan/nsp052

45. Hennink M, Hutter I, Bailey A. Qualitative research methods. WISE; 2020

46. Feldman R, Sanger J. The text mining handbook: advanced approaches to analyzing unstructured data. Cambridge; New York: Cambridge University Press; 2007

47. Bird S, Klein E, Loper E. Natural language processing with {Python}. 1st ed. Beijing; Cambridge [Mass.]: O'Reilly; 2009

48. Saldaña J. Thinking qualitatively: methods of the mind. Thousand Oaks, California: SAGE; 2015

49. Coffey A, Atkinson P. Understanding qualitative data: complementary research strategies. Thousand Oaks: Sage Publications; 1996

50. Glaser J, Laudel G. Life with and without coding: two methods for early-stage data analysis in qualitative research aimed at causal explanations. Forum Qual Sozialforsch. 2013 March;14

51. Gibbs G. Analyzing qualitative data. Repr. Los Angeles: SAGE; 2012. (The Sage qualitative research kit / ed. by Uwe Flick)

52. Lofland J, Snow D, Anderson L, Lofland LH. Analyzing Social Settings: A Guide to Qualitative Observation and Analysis. 4th edition. Waveland Press; 2022

Received: May 1, 2024

Approved: July 1, 2024

Published: July 26, 2024