Journal of the Academy
Asociación de Universidades del Perú, Lima, Perú.
ISSN 2707-0301
Núm. 12. Enero-Junio 2025
http://journalacademy.net/

AGREEMENT BETWEEN HUMAN QUALITATIVE CODING PROCESSES AND AI-BASED AUTOMATED QUALITATIVE CODING

David ÁLVAREZ-MALDONADO [1]
Anna MILANO-MENESES [2]

Received	:	07/01/2025
Approved	:	15/06/2025
Published	:	21/06/2025

ABSTRACT: Artificial intelligence offers significant enhancements to traditional qualitative research methods, particularly in handling large volumes of data and improving result reliability. This study explores the potential of AI-driven automated qualitative analysis by comparing two parallel coding processes of unstructured data composed of open-ended textual responses: one automated using artificial intelligence and the other conducted traditionally through human cognition. An open-ended questionnaire was administered to a sample of 263 Disney fans to understand their perceptions of what the brand represents to them, through a free-response question. The automated coding process employed Python and a language model called Llama 3.2-1b-Instruct. The results showed that while the coding outcomes were highly similar across the dataset, there was only moderate agreement at the individual case level. It is concluded that artificial intelligence demonstrates strong potential in terms of analytical efficiency and scalability, but also reveals limitations by introducing inconsistencies and redundancies in coding, underscoring the need for oversight through human cognitive processes.

Keywords: Qualitative Research, Coding, Artificial Intelligence, Language Models.

INTRODUCTION

Qualitative analysis stands out for its ability to address complex phenomena and unstructured data (Jiménez-Partearroyo et al., 2024; Mees-Buss et al., 2022; Magnani and Gioia, 2023), generating a depth and detail in analysis of the empirical material that other methodological approaches do not achieve (Alvesson and Karreman, 2000; Gioia et al., 2013). However, qualitative methodology has faced valid criticisms and widely recognized historical limitations, such as the difficulty of working with large volumes of information and the

[1] Anthropologist. Universidad Tecnológica Metropolitana, david.alvarez@utem.cl and ORCID: https://orcid.org/0000-0002-6375-0461

[2] Commercial Engineer. Instituto de Capacitaciones J21, contacto@juego21.com and ORCID: https://orcid.org/0009-0006-1548-9330

subjective relativism of the results, which limit the reliability and generalizability of conclusions (Filieri et al., 2022; Marcolin et al., 2023; Schmitt, 2024). In this context, automated qualitative research processes, driven by artificial intelligence, emerge as an alternative to overcome both the limits to cope with large volumes of information efficiently and to improve the reliability of their results (Christou, 2024; Filieri et al., 2022; Marcolin et al., 2023; Schmitt, 2024).

To explore this situation, qualitative research with automated processes based on artificial intelligence is tested by comparing its results with those derived from a traditional qualitative study based on human cognition (Marcolin et al., 2023). Using a sample of 263 people who responded verbatim to an open-ended questionnaire asking them about what the Disney brand represents to them, two parallel coding processes are performed on the diversity of fan responses. This unstructured qualitative data was a test for both artificial intelligence and the human cognition process, to evidence similarities and differences in their achievements.

The scope of this research is exploratory, using Cohen's Kappa concordance index to verify how similar or different the results of both processes are. In this way, it contributes to the reflection on the role and potential of artificial intelligence in qualitative scientific research. First, a review of the specialized literature on the subject of qualitative research using automated processes driven by artificial intelligence is addressed, then the specifications of the applied methodology are detailed, followed by the results obtained from the comparison of the two coding processes, and finally it is concluded that both processes are complementary, being useful to maintain a hybrid approach that combines automation and traditional analysis practices, improving the scalability and reliability of qualitative analysis.

DEVELOPMENT

At present, qualitative analysis is transversally recognized as having advantages and strengths in scientific research (Christou, 2024; Filieri et al., 2022; Gioia, 2021; Jiménez-Partearroyo et al., 2024; Marcolin et al., 2023; Mees-Buss et al., 2022; Magnani and Gioia, 2023; Schmitt, 2024; Lexman et al, 2024), for example, generates information from

How to cire: Álvarez-Maldonado, D., and Milano-Meneses, A. (2025). Agreement Between Human Qualitative Coding Processes and AI-Based Automated Qualitative Coding. *Journal of the Academy,* 12, 147-166.
https://doi.org/10.47058/joa12.8

unstructured data of greater depth and richness compared to other methodological approaches, however, it also has important limitations or weaknesses, as is the case that it requires intensive use of energy and time by researchers, being inefficient research processes (Filieri et al., 2022), in addition to criticisms regarding the subjectivity and unreliable relativism of the results (Schmitt, 2024).

In this context, the indicated weaknesses of the qualitative approach can be addressed and overcome through the inclusion of automated processes based on artificial intelligence in the analysis practices, there being different research that demonstrates the potential for scalability and efficiency for qualitative analysis driven by automated processes (Christou, 2024; Marcolin et al., 2023; Schmitt, 2024) or the possibility of incorporating higher degrees of reliability and validity in the analysis processes, decreasing the relativism derived from the subjectivity of the researcher (Schmitt, 2024).

While there are these opportunities for improvements in qualitative analysis assisted by automated processes driven by artificial intelligence, it is important to note that these potential advances do not replace analyses based on human cognitive capabilities, arguing for the importance of maintaining analysis from a hybrid approach that blends the human cognitive operations of traditional analysis and automated operations driven by artificial intelligence (Filieri et al., 2022; Marcolin et al., 2023; Gao et al., 2023).

The causes of this need for a hybrid approach in incorporating automated processes in qualitative analysis lie in the potential interpretive limits of artificial intelligence-driven automated processes (Filieri et al., 2022), as compared to the interpretive capabilities of human cognitive operations. Furthermore, AI-driven automated analysis has demonstrated problems in recognizing emotions or time sequences in discursive analysis (Filieri et al., 2022; Marcolin et al., 2023). Along with this, automated analysis incorporated in qualitative research is estimated to generate excessive homogenization due to standardization, which limits the variability of analysis (Gao et al., 2023), with a greater potential for diversity in human analysis processes. In this framework, some argue that artificial intelligence does not generate analysis of sufficient quality compared to human analysis, showing that artificial intelligence cannot replicate the capabilities of human expertise in qualitative research (Gibson and Beattie, 2024).

Regardless of the different limits identified, and the criticisms of automated processes for the purpose of qualitative analysis, it is possible to sustain the relevance of the combination of qualitative analysis based on human cognitive processes and qualitative analysis based on automated processes using artificial intelligence, given that in both cases, they improve the results and overcome their own limits about the problems of efficiency and interpretative depth (Christou, 2024; Filieri et al., 2022; Marcolin et al., 2023; Schmitt, 2024). It is necessary to consider that the limits of traditional qualitative analysis are fundamentally its limited scaling possibilities when increasing the sample size and the criticism of the subjective relativism of analysis based on a human researcher (Marcolin et al., 2023; Schmitt, 2024). The hybrid approach combining human cognitive processes and automated processes based on artificial intelligence allows addressing these two gaps, increasing the potential for scalability and indicating possible biases of traditional analysis (Marcolin et al., 2023). Thus, along with overcoming the small, non-generalizable samples of qualitative studies, artificial intelligence can provide validation to the subjective and relative nature of qualitative research, overcoming human intuition (Schmitt, 2024), while human capability can monitor and improve the results of the automated process.

In this context, some argue that traditional qualitative research could become obsolete, being replaced by unstructured qualitative data engineering (Schmitt, 2024), however, some posit that artificial intelligence has limits that must be addressed by a hybrid approach (Christou, 2024; Filieri et al., 2022; Marcolin et al., 2023), given that the critical approach of traditional analysis based on human cognitive capabilities remains irreplaceable, allowing human analysis to be enhanced rather than replaced (Christou, 2023a; Christou, 2023b; Christou, 2024). The preliminary consensus regarding this debate is that there is a need to overcome skepticism and abstinence from the use of automated processes in artificial intelligence, being necessary to combine traditional qualitative research methods with new technologies driven by artificial intelligence (Schmitt, 2024).

In this regard, the noted skepticism has been rapidly reduced, which is evidenced by the fact that artificial intelligence has been widely and increasingly used in scientific research (Christou, 2023a; Christou, 2023b; Gebreegziabher et al., 2023; Jeldes-Delgado et al., 2024), for example, for predictive purposes (Kumbure et al., 2022) or for formative purposes (Palea et al., 2024; Sinha et al., 2024). In addition, the potential of collaborative work in coding processes has been highlighted, which can improve more effective and efficient coding concordance, in less time and with a higher percentage, according to Gao et al. (2023).

In these hybrid coding processes, mixing human cognitive operations and automated operations based on artificial intelligence, show that the results present similarities and differences between the human analysis and the one generated by artificial intelligence, happening that human coding recognizes some contents that artificial intelligence does not recognize, and vice versa, artificial intelligence recognizes contents that human cognitive operation does not recognize (Hamilton et al., 2023). In this context, the present research seeks to contribute to this topic, aiming to compare human and artificial intelligence analysis, with the purpose of identifying their similarities and differences, contributing to exploring this operational area in the analysis of qualitative data.

METHODOLOGY

This research aims to explore the potential of automated qualitative analysis driven by artificial intelligence, by comparing human qualitative analysis and artificial intelligence qualitative analysis (Marcolin et al., 2023), based on the identification of concordances in coding processes (Kull, 2020). For this purpose, an open questionnaire is applied to people who are part of the audience of Disney movies, identifying them in virtual communities of fans of this brand of movies, in which key informants defined as influencers in social networks were contacted, which were associated with Disney, for example, creators of content about Disney, tattoo artists of Disney iconography, opinion leaders of Disney fan communities, among other types of informants. To address this information gathering, a qualitative netnographic approach is used (Kozinets, 1998; Kozinets, 2002; Kozinets, 2006; Kozinets et al., 2018), which is a validated methodology of fieldwork in virtual environments

How to cire: Álvarez-Maldonado, D., and Milano-Meneses, A. (2025). Agreement Between Human Qualitative Coding Processes and AI-Based Automated Qualitative Coding. *Journal of the Academy,* 12, 147-166.
https://doi.org/10.47058/joa12.8

for qualitative data access. These people contacted took on the role of key informants in the research process, and were asked to collaborate by publishing an online questionnaire in Google Forms on their social networks, to have their followers and close friends, part of the virtual community of Disney fans, answer an open questionnaire with the question: What does the Disney brand represent for you? In this context, the participants of the questionnaire were able to express themselves freely in writing, developing their ideas about what the Disney brand represents for them, thus collecting unstructured qualitative data for this research, which requires coding processes.

In this way, the responses constituted unstructured qualitative data on the perception of Disney fans regarding what the brand of movies and series represents for them, from their subjective criteria. In this sense, it is important to highlight that the qualitative data generated were collected in an open and unstructured way, with a single answer for each participant, which could contain a variable extension of content, according to the participant's criteria, without generating a reduced word limit. The final sample consisted of 263 people, 89.4% of whom identified with the female gender, while 10.6% identified with the male gender. The gender difference in the volunteers was not the subject of analysis in this research, but it allows us to assume that the case study community is mostly female, or that the self-selected volunteers have a mostly female profile. The age ranges reported were as follows: 1.1% under 14 years of age; 6.1% between 15 and 17 years; 28.5% between 18 and 24 years; 51% between 25 and 34 years; 12.5% between 35 and 44 years; and 0.8% over 45 years. Thus, it is also possible to conclude that the highest percentage of people interested in answering this voluntary questionnaire are young people or young adults. Finally, informed consent was applied in the same instrument.

The human coding procedure was inductive and emergent, analyzing participants' written responses to assign them to a category representative of the set of meanings of the response. Similar codings were grouped into categories representative of common themes in the responses. The purpose of coding is to structure the information by patterns of common meanings that group different cases. The artificial intelligence-based automated coding procedure was also inductive and emergent, without considering any prior categorization or preliminary instruction on the codes applicable in the analysis. The programming language Python (Bird et al., 2009; Chollet, 2021) was used using Sypder software and the artificial intelligence model llama-3.2-1b-instruct using LM Studio. The prompt used is as follows: You are a text analytics assistant who will use your deep language understanding skills. Your task will be to review comments from Disney fans about what the Disney brand represents to themselves, and find the topics that are mentioned. A comment may contain one or more topics, so please, I need you to identify exclusively the most important one, identifying one general topic per comment.

Finally, the different codings generated are grouped according to themes or topics, based on common elements, structuring the qualitative data generated by the questionnaire by identifying patterns of common meanings. Finally, the Chi-Square test and Cohen's Kappa test are applied to verify the concordance between human and artificial intelligence coding, concluding in an interpretation of the similarities and differences in the results of both coding processes.

RESULTS

Traditional qualitative coding based on human cognitive processes was able to identify six general themes that grouped the different responses of the participants, which are as follows:

1) Nostalgia and childhood memories, associated with participants mentioning that Disney evokes nostalgic and happy memories of their childhood. This theme is understood under the logic that Disney products, their stories and characters, exert a stimulus that generates memories associated with past childhood experiences in which these products were relevant. Some examples of segments of responses from questionnaire participants are as follows: “Connection with the inner child” (Participant 1); ‘Disney means my childhood and very good memories, my closeness to cinema which is what I want to dedicate myself to now therefore my thoughts and feelings are of nostalgia since every time I see an old or new movie I feel again the child I was’ (Participant 5); “I love with my life Disney because it evokes a deep nostalgia and transports me to happy moments of my childhood” (Participant 49); ‘Disney is my childhood and that inner child that I will always carry’ (Participant 96); and,

How to cire: Álvarez-Maldonado, D., and Milano-Meneses, A. (2025). Agreement Between Human Qualitative Coding Processes and AI-Based Automated Qualitative Coding. *Journal of the Academy,* 12, 147-166.
https://doi.org/10.47058/joa12.8

'Disney for me represents going back to meet your inner child, it is a feeling of nostalgia when remembering the good times growing up next to the brand' (Participant 185).

2) Magic and fantasy, a situation in which Disney is seen as a fantastic universe and an escape from daily routine, where fantasy can develop. Some examples of this theme are the following sentences: "It represents my place of escape from reality. Nowadays, I still like to escape from adult reality and enjoy movies and stories" (Participant 40); "Magic and dreams can come true! I was able to fulfill my dream of traveling to Orlando, and it was the most magical and wonderful thing in the world! I still don't get over it, I think that for someone who is a fan it is very difficult to get over it haha, I understood how important it is to believe and dream, so never stop dreaming is forever engraved in my mind" (Participant 66); 'Disney for me represents magic and happiness, it is to feel and see that everything is possible, in every movie it has, you can find wonderful things' (Participant 124); "Disney represents a dream, it is like achieving the unattainable, happiness and excitement" (Participant 156); and, 'It represents the ability to escape reality for a while through its fantasy stories' (Participant 183).

3) Positive emotions, in which Disney is associated with the generation of joy and other positive feelings, such as tranquility, representing a safe and emotionally positive place. The following are examples of relevant segments of this theme: "Peace, it gives me a lot of tranquility. It is very safe place" (Participant 52); 'Feelings of joy, happiness, excitement, beautiful things' (Participant 109); 'Disney for me is happiness, whenever I think of something related to it I am a very happy person' (Participant 176); and, 'It reassures me to know that most of the stories have happy endings' (Participant 249).

4) Teaching and values, in which participants value the life lessons and values conveyed by Disney products, i.e., their stories and characters. In this sense, Disney products generate an impact on the beliefs and behaviors of consumers or fans. This theme can be represented in the following sentences: "It represents who I am today, their emotions, and the teachings of each movie, both Disney and Pixar reach me, and some even move me to the point of tears. They make me reflect on my day to day life or how to cope with some problems in my life" (Participant 57); 'Disney represents values such as those in the Lion King and Mulan'

How to cire: Álvarez-Maldonado, D., and Milano-Meneses, A. (2025). Agreement Between Human Qualitative Coding Processes and AI-Based Automated Qualitative Coding. *Journal of the Academy,* 12, 147-166.
https://doi.org/10.47058/joa12.8

(Participant 77); ‘It teaches you wonderful things, lessons that we have to learn’ (Participant 136); and, ‘It is a world of which I am not a big fan, but there are certain movies that left me some lessons in my life’ (Participant 255).

5) Connection with the family, in which Disney represents the parental union and the memories shared with loved ones from different generations, establishing itself as a mediator and an emotional bridge between the different listeners. This content can be evidenced in the following example sentences: “Enjoying as a family” (Participant 90); “It makes me think of the times when, as a family, my parents would go with us to watch VHS movies. Disney is something I grew up with. Their stories, especially” (Participant 105); and, ‘Family, quality time and laughter’ (Participant 201).

6) A personal lifestyle, where Disney represents more than just entertainment, being a lifestyle and a constituent element of identity. This can be seen in that some participants make reference to collecting products or structuring daily life inspired by what they perceive Disney to be, which means that the brand is strongly integrated into consumers. This theme can be observed in the following relevant sentences: “It represents everything, I love collecting Disney stuff” (Participant 14); ‘It is part of who I am’ (Participant 20); ”It represents a lifestyle, a way of living, my way of thinking. My references in various areas of life” (Participant 21); and ‘Disney is what represents me, in the sense that if you have to associate something with me, it's Disney’ (Participant 22).

Within this framework, the six different categories were coded using an indicative number from 1 to 6. Table 1 below shows the frequency and percentage of each coding in the human-based process:

Table 1

Traditional coding of Disney brand representations

Traditional Coding of Responses to the Question What does the Disney brand represent to you?	**Frequency**	**Percentage**
1) Nostalgia and childhood memories	128	48,7

2) Magic and fantasy	41	15,6
3) Positive emotions	59	22,4
4) Teaching and values	8	3,0
5) Family connection	10	3,8
6) Personal lifestyle	17	6,5
Total	263	100,0

The automated coding process based on artificial intelligence did not generate a homogenization of the codings or reduce variability, as noted in the literature (Gao et al..., 2023), on the contrary, it generated more coding categories than necessary, generating overlaps with lack of discrimination and redundancies, for example: established the topic of childhood memories, separately from the topic of childhood nostalgia, which can be common themes, although one has feelings of nostalgia, and another refers to the cognitive process of remembering, given that nostalgia necessarily involves the cognitive process of remembering, that is, memories are a set that involves the set of nostalgia, generating the need to code them with the same category due to the overlap of meanings; It also generated unnecessary differences in relation to mentioning topics about family memories, in a differentiated way with connection to the family, among other codes with unnecessary variability and redundancies, in which the generalizations of common patterns of meanings were not adequately established.

A possible explanation for these overlaps with lack of discrimination can be found in the literalness of the automated coding process, given that parameters of the model were configured to minimize its creativity, reducing the level of randomness in the responses generated, establishing an artificial intelligence temperature parameter with low values close to zero, making the model more deterministic, choosing the most probable words according to the training. The temperature is a configuration parameter of the artificial intelligence and the language model used. The temperature was set to 0.001, which proved to be quite literal and not very creative. It should have been a more balanced number, for example, 0.5 or similar. This option would have allowed more creative and less literal responses to be generated. Probably the coding work requires creative rather than technically deterministic processes, which would be very attached to the literalness of the word, preventing generalization through interpretation of meanings and patterns. This issue explains the advantage of human cognition processes in these tasks versus automated processes before artificial intelligence, since we are able to work with different temperature levels, in a flexible way, without requiring configuration.

Along with this, coding errors were also found, generating isolated codes with only one response. For example, participant 14, who indicated "It represents everything, I love to collect Disney stuff", was coded under the category of collector, which, being a coding with only one case, loses the technical purpose of coding of grouping different elements and common answers into a broader meaning. In the case of human coding, case 14 was interpreted under the lifestyle and personal culture code. This type of error was infrequent, reaching only 9 cases, that is, 3.4% of the sample, which were recoded with the number zero, to differentiate them from the other codings.

Finally, automated coding also generated approximately six coding categories, but required critical human cognitive processing assistance that regrouped redundant categories, for example, linking nostalgia for childhood with childhood memories. Without human assistance, the high variability with little discrimination would not be useful for qualitative analysis, ultimately establishing that the output generated requires supervision. Probably if the six categories had been indicated to him beforehand, as an input for him to classify the responses deductively, instead of inductively emerging these categories, this problem would have been solved. In this sense, it may be necessary to apply a preliminary human inductive coding process to train automated deductive processing, rather than allowing it to perform a human-like inductive and emergent process. To complement both traditional coding processes with human coding, one could first generate an inductive and emergent human coding stage, and then scale up to a larger sample size by deductively automated coding trained with the first coding.

It is important to highlight that, despite the coding errors generated by the artificial intelligence, the results of the automated coding process were very similar to the human

How to cire: Álvarez-Maldonado, D., and Milano-Meneses, A. (2025). Agreement Between Human Qualitative Coding Processes and AI-Based Automated Qualitative Coding. *Journal of the Academy,* 12, 147-166.
https://doi.org/10.47058/joa12.8

coding process, except for the coding errors of 3.4% and the fact that it was not always coded under the same content as the human process, generating similar categories, but different applications of these at the time of coding. This is because the participants' phrases and interpretations of these can be ambiguous, for example, having nostalgia for childhood can also be interpreted as a moment of connection with the family, if it is considered that childhood happens in a family context generally, there being an ambiguity in phrases regarding whether they are considered one or the other category at the time of coding. Probably a relevant number of responses fit more than one code. Thus, although there is a similarity in the generality of the data at the time of coding, the case-by-case application of these codes may differ due to overlapping patterns of common meanings.

Table 2 below shows the frequencies and percentages of the artificial intelligence-driven automated coding, which reported the same codes. Errors in the coding process were coded with the number zero.

Table 2

Automated coding of Disney brand representations

Automated Coding of Responses to the Question What does the Disney brand represent to you?	Frecuencia	Porcentaje
0. Non-groupable coding errors	9	3,4
1. Nostalgia and childhood memories	112	42,6
2. Magic and fantasy	33	12,5
3. Positive emotions	85	32,3
4. Teaching and values	7	2,7
5. Family connection	10	3,8
6. Personal lifestyle	7	2,7
Total	263	100,0

In this context, it is possible to conclude that there are similar codings between the human process and the automated process, with six codes found in both processes, but requiring supervision and cleaning in the automated process output. This similarity can be verified by Pearson's Chi-Square analysis, which evaluates whether there is a statistically significant association between the data generated by the human process and the data generated by the automated process. The Chi-Square value was 324.875 with 30 degrees of freedom, and a p-value of 0.000. This indicates that the association is highly significant, suggesting that both coding processes are associated. A comparison of coding frequencies and percentages is shown below in Table 3, and the Chi-Square results in Table 4.

Table 3

Comparison of traditional and automated coding

Codes	Frequency Artificial Intelligence	Percentage Artificial Intelligence	Frequency Human	Percentage Human
0	9	3,4	0	0
1	112	42,6	128	48,7
2	33	12,5	41	15,6
3	85	32,3	59	22,4
4	7	2,7	8	3,0
5	10	3,8	10	3,8
6	7	2,7	17	6,5
Total	263	100,0	263	100,0

Note. Codes were represented with a number from 1 to 6 as a categorical variable, where: 1 represents nostalgia and childhood memories; 2 represents magic and fantasy; 3 represents positive emotions; 4 represents teaching and values; 5 represents connection with family; 6 represents a personal lifestyle; and 0 represents non-groupable errors in the coding process.

Table 4

Pearson's Chi-Square test for traditional and automated coding

Chi-square test			
	Value	**df**	**Asymptotic significance (bilateral)**
Pearson's Chi-square	324,875[a]	30	,000
Likelihood ratio	242,561	30	,000
Linear by linear association	62,749	1	,000

How to cire: Álvarez-Maldonado, D., and Milano-Meneses, A. (2025). Agreement Between Human Qualitative Coding Processes and AI-Based Automated Qualitative Coding. *Journal of the Academy,* 12, 147-166.
https://doi.org/10.47058/joa12.8

N of valid cases	263
a. 31 boxes (73.8%) have an expected count of less than 5. The minimum expected count is, 21.	

Regarding the agreement between human coding and automated coding, Cohen's Kappa index is used to measure the agreement between both evaluations, beyond that expected by chance. The Kappa value reached is 0.514 with a p-value of 0.000, indicating a moderate but significant level of agreement. When two humans perform qualitative coding analysis in parallel, differences should be resolved to a Kappa value of 0.8, so this moderate agreement indicates a further need for supervision to resolve coding differences. Table 5 below shows the results of Cohen's Kappa index:

Table 5

Cohen's Kappa index agreement measure for the traditional and automated process.

Symmetrical measurements					
		Value	**Asymptotic standard error[a]**	**Approximate T[b]**	**Approximate significance**
Measure of agreement	Kappa	,514	,038	14,497	,000
N of valid cases		263			
a. The null hypothesis is not assumed.					
b. Use of the asymptotic standard error that assumes the null hypothesis.					

DISCUSSION OF RESULTS

This research conducted a comparison of traditional coding processes based on human cognitive processes and automated coding driven by artificial intelligence (Marcolin et al., 2023), using as a case study the perceptions of Disney fans regarding what the brand represents to them. In this context, both coding processes identified very similar themes in general; however, a moderate level of agreement was evidenced when using these themes to code the responses of the study participants, at the level of the particular from case to case, achieving a significant Kappa index of 0.514. This would represent moderate, but significant, agreement in the interpretation of the data. The differences between the two particular codings can be explained by meaning overlaps, where one response may refer to two or more thematic codes.

This situation of high general similarity allows us to argue that automated qualitative analysis processes have a relevant potential to complement traditional qualitative analysis processes, because of similar capabilities to recognize themes in the responses to open-ended questionnaires (Christou, 2024; Marcolin et al., 2023; Schmitt, 2024). The contribution of the automated process is that it significantly reduces the work time required for large volumes of data, addressing thousands or millions of responses in much less time and with fewer human resources, achieving a similar result to traditional coding processes in terms of thematic identification. However, the automated analysis generated codings with less discrimination, resulting in redundancies, unable to group meanings with subtle differences and similarities, due to the literalness of the language model interpretation and the ambiguity of the qualitative data. This suggests the importance of the automated analysis process being human-supervised. In addition, it is necessary to consider that the language model used was quantized, i.e., reduced for more efficient performance, and that models with better performance might exist, but require greater computing resources.

Thus, the hybrid approach combining human cognitive capabilities with automated tools is supported by the results, discarding the idea that automation would replace human coding activity (Schmitt, 2024). Through this perspective, the limitations of both methods separately can be reduced, improving the reliability and scalability of qualitative analysis, thanks to its increased efficiency. However, it is currently not possible to replace the critical and analytical capabilities of the traditional process based on human cognitive capabilities, as far as its supervisory function is concerned (Gibson and Beattie, 2024).

One improvement to the limits of the automated process is to train the language model with a human coding process beforehand so that, when analyzing large volumes of data, it has deductive orientations when working, rather than inductive, as was the case in this research. However, this would limit the emergence of non-intuitive patterns detected by artificial intelligence. For this reason, it is relevant to generate further explorations on this hybrid approach to establish a scientific standard that allows replicability, meeting the expectations of greater scalability and reliability, which would overcome historical barriers of qualitative analysis related to subjective bias and the difficulty of working with large volumes of data.

CONCLUSIONS

This study compared traditional qualitative coding, based on human cognitive processes, with automated coding using artificial intelligence, using Disney fans' perceptions of the brand as a case analysis. The results show a high overall similarity in the identification of themes, although with moderate agreement at the case-to-case level (Kappa = 0.514). This is evidence that language models can effectively recognize thematic patterns, although with lower discriminative ability in ambiguous or complex responses.

Automation offers clear advantages in terms of efficiency and scalability, especially for large volumes of data. However, observed limitations-such as literal interpretation and generation of redundancies-reinforce the need for human supervision to ensure the validity of the analysis. The findings support a hybrid approach, where artificial intelligence complements, but does not replace, human analytical capabilities.

Finally, it is proposed to move towards models previously trained with human coding to improve their performance in specific tasks, without losing the inductive capacity to discover emerging patterns. This highlights the importance of establishing scientific standards to ensure the replicability, reliability, and scalability of qualitative analysis in the era of artificial intelligence.

BIBLIOGRAPHIC REFERENCES

Alvesson, M., and Karreman, D. (2000). Varieties of discourse: On the study of organizations through discourse analysis. *Human relations*, *53*(9), 1125-1149. https://doi.org/10.1177/0018726700539002

Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* O'Reilly Media, Inc.

Chollet, F. (2021). *Deep learning with Python.* Simon and Schuster.

Christou, P. (2023a). A critical perspective over whether and how to acknowledge the use of artificial intelligence (AI) in qualitative studies. *The Qualitative Report*, *28*(7), 1981-1991. https://doi.org/10.46743/2160-3715/2023.6407

Christou, P. (2023b). How to use artificial intelligence (AI) as a resource, methodological and analysis tool in qualitative research? *Qualitative Report*, *28*(7), 1968-1980. https://doi.org/10.46743/2160-3715/2023.6406

Christou, P. A. (2024). Thematic Analysis through Artificial Intelligence (AI). *The Qualitative Report*, *29*(2), 560-576. https://doi.org/10.46743/2160-3715/2024.7046

Filieri, R., Lin, Z., Li, Y., Lu, X., and Yang, X. (2022). Customer emotions in service robot encounters: A hybrid machine-human intelligence approach. *Journal of Service Research*, *25*(4), 614-629. https://doi.org/10.1177/10946705221103937

Gao, J., Choo, K. T. W., Cao, J., Lee, R. K. W., and Perrault, S. (2023). CoAIcoder: Examining the effectiveness of AI-assisted human-to-human collaboration in qualitative analysis. *ACM Transactions on Computer-Human Interaction*, *31*(1), 1-38. https://doi.org/10.1145/3617362

Gebreegziabher, S. A., Zhang, Z., Tang, X., Meng, Y., Glassman, E. L., and Li, T. J. J. (2023). Patat: Human-ai collaborative qualitative coding with explainable interactive rule synthesis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-19). https://doi.org/10.1145/3544548.3581352

Gibson, A. F., and Beattie, A. (2024). More or less than human? Evaluating the role of AI-as-participant in online qualitative research. *Qualitative Research in Psychology*, *21*(2), 175-199. https://doi.org/10.1080/14780887.2024.2311427

Gioia, D. (2021). A systematic methodology for doing qualitative research. *The Journal of Applied Behavioral Science*, *57*(1), 20-29. https://doi.org/10.1177/0021886320982715

Gioia, D. A., Corley, K. G., and Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational research methods*, *16*(1), 15-31. https://doi.org/10.1177/1094428112452151

Hamilton, L., Elliott, D., Quick, A., Smith, S., and Choplin, V. (2023). Exploring the use of AI in qualitative analysis: A comparative study of guaranteed income data. *International Journal of Qualitative Methods*, *22*. https://doi.org/10.1177/16094069231201504

Jeldes-Delgado, F., Ferreira, T. A., Diaz, D., and Ortiz, R. (2024). Exploring gender stereotypes in financial reporting: An aspect-level sentiment analysis using big data and deep learning. *Heliyon*, *10*(20). https://doi.org/10.1016/j.heliyon.2024.e38915

Jiménez-Partearroyo, M., Medina-López, A., and Rana, S. (2024). Business intelligence and business analytics in tourism: insights through Gioia methodology. *International Entrepreneurship and Management Journal*, *20*, 2287-2321. https://doi.org/10.1007/s11365-024-00973-7

Kozinets, R. V. (1998). On netnography: Initial reflections on consumer research investigations of cyberculture. *Advances in consumer research*, *25*(1), 366-371.

Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of marketing research*, *39*(1), 61-72. https://doi.org/10.1509/jmkr.39.1.61.18935

Kozinets, R. V. (2006). Click to connect: Netnography and tribal advertising. *Journal of advertising research*, *46*(3), 279-288. https://doi.org/10.2501/S0021849906060338

Kozinets, R. V., Scaraboto, D., and Parmentier, M. A. (2018). Evolving netnography: How brand auto-netnography, a netnographic sensibility, and more-than-human netnography can transform your research. *Journal of Marketing Management*, *34*(3-4), 231-242. https://doi.org/10.1080/0267257X.2018.1446488

Kull, K. (2020). Codes: Necessary, but not sufficient for meaning-making. *Constructivist Foundations*, *15*(2), 137-139.

Kumbure, M. M., Lohrmann, C., Luukka, P., and Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*. https://doi.org/10.1016/j.eswa.2022.116659

Lexman, R. R., Krishnan, G., Baral, R., and Thomas, S. C. (2024). Navigating the digital classroom: a qualitative content analysis of MOOC discourses in Indian e-newspapers. *Journal of Information, Communication and Ethics in Society*, *22*(4), 494-516. https://doi.org/10.1108/JICES-03-2023-0027

Magnani, G., and Gioia, D. (2023). Using the Gioia Methodology in international business and entrepreneurship research. *International Business Review*, *32*(2). https://doi.org/10.1016/j.ibusrev.2022.102097

Marcolin, C. B., Diniz, E. H., Becker, J. L., and de Oliveira, H. P. G. (2023). Who knows it better? Reassessing human qualitative analysis with text mining. *Qualitative Research in Organizations and Management: An International Journal*, *18*(2), 181-198. https://doi.org/10.1108/QROM-07-2021-2173

Mees-Buss, J., Welch, C., and Piekkari, R. (2022). From templates to heuristics: How and why to move beyond the Gioia methodology. *Organizational Research Methods*, *25*(2), 405-429. https://doi.org/10.1177/1094428120967716

Palea, D., Vadhul, G., and Lee, D. T. (2024). Annota: Peer-based AI Hints Towards Learning Qualitative Coding at Scale. En *Proceedings of the 29th International Conference on Intelligent User Interfaces* (pp. 455-470). https://doi.org/10.1145/3640543.3645168

Schmitt, B. (2024). Transforming qualitative research in phygital settings: the role of generative AI. *Qualitative Market Research: An International Journal*, *27*(3), 523-526. https://doi.org/10.1108/QMR-08-2023-0107

Sinha, R., Solola, I., Nguyen, H., Swanson, H., and Lawrence, L. (2024). The Role of Generative AI in Qualitative Research: GPT-4's Contributions to a Grounded Theory Analysis. En *Proceedings of the Symposium on Learning, Design and Technology* (pp. 17-25). https://doi.org/10.1145/3663433.3663456