

Artificial Intelligence-Assisted Thematic Analysis: A Contribution to Architectural Research

Estelle Duval¹, Catherine Chouinard¹, Pablo Garcia de Paredes¹, Carole Després¹^[0000-0002-5624-8270]

¹ École d'architecture de l'Université Laval, Québec, Canada

ABSTRACT: Architectural research involves comprehensive analysis of diverse data, including user satisfaction data, interview excerpts, plans, and photographs. Analyzing such qualitative material is costly and time-consuming. While architectural programming firms often use qualitative data analysis software, many practitioners rely on informal methods. This article explores AI-assisted content analysis by comparing it with human-centered analysis in two case studies. Efficiency, accuracy, and replicability were evaluated as key criteria. Case Study 1 utilized open-ended responses from a survey of 356 school staff about school premises, informing design guideline development for public school renovations in Quebec, Canada. Case Study 2 analyzed 33 interviews with teleworking mothers involving architectural plans, dwelling photographs, and participant profiles aimed at enhancing home office environments. The human-centered analysis included manual transcript splitting and coding, alongside reviewing photographs and plans. Our findings show that AI-assisted analysis excels in efficiently disassembling and reassembling data, while having some shortcomings in compiling and interpreting data. It also demands considerable expertise in prompt formulation, which hinders accuracy and reliability. Human-centered analysis, although inefficient, requires data familiarity, knowledge of evidence, and research experience. All three are absent from the models' training. In conclusion, the potential of artificial intelligence to streamline qualitative analysis is discussed.

KEYWORDS: School renovation, POE, Thematic analysis, AI-Assisted Qualitative Analysis, ChatGPT-4.0®

INTRODUCTION

Qualitative content analysis in architecture, whether in academia or architectural practice, presents significant challenges in terms of resource allocation and costs despite the availability of specialized software packages. This form of analysis, applied to data of various kinds, from focus groups, unstructured interviews, or open-ended surveys, supplemented by various forms of imagery, is particularly valuable in architectural programming and post-occupancy evaluation. Such expertise is crucial for design projects tailored to meet user needs within specific geographical, cultural, and political contexts. While qualitative analysis software like NVivo exists, its adoption in architectural firms remains limited due to the required know-how and time constraints. Can artificial intelligence (AI) alleviate these challenges? Are AI-driven analyses efficient, accurate, and reliable? How do AI-generated interpretations compare with those derived from human judgment? This paper explores these questions through a comparative study of traditional human-driven content analysis and AI-assisted methods, evaluating ChatGPT's efficacy in processing text and images and aligning its categorizations with human-derived insights.

AI has been available for several years, and extensive literature is emerging to determine how AI tools like Large Languages Models¹ (LLM) and machine learning can be proficiently used in various disciplines (Aslam & Nisar 2023; Ayoub et al. 2023; Chubb 2023; Dergaa et al. 2023; Hitch 2024; Hyun Baek & Kim 2023; Lixandru 2024). Furthermore, the integration of LLM like ChatGPT-4.0® with image processing capabilities, e.g., DALL-E² in the case of ChatGPT-4.0®, is beginning to enhance qualitative analysis for professions heavily reliant on visual data, such as architecture. Epistemologically, qualitative analysis is prone to replicability issues since text coding is inherently influenced by cultural perspectives (Aiello & Sorde-Marti 2021; Pérez & Cannella 2013; Uri 2015). Similar concerns apply to ChatGPT's training, which is based on internet-available data, reflecting a Westernized worldview. Nonetheless, its development has advanced our understanding, mirroring insights into human cognitive processes (Waldrop 2024). For example, at the start of this project, we expected ChatGPT to operate like a black box, stable, and influenced only by software updates to the model. However, it turned out that AI-assisted analysis also has its own biases.

The limitations of AI usage raise questions about overlooking significant errors or failing to grasp important discoveries, as well as offering an allegedly replicable analysis. How efficient is its output? Is it truly replicable? How reliable and accurate is it for researchers, and what precautions are necessary? Premature reliance on AI may obscure essential analytical depths crucial for rigorous inquiry, prompting careful consideration both in architectural educational settings and professional practice.

This article stems from a workshop presented at the EAAE/ARCC Conference 2024 *Architecture into the unknown* in Aarhus, Denmark. Audience feedback highlighted key aspects of qualitative research with ChatGPT, aiding in the development of an evaluation framework based on its application in psychology, where accuracy and replicability emerge as primary concerns (Orrù et al. 2020). Among the comments from the workshop's audience that enriched our framework was the model's inherent versatility in assigned roles. For instance, it can prioritize elements in images or text or simulate the architect-client relationship in a classroom setting after being told, for

example, to "become a client for a residential project." These capabilities complicate assessing tool accuracy, as slight adjustments can disrupt the study or require restarting interactions, thereby diminishing efficiency. Therefore, we emphasize the complexity and potential pitfalls of nuances during prompt writing, which entails a steep learning curve and varies with the specific characteristics of each project. For this study, we concentrate on efficiency, accuracy, and replicability.

Efficiency is defined as requiring less energy and time than human analysis to perform the following tasks involved in qualitative thematic analysis (Braun & Clarke 2019): compilation, disassembly, reassembly, interpretation, and conclusion (Castleberry & Nolen 2018). Accuracy is assessed by the presence of errors or the confirmability of the analysis, which aims to establish that the researcher's interpretations and findings are clearly derived from the data, requiring the researcher to demonstrate how the conclusions and interpretations were arrived at as explained by (Nowell et al. 2017, 3), while replicability is defined as a measure of accuracy when repeating analyses on different types of data, influenced by inductive and deductive iterations within thematic analysis (Fereday & Muir-Cochrane 2006).

Case study 1: the first case study is drawn from Schola, a five-year action research program funded by the Quebec Ministry of Education (Canada), to develop a platform of expertise in school architecture supporting professionals involved in the renovation of public schools (Després et al. 2023). These buildings, built mostly between 1948 and 1973, have reached the end of the first architectural life cycle and require considerable investment. In the specific case of secondary schools, staff working in a sample of 75 school properties, randomly selected among the 391 buildings in the province of Quebec, were invited through their administration to take part in an online survey about their perceptions of each school's work environments: a total of 356 employees from 41 school properties completed the 140-item questionnaire. AI-assisted and human-driven content analyses were performed on the 25 open-ended questions present in the survey. In the human-centered one, responses were manually split into meaningful segments, assigned to one or more labels, and later grouped into thematic categories. In the AI analysis, the textual responses were submitted to the Chat Generative Pre-Trained Transformer (ChatGPT).

Case study 2: The second case study is part of a research project that examines the impact of housing on mental health during the COVID-19 pandemic. This qualitative study focuses on how housing conditions influenced the material adaptations of teleworking mothers during the pandemic. It investigates the interplay between housing conditions, daily routines, and telework setups during stay-at-home orders and progressive adaptations carried out by residents. Narrative accounts were gathered through 33 semi-structured Zoom interviews with teleworking mothers, supplemented by annotated architectural plans. AI-assisted discourse analysis using ChatGPT-4.0® was conducted on floor plans, dwelling pictures, participant profiles, and interview extracts for thorough architectural assessments and to uncover each group's material adaptations. The findings were compared to manual qualitative thematic analysis of assemblies using images and interviews. We employed codes obtained from live interview notes and a two-month reflection process.

The following results section presents the content of our human-driven analysis, followed by the ChatGPT-4.0® analysis for both case studies, and then moves on to a comparative discussion of the efficiency, accuracy, and replicability of these methods.

1.0 METHODOLOGY AND OUTPUT

1.1 Human-driven content analysis of cases 1 and 2

The human-driven content analysis was carried out in several stages, which were identical for both case studies, with the sole exception that in **Case Study 1**, with the aim of avoiding personal biases, inter-coder reliability checks were performed between two research assistants. In contrast, for **Case Study 2**, the analysis was performed in text files, not tables, and images were also sorted out and categorized by one single coder.

Firstly, the files were anonymized. Comments to open-ended questions were extracted from the original transcriptions and cleaned up for grammar and spelling. A careful reading of all responses to each question followed next for the coders (or coder in study 1) to acquire a good sense of the content. These comments were then transferred to an Excel spreadsheet in the case of the first study and to text files in the case of the second one and split up into meaningful segments. The coding labels were defined as the responses to each question were read, which requires to go back and forth to already analyzed segments when a new code needed to be added. These labels were then grouped according to thematic categories.

A closer look at respondents' open-ended comments in **Case Study 1** illustrates the coding process. For example, a question asked school staff about the class(es) they teach: "*To what extent is the furniture in your main workroom adapted to your needs?*". Their response options were "*very much*," "*fairly well*," and "*poorly adapted*." If respondents checked "poorly" or "fairly well", they were asked to explain their choice. Over one-third of all staff who identified a class as their main workspace provided an open-ended comment (83/233) for a total word count of 1246. The recurrent themes identified in 154 text segments are, by decreasing order of importance, lack of storage space (56), worn-out furniture (27), lack of comfort (18), unsuitable for use (14), inappropriate dimensions (13), lack of security (12), insufficient furniture (5), lack of space (5) and difficulty to move furniture around (4). The answer provided by a special needs teacher covers at once several of these categories, addressing the wear and tear of the furniture, its lack of ergonomic comfort, and the impossibility of securing his equipment:

My desk drawers do not open properly, and my filing cabinet does not lock. My chair is torn, and my

working position is not ergonomic. I have no keyboard for my laptop, and my screen is low. The bookcase provided for storing equipment lacks sturdiness.

(Teacher, special education, ID 114339694070)

For **Case Study 2**, all participant comments and various data from a quantitative database on the participants were grouped into text files, while images were organized into folders for each participant. The images consisted of two types: architectural plans of the dwellings of each of the 33 interview participants, for which we had extensive quantitative data and photographs of the participants' teleworking environments. Manually cross-referencing these three types of information proved to be very onerous due to the material's complexity, involving constant back and forth between text files and folders with images. As a result, we tried different approaches, first by working on text and images separately and then creating assemblies of both text and images to familiarize ourselves with the data.

The information was then manually coded using categories derived from these various approaches and along combinations of housing flexibility and participant adaptability: 1. Low Housing Flexibility - Low Participant Adaptability, 2. Low Housing Flexibility - High Participant Adaptability, 3. High Housing Flexibility - Low Participant Adaptability, 4. High Housing Flexibility - High Participant Adaptability.

1.2 Artificial intelligence-assisted content analysis for cases 1 and 2

For the AI-assisted content analysis, ChatGPT-3.5 © was first tested for both case studies. This free-to-use version requires that input text be of a maximum of 1024 words, that is, 4096 tokens. The language models used by ChatGPT process text in tokens rather than words, enabling them to consider the syntactic and semantic structure of a language. A token can be as short as a unique character or space or, else, a word. This maximum length includes the queries and the textual material to be analyzed as well. Since the 3.5 version does not accept tables, textual content had to be submitted in paragraph form, each line corresponding to one respondent. To overcome these limits, we turned to the paid version of ChatGPT-4.0©. Here, the maximum word count is 25,000 or 32,000 tokens, including prompts and answers. This version also allows for textual material to be input as tables that can be directly imported from a spreadsheet. In the case of the first study, the outputs can also be transposed into tables and exported back to a spreadsheet. Before proceeding with the analysis, participants' names and addresses were carefully removed from the answers to avoid uploading sensitive or identifying information to the ChatGPT database since it then became public domain.

Finding the correct prompt formulation was next, a task of paramount importance. It required several successive tests and adjustments before a sequence of prompts would generate the intended output. In addition to context, instructions, and content, the model requires what we preliminarily term category orientation. These inputs provide the model with a framework within the field and along the scientific literature, upon which deeper principles and research hypotheses are based. In the case of the first study, it was crucial to input sufficient context, incorporating five years of research experience in evaluating school environments in the province. Furthermore, prompts were systematically reviewed and refined using the model itself. In essence, the model can function both as a tool and as an expert assistant, helping with the sensitive task of formulating thorough instructions. In **Case Study 1**, for example, the initial query comprises three parts: first, a description of the study context; second, the specific survey question and answer; third, the information requested for the textual material to be analyzed, to which a fourth one was added when the 4.0 version was adopted, asking to format the results into a table (Figure 1). This last point was not included in **Case Study 2**, instead asking the model to evaluate images for each participant (including plans and dwelling pictures) as additional data.

2.0 RESULTS

In **Case Study 1**, once the query was launched along with the textual file, the Chatbot almost instantaneously generated the output, ranking themes from most to least frequently mentioned (Figure 2). In decreasing order of importance, they lack storage space (20), inadequate and old furniture (18), insufficient space (15), need for more modern equipment (12), ergonomics issues (8), and sharing and security concerns (5). Even though AI generated fewer categories, five instead of 8, they cover the same issues (Figure 2). The first two ones are about storage and furniture, the same as the human-coded ones. The two categories having to do with inappropriate dimensions and lack of space when codified manually were considered as one by the AI and labeled as "insufficient space."

Case Study 2 was interested in understanding the relationship between the materiality of the dwelling and the progressive adaptations of the participants, including cognitive, material-architectural, and time-use adaptations in the home. The thematic analysis was conducted using the model, but it was guided by previously established categories, some of which had been refined during the human-driven manual analysis. Specifically, we knew that participant's adaptations had been primarily of three types: material, cognitive and time use. We also had four typical profiles extracted through quantitative methods in a previous project, which roughly corresponded to the categories used during manual coding, combining housing flexibility and participant adaptability scores. The model was tasked with finding the themes describing participants' adaptations present in each text file and corroborating their content with the images provided in an iterative sequence. This triangulation between existing categories, information in text format with data on each participant, and images, i.e., floor plans and photographs of work areas within the dwellings, allowed for a complex thematic analysis yielding common and singular categories for adaptation strategies (Table 1).

Context of the query	These comments were collected via a survey conducted as part of a research project on the renovation of secondary schools. The aim is to analyze the qualitative responses to identify and quantify recurring themes.
Survey questions to be analyzed	The following question was asked to respondents: "Does the IT equipment you have at your disposal meet your pedagogical needs? Please justify your answer."
Content analysis prompts	For each category, the following information is required: 1. The title of the category, 2. The number of respondents mentioning it, 3. A summary in a few sentences describing the category, 4. An exemplary quotation supporting the summary, accompanied by the respondent's identification number and function. It is necessary to separate positive answers from negative ones."
Output format requested (in 4.0 version only)	Based on the results presented, can you create a table with four columns?

Figure 1: Prompts to ChatGPT for textual content analysis (translated in English but initially submitted in French).

<p>1. Category: Lack of Storage Space</p> <ul style="list-style-type: none"> Number of Respondents: 20 Summary: Many respondents expressed a need for more storage space, including shelves for books and materials, drawers, and secure cabinets. The lack of adequate storage hampers organization and limits the materials teachers can keep in the classroom. Exemplary Quotation: "Lack of space for a shelf for novels." - French Teacher, 114357361361 <p>2. Category: Inadequate and Old Furniture</p> <ul style="list-style-type: none"> Number of Respondents: 18 Summary: Teachers are dealing with outdated, sometimes damaged, and non-ergonomic furniture. This includes old desks, chairs, and cabinets that are not suitable for modern teaching needs. Exemplary Quotation: "The teacher's desk is a real antique: it lacks a drawer, the center drawer regularly jams, and you have to beware of splinters." - Social Sciences Teacher, 114343102421 <p>3. Category: Insufficient Space</p> <ul style="list-style-type: none"> Number of Respondents: 15 Summary: Respondents highlighted the issue of insufficient space in classrooms to accommodate additional or more suitable furniture. This limitation affects the ability to create a dynamic and flexible learning environment. Exemplary Quotation: "Lack of space." - French Teacher, 114354910122 <p>4. Category: Need for More Modern Equipment</p> <ul style="list-style-type: none"> Number of Respondents: 12 Summary: Several teachers pointed out the need for more modern and adaptable furniture to meet current educational needs and support different teaching methods. Exemplary Quotation: "Most date from 1974. Not adapted to today's orders and needs. Outdated and in poor condition." - French Teacher, 114351267561 <p>5. Category: Ergonomic Issues</p> <ul style="list-style-type: none"> Number of Respondents: 8 Summary: Ergonomic issues with existing furniture, such as non-adjustable desks and uncomfortable chairs, were mentioned. These issues can lead to discomfort and hinder effective teaching. Exemplary Quotation: "Desk non-adjustable, non-ergonomic at the height level." - Ethics and Religious Culture Teacher, 114349079520 <p>6. Category: Sharing and Security Concerns</p> <ul style="list-style-type: none"> Number of Respondents: 5 Summary: Teachers who share classrooms or storage spaces expressed concerns about the lack of personal storage and security for their materials. Exemplary Quotation: "Not much storage, especially when sharing." - English Teacher, 114354898283 <p>7. Positive Feedback</p> <ul style="list-style-type: none"> Number of Respondents: 2 Summary: A small number of teachers expressed satisfaction or acceptance of the current situation, indicating that they manage with the available resources. Exemplary Quotation: "It's okay." - Ethics and Religious Culture Teacher, 114348969072 	<table border="1"> <thead> <tr> <th>Category</th> <th>Summary</th> <th>Exemplary Quotation</th> <th>Number of Respondents</th> </tr> </thead> <tbody> <tr> <td>Lack of Storage Space</td> <td>Need for more storage space including shelves for books and materials, drawers, and secure cabinets.</td> <td>"Lack of space for a shelf for novels." - French Teacher, 114357361361"</td> <td>20</td> </tr> <tr> <td>Inadequate and Old Furniture</td> <td>Dealing with outdated, sometimes damaged, and non-ergonomic furniture including old desks, chairs, and cabinets.</td> <td>"The teacher's desk is a real antique: it lacks a drawer, the center drawer regularly jams, and you have to beware of splinters." - Social Sciences Teacher, 114343102421"</td> <td>18</td> </tr> <tr> <td>Insufficient Space</td> <td>Issue of insufficient space in classrooms to accommodate additional or more suitable furniture.</td> <td>"Lack of space." - French Teacher, 114354910122"</td> <td>15</td> </tr> <tr> <td>Need for More Modern Equipment</td> <td>Teachers pointed out the need for more modern and adaptable furniture to meet current educational needs.</td> <td>"Most date from 1974. Not adapted to today's orders and needs. Outdated and in poor condition." - French Teacher, 114351267561"</td> <td>12</td> </tr> <tr> <td>Ergonomic Issues</td> <td>Ergonomic issues with existing furniture, leading to discomfort and hindering effective teaching.</td> <td>"Desk non-adjustable, non-ergonomic at the height level." - Ethics and Religious Culture Teacher, 114349079520"</td> <td>8</td> </tr> <tr> <td>Sharing and Security Concerns</td> <td>Teachers sharing classrooms or storage spaces expressed concerns about personal storage and security.</td> <td>"Not much storage, especially when sharing." - English Teacher, 114354898283"</td> <td>5</td> </tr> <tr> <td>Positive Feedback</td> <td>A small number of teachers expressed satisfaction or acceptance of the current situation.</td> <td>"It's okay." - Ethics and Religious Culture Teacher, 114348969072"</td> <td>2</td> </tr> </tbody> </table>	Category	Summary	Exemplary Quotation	Number of Respondents	Lack of Storage Space	Need for more storage space including shelves for books and materials, drawers, and secure cabinets.	"Lack of space for a shelf for novels." - French Teacher, 114357361361"	20	Inadequate and Old Furniture	Dealing with outdated, sometimes damaged, and non-ergonomic furniture including old desks, chairs, and cabinets.	"The teacher's desk is a real antique: it lacks a drawer, the center drawer regularly jams, and you have to beware of splinters." - Social Sciences Teacher, 114343102421"	18	Insufficient Space	Issue of insufficient space in classrooms to accommodate additional or more suitable furniture.	"Lack of space." - French Teacher, 114354910122"	15	Need for More Modern Equipment	Teachers pointed out the need for more modern and adaptable furniture to meet current educational needs.	"Most date from 1974. Not adapted to today's orders and needs. Outdated and in poor condition." - French Teacher, 114351267561"	12	Ergonomic Issues	Ergonomic issues with existing furniture, leading to discomfort and hindering effective teaching.	"Desk non-adjustable, non-ergonomic at the height level." - Ethics and Religious Culture Teacher, 114349079520"	8	Sharing and Security Concerns	Teachers sharing classrooms or storage spaces expressed concerns about personal storage and security.	"Not much storage, especially when sharing." - English Teacher, 114354898283"	5	Positive Feedback	A small number of teachers expressed satisfaction or acceptance of the current situation.	"It's okay." - Ethics and Religious Culture Teacher, 114348969072"	2
Category	Summary	Exemplary Quotation	Number of Respondents																														
Lack of Storage Space	Need for more storage space including shelves for books and materials, drawers, and secure cabinets.	"Lack of space for a shelf for novels." - French Teacher, 114357361361"	20																														
Inadequate and Old Furniture	Dealing with outdated, sometimes damaged, and non-ergonomic furniture including old desks, chairs, and cabinets.	"The teacher's desk is a real antique: it lacks a drawer, the center drawer regularly jams, and you have to beware of splinters." - Social Sciences Teacher, 114343102421"	18																														
Insufficient Space	Issue of insufficient space in classrooms to accommodate additional or more suitable furniture.	"Lack of space." - French Teacher, 114354910122"	15																														
Need for More Modern Equipment	Teachers pointed out the need for more modern and adaptable furniture to meet current educational needs.	"Most date from 1974. Not adapted to today's orders and needs. Outdated and in poor condition." - French Teacher, 114351267561"	12																														
Ergonomic Issues	Ergonomic issues with existing furniture, leading to discomfort and hindering effective teaching.	"Desk non-adjustable, non-ergonomic at the height level." - Ethics and Religious Culture Teacher, 114349079520"	8																														
Sharing and Security Concerns	Teachers sharing classrooms or storage spaces expressed concerns about personal storage and security.	"Not much storage, especially when sharing." - English Teacher, 114354898283"	5																														
Positive Feedback	A small number of teachers expressed satisfaction or acceptance of the current situation.	"It's okay." - Ethics and Religious Culture Teacher, 114348969072"	2																														

Figure 2: Results of textual analysis of open-ended comments to a question about school furniture, with and without table query to ChatGPT-4.0@ . Source: (ChatGPT, 2024)

Table 1: Output for adaptations across the four profiles (n=33). Source: (Authors, 2024)

Common Adaptations	Profile 1: Low Housing Flexibility - Low Participant Adaptability	Profile 2: High Housing Flexibility - Low Participant Adaptability	Profile 3: High Housing Flexibility - High Participant Adaptability	Profile 4: Low Housing Flexibility - High Participant Adaptability
Material	Creating Dedicated Workspaces. Adapting various areas into makeshift offices or workspaces. Acquiring Furniture and Equipment. Purchasing desks, ergonomic chairs, and digital devices for teleworking and online education. Spatial Reorganization. Reconfiguring living spaces to accommodate work, study, and leisure activities.			
Time	Adjusting Work Schedules. Aligning work hours with family needs or children's routines. Establishing Routines. Creating structured daily routines for work, childcare, and education. Balancing Work and Childcare. Actively participating in children's education and managing work-life balance.			
Cognitive	Reconceptualizing Home. Shifting the perception of home from a purely private space to a multifunctional environment. Altered Views on Work and Technology. Adjusting attitudes towards the role of work and the use of technology in daily life. Awareness of Spatial Needs. Recognizing the importance of personal space and the impact of living conditions on mental health.			
Unique adaptations	Profile 1	Profile 2	Profile 3	Profile 4
Material	High reliance on technology for space management use of communal spaces.	Conversion of home areas into multifunctional spaces, with emphasis on technological integration.	Creation of specific work and study areas, investment in noise-canceling devices.	Adapting smaller spaces for work, seeking external spaces for relaxation.
Time	Flexible scheduling to allocate individual time and space within limited living areas.	Creative engagement of children in educational activities, the establishment of professional appearance during remote work.	Delegation between partners for childcare, scheduling personal time for relaxation.	Establishing strict work hours and personal boundaries.
Cognitive	Heightened concern for personal privacy and space optimization due to confined living conditions.	Valuation of outdoor spaces for recreation, reflection on the dual roles of work and parenting.	Increased focus on physical and mental health management.	Recognition of the limitations of apartment living and desire for more open environments.

In a second step, the model was asked to provide the themes describing housing flexibility and participant adaptability for each participant and each of the four profiles, assigning points and providing a brief explanation to justify why some participants were enduring adaptation difficulties, and others were not. This procedure allowed us to understand the rationale behind the adaptation strategies that the model reported in its list of themes by category, effectively giving a sense of the models' accuracy/confirmability (Nowell et al. 2017). We were also able to see what the model was looking for when performing its analysis and how and with what precision. For example, when justifying why a group of participants had endured few challenges, the model writes on three main themes: "1. Emotional Resilience: Despite the obstacles, the participants remained positive and adaptive. Their ability to reorganize and set boundaries reflects an emotional strength and fortitude. 2. Flexibility: These participants exhibited a great deal of flexibility, whether it was moving homes, reorganizing spaces, or setting up new routines. 3. Seeking External Solutions: Whether it was moving to a new place, going outside more, or seeking therapy, they looked for solutions outside their immediate environment".

3.0 COMPARING THE ADVANTAGES AND DRAWBACKS

The results of applying human-driven analysis and AI-assisted analysis methods enabled us to weigh the advantages and disadvantages of each one and to take a critical look. The aim was to test AI for 1) *efficiency* by estimating the time and energy required to complete each type of analysis, 2) *accuracy* by comparing the grouping and labeling of comments with the ones derived from manual coding, and 3) *replicability* or the stability of AI analyses by having the same queries run multiple times or by different people and content (Table 2).

TABLE 2: Advantages and Limitations of Human-Driven and IA-Assisted Content Analyses. Source: (Authors, 2024)

	Human-driven analysis	AI-assisted analysis
Efficiency	<ul style="list-style-type: none"> – Extensive grammar and linguistic corrections – Coding task repetitive – Time consuming – High costs-benefits ratio 	<ul style="list-style-type: none"> – Anonymization to remove sensitive data + Minimal grammar corrections + Fast execution + Simultaneous analysis of all data + Tabulation of analysis results + Low costs-benefit ratio
Accuracy	<ul style="list-style-type: none"> + Adequate background knowledge + Knowledge of the database + Overall understanding of coding material – Omission of important elements 	<ul style="list-style-type: none"> + Correct vocabulary + Effective categorization – Simulated quotation – Invented persona – Difficulty in understanding colloquial expressions

Efficiency: The content analysis performed in **Case Study 1** by Schola's two research assistants required up to four hours of work per question, depending on the number of responses to be processed. The analysis grid was built up progressively as the responses were read, requiring coders to frequently either return to former text segments to add categories or carry out verifications of the results. Completing the task for the 25 questions took the equivalent of three 32-hour work weeks per employee. In contrast, the ChatGPT did the same job in a matter of seconds. While the Chatbot is capable of handling multiple queries simultaneously, slowdown or interruption in the workflow was experienced when doing so: submitting one query at a time for best results is recommended. Overall, AI-assisted content analysis offers unrivaled speed and the ability to process large volumes of data. In large-scale surveys, it facilitates the efficient synthesis of textual material and presents the results in an organized manner, namely analytical tables, increasing the capacity to use the findings in shorter timescales to support decision-making.

The difficulty of processing different types of information at once became evident when manually coding the data in **Case Study 2**. The sheer volume of text to be analyzed led us to set a limit on the coding work. This workforce limitation permitted us to analyze only 20% of the content for each participant by the end of one week of work. Despite this complexity, the manual coding process allowed us to identify criteria for synthesizing information that might not have emerged as easily if we had used the model. During the familiarization process, human-centered analysis can act as a hidden process where the researcher or coder unexpectedly discovers themes and categories reminiscent of scientific literature through inductive and deductive sequences, capturing "the qualitative richness of the phenomenon" (Fereday & Muir-Cochrane, 2006). In contrast, the model currently lacks sufficient memory and specialized knowledge. Nevertheless, once these classification criteria were established, identifying recurring themes proved to be an enormous task due to the 350 pages of text involved. The rest of the work was carried out using the model, and the manual work served to verify the accuracy of the thematic analysis.

Compiling, disassembling, reassembling, interpreting, and concluding are considered the five main tasks that account for the efficiency of thematic analysis. When comparing human-centered analysis with AI-assisted analysis, we observed that the model is particularly strong in disassembling and reassembling large amounts of information, but it is not recommended to prioritize its use for compiling, interpreting, and concluding. The process of compiling requires a certain order and direction and occurs in contact with the information but also with the experience and knowledge of the researcher. The same can be said for interpreting and concluding. For these tasks the model is better considered as an assistant than as a stand-alone tool. Its use does not replace familiarization with the data and iterative induction/deduction processes.

Accuracy: AI has not only the capacity to group segments of textual content around recurring themes, but it can also label the categories with explicit descriptive terms. This precise vocabulary is often lacking to architects, even more so students, while interpreting qualitative material. Even when compared to the labels chosen by coders, both trained in architecture and involved in the research project for over a year, the labels attributed by the Chatbot were often more accurate, the model having access to public web sources worldwide, including dictionaries. These AI-assigned labels could be used for further queries to ChatGPT to search for other knowledge types, such as evidence or architectural precedents. Furthermore, as evidenced by case study two, once researchers have identified the strategic axes for the types of categories or classifications they are seeking, the model can efficiently and accurately create labels and identify themes. For both case studies, during this labeling process, the Chatbot grouped certain segments of responses into themes that had slipped under the researchers' radar, raising useful hypotheses for design. The platform sometimes generated irrelevant groupings or else, categories that later had to be combined. In this last case, it was possible to ask the software to perform mergers, even though this was sometimes completed manually. Interpreting culture-specific linguistic expressions also proved to be sometimes difficult for ChatGPT (analogies, metaphors, colloquial expressions). Indeed, although the model works well in French, some expressions had to do with current events or popular language and were misinterpreted; corrections were then made manually.

One of ChatGPT's main limitations is highlighted by its response to one of **Case Study 1**'s requests to provide an exemplary quotation supporting the summary of each thematic category, along with the respondent's identification number and employment type. Manual checks in the original text were carried out for all proposed quotations. It appears that some of those, and even some respondents, had been invented based on the overall content of the target category. The Chatbot was thus quoting a "persona," a sort of robot portrait of respondents in the category. The veracity of all quotes was also checked by asking ChatGPT: "Are all the quotes you provided in your previous answer true and taken from the data I provided in my question?" The results were not always conclusive, but the Chatbot occasionally replied, "I apologize for the previous confusion. The examples I have provided are not real but have been mentioned for illustrative purposes". Finally, on complex data from multiple sources in **Case Study 2**, the model's output was verifiable but had significant limitations when it comes to counting cases.

Replicability: An important quality criterion for quantitative research is replicability. In qualitative research, it is, however hard to achieve. Indeed, any coding process is not exempt from bias, even if reliability checks between coders contribute to minimizing the risks. To check if AI would be more consistent than human judgment on the replicability of results, we ran the same queries multiple times. If the general themes remained, labels changed to some extent. In all cases, the qualifiers used to label the categories remain relevant, but the frequency of comments referring to them varied considerably. It may be sufficient to ask the Chatbot to distinguish between a "very high," "medium," or "very low" number of text segments associated with each thematic category to quantify data from a non-random sample. It is also important to understand that qualitative research proceeds from multiple sources, including floor plan analysis, on-site observation, or measurements.

By triangulating sources and by adding scientific evidence provided by the literature review, the subjective involvement of researchers is paired with that of many others. In this regard, results obtained by AI are generally considered to be more reliable, as the analysis is carried out without the potential biases inherent in human analysis, although small changes in the prompts can alter the output in unforeseen ways.

CONCLUSION

Our research evidences many advantages of AI-assisted content analysis. It significantly outperforms human-driven methods in terms of speed and efficiency, processing large volumes of data in seconds and facilitating faster decision-making for the analysis of large-scale surveys, specifically when thematic analysis outputs are not required to include verbatim or frequencies. AI demonstrates high accuracy in grouping and labeling textual content, often providing more precise terms than human coders, although it has limitations with colloquial and feeling expressions, as evidenced by **Case Study 1**. AI shows greater consistency in replicating results compared to human judgment. However, there are also significant disadvantages. One of these is the challenge of knowing when to utilize these tools, as premature adoption can hinder familiarity with the content and potentially overlook crucial information, as was evidenced in **Case Study 2**.

Specialized knowledge of the literature and the data at hand, cultural awareness, and the researcher's life experience are indispensable in providing direction and guidance for the analysis in ChatGPT-4.0®. These elements complement the understanding of the content being analyzed. Together, these three components form a triad that can be augmented by AI but not replaced by it. Neglecting the advantages of human interpretation and culture would be a serious mistake, likely thwarting numerous potential discoveries in the social sciences. As AI tools continue to evolve, they promise to transform architectural research and design by enhancing efficiency, accuracy, and replicability and above all, let humans overcome memory limitations while complementing the strengths of human-centered analysis. Researchers need to understand the limitations and non-linear nature of these tools. Like in manual coding, expert knowledge is still crucial for uncovering key ideas, ensuring humans remain integral to the field's advancement.

Several insights from the workshop's participants focused on the original instructions given before starting the analysis in ChatGPT. For instance, participants noted the ability to instruct the model to take the perspective of a client evaluating a project or to adopt the role of a housing ministry critic before the prompt was inputted. These perspectives mirrored the nuanced directions we gave while creating prompts. They were perhaps the most intriguing aspect of using the tool due to the important variations in replicability from such changes, soliciting considerable sensibility when choosing references or providing examples and finesse in the wording of prompts.

ACKNOWLEDGEMENTS

We thank workshop participants for their useful comments and contributions. Also, we thank Schola, whose research is made possible through the financing of Quebec's Ministry of Education (Canada).

REFERENCES

- Aiello, E., & Sorde-Marti, T. 2021. "Capturing the Impact of Public Narrative: Methodological Challenges Encountered and Opportunities Opened." *International Journal of Qualitative Methods* 20: 160940692110501. <https://doi.org/10.1177/16094069211050160>.
- Aslam, M. S., & Nisar, S. 2023. *Artificial Intelligence Applications Using ChatGPT in Education: Case Studies and Practices*. IGI Global. <https://doi.org/10.4018/978-1-6684-9300-7>.
- Ayoub, N. F., Lee, Y.-J., Grimm, D., & Balakrishnan, K. 2023. "Comparison Between ChatGPT and Google Search as Sources of Postoperative Patient Instructions." *JAMA Otolaryngology–Head & Neck Surgery* 149 (6): 556. <https://doi.org/10.1001/jamaoto.2023.0704>.
- Braun, V., & Clarke, V. 2019. "Reflecting on Reflexive Thematic Analysis." *Qualitative Research in Sport, Exercise and Health* 11 (4): 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>.
- Castleberry, A., & Nolen, A. 2018. "Thematic Analysis of Qualitative Research Data: Is It as Easy as It Sounds?" *Currents in Pharmacy Teaching and Learning* 10 (6): 807–815. <https://doi.org/10.1016/j.cptl.2018.03.019>.

- Chubb, L. A. 2023. "Me and the Machines: Possibilities and Pitfalls of Using Artificial Intelligence for Qualitative Data Analysis." *International Journal of Qualitative Methods*, 22, 16094069231193593. <https://doi.org/10.1177/16094069231193593>.
- Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. 2023. "From Human Writing to Artificial Intelligence Generated Text: Examining the Prospects and Potential Threats of ChatGPT in academic writing." *Biology of Sport* 40 (2): 615–622. <https://doi.org/10.5114/biolsport.2023.125623>.
- Fereday, J., & Muir-Cochrane, E. 2006. "Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development." *International Journal of Qualitative Methods* 5 (1): 80–92. <https://doi.org/10.1177/160940690600500107>.
- Hitch, D. 2024. "Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future?" *Qualitative Health Research* 34 (7): 595–606. <https://doi.org/10.1177/10497323231217392>.
- Hyun Baek, T., & Kim, M. 2023. "Is ChatGPT Scary Good? How User Motivations Affect Creepiness and Trust in Generative Artificial Intelligence." *Telematics and Informatics* 83, 102030. <https://doi.org/10.1016/j.tele.2023.102030>.
- Lixandru, I.-D. 2024. "The Use of Artificial Intelligence for Qualitative Data Analysis: ChatGPT." *Informatica Economica* 28 (1/2024): 57–67. <https://doi.org/10.24818/issn14531305/28.1.2024.05>.
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. 2017. "Thematic Analysis: Striving to Meet the Trustworthiness Criteria." *International Journal of Qualitative Methods* 16 (1): 160940691773384. <https://doi.org/10.1177/1609406917733847>.
- Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. 2020. "Machine Learning in Psychometrics and Psychological Research." *Frontiers in Psychology* 10: 2970. <https://doi.org/10.3389/fpsyg.2019.02970>.
- Pérez, M. S., & Cannella, G. S. 2013. "Situational Analysis as an Avenue for Critical Qualitative Research: Mapping Post-Katrina New Orleans." *Qualitative Inquiry* 19 (7): 505–517. <https://doi.org/10.1177/1077800413489514>.
- Uri, T. 2015. "The Strengths and Limitations of Using Situational Analysis Grounded Theory as Research Methodology." *Journal of Ethnographic & Qualitative Research* 10: 135–151.
- Waldrop, M. M. 2024. "Can ChatGPT Help Researchers Understand How the Human Brain Handles Language?" *Proceedings of the National Academy of Sciences* 121 (25): e2410196121. <https://doi.org/10.1073/pnas.2410196121>.

ENDNOTES

¹ A large language model is an AI system designed to understand and generate human language using vast amounts of text data and neural networks.

² DALL-E within ChatGPT-4 is an integration of a visual language model. It was developed by OpenAI to generate images from textual descriptions, although it can also perform the opposite task, analyzing images and describing their content.