Comparative Analysis of AI models – Using AI-supported Qualitative Data Analysis for Interview Analysis

Gábor Sterczl, Ágnes Csiszárik-Kocsir

Óbuda University, Bécsi út 96/b, 1034 Budapest, Hungary;
gabor.sterczl@stud.uni-obuda.hu; kocsir.agnes@uni-obuda.hu

Abstract: This study aims to comparatively evaluate the performance of currently popular Artificial Intelligence (AI) models in supporting qualitative data analysis, specifically focusing on the coding and hypothesis validation of interview transcripts. We investigate how models from OpenAI, Google Gemini, and Anthropic perform in these tasks compared to traditional manual analysis and established CAQDAS tools. Utilizing transcripts from three exploratory interviews, the methodology involved applying each AI model and selected CAQDAS tools to generate codes and quantify references based on predefined research objectives and a set of established codes. Key findings reveal significant variability in the ability of different AI models to accurately identify and quantify relevant data segments, with some models demonstrating greater efficiency and the capacity to suggest novel, relevant categories not initially identified through manual analysis (e.g., external influences, roles, and responsibilities). Conversely, instances of inaccuracies, such as hallucinated quotes, were observed in other models. The study highlights that while AI offers substantial potential for increasing the efficiency and objectivity of qualitative analysis, its effectiveness is highly dependent on the specific model used and necessitates critical human oversight and validation. The implications underscore the importance of a hybrid human-AI approach in qualitative research, emphasizing careful model selection, robust data management protocols, and continuous attention to ethical considerations, particularly regarding data privacy and algorithmic bias.

Keywords: artificial intelligence; qualitative research; GPT 9, CAQDAS

1 Introduction

AI tools are having a significant impact on academic writing, research, and teaching. This creates new opportunities, such as increasing efficiency, but also raises serious ethical and quality challenges, such as scientific integrity, plagiarism, and professional responsibility [1]. For this reason, it is necessary to develop as many studies as possible that can contribute to supporting academic use and maintaining quality.

Based on our experiences to date, we hypothesize that AI tools can effectively support the evaluation of qualitative research. However, to ensure objectivity and minimize hallucinations, it is recommended to use the support of multiple models.

Qualitative research typically aims to gain a deeper understanding of human behaviour, thinking, and experience. Typical methods include interviews, focus groups, document analysis, and observations. They are characterised by the fact that they are usually based on a small sample and focus on subjective, individual experiences, whereby the context can be taken into account. One of their advantages is that they are flexible and allow complex phenomena to be examined in-depth while taking the context into account. Generalisability and subjective interpretation depending on the researcher's interpretation are the main concerns in applying this method. Qualitative data is information that cannot be quantified directly or has not yet been quantified. These data are usually textual, descriptive in nature, and can be collected using a variety of research methods [2]. The primary aim of qualitative research is to collect non-quantifiable information that can be used not only in scientific research but also in decision making at various levels [3]. In research, we are primarily interested in people's experiences, feelings, and subjective opinions. In general, statistical generalisation is not the aim, so we typically work with small sample sizes. The main questions in qualitative research are why, how, and what. Qualitative synthesis is a methodological approach that combines the results of several qualitative studies to generate new knowledge, theories, and applications. This approach can provide more comprehensive explanations and greater generalisability than individual studies [4] [5]. Qualitative data analysis, therefore, involves the interpretation of non-numerical data. Qualitative research traditionally relies on methods such as interviews, focus groups, and participant observation (shadowing) to gain insights into human experiences and social phenomena. Typically, they can collect rich and detailed data that quantitative methods might overlook [6], and synthesis aims to provide context-specific insights [5]. Among the types of qualitative synthesis, we distinguish between clustering models, which focus on summarizing the results of studies, and explanatory models, which seek to generate new theories or insights by interpreting the data [5].

Common methods, along the lines of Bearman and Dawson [6], include thematic analysis (identifying and analysing patterns or themes in qualitative data), meta-ethnography (translating concepts across studies to develop new interpretations), and realist synthesis (examining how and why complex interventions work in particular contexts). Also, worth mentioning is qualitative meta-synthesis (QMS), where the results of qualitative studies are integrated to create higher-level abstractions [7] [8].

It is also important to mention the challenges and considerations that arise in the course of the work:

- Methodological ambiguity: there are many different methods, no single standard approach [5] [9].

- Quality vs. quantity: the quality of the studies included is key [10].
- Context-dependence: qualitative results are often context-specific [11]
- Philosophical and methodological tensions: integrating qualitative synthesis into systematic reviews [12].
- Resource factor: time, money, and staff constraints (the individual skills and experience of the researcher) can influence the methods that can be implemented [13].

The various generative AI solutions have had a far-reaching impact on technical and social developments around the world, and their potential uses and implications are still subject to exploration, debate, and criticism. Most importantly, there is a serious lack of reliable criteria in the academic research community for researching the use of AI [14].

Currently, several significant limitations of current generative AI, include its tendency to produce inaccurate information or 'hallucinations,' its potential to operate with outdated knowledge, and the inherent biases it can inherit from its training data [15] [16] [17]. The lack of transparency in how these complex models arrive at decisions and the considerable data security risks associated with handling large volumes of sensitive information are also major concerns. Generative AI solutions indicate many ethical issues, such as the perpetuation of bias leading to unfair outcomes, the creation and spread of misinformation through realistic fake content, and the potential for privacy violations due to extensive data usage, are at the forefront [15] [16] [18].

Integrating MI into qualitative research can be approached from several angles. The integration of AI into every day work is redrawing the future labour market, which will put human-machine collaboration on a new footing [19]. One approach is to approach the issue from the perspective of optimising the human and time requirements of data collection. In this case, AI algorithms are used to transcribe the collected audio or audiovisual material into text, and then this textual content is analysed using traditional methods. This has already led to significant efficiency gains in recent years. Another approach is to outsource data collection to AI applications, for example, by conducting interviews. And a third approach is to analyse qualitative data using AI to identify themes, emotions, coding related to feelings, and patterns in large data sets, allowing researchers to gain deeper and more nuanced insights [20].

2 Literature Review

2.1. Computer-assisted Qualitative Analysis

Computer-assisted qualitative synthesis uses software tools to facilitate the synthesis of qualitative research findings [21]. The history of the methodology (CAQDAS - Computer Assisted Qualitative Data Analysis Software) dates back to the 1980s, when several software tools were developed independently to process qualitative data. The concept of computer-assisted qualitative data analysis was introduced to the scientific community by Nigel Fielding and Raymond Lee [22]. In 1994, the CAQDAS Networking Project was officially established and has played an important role in the development of the field. A major breakthrough came in the 2000s, with the emergence of ATLAS.ti, MAXQDA and NVivo, which are still widely used today [23] [24]. These applications can now analyse not only text but also images, audio and video data. These applications are essentially focused on structuring data into categories, but they also support comparative analysis of categories. The development of CAQDAS over the past decades has significantly transformed the practice of qualitative research, allowing for more efficient processing and analysis of larger data sets, and the creation of the REFI standard has simplified data exchange. However, the use of software remains a challenge for researchers and it is important to maintain the role of human interpretation in the analysis process.

The advantages of CAQDAS in qualitative synthesis:

1. Increased rigour and repeatability CAQDAS tools provide structured methods for data analysis. They also allow for the use of inter-coder reliability statistics (e.g. Cohen's kappa) and ensure consistency and reliability of coding [21] [25].
2. Efficient data management: support transcription analysis, assist text encoding and annotation, and simplify data mapping. In addition, they significantly simplify the handling of large amounts of qualitative data [26] [27] [28].
3. Support for multiple analysis techniques: they allow discourse and content analysis, suitable for grounded theory method. It can also be adapted to different research needs [26] [27] [28].
4. Advanced collaboration: facilitates teamwork Multiple researchers can work on the same project at the same time, making it particularly useful for large research teams [29] [30].

There are of course challenges associated with CAQDAS.

– Learning curve: learning to use supporting software is typically time-consuming and researchers need to adapt their analytical methods so that they can be easily processed later. Moreover, rapid changes in tools require continuous training [31].

- Software choice: Although the three major companies in the field (NVivo, Atlas.ti, MAXQDA, or open-source products such as RQDA) are the most prevalent, there are many options available, which differ substantially in their architecture, cost, and use. The choice should fit the methodological approach [31, 26, 28].
- Methodological debates: there is a broad and ongoing discourse on the role of technology in CAQDAS, for example, how to preserve the interpretive nature of qualitative research and the impact of the structured nature of CAQDAS on analysis [32].

2.2. Qualitative Analysis Supported by Artificial Intelligence Today

The purpose of using MI in qualitative research is, therefore, multifaceted. However, it can be argued that one of the main aims of using the technology is to make the growing amount of data currently available and used more manageable. AI tools can help us to do this by coping with large amounts of data by providing efficient ways to process and analyse them in ways that would be labour and time intensive to do manually [33]. By automating certain aspects of data processing, AI allows us to turn our primary focus to understanding the themes and patterns. Human-machine collaboration in qualitative analysis allows for iterative refinement and validation of thematic analysis, ensuring that the insights generated are robust and robust [34]. And by leveraging this so-called 'hybrid intelligence', the objectivity of research can be better guaranteed.

AI-based tools and methods are increasingly being incorporated into research across different disciplines, including qualitative research [14]. However, the use of AI in qualitative research also raises several ethical and methodological challenges [35][36]. In many areas, additional solutions are still needed, such as data archiving, data security, or bias factors (biases) in the data and historical information used to train models. Comprehensive solutions are needed to address problems related to data archiving, data security, and computational capacity [37].

There are many caveats and criticisms associated with the use of AI models in theoretical discussions and literature reviews. Main challenges include ensuring the transparency of MI methods or, for example, the comprehensibility and reliability of the decision making process.

Furthermore, because they are based on statistical patterns in large amounts of data, rather than expert knowledge or critical analysis, they do not always provide accurate or reliable information [38]. In addition, GPTs may not be able to detect errors or inconsistencies in the text, and may even generate false information [39] [40] [41] [42]. The proliferation of smaller, more localised language models is suggested to address these privacy concerns. Researchers should consider these issues carefully and ensure that the use of AI tools is used to ensure the integrity and reliability of research. Researchers should ensure robust digital data security and adhere to fundamental research ethics principles [43].

Based on previous research, AI solutions can help generate new theoretical concepts. For example, in some studies, AI has been used to generate textual data (i.e., generating output) for qualitative research in the social and behavioural sciences [44]. GPTs can mimic human intelligence in their responses, but can struggle with more subtle information or concepts that require human (i.e., cognitive and emotional) understanding. This limitation can be particularly problematic in conceptual research, which generally seeks to generate new theoretical frameworks or conceptual models [14]. AI as a tool for qualitative data analysis can significantly increase the effectiveness of qualitative data analysis. Large Language Models (LLMs) can conduct up to thousands of interviews per hour, significantly reducing the human resource requirements of qualitative research.

AI technologies, in particular natural language processing (NLP) tools, play an important role in supporting qualitative text analysis. These tools process and analyse data faster than traditional methods. For example, an initial modelling technique using LLM software can generate topic categories (codes) from qualitative survey responses that researchers can further refine to identify general themes, similar to those identified by human analysis. This approach not only speeds up the analysis process, but also increases the reliability of the results. Integrating AI into qualitative research involves another key methodology. In addition to coding, it is possible to organise codes into categories and compare these categories for thematic, conceptual, emotional, and theoretical analyses.

The integration of MI into qualitative research can also support the re-evaluation of traditional methodological assumptions/hypotheses. Collins and Stockton [45] highlight the importance of the concept of the theory of evidence, which refers to the idea that observations are influenced by the theoretical perspectives of researchers. This concept sends a message that qualitative researchers should be aware of their own biases and critically engage with data, even when using AI tools. In this way, researchers can ensure that their results are not only accurate but also reflect the complex social realities they seek to understand.

However, it is important to recognise that the use of AI in qualitative research is still a developing field, especially in academia. For this reason, it has many uncertainties and challenges to address. Feuston and Brubaker [33] emphasise that further experience and clarification work is needed on how AI can be integrated into qualitative research in a way that foregrounds and highlights qualitative ways of knowing the world. This will require the development of new methods and approaches, resulting in a more effective way of analysis.

Current trends at the intersection of AI and qualitative research point to a dynamic shift towards integrating AI as a collaborative tool rather than as the primary analyst [33]. Researchers are using AI to complement their inductive coding processes to uncover more accurate, efficient coding and relationships. This collaborative approach ensures that people have control over the results and the final interpretation [46] [47]. Despite all this support, however, researchers still need to know their data to ensure that their analysis is reliable. A deep knowledge of basic research data and results is a guarantee that they can identify biases and biases that may affect their results [48] [49] [50].

Traditionally, researchers manually transcribed the recordings, categorised the themes, and analysed the results using qualitative methods such as grounded theory or thematic analysis. While these methods provide rich, in-depth insights, they are often slow, prone to human error, and difficult to scale. The 3 services mentioned earlier already have automatic transcription capabilities for video and audio files, but encoding is still primarily done manually. The interaction skills of large language models allow them to conduct qualitative research interviews quickly, efficiently, and quantitatively [51]. The use of AI for qualitative data analysis opens up new avenues for researchers in terms of both data collection and the efficiency and accuracy of data analysis. This has been recognised by the three development companies specialising in qualitative analysis, which have recently integrated AI solutions into their products. AI solutions have already had a significant impact on the effectiveness of qualitative research by enabling a wide range of data transcription processes, as well as providing the opportunity for sentiment analysis and context analysis. In addition to increasing efficiency, AI tools also create opportunities to eliminate bias.

Artificial Intelligence (AI) has the potential to have a significant impact on qualitative research by increasing its insight, efficiency, and accuracy. By using AI tools, we can automate time-consuming tasks such as data collection, one of the effects of which is to make research more cost-effective, and also potentially create the opportunity to collect larger amounts of data, regardless of language barriers. The increase in efficiency is also reflected in the way the data collected is managed and processed. As it is no longer necessary to convert video and audio material into text [52] [53] In addition to efficiency, AI solutions can also support more accurate and objective analysis. In addition to identifying patterns that human researchers might ignore, they can also reduce the likelihood of human error by automating repetitive tasks [53].

A common concern with qualitative research is ensuring objectivity. If research and evaluation are conducted within an ethical framework, AI-enabled processing can support researchers in objective data analysis and decision making. This will make the results extracted more reliable and accurate, even for more complex research [53] [54]. Mitigating bias is key to fair and inclusive research [52].

However, it is also important to highlight the ethical framework mentioned above and the potential inaccuracies in the functioning of AI solutions. When designing data collection using AI tools, it is key to navigate the process carefully and ensure that ethical standards and rigorous methodology are adhered to [55]. By incorporating critical aspects into the data collection plan (clear objectives, careful data selection, ethical considerations, quality control, and mitigating bias), we can ensure that our AI research is ethical and thorough, producing reliable results [55] [56].

AI offers us a lot of potential benefits, but it does matter what happens to the information we upload, so it is important to keep in mind the data protection guidelines and choose the right AI tool in line with them. It is also important to manage the anonymisation that often occurs. And AI solutions can still be an effective solution, not for us as researchers, but in support of us, always requiring knowledge of the research data and review of the analyses and responses to ensure that errors due to model bias are filtered out and to ensure a real, quality outcome.

3 Methodology

3.1. Market Analysis Software vs. Generative AI Services

As mentioned earlier, Atlas.ti, MAXQDA, and NVivo have integrated AI into their products. Therefore, we have also considered these solutions to see if they are able to deliver more than if they were just using a publicly available generative AI solution. We had the opportunity to test two of the three products, as we could not find any freemium period offering for NVivo, it is not part of the analysis. In the case of Atlas.ti, its built-in generative AI solution is based on Open AI's solution. During the analysis of uploaded documents, it offers the possibility of AI coding. As a first step in using this feature, it is possible to define a research objective to improve the coding. The program then generates suggestions that can be accepted or rejected. The text is then manually marked in the "traditional" way to indicate which text fragment you want to assign to the code. In addition, a built-in AI-based chat service is available, as well as a sentiment analysis. Although the application handles documents uploaded in unsupported languages (e.g. in this case, Hungarian), it can only provide suggestions and answers in English.

In the case of MAXQDA, you have to select a text section in the uploaded text, and the application will give you suggestions for the selected text section, which you can accept or reject. There is also a chat function to ask questions about the document and the possibility to generate an AI-based summary. Although the application handles documents uploaded in unsupported languages (such as Hungarian in this case), it can only provide suggestions and answers in a limited language.

3.2. Research Context and Methodology

We recently conducted a study on agile maturity. One part of this study that we used as a basic database for the actual measurement. In this interview series, Team Leaders, Product Owners, and Scrum Masters participated in the research. As a starting point for the research, we hypothesized about the relationship between roles and role maturity. For this study, the three Scrum Master interviews were used. These interviews were 1,5 hours long, we asked more than 80% open questions. The analysis was made between 2024 December 1-12. The interview analysis was first conducted using the traditional manual method, followed by Atlas.ti, and finally using Open AI Chat GPT 4o, Google Gemini 2.0, Anthropic Claude Sonnet 3.5 and Haiku V3, and Llama3.1405B Instruct models, as well as manual evaluation. In all cases, we went through the same set of questions. Since the MAXQDA application gives suggestions from which the researcher has to choose, this was not included in the testing reported. The sample (3 subjects) consisted of Scrum Masters working on the development of a product. During the research, the participants were asked to share their experiences and thoughts about the team they support and its environment, and to limit themselves to the events of the last 6-12 months.

4 Results

We started the investigation by designing the basic code system. Coding is an integral part of the qualitative research process [57]. In the manual solution, we used an affinity map method to determine the codes. For Atlas.ti and MAXQDA, we used the built-in AI capabilities and copied the starting points and questions described above as our research objective. For the AI models, we used the following text for the bot configuration:

"*You are an experienced academic researcher specialising in the analysis of interview transcripts. You will help to produce well-structured, objective, high-quality research analyses and summaries. You will also organise findings, synthesise insights, and make recommendations. While you do not have access to confidential data, you will support the research process by increasing efficiency and providing guidance for both transcript analysis and general research tasks*."

And at the beginning of each interview, we uploaded the interview transcript and pasted in the same text as for Atlas.ti. For the GPT solutions, the same instructions were used for all models.

1. "*Based on the stated research objectives, examine the uploaded interview results and suggest codes for further qualitative analysis based on the interview transcripts.*"

2. "*For each code, examine all 3 interviews and list the total number of references to each code, supported by quotes. Note that a reference may belong to more than one code.*"

The results are summarised in the table below. The number of codes is shown in brackets.

Table 1

Comparison of the coding of qualitative results with AI and the number of references to the codes

Own result	**Chat GPT o1**	**Chat GPT o4**	**Google Notebook LM, Gemini 2.0**	**Llama 3.1 405B Instruct**	**Sonnet 3.5**	**Haiku V3**	**Atlas ti (chat GPT 4 based)**
Agile principles (23)	Agile values and goals (7) Agile practices and challenges (8)	The emergence of Agile core values (5)	Agile practices and processes (22)	Agile core values (12)	Agile operations (31)	Agile values in everyday work (8)	Agile values
Teamwork and communication (22)	Team dynamics and cooperation (6)	Communication and information sharing (5)	Teamwork and communication (17)	Team dynamics (9) Communication and cooperation (14)	Team dynamics (25)	Common goals within the agile team (5)	Common agile goals SM communication
Performance and metrics (6)	Measure team progress and improve efficiency (6	Measurement and efficiency (3)	Measurement and efficiency (6)	Improving measurement and efficiency (9)	Measurement and evaluation (12) Process development (18)	-	-
Product delivery, customer and user satisfaction (12)	User feedback and a user-centred approach (5)	User feedback (3)	Stakeholders and user focus (17)	User-centred approach (10)	User value creation (22)	-	User-centric support Product usability testing

							Product feedback source
Scaling and alignment (12)	Communication between SMs (8)	Support between SM and TM (6)	The role of Scrum Master in development (15)	Cooperation between teams (22)	Organisational cooperation (20)	Knowledge sharing and coordination between SMs (5) Cooperation between SM and TM (6)	TM support and cooperation Information exchange Redundant topics for scrum teams
Development and knowledge sharing (7)	SM development and support (6) Professional development and learning (7)	-	-		Professional development (15)	SM professional development (7)	Find professional support Find Agile support
-	Role boundaries and overlaps (9)	-	Roles and responsibilities (36)	Roles and responsibilities (15)	Role responsibilities (28)	-	-
-	External influences on the implementation of agile (6)	-	Challenges and obstacles (16)	-	-	-	-

Source: own research, 2025

When identifying the codes, we have tried to put similar categories next to each other in the table to make the results easier to read. Some models generated more or even fewer codes than we did, in each case we looked at the associated citations to see how they related to our codes and if they were related to the same topic, we listed them in a common cell. As can be seen from the table, each model brought several codes were similar to our own, such as in the case of agile values, or even customer and user satisfaction. In some cases, we observe that the only difference with the code we defined is that we have grouped a topic into one category, which the model has split into two (or more) codes, such as in the case of teamwork and communication. In another case, for example, for scaling and alignment, the models also named categories based on quotes similar to ours, but they just gave them all different code names.

In terms of further results, two categories emerged that we did not define, one for externalities and challenges, defined by o1 and NotebookLM, and one for roles and responsibilities, published by Sonnet and Llama in addition to the two former models. Reviewing the quotes assigned to the codes, the latter is due to the fact that the text mentions role names and responsibilities several times, yet the former does not contribute to the research purpose, the latter texts we have placed in the categories of teamwork and/or scaling, so in our opinion this coding has no added value for the processing of the research. On the other hand, I believe that the separate categorisation of external influences and challenges is a suggestion that is relevant to the research objective and adds to objectivity if it is presented as a separate category.

We then carried out another study. This involved asking the AI models to use the codes we had defined and testing their numerosity. The results are shown in the table below. The instruction was as follows in this case:

"Using the following codes, revisit all three interviews and count how many references can be linked to these keys: teamwork and communication, Agile principles, performance and metrics, product delivery, customer and user satisfaction, scaling and alignment, development and knowledge sharing. Don't forget one reference can belong to more than one code."

Table 2

Identification and quantification of qualitative data assigned to our defined codes for different AI models

	Own result	**Chat GPT o1**	**Chat GPT o4**	**Google Notebook LM, Gemini 2.0**	**Llama 3.1 405B Instruct**	**Sonnet 3.5**	**Haiku V3**
Teamwork and communication	22	15	10 (5)	13	23	23	9 (7)
Agile principles	23	13	8 (4)	9	17	18	10 (7)
Performance and metrics	6	5	6 (3)	5	12	12	5 (3)
Product delivery, customer and user satisfaction	12	9	5 (3)	10	15	20	7 (5)

Scaling and alignment	12	9	4 (2)	6	8	15	6 (5)
Development and knowledge sharing	7	15	7 (4)	6	13	16	7 (5)

Source: own research, 2025

We then looked at which text in each category was included in which model.

a. Llama results

We started with the Llama model, which quickly showed that most of the quotes it listed were not in the text. When we fed this back and asked it to recheck the results, it found only 3 for the teamwork and communication code, for example, and a similar proportion for the other categories. This could be considered a serious error.

b. Sonnet results

In the case of the Sonnet model: however, we were immediately provided with precise quotations, which were invariably found in the original text. All of these were immediately grouped into subcategories, which helped to make them clearer. First, we looked at teamwork and communication. Compared with the texts we had also classified here, we found minimal differences. For example, in this code, the Sonnet model for this question classified cases such as "function team coordination" that appeared in our analysis only in the scaling and alignment group, but did not take into account, for example, when the subject talks about the agile maturity of the team or when one of the interviews was about the involvement in retro. For agile principles, the parts that Sonnet flagged were also included in our analysis. However, it did not include themes related to customer engagement, hence the minor discrepancy. In the case of performance and metrics, the discrepancy was due to the fact that we did not include certain items, such as UI tests or information on a retrospective, so in this case, we found items that added to the analysis. However, it also included 2 pieces of information on process improvement, which are not included here, such as "I am trying to improve my processes". Sonnet found all the elements that we identified, plus additional findings that we considered completely valid, so here too the search was more accurate. In terms of scaling and alignment, Sonnet's extra hits included those that could be classified under the issue of collaboration with the PO only, for example, "Synchronisation of PO and team interests". Here again, there were more hits for AI than we identified. Overall, Sonnet 3.5 provided results very similar to our codes, complementing our assessment on several points.

c. Haiku and GPT 4o results

Haiku V3 identified significantly fewer items in each group than we did (or as did Sonnet 3.5, also Calude). The citations were accurate, and the hits were all the same as Sonnet found. What we did notice was that Haiku included some of the interview questions found in the text in the results. Of course, this is a correctable error, but in this case, the hit rate is even lower (a problem that also occurred with the GPT models). However, we still found two anomalies. Overall, it can be concluded that this Claude model proved to be significantly less effective than its sister product.

d. Chat GPT o1 and NotebookLM (Gemini 2.0)

Open AI's o1 model performed significantly better than o4, and did not include questions for references to code. The results were also automatically subdivided into subgroups by this model, which made it much easier to review. And the results were all ones that we identified. Numerosity was lower in most categories with some cases not classified there that we would have put there. However, more hits were recorded for scaling and alignment and development and knowledge sharing. We compared the o1 model and NotebookLM because the results were very similar, although not the same number of citations identified, a similar "thinking" was observed in the results.

The results show that it makes a difference which AI model we work with. Although models are constantly evolving and new models are emerging, it is important to check and validate how the chosen assistant performs before using it. From the results of the experiment, it has also become apparent that on a "one measurement, no measurement" basis, it is practical to include more than one model in the data analysis, thus further reducing the possibility of unwanted bias and taking a further step towards objective analysis.

Table 3
Research experience strengths, advantages and weaknesses for different AI models

Model	Strengths	Pro's	Contra's
GPT O1	Fast, prioritises responses	Very good creative responses	Project configuration not available
GPT O4	Versatile, open-ended nuanced responses	Almost excellent in creative responses	Slower responses
Gemini 2.0	Fast, structured, detailed answers	Structured responses	Project configuration not available
Llama 3.1 405B	High quality instruction tracking	Structured responses, better quantitative data management	Produce more hallucinations in a larger database
Sonnet 3.5	Structured, moderately creative output, good explanations, advanced understanding	Very good creative responses	High latency, sometimes overthinking simple tasks
Haiku V3.5	Quick simple answers, good for creative but concise outputs	Ideal for simple tasks	Limited depth and detail, not suitable for complex prompts

Source: own research, 2025

Nevertheless, the test demonstrated that AI solutions can contribute to better quality research results. As observed in the categorization, in addition to our categorization, several models were able to introduce aspects that positively contribute to the quality of the analysis. The inclusion of role boundaries and externalities as separate categories is important and meaningful to the study and supports the effectiveness of the presentation of the results.

4.1. Limitations

The report primarily concentrates on the period toward the end of 2024. While this provides a current snapshot of the situation, due to the rapid pace of technological development, subsequent developments may no longer be accurately reflected. This limitation is particularly important in the context of qualitative data analysis using AI, as AI models and tools are rapidly evolving. Therefore, the findings of the study, while relevant at the time of writing, may need to be considered in light of future AI advancements.

Artificial intelligence models are trained on vast amounts of data to recognize patterns. If the input data (such as interview transcripts) lacks diversity, the AI's ability to handle the complexity and nuances of language may not be fully tested.

5 Summary

The integration of Artificial Intelligence (AI) models into qualitative research brings significant advances in the efficiency and depth of data analysis. Based on the results of the study, it can be stated that large language models (LLMs) such as OpenAI GPT-4o, Google Gemini 2.0, Anthropic Claude Sonnet 3.5 and Haiku V3 can speed up coding processes and identify complex patterns in text data. Automated analysis tools such as Atlas.ti and MAXQDA's built-in AI functionality also help to optimise research workflows, although they face language limitations for non-English data. However, due to the performance differences between models, it is advisable to be careful in model selection and to perform the analysis with several models, using the "one measurement not measurement" principle. If the scientific analysis study is supported by generative solution, it is worth using 2-3 models, this way the objectivity can be better guaranteed and the number of erroneous conclusions arising from AI bias can be reduced In qualitative analysis, the use of AI not only saves time, but also allows the identification of themes and relationships that would be difficult to identify using traditional methods. For example, the Claude Sonnet 3.5 model showed more accurate coding results for scaling and alignment, while GPT-4o showed more frequent errors due to misinterpretation of context. However, the models' abilities remain limited without critical reflection of human analysis, especially in areas such as the interpretation of roles and responsibilities, where the subtle nuances of the text are crucial. Ethical challenges remain central. The paper highlights privacy risks, particularly when public AI models have access to confidential interview data. Researchers need to ensure anonymity and strict data management protocols, while continuously monitoring model biases (e.g. language biases). A hybrid approach, where AI assists but does not replace the researcher, is paramount to maintain the integrity of qualitative research. As previous research has highlighted, there are concerns about the potential misuse of AI in the field of academic dishonesty and the need to develop ethical guidelines for its use [58[59]. Of particular importance for future research is the development of transdisciplinary methodologies that link AI technologies with traditional frameworks in the social sciences. There is also a need to develop standardised evaluation frameworks that can objectively measure the effectiveness of models at different stages of qualitative analysis. Overall, the available large language models are already an indispensable tool for qualitative research, but their effectiveness relies on critical thinking and methodological rigour. With the rapid pace of technological development, the research community must remain in constant dialogue with ethical, methodological, and practical challenges to ensure that AI is truly at the service of scientific reliability and societal utility.

References

[1] Cano, Y. M., Venuti, F., & Martinez, R. H. (2023) ChatGPT and AI text generators: Should academia adapt or resist? Harvard Business Publishing. https://hbsp.harvard.edu/inspiring-minds/chatgpt-and-ai-text-generators-should-academia-adapt-or-resist

[2] Saunders, M., Lewis, P., & Thornhill, A. (2023) Research Methods for Business Students ISBN: 978-1-292-40273-4, 650-718

[3] Séllei, B., Krén, H., Molnár, G., & Stumphauser, N. (2024) Gaps and Bridges between Future Managers and IT People, *Acta Polytechnica Hungarica, 21*(3), 53-72

[4] Skinner, R., Ryan Nelson, R., & Chin, W. W. (2022) Synthesizing Qualitative Evidence: A Roadmap for Information Systems Research. *Journal of the Association for Information Systems, 23*(3), 639-677, https://doi.org/10.17705/1jais.00741

[5] Drisko, J. W. (2020) qualitative research synthesis: an appreciative and critical introduction. *qualitative social work, 19*(4), 736-753, https://doi.org/10.1177/1473325019848808

[6] Sasvári P. (2024) Interjúk mesterséges intelligenciával: A kvalitatív kutatás átalakulása https://doi.org/10.13140/RG.2.2.25156.64647

[7] Bearman, M., & Dawson, P. (2013) qualitative synthesis and systematic review in health professions education. *medical education, 47*(3), 252-260. https://doi.org/10.1111/medu.12092

[8] Moeller, K. (2021) Qualitative meta-synthesis. In *The Encyclopedia of Research Methods in Criminology and Criminal Justice: Volume II: Parts 5-8* (pp. 417-422) https://doi.org/10.1002/9781119111931.ch86

[9] Finlayson, K. W., & Dixon, A. (2008) Qualitative meta-synthesis: a guide for the novice. *Nurse Researcher, 15*(2), 59-71, https://doi.org/10.7748/nr2008.01.15.2.59.c6330

[10] Xu, Y. (2008) Methodological issues and challenges in data collection and analysis of qualitative meta-synthesis. *Asian Nursing Research, 2*(3), 173-183, https://doi.org/10.1016/S1976-1317(08)60041-9

[11] Bearman, M., & Dawson, P. (2013) qualitative synthesis and systematic review in health professions education. *medical education, 47*(3), 252-260. https://doi.org/10.1111/medu.12092

[12] Saini, M., & Shlonsky, A. (2012) *Systematic Synthesis of Qualitative Research* (Vol. 9780195387216, p. 224) https://doi.org/10.1093/acprof:oso/9780195387216.001.0001

[13] Belotto, M. J. (2018) Data analysis methods for qualitative research: managing the challenges of coding, interrater reliability, and thematic analysis. *The Qualitative Report, 23*(11), 2622-2633

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

[15] Jiang, X., Tian, Y., Hua, F., Xu, C., Wang, Y., & Guo, J. (2024) A survey on large language model hallucination via a creativity perspective. *arXiv preprint* arXiv:2402.06647, https://doi.org/10.48550/arXiv.2402.06647

[16] Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., ... & Shi, S. (2023) Siren's song in the AI ocean: a survey on hallucination in large language models. *arXiv preprint* arXiv:2309.01219. https://doi.org/10.48550/arXiv.2309.01219

[17] O'Brien, M. (2023, August 1) Chatbots sometimes make things up. Is AI's hallucination problem fixable? *AP News.* https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4

[18] Etli, D. (2024) Generative AI and Patient Care: A Systematic Review Examining Applications, Limitations, and Future Directions for ChatGPT in Healthcare. Journal of Quality in Health Care & Economics. https://doi.org/10.23880/jqhe-16000361

[19] Nagy, E., Karl, É., & Molnár, G. (2024) Exploring the Role of Human-Robot Interactions, within the Context of the Effectiveness of a NAO Robot. *Acta Polytechnica Hungarica, 21*(3), 177-190

[20] Pérez Gamboa, A. J., & Díaz-Guerra, D. (2023) Artificial Intelligence for the development of qualitative studies. LatIA

[21] Costa, K., & Costa, A. P. (2024) Qualitative Evidence Synthesis Using COSTAQDA. *Lecture Notes in Networks and Systems, 1061 LNNS,* 51-77. https://doi.org/10.1007/978-3-031-65735-1_4

[22] Fielding, N. G., & Lee, R. M. (Eds.). (1991) Using computers in qualitative research. Sage Publications, Inc.

[23] C. A. C., Cisneros Puebla, C. A. & Davidson, J. (2012) Qualitative Computing and Qualitative Research: Addressing the Challenges of Technology and Globalization. Forum Qualitative Sozialforschung / Forum Qualitative Social Research 13(2). Retrieved 24.11.2017, from http://nbn-resolving.de/urn:nbn:de:0114-fqs1202285

[24] Seale, C. (2010). Using computers to analyse qualitative data. In Silverman, D. (Eds.): Doing qualitative research (pp. 251-265) Sage, London

[25] Dalkin, S., Forster, N., Hodgson, P., Lhussier, M., & Carr, S. M. (2020) Using computer assisted qualitative data analysis software (CAQDAS; NVivo) to assist in the complex process of realist theory generation, refinement and testing. International Journal of Social Research Methodology, 24(1), 123-134, https://doi.org/10.1080/13645579.2020.1803528

[26] Mellado, B. H., Brandão, C., Pilger, T. L., & Dos Reis, F. J. C. (2020) using rqda in qualitative data analysis: a tool for researching chronic pelvic pain using thematic analysis. *New Trends in Qualitative Research, 1*, 153-170, https://doi.org/10.36367/ntqr.1.2020.153-170

[27] Reis, L. P., Costa, A. P., & De Souza, F. N. (2016) A survey on computer assisted qualitative data analysis software. 2016-July, https://doi.org/10.1109/CISTI.2016.7521502

[28] Freitas, F., Ribeiro, J., Brandão, C., Reis, L. P., De Souza, F. N., & Costa, A. P. (2017) Learn by yourself: the self-learning tools for qualitative analysis software packages. *Digital Education Review, 32*, 97-117

[29] Jamison, T., Kemp, C. L., Speirs, K. E., Swenson, A., & Vesely, C. K. (2018) Dialogue About Qualitative Data Analysis Software. In *How Qualitative Data Analysis Happens: Moving Beyond "Themes Emerged"* (pp. 221-227) https://doi.org/10.4324/9781315171647-16

[30] Andrade, L. R., Costa, A. P., Linhares, R. N., de Almeida, C. A., & Reis, L. P. (2019) Qualitative data analysis software packages: an integrative review. *Advances in Intelligent Systems and Computing, 861*, 279-290, https://doi.org/10.1007/978-3-030-01406-3_24

[31] Niedbalski, J., & Ślęzak, I. (2021) exploring CAQDAS-How to Support a Novice User of Computer Aided Qualitative Data Analysis Software. *New Trends in Qualitative Research, 6*, 1-14, https://doi.org/10.36367/ntqr.6.2021.1-14

[32] Silver, C. (2018) CAQDAS at a crossroads: choices, controversies and challenges. *Advances in Intelligent Systems and Computing, 621*, 1-13, https://doi.org/10.1007/978-3-319-61121-1_1

[33] Feuston, J. L., & Brubaker, J. R. (2021) Putting Tools in Their Place: The Role of Time and Perspective in Human-AI Collaboration for Qualitative Analysis. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1-25, https://doi.org/10.1145/3479856

[34] Hamilton, L., Elliott, D., Quick, A., Smith, S., & Choplin, V. (2023) Exploring the Use of AI in Qualitative Analysis: A Comparative Study of Guaranteed Income Data. International Journal of Qualitative Methods, 22, https://doi.org/10.1177/16094069231201504

[35] Hunt, C. S., Crick, T., Groves, M. J., & Hanna, P. (2022) Artificial Intelligence and the Changing Nature of Work. In *Handbook of Research on Artificial Intelligence and Productivity* (pp. 1-22) IGI Global

[36] Khaliq, A., Rehman, A. U., Akram, M. U., Iqbal, M. A., & Sarwar, M. Z. (2022) Impact of Artificial Intelligence on Future of Work: A Systematic Literature Review. *Journal of Open Innovation: Technology, Market, and Complexity, 8*(2), 87

[37] Lennon, R. P., Fraleigh, R., van Scoy, L. J., Keshaviah, A., Hu, X. C., Snyder, B. L., Miller, E. L., Calo, W. A., Zgierska, A. E., & Griffin, C. (2021) Developing and testing an automated qualitative assistant (AQUA) to support qualitative analysis. *Family Medicine and Community Health, 9*, https://doi.org/10.1136/fmch-2021-001287

[38] Grimmer et al, 2021 Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021) Machine Learning for Social Science: An Agnostic Approach. Annual Review of Political Science, 24(1), 395-419, https://doi.org/10.1146/annurev-polisci-053119-015921

[39] Buruk, 2023 Buruk, O. “Oz.” (2023) Academic Writing with GPT-3.5 (ChatGPT): Reflections on Practices, Efficacy and Transparency. 26th International Academic Mindtrek Conference, 144-153, https://doi.org/10.1145/3616961.3616992

[40] Liu, M., Ren, Y., Nyagoga, L., Stonier, F., Wu, Z., & Yu, L. (2023) Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools. Future in Educational Research, 1, https://doi.org/10.1002/fer3.10

[41] Saliba T, Boitsios G. ChatGPT, a radiologist's perspective. Pediatr Radiol. 2023 May;53(5):813-815. doi: 10.1007/s00247-023-05656-z. Epub 2023 Apr 5. PMID: 37017719

[42] Perkins, M., & Roe, J. (2024) The use of Generative AI in qualitative analysis: Inductive thematic analysis with ChatGPT. *Journal of Applied Learning and Teaching, 7*(1) https://doi.org/10.37074/jalt.2024.7.1.22

[43] Bryda, G., & Costa, A. (2023) Qualitative Research in Digital Era: Innovations, Methodologies and Collaborations. social sciences. https://doi.org/10.3390/socsci12100570

[44] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020) Don't Stop Pretraining: Adapt Language Models to Domains and Tasks (No. arXiv:2004.10964) arXiv. https://doi.org/10.48550/arXiv.2004.10964

[45] Collins, C. S., & Stockton, C. (2022) The Theater of Qualitative Research: The Role of the Researcher/Actor. International Journal of Qualitative Methods, 21, https://doi.org/10.1177/16094069221103109

[46] Atkinson, C. F. (2024) Cheap, Quick, and Rigorous: Artificial Intelligence and the Systematic Literature Review. *Social Science Computer Review, 42*(2), 376-393, https://doi.org/10.1177/08944393231196281

[47] Jiang, J. A., Wade, K., Fiesler, C., & Brubaker, J. R. (2021) Supporting Serendipity: Opportunities and Challenges for Human-AI Collaboration in Qualitative Analysis. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1) https://doi.org/10.1145/3449168

[48] Ketak, R., Mittal, S., Gupta, V., & Gupta, H. (2024) Online Edtech Platform with AI Doubt Assistance. 1395-1399, https://doi.org/10.1109/ICDT61202.2024.10489310

[49] Kilz, L. (2024) Student perspective on the use of ai in higher education: analysis of the students' podcast. 133-140, https://doi.org/10.35011/IDIMT-2024-133

[50] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020) language models are few-shot learners. *advances in neural information processing systems, 33*, 1877-1901

[51] Geiecke, F., & Jaravel, X. (2024, October 30) *AI can carry out qualitative research at unprecedented scale*. LSE Impact of Social Sciences Blog. https://blogs.lse.ac.uk/impactofsocialsciences/2024/10/30/ai-can-carry-out-qualitative-research-at-unprecedented-scale/

[52] Srivastava, A. P. (2024) AI-powered data collection and analysis. in *Utilizing AI Tools in Academic Research Writing* (pp. 114-140) scopus. https://doi.org/10.4018/979-8-3693-1798-3.ch008

[53] Ntsobi, M. P., Costa, A. P., Kasperiuniene, J., Brandão, C., & Ribeiro, J. (2024) Digital Tools and Techniques in Qualitative Research: Digital Skills and Research Optimisation. *Lecture Notes in Networks and Systems, 1061 LNNS,* 1-25, https://doi.org/10.1007/978-3-031-65735-1_1

[54] Williams, R. T. (2024) Paradigm shifts: exploring AI's influence on qualitative inquiry and analysis. *Frontiers in Research Metrics and Analytics, 9,* https://doi.org/10.3389/frma.2024.1331589

[55] Fyfe, P. (2023) How to cheat on your final paper: assigning AI for student writing. *AI & Society, 38*(4), 1395-1405, https://doi.org/10.1007/s00146-022-01397-z

[56] Schmohl, T., Watanabe, A., Fröhlich, N., & Herzberg, D. (2020) *How can Artificial Intelligence Improve the academic writing of students?* The Future of Education, 12-14, https://thesiswriter.zhaw.ch/

[57] Elliott, V. (2018) thinking about the coding process in qualitative data analysis. *The Qualitative Report, 23*(11) 2850-2861

[58] Soulage, C. O., Van Coppenolle, F., & Guebre-Egziabher, F. (2024) The conversational AI "ChatGPT" outperforms medical students on a physiology university exam. *Advances in Physiology Education, 48*(4), 677-684, https://doi.org/10.1152/advan.00181.2023

[59] Frangieh, C., & Abdallah, M. (2024) A Linguistic Comparison between Human-and AI-Generated Texts in Lebanon. *Centre for Applied Linguistics Research Journal, 2024*(15) https://doi.org/10.60149/TMSY9941