Sage Research Methods Data and Research Literacy: How-to Guide

Authors: **Please complete the white fields below.** Direct any questions to your editorial contact.

Title. ▪ *Maximum of 20 words.* ▪ *All principal words capitalized.*		*[How to use Generative AI to Assist the Analysis of Qualitative Data]*
Authors. ▪ *The order of authorship in the publication will follow the order below.* ▪ *Please add additional rows for co-authors if necessary.*		Please specify the total number of authors: **1**
1	Name	Kathryn Wheeler
	Author email	[katy.wheeler@essex.ac.uk]
	Affiliation, country	[University of Essex, UK]
	SAGE Author ID	[office use only]
Author biography. *Bios will not be copy-edited; please ensure they are correct.*		[Maximum of 200 words] Dr Katy Wheeler is a Senior Lecturer in Sociology at The University of Essex. Her research interests are in the fields of sustainability and consumption, and she has published widely within leading academic journals and two authored monographs, Fair-Trade and the Citizen-Consumer: Shopping for Justice (Palgrave: 2012) and Recycling and Consumption Work: Social and moral economies (Palgrave: 2015). She has also co-authored the textbook 'How to do Qualitative Interviewing' (SAGE, 2022 – with Morgan Brett). She has taught social science methodology courses for many years (both qualitative and quantitative), from introductory undergraduate to advanced postgraduate. *I have published with Sage and want to use my existing bio* ☐ *I have published with Sage and want to update my bio* ☒ *I have not published with Sage before* ☐
2	Name	
	Author email	[Insert contact email]
	Affiliation, country	[Insert institution and country]
	SAGE Author ID	[office use only]
Author biography. *Bios will not be copy-edited; please ensure they are correct.*		[Maximum of 200 words] *I have published with Sage and want to use my existing bio* ☐ *I have published with Sage and want to update my bio* ☐ *I have not published with Sage before* ☐
3	Name	

	Author email	[Insert contact email]
	Affiliation, country	[Insert institution and country]
	SAGE Author ID	[office use only]
Author biography. *Bios will not be copy-edited; please ensure they are correct.*		[Maximum of 200 words] *I have published with Sage and want to use my existing bio* ☐ *I have published with Sage and want to update my bio* ☐ *I have not published with Sage before* ☐
For which **student level** is this guide most suitable?		**Advanced Undergraduate**
Methodology categorization		**Qualitative**
Discipline		Guides should be **cross-disciplinary**. If the content of your guide is specific to a discipline(s), please let your editorial contact know.
Series Editor. *For office use only:*		
	Series Editor Name	
	SAGE Author ID	[office use only]

Sage Research Methods: Data and Research Literacy is an educational resource which will equip undergraduate and postgraduate students and researchers with the tools to **understand and critically evaluate research methods and methodologies**, **manage and interpret data**, and **conduct robust social research with integrity** and confidence.

Guides will be **authoritative and accessible** resources which **combine research principles with research practise**, incorporating practical and ethical considerations, to help prepare students and researchers for working with data, evaluating research, and conducting their own research.

When writing your guide, we recommend using **real-world research examples** to keep the reader engaged. You may choose to use one consistent example throughout the guide, or multiple examples.

Each how-to guide is limited to **4000 words**, with a 10% leeway. For topics which require more than 4000 words there may be the option to write multiple guides; please raise this with your editorial contact if required. Guides may include direction to further resources through which the reader can explore each topic in more depth.

You can view two how-to guides from previously published collections here:

- From *Diversifying and Decolonizing Research*

- From *Doing Research Online*

Please ensure you have read the **manuscript guidelines** before you begin writing your guide.

References should conform to American Psychological Association (APA) style, 7th edition, and should contain the digital object identifier (DOI) where available. Sage will not accept guides that are incorrectly referenced; please ensure accuracy before submission. For help on reference styling see https://apastyle.apa.org/style-grammar-guidelines.

Abstract

The abstract should be a concise summary of your how-to guide. What aspect of the research process, working with data, or specific methodological and practical challenges will your guide address? It should be succinct and enticing, and should incorporate key words and concepts discussed in the body of the text. Please do not cite references within the abstract.

[Insert here: Maximum of 250 words]

This guide discusses the use of Generative Artificial Intelligence (GenAI) tools in qualitative analysis. There is a growing body of research which has used programmes like ChatGPT and other OpenAI software to analyse qualitative data. This has been justified because GenAI can produce summaries of large amounts of data which resemble human-created output. This guide offers an overview of the options available to researchers and discusses some of the implications of their use. There is an active debate about whether such tools are appropriate and if so how to ethically and practically implement them into a qualitative analysis workflow. This guide argues that the human analyst remains central, with GenAI acting as a useful assistant. Technology will shape 'how to do' qualitative analysis and it is important for researchers to actively reflect upon the opportunities and challenges of using these tools within their practices.

Learning Outcomes

Learning outcomes must explain what the reader will learn from reading your guide. How will the reader be able to apply what they have learned to their own research practice?

Consider what the **most important aspects of this topic** are. Bear in mind the guide is limited to 4000 words. **The content and structure of your guide should explicitly correspond with these learning outcomes**.

See the links below for guidance on writing effective learning outcomes:

- *Writing learning outcomes*

- *Blooms Taxonomy Action Verbs*

Insert 3–5 learning outcomes, **beginning with an action verb**, completing this statement:

Having read this guide, readers should be able to . . .

Learning outcomes

- Explain the use of Generative AI (GenAI) in qualitative analysis.
- Discuss the ethical and practical implications of AI tools in qualitative workflows.
- Debate the appropriateness of integrating GenAI into qualitative analysis.
- Develop a reflexive protocol for combining GenAI and human analysis in qualitative research.

Introduction

Build on the abstract to further describe what methodological issues will be discussed in this guide; what the student reader will gain from reading the guide; how the guide will be structured; which real-life research examples will be drawn upon, etc. You may wish to begin with a brief positionality statement.

Following the launch of ChatGTP to the public in November 2022, I noticed many researchers on my Twitter/X feed speculating about whether this marked a new era in qualitative analysis. Within a year, two of the major qualitative analysis software products (MAXQDA and ATLAS.ti) introduced Beta versions which had embedded Generative Artificial Intelligence (GenAI) features to assist with analysis, such as summaries and code suggestions. At the same time, I was working with an organisation (SEEd) which had collected data from open-ended survey questions and was wondering how I would analyse this. This survey asked young people about their understandings of and attitudes towards sustainability and included around 15 open-ended questions which were all answered (to variable degrees) by over 1500 respondents. I started to play with GenAI features in MAXQDA, my preferred/usual analysis package, and in preparing this guide I have also experimented with ChatGTP, ATLAS.ti and NVivo. I have used this technology to assist analysis of qualitative survey data and I believe it offers many opportunities to the qualitative researcher. But we must use GenAI in a human-centred way and be attentive to the possible ethical and practical challenges it presents.

Utilising technology for qualitative analysis (even without GenAI) has been controversial amongst some qualitative researchers who feel that computer tools with prescribed functions do not align with the flexible and iterative nature of qualitative inquiry (Silver & Woolf, 2020). Though there are strategies for harnessing software tools effectively for qualitative analysis, training in CAQDAS (computer assisted qualitative data analysis software) is patchy, with few undergraduates programmes offering guidance (Silver & Woolf, 2015). Having integrated NVivo training into an undergraduate methods programme, I understand the challenges of convincing both students and staff of its relevance. In the wake of GenAI, we have a strong duty to equip qualitative researchers with appropriate tools for the intentional and reflexive analysis of often messy and complex data. This guide offers an overview of some GenAI tools qualitative researchers are using and considers how these might influence and shape a qualitative analysis workflow. It addresses ethical questions and established understandings of meaning-making within qualitative analysis. It is hoped that after reading this guide, readers should be able to design their own protocols for using GenAI in their qualitative projects.

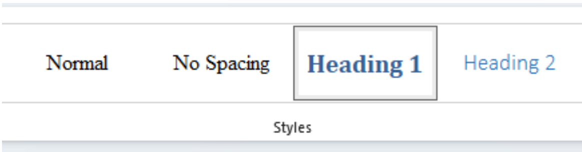
Main Body of Text

Write the body of your guide below. The text should be between **2,000 - 4,000 words.**

We encourage the use of **headings** and sub-headings to **structure your guide into sections.**

We recommend using **800 words or fewer for each section**.

For section headings please use Word Style 'Heading 1'. For any sub-headings within sections use Word Style 'Heading 2'. To use Word styles in Microsoft Word, select the text you want to format, click the "Home" tab and then use the "Styles" pane.



Every section must be followed by a **Section Summary.**

Each Section Summary should consist of 3-5 bullet points, written out as full sentences, which summarize the key information in the section.

What is Generative AI and how are qualitative researchers using it?

Let's start with some basics – what is Artificial Intelligence (AI) and what is Generative Artificial Intelligence (GenAI)? IBM offer two relatively simple definitions for these terms:

> *Artificial intelligence, or AI, is technology that enables computers and machines to simulate human intelligence and problem-solving capabilities. On its own or combined with other technologies (e.g., sensors, geolocation, robotics) AI can perform tasks that would otherwise require human intelligence or intervention.*
>
> *[…]*
>
> *Generative AI refers to deep-learning models that can take raw data […] and "learn" to generate statistically probable outputs when prompted. At a high level,*

> *generative models encode a simplified representation of their training data and draw from it to create a new work that's similar, but not identical, to the original data.* (IBM, 2024)

These definitions highlight AI is about 'simulating' human capabilities and through deep learning of multiple data sources, GenAI can create new content based on patterns it has learned from these existing data sources. As Silver (2023) notes, many AI tools currently being discussed, like machine learning and natural language processing (NLP), have an established history within CAQDAS packages, but they have not always been embraced by qualitative researchers. The public release of ChatGTP marked a significant change and has caused quite a 'hullabaloo' (to quote Silver).

Reviewing available technology is risky because it evolves so quickly so what follows may become outdated. For the latest on CAQDAS AI packages and features, visit the Surrey CAQDAS networking project run by Dr Christina Silver and check her regularly updated blog and podcast series (see further resources at end of this guide).

Overview of AI qualitative analysis features

As of July 2024, two major CAQDAS packages (MAXQDA and ATLAS.ti) have adopted AI large language models – (either partnering with OpenAI the company behind ChatGPT – or another provider) and are increasingly embedding more GenAI features. As a regular user of MAXQDA, I am aware the software has updated itself about 5 times already in 2024, and each time some tweak is made to the *AI Assist* feature. The other market leader, NVivo, has not partnered with OpenAI (yet) but has enabled GenAI features to autocode data. There is also a company called CoLoop which specialises in GenAI-driven qualitative analysis. We are starting to see a growing number of published and preprint articles discussing using ChatGPT and ChatPDF to assist qualitative data analysis (Chubb, 2023; Hamilton et al., 2023; Hitch, 2024; Morgan, 2023). There are also papers which question the appropriateness of using such tools (Paulus & Marone, 2024), and practical guidance on how to embed them within established workflows (De Paoli, 2023; Nguyen-Trung, 2024; Paulus & Lester, 2023; Zhang et al., 2023).

The current GenAI tools offered to qualitative researchers (whether via an add-on or bespoke CAQDAS, or via web-based companies and interactive chatbot providers) are:

- **Automatic transcription:** This became common during the pandemic, with data collection on Zoom and Teams, where cloud recordings generated transcripts. There are bespoke automatic transcription companies (like OtterAI), transcription features in Microsoft Word and Dropbox, as well as automatic transcription embedded within CAQDAS programmes, MAXQDA and NVivo. Researchers upload their audio files and receive a transcript which will still require some editing, but it is *much* more accurate than those produced even five years ago. From conversations with peers, this is the tool that been most embraced by qualitative researchers, perhaps without them realising GenAI is involved.

- **Summarizing tools:** This feature allows researchers to send parts of their data to a GenAI assistant to receive summaries of the content. These can be applied to short data segments, entire interview transcripts, a code (to provide an overview of already coded data), or a series of open-ended survey questions. Summaries can be customised by length and format. This feature may help researchers familiarise themselves with their data at the start of a project or be harnessed during the analysis stage to generate themes. Chubb (2023) describes using summaries to create vignettes of interview data for analysis and data presentation purposes.

- **Coding suggestions:** Closely related to summarising tools is the possibility of using GenAI to suggest possible ways to code your data. This can again be performed on a range of 'chunks' of data, such as a list of coding ideas for a whole transcript or survey question responses, or on a small segment of data which the researcher has identified as a coding segment. In my survey project, I have found using the coding suggestions on smaller segments of text more useful, as suggestions were more targeted to the unit of meaning. This feature was beneficial in the early stages of coding before my coding list was fixed. However, coding lists for all survey responses mainly reaffirmed summaries, useful for familiarization but not much beyond that.

- **GenAI Automatic coding:** Whilst auto-coding based on key-word searches has been established in CAQDAS for years, GenAI auto-coding offers something different. The CAQDAS packages have implemented automatic coding features in different ways, with sentiment analysis, open coding and intentional coding possibilities. All offer sentiment analysis – which uses NLP to detect positive, negative or neutral sentiments within data. MAXQDA auto-codes only survey data with sentiments, while NVivo and ATLAS.ti code all data on request. I have found sentiment analysis unhelpful for my research purposes, not least because the sentiments can be incorrectly labelled and the nuances in the language are not always picked up. For example, 'climate change' was detected as 'slightly positive' by the feature in MAXQDA.

Researchers (or perhaps companies) are most excited by open and intentional coding features, leading to rumblings that the need to spend hours coding your data are gone (more on this below). Open coding features are available in ATLAS.ti and NVivo and return a list of 'codes' and chunked data associated with those labels. For ATLAS.ti what tends to come back is a very long list of codes which will require a lot of sorting, whereas for NVivo the list was shorter but again not terribly useful because codes did not apply to all survey responses. The intentional coding, currently offered by ATLAS.ti, MAXQDA (currently just for single documents) and CoLoop (which I haven't trialled yet) is the one that promises the end of coding. These systems ask researchers a series of questions about their research projects to train GenAI what to look for, and then codes the data based on those answers. The intentional coding feature is the space to watch over coming months.

- **Interactive 'chat' with your data:** Chatbots like ChatGPT, Copilot, and Bard exemplify the GenAI era, enabling dialogue with your data and intentional coding. Many published/pre-print papers have used ChatGPT (or similar) as the GenAI assistant in their projects. Researchers upload their data and ask ChatGTP to analyse it for key ideas or themes, receiving lists of codes and training the GenAI to achieve more relevant results. MAXQDA and ATLAS.ti have a 'chat' feature to ask customisable questions of specific documents – with warnings that the resulting content should be carefully checked for accuracy as the chat may make mistakes.

Using GenAI in your qualitative analysis projects is a choice – and it will be a choice that many qualitative researchers will not make. For those who are interested in possible ways to implement these tools, the next section explores how to integrate them into existing qualitative analysis workflows. This is followed by a discussion of the ethics and possible pitfalls of using these features.

Section Summary

- *Breakthroughs in AI technology have opened the possibility of using GenAI tools to assist qualitative data analysis.*
- *Researchers can use existing CAQDAS packages, new GenAI-Based CAQDAS programmes or free chatbots, like CHATGPT, to assist their analysis.*
- *The qualitative analysis tasks GenAI can currently assist with are, generating transcripts, summarising data, suggesting ways of coding data and engaging in dialogues around the content of the data.*

Implementing AI into your qualitative analysis workflow

Before jumping headfirst into the opportunities these new tools present, consider your methodological position on data construction and interpretation. Silver and Woolf's (2015) five-level QDA emphasises understanding what your research goals are and what you want to accomplish through your analysis before turning to technology. For example, Braun and Clarke's (2019, 2022) 'reflexive thematic analysis' has been influential and the assumptions behind this approach are that human interpretation and positionality will always shape every stage of the analysis process. Such a position does not sit easily with GenAI-assisted analysis – though there are those who have argued that GenAI can offer a useful second opinion and deepen the researcher's awareness of their own reflexivity (Hitch, 2024). For those whose approach to analysis takes a less constructivist stance, such as more realist versions of

Qualitative Content Analysis (QCA) (Schreier, 2013), GenAI-assisted approaches may aid inter-coder reliability which an individual researcher can find hard to achieve.

This guide is not the place to outline different types of qualitative data analysis (good overviews are provided by Braun & Clarke, 2021; Gibbs, 2018), but most approaches to qualitative data analysis involve at least four stages in a workflow model - though the importance of each stage may differ and these are often iterative rather than linear. The four stages are familiarisation with the collected data, description or categorisation of that data, systematic sorting (often called 'coding' but not for all analysts) in search of patterns, interpretation of the preceding stages through writing about the data. The section that follows this discusses the ethical and analytical challenges associated with using GenAI for qualitative analysis tasks. Readers should consider how they might integrate GenAI into their research protocols.

For the rest of this section, I explore how GenAI-tools could be used at each stage, based on my experiences of analysing open-ended survey data with some GenAI-assistance (Wheeler, 2024) and my awareness of analysing qualitative interviews without it (Morgan Brett & Wheeler, 2022). Much of the 'hullabaloo' surrounding GenAI-tools relate to their time-saving properties, though I'm with Paulus and Malone (2024) when they point out that spending time on analysis is part of the joy and necessity of qualitative research. So, in what follows, timesaving is not my key focus!

1. Familiarisation

To move from raw data to interpretation, it is essential you know what your data is about. This involves reading and immersing yourself in the data. With interview data, transcribing helps with familiarisation, but GenAI (or a human transcriber) can ease some of the time-burden by providing a transcript. However, you should re-listen to the recording alongside the transcript to re-situate yourself and judge the accuracy of the transcription. If the transcript contains non-relevant data (like the consent to tape record for instance), these sections should be removed to avoid skewed AI summaries. GenAI summaries of interviews or survey responses provide useful snapshots of the data, and asking for these in different lengths can aid familiarisation. This can be particularly helpful for spotting key features of the data which may have been missed by a single researcher who has a tightly defined focus. For example, when reading thousands of survey responses, it was difficult to hold onto the breadth and diversity of issues in the early stage of analysis, so having a second opinion from GenAI was useful. Researchers should judge the accuracy of GenAI summaries as they become familiar with the raw data.

2. Description or categorisation of the data

Initial data summaries can be used as a springboard for developing a list of key concepts or code lists. Entering into dialogue with a transcript or set of survey responses, might help you pinpoint the important features of your data, and relevant quotations to explore further. CAQDAS with embedded GenAI might limit the amount of data processed at once within their 'free' features, and ChatGPT also has limits, so results need to be carefully checked for applicability to the whole corpus of responses. I have found the 'code suggestions' feature on smaller data chunks useful for sparking ideas about how to categorise data. The GenAI will usually suggest more ways to code data than I would have thought of (perhaps 12 labels for the same piece of text) – which have helped me to think about my data in a range of ways. Here the AI-assistant is acting like a junior researcher who has lots of ideas but needs to be reined in by the more experienced researcher. Those using prompt-based commands in

ChatGPT for this categorisation stage highlight that getting the right prompt can take time and several iterations, and qualitative researchers will likely have to adapt suggestions made by GenAI-tools (Hitch, 2024; Nguyen-Trung, 2024).

3. Systematic sorting to look for patterns

Once you have decided how you will categorise your data, you can ask GenAI-tools to code the data for you. I'll be up-front and say I have not done this – even though trawling through thousands of responses was time-consuming, I needed to be sure that my coding was applied systematically to all answers. Part of my decision to not use this feature was shaped by the fact that MAXQDA do not (yet) offer intentional AI-Based coding for survey responses (currently limited to single documents) though it can make suggested for single responses . But experimenting with ChatGPT, I have been impressed with the possibilities, though mindful that getting the prompts right and engaging actively with what GenAI has produced is important. I wanted some of my codes to be mutually exclusive (only one category to apply to each answer) following QCA principles – I had coded survey definitions of sustainability as predominantly environmental, social or economic, more than one of these, or none of these. My attempts to get ChatGTP to do this initially resulted in it miscategorising over 1000 responses as 'none of these', and though the broad trends confirmed my manual coding that vague definitions dominated, I coded many responses differently than ChatGPT. I asked ChatGPT for a CSV file to inspect how it had coded the responses and to further train the system to get more relevant answers. The GenAI-coding was an important check on my own coding decisions (much as inter-coder reliability might be), forcing me to evaluate why I had coded under some categories rather than others. This is something hard to achieve as a single researcher and for QCA, I see the value of using GenAI to perform this function. For thematic analysis, I need to experiment more – the dialogue feature is helpful for locating key quotations but data length restrictions and GenAI's forgetfulness are concerns. Nguyen-Trung (2024) describes the process of getting ChatGPT to consolidate and cluster her coding list but found that it was often not up to the task of spotting overlapping codes, so subjective researcher decisions had to be made.

4. Interpretation through writing

Experiments with GenAI show it performs poorly at generating interpretive themes, offering more literal and descriptive summaries instead (Morgan, 2023; Nguyen-Trung, 2024). While prompts can be adapted to get more relevant results, having the ‘human in the loop’ to intervene and make judgements around interpretation is crucial (De Paoli, 2023). Already coded data can be summarised by GenAI tools and what I found useful here was the ability to produce code summaries for different groups of responses – such as asking for summaries by gender, age group and levels of income deprivation (features that were coded as document variables in MAXQDA). Pulling out the nuances between these groups in the summaries helped me to interpret key differences and similarities across survey respondents. Insights are further connected through CAQDAS features like visualization and memo writing. Storing everything in one project file within CAQDAS is a key benefit over ChatGTP, where information can easily be lost if the researcher does not store its outputs systematically.

Section Summary

- *Think carefully about what your key analytic tasks and assumptions are before turning to technology and GenAI tools to assist you.*
- *GenAI can be embedded at each stage of the qualitative analysis workflow, but its output must be carefully inspected and reflected upon to ensure its credibility and trustworthiness.*

- *GenAI could be viewed as a junior assistant whose work is descriptive and sometimes misses the point but who challenges the lead researcher to think about their analysis in different ways.*

Key considerations for human analysis

Though GenAI could be used at each stage of a qualitative analysis workflow, this section asks should it be? The human analyst is central for determining whether AI-generated output is credible and appropriate. Claims that GenAI will save the researcher hours have already been questioned, and many studies using ChatGTP have used data that researchers had analysed manually meaning they had a good grounding in the data before turning to GenAI. This section considers three major challenges associated with using these tools – ethics, bias, and the construction of meaning through qualitative analysis. Challenges around needing to learn how to use these systems practically and decision trail recording with chatbots have been discussed in the previous section and will not be returned to here.

Ethics

A key set of concerns when using GenAI relates to ethical research practice. If interview or sensitive research data is analysed by these systems, questions are raised about the security of this data. Given these systems are trained on data entered, does this mean any data uploaded becomes part of the model, putting the privacy of our participants under threat? OpenAI admits collecting personal data and possibly selling it. Users can choose not to store their chat history with conversations deleted after 30 days – though this does mean saving your chats for your decision trail elsewhere. CAQDAS packages with GenAI assure that user data is secure, not used for training, and deleted within 30 days. The CAQDAS user must opt-in and is in control of what data is sent to AI. At a minimum, data should be de-identified (being mindful that changing names alone is not always sufficient to achieve this) before it is processed. With my survey data (where personal information was not requested) there were still cases where respondents included information like their school/teacher's name which was removed before using GenAI. Given the widespread use of transcription services for raw data, anonymisation is not always happening before processing. Therefore, researchers need to start including statements about GenAI processing on their consent forms, so participants are informed about data use and associated risks. From the participant's perspective, this is similar to giving consent to archive their data for re-use, though researchers cannot add restrictions about who will access it.

There are also some broader ethical considerations about how GenAI systems have been trained on exploitative data practices, with publishing companies selling content to them without author's consent (Morreale et al., 2024; Potter, 2024). Sustainability concerns regarding high use of energy and water to power processors might also be a factor to consider before using this technology (Crawford, 2024).

Bias and research integrity

Because GenAI systems have been trained on human data which contains biased language and culturally offensive values, there have been cases where its output has reproduced these. Indeed in a study of three GenAI tools, Zhou et al (2024) found these tools exhibited systematic gender and racial bias, under-representing women and ethnic minorities beyond real-world disparities and assigning submissive characteristics to these groups. This, coupled with 'hallucination' – the production of false though seemingly credible responses by GenAI – raises concerns over its use in research. If inconsistencies and biases in output are not picked up by the researcher, this compromises research integrity. These issues underscore the necessity to engage critically with AI-generated content to ensure accuracy and credibility. This comes down to questions of authorship and researcher responsibility – if we are using results from GenAI, we should declare this in our writing and take full responsibility for anything we publish. GenAI is the assistant/junior researcher who should be acknowledged but 'the buck stops' with the lead researcher if what they produce is incorrect or biased. Many academic journals now require authors to disclose GenAI use and explain its contribution.

Making meaning through analysis

The final consideration addressed in this guide returns to an earlier point about the epistemology of qualitative research, which values the co-constructed nature of knowledge produced by this collection of methods. Many qualitative researchers make their knowledge claims based on deep relationships with their participants which have been gathered through spending time in specific research contexts and interpreting these social worlds through the lens of their unique positionality. GenAI is no replacement for human-centred analysis and this guide has stressed throughout that if it is used it must be done actively and reflexively. Claims that GenAI will save the researcher time and that analysis can be a quick process are not compatible with the practices of experienced qualitative researchers who know that iterative cycles of data collection, analysis and reflexivity are central to how meaning is made. The novice researcher may be tempted by quick results, and the danger of GenAI is that it can produce seemingly credible analysis at the touch of a button. Silver (2024) says the 'qualitative deepfake' – where technology generates fake data and analysis – is now a reality. Therefore, experienced researchers must train the next generation what it means to produce credible qualitative research which is supported by a clear decision-trail, deep analysis and reflexive awareness of the impact technology has on practices of research.

Section Summary

- *Using GenAI in your research carries data security and data privacy risks so only anonymised data ought to be uploaded to these systems.*
- *It is recommended that researchers include a statement on participant consent forms, informing them of the risks of data being processed by GenAI.*
- *Researchers need to be aware that GenAI can produce biased and inaccurate outputs and the responsibility for finding and correcting those errors lies with the researcher.*
- *Good qualitative research prioritises human analysis and uses GenAI to augment rather than replace this.*

Conclusion

Includes a **summary of the key lessons** discussed within each section of your guide.

What can readers learn from this guide and apply when conducting their own research and evaluating the research of others?

> *'We believe that society and culture influence the construction of technologies and that these technologies also influence society and culture. To accept this view means accepting that our qualitative work is sometimes influenced by NVivo'* (Jackson & Bazeley, 2019, p. 35)

Writing about CAQDAS more generally, the quotation above capture what I hope readers will take from this guide. Technology is shaped by society and it in turn shapes us. If you choose to use GenAI features in your qualitative analysis, do so mindfully and with careful consideration of how this may affect your research protocols and workflows. The speed of change has taken many researchers by surprise, and as CAQDAS did previously, I feel sure qualitative analysis will be transformed by these technologies – at least for some researchers. Rather than use a technologically determinist lens, I hope this guide has given you some tools to experiment with and imagine how research could be augmented through GenAI. I believe the human analyst must remain in the driving seat of these technologies because not all tasks can be delegated to it. The technology has flaws which could lead to inaccurate outputs, data privacy breaches and unethical research practice. Incidentally, these are flaws that a human analyst can also have. It is possible to see how GenAI can become a critical friend that shows the researcher alternative ways of summarising, categorising and understanding their data. I do not see GenAI as a threat to qualitative research practices but an opportunity to make our data analysis more robust. But this opportunity can only be realised if researchers are trained in critical, well-documented and reflexive research practices.

Multiple Choice Quiz Questions

Multiple Choice Quiz Questions should:

- Test readers' understanding of your guide.

- Focus on relevant aspects of data and research literacy.

- Not require any information that is not included in this guide.

Multiple Choice Quiz Questions should not:

- Include 'all of the above' or 'none of the above' options, or implausible responses.

- Require information not included in the guide.

Example:

1. What is critical reflexivity?

a. An understanding of how a researcher relates to and actively engages with the complex contexts and dynamics within which the research is embedded. ***[CORRECT]***

b. An understanding of how over-researched populations can experience research fatigue when directly engaged by researchers.

c. An understanding of anonymity and confidentiality in research.

Guidance for writing MCQs can be accessed using these links:

- *Tips for writing effective multiple-choice questions*

- *The process of writing a multiple-choice question*

[Insert three to five multiple choice quiz questions below. **Each MCQ must have three possible answers (A, B, or C), with one correct answer.** Please indicate the correct answer by writing [CORRECT] after the relevant answer.]

1. What is a key ethical concern when using GenAI for qualitative analysis?
    a. High cost of the technology
    b. Data security and privacy risks [CORRECT]
    c. Lack of available software packages

2. What role should GenAI play in the qualitative research process?
    a. Fully replace human analysis
    b. Augment and support human analysis (CORRECT)
    c. Automate the entire research workflow

3. What is a potential benefit of using GenAI tools in qualitative research?
    a. Providing alternative ways to summarize and categorize data (CORRECT)
    b. Automatically generating research hypotheses
    c. Ensuring complete accuracy of research findings

4. Which approach should researchers take to effectively combine GenAI and human analysis in qualitative research?

    a. Rely solely on GenAI for all stages of analysis.
    b. Use GenAI only for data collection and not for analysis.
    c. Actively engage with GenAI outputs and critically evaluate them. (CORRECT)

Further Reading

Please ensure that the recommended readings, web resources, and cited references in the guide are inclusive, and represent a diversity of people. Given our global readership, we aim for content that allows individuals with a broad range of perspectives to see themselves reflected in our published resources.

[Insert list of up to six further readings here]

- 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
- Morgan, D. L. (2023). Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. *International Journal of Qualitative Methods*, 22, 16094069231211248. https://doi.org/10.1177/16094069231211248
- Paulus, T. M., & Marone, V. (2024). "In Minutes Instead of Weeks": Discursive Constructions of Generative AI and Qualitative Data Analysis. *Qualitative Inquiry*, 10778004241250065. https://doi.org/10.1177/10778004241250065

Web Resources

[Insert links to up to six relevant web resources here]

- CAQDAS Chats with Christina, Podcast series, available at https://open.spotify.com/show/28usVeqag9q7irrrAgJTBh?si=7e44fcf1362c441d (accessed 28/07/2024)
- Computer Assisted Qualitative Data Analysis (CAQDAS) networking project, available at https://www.surrey.ac.uk/computer-assisted-qualitative-data-analysis (accessed 28/07/2024)
- Qualitative Data Analysis Services, Blog posts, available at https://www.qdas.co.uk/blog (accessed 29/07/2024)

References

[Insert bibliography of references cited in text here]

References should conform to American Psychological Association (APA) style, 7th edition, and should contain the digital object identifier (DOI) where available. Sage will not accept guides that are incorrectly referenced. Please ensure accuracy before submission. For help on reference styling see https://apastyle.apa.org/style-grammar-guidelines.

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

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, *21*(1), 37–47. https://doi.org/10.1002/capr.12360

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

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

Crawford, K. (2024). Generative AI's environmental costs are soaring—And mostly secret. *Nature*, *626*(8000), 693–693. https://doi.org/10.1038/d41586-024-00478-x

De Paoli, S. (2023). Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach. *Social Science Computer Review*, 08944393231220483. https://doi.org/10.1177/08944393231220483

Gibbs, G. R. (2018). *Analyzing Qualitative Data*. SAGE Publications Ltd. https://doi.org/10.4135/9781526441867

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*, 16094069231201504. https://doi.org/10.1177/16094069231201504

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

IBM. (2024). *What is artificial intelligence (AI)?* IBM.Com. https://www.ibm.com/topics/artificial-intelligence

Jackson, K., & Bazeley, P. (2019). *Qualitative data analysis with NVivo* (3. edition). Sage.

Morgan Brett, B., & Wheeler, K. (2022). *How to do Qualitative Interviewing*. Sage.

Morgan, D. L. (2023). Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. *International Journal of Qualitative Methods*, *22*, 16094069231211248. https://doi.org/10.1177/16094069231211248

Morreale, F., Bahmanteymouri, E., Burmester, B., Chen, A., & Thorp, M. (2024). The unwitting labourer: Extracting humanness in AI training. *AI & SOCIETY*, *39*(5), 2389–2399. https://doi.org/10.1007/s00146-023-01692-3

Nguyen-Trung, K. (2024). *ChatGPT in Thematic Analysis: Can AI become a research assistant in qualitative research?.* OSF PrePrint. https://files.osf.io/v1/resources/vefwc/providers/osfstorage/65b59ddd99d0100905626737?action=download&direct&version=3

Paulus, T. M., & Lester, J. N. (2023). Digital qualitative research workflows: A reflexivity framework for technological consequences. *International Journal of Social Research Methodology*, 1–14. https://doi.org/10.1080/13645579.2023.2237359

Paulus, T. M., & Marone, V. (2024). “In Minutes Instead of Weeks”: Discursive Constructions of Generative AI and Qualitative Data Analysis. *Qualitative Inquiry*, 10778004241250065. https://doi.org/10.1177/10778004241250065

Potter, W. (2024, July 23). An academic publisher has struck an AI data deal with Microsoft – without their authors’ knowledge. *The Conversation*. https://theconversation.com/an-academic-publisher-has-struck-an-ai-data-deal-with-microsoft-without-their-authors-knowledge-235203

Schreier, M. (2013). Qualitative Content Analysis. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis* (pp. 170–183). SAGE Publications Ltd.

Silver, C. (2023, May 5). What’s a foot in the Qualitative AI space? *Qualitative Data Analysis Services*. https://www.qdas.co.uk/post/what-s-a-foot-in-the-qualitative-ai-space

Silver, C. (2024, June 10). The risk of qualitative deepfake is a reality. *Qualitative Data Analysis Services*. https://www.qdas.co.uk/post/ai-deepfake-qualitative-analysis

Silver, C., & Woolf, N. (2020). Five-Level QDA Method. In P. Atkinson, A. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), *SAGE Research Methods Foundations*. SAGE Publications Ltd. https://doi.org/10.4135/9781526421036818833

Silver, C., & Woolf, N. H. (2015). From guided-instruction to facilitation of learning: The development of *Five-level QDA* as a CAQDAS pedagogy that explicates the practices of expert users. *International Journal of Social Research Methodology*, *18*(5), 527–543. https://doi.org/10.1080/13645579.2015.1062626

Wheeler, K. (2024). This Is Not a 'Drill': Young People's Understandings of and Hopes for Sustainability Education in England. *Sustainability*, *16*(17), 7614. https://doi.org/10.3390/su16177614

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2023). *Redefining Qualitative Analysis in the AI Era: Utilizing ChatGPT for Efficient Thematic Analysis* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2309.10771

Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). *Bias in Generative AI* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2403.02726