The Ethics of Using Generative AI for Qualitative Data Analysis

Authorship

The following authors contributed equally to this editorial:

Robert M Davison, Dept of Information Systems, City University of Hong Kong, Hong Kong, isrobert@cityu.edu.hk

Hameed Chughtai, Department of Organisation, Work, and Technology, Lancaster University Management School, UK, h.chughtai@lancaster.ac.uk

Petter Nielsen, HISP centre and Department of Informatics, University of Oslo, Norway, pnielsen@ifi.uio.no

Marco Marabelli, Computer Information Systems Department, Bentley University, USA, Mmarabelli@bentley.edu

Federico Iannacci. University of Sussex Business School, Falmer Brighton, UK, F.Iannacci@sussex.ac.uk

Marjolein van Offenbeek, Dept of Innovation Management & Strategy, University of Groningen, Netherlands, m.a.g.van.offenbeek@rug.nl

Monideepa Tarafdar, Isenberg School of Management, University of Massachusetts Amherst, USA, mtarafdar@umass.edu

Manuel Trenz, University of Goettingen, Germany, trenz@uni-goettingen.de

Angsana A. Techatassanasoontorn, Auckland University of Technology, New Zealand, angsana@aut.ac.nz

Antonio Díaz Andrade, University of Agder, Norway, antonio.diaz@uia.no

Background to This Editorial

It is important to note that the text of this editorial is entirely written by humans without any Generative Artificial Intelligence (GAI) contribution or assistance. The Editor of the ISJ (Robert Davison) was contacted by one of the ISJ's Associate Editors (AE) (Marjolein van Offenbeek) who explained that the qualitative data analysis software ATLAS.ti was offering a free-of-charge analysis of research data if the researcher shared the same data with ATLAS.ti for training purposes for their Generative AI (GAI[1]) analysis tool. Marjolein believed that this spawned an ethical dilemma. Robert forwarded Marjolein's email to the ISJ's Senior Editors (SEs) and Associate Editors (AEs) and invited their comments. Nine of the SEs and AEs replied with feedback. We (the eleven contributing authors) then engaged in a couple of rounds of brainstorming before amalgamating the text in a shared document. This was initially created by Hameed Chughtai, but then commented on and edited by all the members of the team. The final version constitutes the shared opinion of the eleven members of the team, after several rounds of discussion. It is important to emphasise that the eleven authors have contrasting views about whether GAI should be used in qualitative data analysis, but we have reached broad agreement about the ethical issues associated with this use of GAI. Although many other topics related to the use of GAI in research could be discussed, e.g., how GAI could be effectively used for qualitative analysis, we believe that ethical concerns overarch

[1] By GAI, we are primarily referring to the Large Language Model (LLM) Chatbots, such as ChatGPT, Bard and LLaMa.

many of these other topics. Thus, in this editorial we focus on the ethics associated with using GAI for qualitative data analysis.

INTRODUCTION

The emergence and ready availability of GAI has profound implications for research. This powerful technology, capable of generating human-like text, has the potential to create many opportunities for researchers in all disciplines. However, the technology brings ethical challenges and risks. We unearth and comment on many facets of qualitative data-related ethics. Our goal is to engage with and inform the many stakeholders of the ISJ, including other editors, (prospective) authors, reviewers and readers.

We intend that this discussion serves as a starting point for a broader conversation on how we can responsibly navigate the evolving landscape of GAI in research. It is important to point out that we are not advocating for or against the use of GAI in research, nor are we attempting to find ways to make it easier (or harder) for researchers to incorporate GAI in their research designs and practices. Our focus relates to the ethical issues associated with GAI use in analysing qualitative data that scholars in the conduct of academic research may encounter and should consider.

One of the allures of GAI lies in its capability to discover patterns to produce new codes in a data corpus faster and more comprehensively than humans, by drawing from its trained data. This capability implies that GAI may identify patterns missed by humans. However, speed and comprehensiveness do not necessarily translate to appropriateness, substantive helpfulness, or insightful understanding. More fundamentally, they should not be achieved at the cost of unethical research practices or of the commitment to “do no harm” to individuals, communities, organisations and society from research participation (Iphofen and Tolich, 2018). Thus, in our view, a utilitarian argument (the ends justify the means) constitutes an inadequate justification for GAI use for qualitative data analysis. Such a utilitarian argument would permit the unscrupulous or unprincipled researcher to use GAI, however they liked, in pursuit of a goal that might superficially benefit that individual researcher yet that would also violate the codes of ethics or behaviour that we esteem. Thus, the means that we employ to undertake research must be ethical, and must be seen to be ethical by our peers via the peer review system.

The use of GAI in research is not merely a question of tool selection, but also a matter that touches on the essence of research integrity, conduct and value. It challenges us to redefine what we consider as ‘doing research’ and pushes us to revisit how we maximise the benefits of research and minimise risks and harms for individuals and society (Gibbs, 2018). It also challenges us to consider our understanding and the implications of authorship, data ownership and rights, responsibility, privacy, and transparency. Thus, the overarching question that we ask is “What are the ethical issues that might surface when using GAI for the analysis of qualitative data”? (cf. UNESCO, 2021).

To address this question, we focus on five areas: 1) **data ownership and rights**; 2) **data privacy and transparency**; 3) **interpretive sufficiency**; 4) **researcher responsibilities and agency;** and 5) **biases manifested in the technology**. We foresee that this exploration would eventually enable us to inform the development of living guidelines for qualitative data analysis, pertaining to ISJ and the Information Systems field more generally, in the context of GAI. Our hope is that such living guidelines will align with broader discussions in scholarship and emerging AI policies around the world[2],[3]. In addition, this editorial reacts to GAI-related

[2] US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

[3] EU AI Act: first regulation on artificial intelligence https://artificialintelligenceact.eu/the-act/

policies that other journals have already set, for instance, the recent Academy of Management Review's editorial (Grimes et al. 2023), which are indicators of the siloed approach that not just academic fields but also individual journals are taking, for making sense of the use of GAI in scholarly settings. We instead aim at developing a fluid document that simply points to potential ethical implication of using GAI, and we do so by focusing specifically on qualitative data analysis.

DATA OWNERSHIP AND RIGHTS

We are concerned about surrendering research data to commercial entities, e.g., sharing data with a GAI tool in exchange for automated analysis because that could violate data rights and confidentiality. While automated data analysis is standard for quantitative data, using Language Learning Models such as ChatGPT for qualitative data is different as they require data for training their models. Qualitative research is typically an in-depth inquiry that uses "relatively unstructured forms of data, whether produced through observation, interviewing, and/or the analysis of documents" (Hammersley and Traianou, 2012, p.1). As such, the "production of such data can involve researchers in quite close, and sometimes long-term, relationships with people" (ibid.), suggesting that the participant should not simply be viewed as a data machine. Instead, participants should be highly valued as collaborators in our research endeavours (Oakley, 2013). Therefore, researchers can be considered as violating the commitment of non-maleficence given to our participants by volunteering their data to train Language Learning Models. This would happen mainly because we cannot guarantee that robust precautions are in place to avoid possible harm stemming from sharing their data with these models.

A related concern stems from providing research data to a profit-driven entity to enhance the quality of its product. We acknowledge that research organisations and universities may have internal GAI platforms for such purposes. However, unless explicitly included in Institutional Review Board (IRB) applications, providing research data to GAI platforms, whether hosted by internal or profit-driven entities, could conflict with data privacy protection laws and most IRB approvals[4]. It is suggested that IRBs may soon start considering the use of GAI in their approval processes, i.e. they will (and probably should) become more sensitive to the application of GAI by researchers, and may even proscribe or restrict such application.

DATA PRIVACY AND TRANSPARENCY

One of the ethical principles agreed by the UK Academy of Social Sciences in 2015 is "[a]ll social science should respect the privacy, autonomy, diversity, values, and dignity of individuals, groups, and communities" (UK Academy of Social Sciences, 2015). Therefore, the application of GAI raises privacy concerns, particularly when sensitive research data is shared with an AI tool. When a research study involves organisations, privacy issues surface when they require the signing of Non-Disclosure Agreements (NDAs) as a condition for data collection but are not made aware that the researchers may share the same data with an AI tool and its owners (Pearlman, 2017). Researchers will need to ensure that the involved organisations affirm their consent to this data exchange, if they plan to use a GAI analysis tool that uses the data they collected for further training. In addition, individual data is also subject to privacy protection. NDAs are negotiated between the researcher and the organisation's leadership or their legal office. However, the data collected relates to employees, clients and potentially other stakeholders, who might not be aware that their conversations with researchers will be handed over to third parties that do not fully disclose to which ends and

[4] An IRB, "also known as an independent ethics committee, ethical review board or research ethics board, is a committee at an institution that applies research ethics by reviewing the methods proposed for research involving human subjects to ensure that the projects are ethical" [https://en.wikipedia.org/wiki/Institutional_review_board].

how they will use these data and whether they will operate appropriate, and for future re-combinations, sufficient deidentification practices. Thus, the researcher must arguably also obtain consent from each employee and any other stakeholder, to transfer their data to each and every specific third party, such as a qualitative GAI-powered software. Participants who decline to provide such consent, or who simply fail to respond to the request for consent, must have their data excluded from what is shared. Steps to ensure data privacy of individuals must be included in the IRB approval documents. In the European Union and the US, for instance, human subjects participating in research must consent to how data about them is processed and have the right to withdraw any consent previously given with the consequence that this person's personal data must be deleted.

INTERPRETIVE SUFFICIENCY

We highlight two intertwined concerns about interpretive sufficiency in qualitative analysis. Software for qualitative data analysis has been in use since at least the early 1980s[5]. For instance, both ATLAS.ti and NVivo (among others) allow automated coding by structure, style or existing coding patterns. However, the incorporation of GAI into this process has new implications because it allows coding 'from scratch' based on the external datasets with which the GAI was trained. There is an inherent value associated with manual (i.e. not software assisted) data analysis, especially in qualitative interpretive and ethnographic research (Davison and Ravishankar, 2024). Analysis of qualitative data often relies on the researcher's creative and conceptual ability to discern meaning, salience and interconnectedness of logic in emerging themes (Amis and Silk, 2008). GAI tools, while powerful and efficient, only have access to the text. As of today, they cannot capture the nuances of the research environment, body language, facial expressions, tone of voice, interactions between the researcher and the research subjects, and the researcher's own accumulated understanding of the domain. Moreover, automated systems detecting these "soft" characteristics of human interactions are highly questioned because of their poor reliability and potential for discriminations (Crawford 2021). Relying on such tools thus constitutes both an abdication of our responsibilities as researchers and a voluntary diminishment in our own agency and human consciousness required in the construction of knowledge (Amis and Silk, 2008).

A deeper and more contextually nuanced analysis of qualitative data is important because knowledge encompasses syntactic, semantic, and pragmatic layers (Mingers, 2008). Automated qualitative coding can only examine syntax, but cannot genuinely grasp data's semantic and pragmatic aspects. As researchers, we do not claim to be neutral or value-free in our analyses, and indeed being value-free may be inappropriate; for instance, with respect to critical interpretive studies, an automated coding process could lead to a banal and neutral analysis that fails to identify or disclose hidden aspects in the qualitative data. The output analysis will then incorporate an incomplete and potentially superficial reading of the data. Further, mainstream (or neutral) chunking and coding could influence and limit our potential learning from the data analysis. For instance, it is important that the researcher be aware of the risks associated with introducing and reinforcing existing biases, especially in research on marginalisation, oppression, activism, conflicts, and decolonisation. In addition, in an investigation of socially relevant problems stemming from digital technology, the researcher bounded by personal and community responsibility joins forces with the researched to produce understanding that empowers those disadvantaged by the technology (Amis and Silk, 2008). These are areas where human insight, empathy, and understanding (verstehen) are crucial. For these reasons, the use of GAI as a data analysis assistant becomes ethically questionable because it will necessarily exclude some aspects that are central to the analysis and that would be central if it was undertaken by humans.

[5] https://en.wikipedia.org/wiki/Computer-assisted_qualitative_data_analysis_software

RESEARCHER RESPONSIBILITIES AND AGENCY

The emergent and interactional nature of most qualitative research requires more scrutiny on the researcher and their conduct (Iphofen and Tolich, 2018). Drawing on the discussion on data ethics in the age of algorithms, some authors argue that "the gradual reduction of human involvement or even oversight over many automatic processes, pose pressing issues of fairness, responsibility and respect of human rights, among others" (Floridi and Taddeo, 2016, p. 2). In addition to the researcher's responsibilities and obligations to participants, the researcher also has to take epistemic responsibility, which involves being accountable to the evidence where evidence is relationally constituted between the researcher and the researched and to assume responsibility for what the researcher claims to know (Code, 2001). Some researchers may argue that GAI could help them identify preliminary patterns from large datasets, providing them with initial insights. However, it is recognised that GAI is not infallible. For instance, it is prone to what are known as 'hallucinations', where it 'lies' and 'fabricates facts' (Ji et al. 2023). Thus, the veracity of any 'preliminary patterns' identified through technology (such as GAI) must be checked by the researcher who must both claim authorship of them and so take responsibility for the text; accountability remains with the researcher (Gregor, 2024). GAI cannot be listed as a co-author (this is publisher policy at ISJ), and thus cannot be permitted to have any agency in the research or its outputs. To wit, we consider blind, automated applications of GAI for data analysis without human agency unethical in every aspect of the research process.

BIASES MANIFESTED IN GAI

OpenAI's ChatGPT[6] service openly acknowledges that their results are "not free from biases and stereotypes", are "skewed towards Western views" and can "reinforce a user's biases". As an example, ATLAS.ti (qualitative analysis software) uses OpenAI's GPT model and clearly acknowledges that their results may encode "social biases, e.g. via stereotypes or negative sentiment towards certain groups"[7]. Thus, GAI may produce analyses that are biased, unjust or discriminatory to certain groups or individuals, based on the data they are trained on or the criteria they use to analyse qualitative data. GAI may generate patterns (text, images, and so on) that reinforce stereotypes, biases, or prejudices against people of different race, gender, culture, or background. GAI may not account for the different needs, preferences, and values of diverse stakeholders and communities, and may impose a dominant or hegemonic perspective on the data analysis process. GAI-based analysis can also perpetuate or exacerbate the colonisation or marginalisation of other modes of knowledge, cultures, or values, by privileging a certain perspective on the data analysis process, for instance, one reflecting Western cultures, because of training data prevalently collected online (Bender and Friedman 2018). GAI may rely on data sources, methods, or frameworks that are derived from or influenced by colonial or imperial histories, ideologies, or power structures. GAI may not acknowledge or address any of these ethical, social, or political implications of its data coding.

In the paragraphs above, we have noted the difficulties associated with developing a fair and objective analysis with GAI. As a result, an interpretation developed partially or fully through GAI-based data analysis may be difficult to critically explain, as the algorithms and models that underlie the data analysis process may be complex, opaque, or black-boxed. For example, GAI tools often use a combination of neural networks, genetic algorithms, and machine learning techniques that are not easily interpretable or transparent to human users or researchers. GAI also does not provide clear and coherent rationales for the outputs it produces and may not allow for feedback or correction.

[6] OpenAI ChatGPT, Is ChatGPT Biased? https://help.openai.com/en/articles/8313359-is-chatgpt-biased
[7] ATLAS.ti AI Coding power by OpenAI https://atlasti.com/ai-coding-powered-by-openai

Given all these characteristics of GAI, we suggest that researchers should engage in critical reflexivity and vigilance to identify, understand and robustly address the ethical issues regarding the use of GAI in their research practices involving qualitative data analysis. We do not wish to see a situation where we are lulled into thinking that GAI use is 'normal' and that researchers do not need either to pay particular attention to it, or to report their use of it.

TOWARDS LIVING GUIDELINES AND QUALITATIVE ANALYSIS

We acknowledge that some ethical issues are specific to particular GAI implementations, which change over time, emphasising the need for clear quality criteria. GAI implementations could also be private. For example, some research organisations have established their own GAI service, enabling students and researchers to use OpenAI's GPT models within university and national data privacy requirements.[8] In this view, the ethical concerns, such as privacy, are specific to implementations of GAI and are not necessarily general issues with the technology class. However, private Language Learning Models are not necessarily expected to improve the quality of coding; they might still be too generic to address specific research questions. While they can, to a certain extent, address privacy issues, they cannot unequivocally improve analysis quality and their biases may still be present.

Existing guidelines both at the general level (Susarla et al., 2023) and more specifically suggesting that qualitative coding and data analysis could benefit from GAI[9] are ambiguous. This ambiguity could lead to misusing these tools or using them unethically, resulting in flawed outcomes, which can compromise research quality and the creation of cumulative knowledge, let alone the fact that research outcomes might be seen by practitioners with scepticism. To preserve research integrity, some research outlets, such as Nature, are disallowing the use of GAI for specific purposes. Wiley, the publisher of the ISJ, has formulated its own guidelines as follows[10]:

> Artificial Intelligence Generated Content (AIGC) tools, such as ChatGPT and others based on large language models (LLMs), cannot be considered capable of initiating an original piece of research without direction by human authors. They also cannot be accountable for a published work or for research design, which is a generally held requirement of authorship (as discussed in the previous section), nor do they have legal standing or the ability to hold or assign copyright. Therefore, in accordance with COPE's position statement on AI tools[11], these tools cannot fulfil the role of, nor be listed as, an author of an article. If an author has used this kind of tool to develop any portion of a manuscript, its use must be described, transparently and in detail, in the Methods or Acknowledgements section. The author is fully responsible for the accuracy of any information provided by the tool and for correctly referencing any supporting work on which that information depends. Tools that are used to improve spelling, grammar, and general editing are not included in the scope of these guidelines. The final decision about whether use of an AIGC tool is appropriate or permissible in the circumstances of a submitted manuscript or a published article lies with the journal's editor or other party responsible for the publication's editorial policy.

That we should resist developing fixed guidelines with a 'one size fits all' character does not mean that for the time being 'anything goes'. Instead, we are suggesting building an ever

[8] For example, University of Oslo, Norway https://www.uio.no/english/services/it/gpt-uio/index.html
[9] https://indeemo.com/blog/generative-ai/qualitative-data-analysis-tool
[10] https://authorservices.wiley.com/ethics-guidelines/index.html#5
[11] https://publicationethics.org/cope-position-statements/ai-author

evolving ‘living guideline’ (Bockting et al., 2023). As GAI’s role in research continues to evolve, such guidelines are crucial to help qualitative researchers navigate this changing landscape. With this editorial, we open a space for dialogue, debate and continuing progress, in formulating ethical principles and guidelines that relate to whether and when GAI is used for qualitative data analyses in papers published in the ISJ. We suggest that the unchecked use of GAI could result in flawed and biased analyses. We emphasise the need for specific but constantly evolving guidance on using GAI tools for qualitative data analysis.

REFERENCES

Amis, J.M. & Silk, M.L. (2008). The philosophy and politics of quality in qualitative organizational Research. *Organizational Research Methods,* **11**(3), 456-480.

Bender, E. M., and Friedman, B. 2018. "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science," *Transactions of the Association for Computational Linguistics* (**6**), pp. 587-604.

Bockting, C. L., van Dis, E. A., van Rooij, R., Zuidema, W., & Bollen, J. (2023). Living guidelines for generative AI: Why scientists must oversee its use. *Nature,* **622**(7984), 693-696.

Code, L. (2001). Rational Imaginings, Responsible Knowings: How Far Can You See From Here? In: Tuana, N. and Morgen, S. (eds) *Engendering Rationalities.* Albany, NY: State University of New York, 261–282.

Crawford, K. (2021). Time to regulate AI that interprets human emotions, *Nature* **592**, 8, 167.

Davison, R.M. & Ravishankar, M.N. (2024). Theory-driven Perspectives on Generative Artificial Intelligence in Business and Management, *British Journal of Management*

Floridi, L., & Taddeo, M. (2016). What is data ethics? Philosophical Transactions A, 374: 20160360. http://dx.doi.org/10.1098/rsta.2016.0360

Gibbs, G.R. (2018). *Analyzing qualitative data.* Sage. https://doi.org/10.4135/9781526441867

Gregor, S. (2024). Responsible artificial intelligence and journal publishing. *Journal of the Association for Information Systems, 25*(1), 48-60. DOI: 10.17705/1jais.00863

Grimes, M., von Krogh, G., Feuerriegel, S., Rink, F., and Gruber, M. 2023. "From Scarcity to Abundance: Scholars and Scholarship in an Age of Generative Artificial Intelligence," *Academy of Management Journal* (66:6), pp. 1617-1624.

Hammersley, M. & Traianou, A. (2012). *Ethics in Qualitative Research: Controversies and Contexts.* London: Sage.

Iphofen, R. & Tolich, M. (2018). Foundational issues in qualitative research ethics. In: Iphofen & Tolich (Eds) *The Sage Handbook of Qualitative Research Ethics.* Sage. https://doi.org/10.4135/9781526435446.

Ji. Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A. & Fung, P. (2023). Survey of hallucination in natural language generation, *ACM Computing Surveys* **55**(12), 1-38.

Mingers, J. (2008). Management knowledge and knowledge management: realism and forms of truth. *Knowledge Management Research & Practice*, *6*, 62-76.

Oakley, A. (2013). ‘Interviewing women: a contradiction in terms'. In: Roberts, H. (ed.) *Doing Feminist Research* (pp. 30–61). London: Routledge.

Pearlman, R. (2017). Recognizing artificial intelligence (AI) as authors and investors under US intellectual property law. *Richmond Journal of Law & Technology*, **24**, 1-38.

Susarla, A., Gopal, R., Thatcher, J.B. and Sarker, S. (2023). The Janus effect of generative AI: Charting the path for responsible conduct of scholarly activities in information systems, *Information Systems Research* **34**(2), 399-408.

UK Academy of Social Sciences (2015). "Developing generic ethics principles for social science research" https://acss.org.uk/wp-content/uploads//Generic-Ethics-Principles-in-Social-Science-Research-2013.pdf.

UNESCO (2021). Recommendation on the ethics of artificial intelligence, https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.