ISSN: 2951-6641 (paper) 2951-8156 (online)
https://doi.org/10.54790/rccs.176

DEBATE/DEBATE: BEYOND BIG DATA: GENERATIVE AI AND LLMS AS NEW DIGITAL TECHNOLOGIES FOR ANALYSING SOCIAL REALITY/
MÁS ALLÁ DEL *BIG DATA*: IA GENERATIVA Y LLMS COMO NUEVAS TECNOLOGÍAS DIGITALES PARA EL ANÁLISIS DE LA REALIDAD SOCIAL

LLMs and Coding in Qualitative Research: Advancements and Opportunities for Social Verbatim as an Integral Qualitative Tool

Los LLM y la codificación en la investigación cualitativa: avances y oportunidades para Social Verbatim como herramienta integral cualitativa

Juan Miguel Gómez Espino
Universidad Pablo de Olavide (UPO), Spain
jmgomesp@upo.es

Received/Recibido: 9-6-2025
Accepted/Aceptado: 14-10-2025

ABSTRACT

This article examines the use of large language models (LLMs) in qualitative coding, highlighting key advances and emerging opportunities for the Social Verbatim tool. It reviews the theoretical foundations of LLMs, their underlying architecture and the role of hardware developments in their evolution. The article explores specific applications of LLMs in qualitative research, including thematic coding and comparative analysis. It also addresses methodological, ethical and epistemological challenges, proposing strategies to mitigate these risks. Finally, it considers the implications of integrating LLMs into platforms such as Social Verbatim, underscoring the importance of transparency and human–machine collaboration in the context of qualitative inquiry.

KEYWORDS: Large language models (LLMs); qualitative coding; generative artificial intelligence (GenAI); qualitative research; open science; AI-assisted qualitative analysis.

HOW TO REFERENCE: Gómez Espino, J. M. (2026). Los LLM y la codificación en la investigación cualitativa: avances y oportunidades para Social Verbatim como herramienta integral cualitativa. *Revista Centra de Ciencias Sociales*, *5*(1), 195-218. https://doi.org/10.54790/rccs.176

Spanish version can be read at https://doi.org/10.54790/rccs.176

RESUMEN

Este artículo explora el uso de Modelos de Lenguaje de Gran Escala (LLM) en la codificación cualitativa, destacando avances y oportunidades para la herramienta Social Verbatim. Se revisan los fundamentos de los LLM, su arquitectura y el impacto del hardware en su desarrollo. Además, se analizan aplicaciones específicas de los LLM en la investigación cualitativa, incluyendo la codificación temática y el análisis comparativo. Se abordan los desafíos metodológicos, éticos y epistemológicos, y se proponen estrategias para mitigar estos problemas. Finalmente, se discuten las implicaciones de la integración de LLM en herramientas como Social Verbatim, subrayando la importancia de la transparencia y la colaboración humano-máquina en la investigación cualitativa.

PALABRAS CLAVE: modelos de Lenguaje de Gran Escala (LLM); codificación cualitativa; Inteligencia Artificial Generativa (IAG); investigación cualitativa; ciencia abierta; análisis cualitativo.

1. Introduction

Over two years of work on a project known as CS-Transcribe,[1] one of the principal outcomes has been the development of an online tool registered as Social Verbatim. This application has been conceived as a digital solution for various stages of the qualitative social research process, and it is progressing towards becoming a comprehensive support tool for this type of research. In addition to transcription functionalities, it incorporates features related to data management, and to analysis and coding support. As detailed on its website (www.socialverbatim.com[2]), Social Verbatim not only allows automatic or manual transcription but also the integration of non-verbal and contextual communication through icons, the review and correction of transcripts, team collaboration, project and interview organisation, the structuring of focus groups or other analytical elements, anonymisation of excerpts and the insertion of comments, bookmarks and analytical notes. It also enables the use of verbatim transcripts for publication purposes, among other functionalities.[3]

In line with efforts to expand the tool's capabilities, this article explores the potential application of large language models (LLMs) in the qualitative coding process. This work aims to assess the recent advances in AI-driven qualitative coding, as well as the challenges and opportunities posed by this type of technology. Special attention is given to the role of researchers as active and necessarily reflexive agents within the process.

Before reviewing the most significant contributions in this field, the concept of LLMs is first introduced. LLMs are artificial intelligence models designed to process and generate natural language text on a large scale. According to Mitchell (2024):

> A large language model (LLM) is a computational system, typically a deep neural network with a large number of tunable parameters [...], that implements a mathematical function called a language model. [...] The neural networks underlying LLMs are trained using broad collections of text typically obtained from websites, digitized books, and other digital resources.

These neural networks constitute a computational model inspired by the human brain, comprising "neurons" (processing units) arranged in layers, which transform inputs (such as text or numbers) into outputs (such as predictions or responses).

In recent years, this technology has seen extraordinary advances, particularly following the introduction of the Transformer architecture, as proposed in a seminal paper by Vaswani *et al.* (2017). This architecture is capable of capturing long-range dependencies in text far more efficiently than earlier models such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs). Prior to this innovation, language models were trained from scratch for each specific task. By contrast, Transformer models are pre-trained on large volumes of unlabelled data and subsequently fine-tuned for particular tasks.[4]

This new architecture is based on the self-attention mechanism (*Attention Is All You Need*), which enables the efficient processing of large volumes of text and the identification of relational patterns by assigning different weights to words within a sentence. LLMs convert words into numerical representations known as embeddings. These representations allow the model to associate similar concepts within a mathematical space, comparing each word with others to determine its relevance within a given sentence. The model assigns different weights to each word in order to better understand the overall meaning of the prompt – that is, the textual input or instruction provided by the user to elicit a response. Moreover, the Transformer architecture enables the simultaneous processing of all words through *parallelisation*. This innovation made it possible to train models containing billions of parameters without excessive increases in training time, thereby significantly enhancing scalability.

In addition to the development of the Transformer architecture, progress in hardware has played a critical role in the advancement of LLMs, particularly in three areas: a) the development of graphics processing units (GPUs), which accelerate the matrix and tensor computations essential in Transformer models; b) increases in RAM, since larger models require terabytes of memory to process data; and c) the emergence of AI-specific chips – such as tensor processing units (TPUs). Wang *et al.* (2019), for instance, demonstrated that TPUs offer significant advantages in terms of performance and energy efficiency compared to traditional GPUs, particularly for deep learning models such as Transformers. Without high-performance hardware, training times would be prohibitive, and real-time model deployment would be infeasible.

2. Generative Artificial Intelligence (GenAI) and LLMs in Qualitative Coding

All these advances have facilitated the introduction of methodological innovations that are capable of transforming – and indeed are already transforming – the ways in which social scientists engage with qualitative data (Hayes, 2025; Van Dis *et al.*, 2023). For Hayes (2025), this entails inhabiting a new "hybrid space" characterised by dynamic interaction with large-scale data and a conversational engagement with it through LLMs – a new relational model that lies somewhere between established qualitative traditions and the possibilities afforded by advanced computational capacities.

In any case, there is little doubt about the far-reaching impact this technology will have on how we conceptualise our world – and science itself. Nonetheless, in disciplines such as sociology, political science, economics and other social sciences, as Bail notes, the transformative potential of GenAI in research remains largely underexplored, despite the fact that "these tools may advance the scale, scope, and speed of social science research—and may enable new forms of scientific inquiry as well" (2024, p. 1).

The contributions of GenAI are already being extensively investigated in both *experimental or quasi-experimental* settings and, naturally, within qualitative research. Ziems *et al.* (2024) evaluated 13 LLMs and found acceptable levels of agreement with human coders. They concluded that, in contrast to supervised and manual text coding – which requires large volumes of human-annotated training data – LLMs present substantial opportunities, without the limitations of other unsupervised methods that often yield unintelligible results. Wu *et al.* (2023) analysed public statements made by elected officials and demonstrated that ChatGPT-3.5 can produce ideological classifications, with results closely aligning with the widely used DW-NOMINATE method for measuring ideology.

Hayes (2025) identifies several possible uses of LLMs for qualitative research, in addition to *basic orientation* within extensive and complex datasets. These include: a) *thematic coding*; b) *comparative analysis* across different texts, by highlighting differences in tone, emphasis or conceptual framing; c) *identifying internal dynamics within the data*, such as contradictions, tensions or evolving narratives within the corpus; d) *scenario testing and hypothetical exercises*; e) creative synthesis and stimulation of further inquiry; f) *reflexive engagement*; and g) less conventional uses of LLMs, such as the generation of storyboards, instructions or descriptive outlines.

The discussion below will focus on the first three of these, as they are embedded in the core nature of the coding process. To this end, we will first refer to the process in its more conventional sense. In the context of qualitative research, coding has occupied a central role, serving as a bridge between raw data and the construction of analytical meaning.

A code is thus a construct generated by the researcher that symbolises – and thereby assigns interpreted meaning to – individual data elements, for the purposes of pattern detection, categorisation, theoretical development and other analytical processes (Miles *et al.*, 2015, p. 78). Coding, in this sense, is the systematic process by which qualitative data (such as interviews, observations or texts) are organised, labelled and grouped in order to identify relevant patterns, themes or categories. Rather than conceiving of coding merely as a technical or preparatory task, there is considerable consensus around its character as a process of deep reflection – one involving the analysis and interpretation of data meaning (González-Veja, 2022; Deterding and Waters, 2021). Codes are used primarily – though not exclusively – to retrieve and categorise similar fragments of data, enabling the identification, extraction and clustering of segments relevant to a research question, hypothesis, construct or specific theme.

Coding, therefore, is a heuristic process that assists the researcher in exploring, discovering and understanding underlying patterns and themes within a dataset. In other words, coding not only structures information but also activates a process of reflection and analysis that leads to new interpretations or findings, functioning as a guide or discovery strategy in qualitative analysis.

Two types of coding may be distinguished: inductive (Glasser and Strauss, 1967) and deductive (Crabtree and Miller, 1999). The inductive approach consists of constructing patterns and themes from the bottom up, organising data into increasingly abstract units of information. By contrast, within a deductive logic, existing patterns and theories are compared with the data (Jiang *et al.*, 2021, p. 94). Although some approaches recommend avoiding prior conceptual frameworks when engaging with data, this does not appear to be a realistic proposition. In practice, a combination of both methods almost invariably occurs (Lindbergh and Korsgaard, 2019), which is plausible given the limitations of each approach when applied in isolation.

Social scientists have begun to use LLMs for text classification and, within this group, researchers in sociology in particular. LLMs can assist them in moving rapidly from an overarching view of thematic patterns to more specific aspects of human communication (Hays, 2025). Overall, there is a high degree of consensus that these models can be highly useful for data coding in qualitative research, although there are clear warnings regarding the importance of using them judiciously and with an awareness of their limitations, which will be addressed later.

LLMs perform natural language processing tasks, that is, without the need for prior task-specific training data. Unlike other GenAI models such as supervised machine learning, in which pre-labelled training data are provided (Molina and Garip, 2019), LLMs operate in a zero-shot mode – that is, without prior training for a specific task (Ziems *et al.*, 2024). In the former case, a "label" or category is assigned to each document (for example, an annotated tweet, a paragraph from a news article or a fragment of a speech), and a model is then trained to automatically predict labels using textual features. Once trained, the model can predict labels for other similar texts, thereby automatically coding new documents.

In any case, LLMs are more than a promise. Indeed, they represent a tangible reality that is contributing to a renewed use of computational techniques in qualitative research by providing: a) *efficiency*, as they help to accelerate the coding process, particularly when working with large datasets; b) *consistency*, insofar as they can ensure uniform coding criteria, thereby reducing human bias and error; and c) *pattern analysis*, by identifying relationships within the data that may be difficult or impossible to detect through manual analysis.

3. LLM-Based Tools Used for Coding in Qualitative Analysis

Table 1 below presents several recent examples of studies (*first* column) that examine the use of LLMs as a means of qualitative coding. The table includes the type of coding applied – whether inductive or deductive (*second* column); the LLM configuration, specifically whether fine-tuning was employed, starting from a zero-shot approach (*third* column); and whether the study involved comparisons between different LLMs or between LLMs and human coders, typically experts (*fourth* column). As previously noted, prompt engineering refines the outcomes of coding by designing and optimising prompts to elicit more accurate, relevant and useful responses from language models. The table also provides information on whether a specific tool or methodology was developed (*fifth* column) and the specific model used for coding (*sixth* column).

Table 1

Use of LLMs in qualitative coding by coding approach, comparison method, tools and model applied

Source	Type of coding	LLM configuration	Comparison	Tool (app) or methodology	LLM used
Chew *et al.* (2023)	Deductive	Zero-shot (fine-tuning via prompt engineering)	With human coders	Methodology. LACA (LLM-Assisted Content Analysis). Code available on Figshare: https://figshare.com/articles/dataset/LLM-Assisted_Content_Analysis_LACA_Coded_data_and_model_reasons/23291147	GPT-3.5
Ziems *et al.* (2024)	Inductive	*Zero-shot*	Between models	Not specified	FLAN -5 (Small, Base, Large, XL, XXL), FLAN UL-2, GPT (3.5, 4), ada-001, babbage-001, curie-001, davinci-001, 002, 003

Source	Type of coding	LLM configuration	Comparison	Tool (app) or methodology	LLM used
Meng *et al*. (2024)	Deductive/ inductive	*Zero-shot*	Comparison between model-only and human-assisted outputs (in both coding approaches)	Methodology. CHALET (Collaborative Human–LLM Analysis for Empowering Conceptualisation in Qualitative Research). No dedicated software tool	GPT-4-1106-preview
Dunivin (2024)	Deductive / content analysis	*Zero-shot*	Between models / with human coders	Not specified	GPT-3.5 and 4
Xiao *et al*. (2023)	Deductive	Zero-shot (fine-tuning via prompt engineering)	With human coders	Not specified	GPT-3
Zhang *et al*. (2024)	Deductive/ inductive	*Zero-shot*	With humans / between models	Software. QualiGPT (open-access): https://github.com/KindOPSTAR/QualiGPT	GPT-4 and Claude 3.5
Zhao *et al*. (2024)	Inductive	Zero-shot (fine-tuning via prompt engineering)	Between models with and without fine-tuning	Software. A2C (Argument2Code) – proprietary software designed to enhance qualitative data analysis using LLM capabilities (not open-source)	Llama-2-13B-Chat
Tai *et al*. (2024)	Deductive	Zero-shot (iteration effects following prompts)	Comparison of outputs after 160 prompt-based iterations using the same input data	Not specified	GPT-3.5
Arlinghaus *et al*. (2024)	Inductive	*Zero-shot*	With humans / between models	Not specified	GPT-3.5 Turbo and GPT-4o
Dai *et al*. (2023)	Deductive/ inductive (thematic analysis)	Prompt-based frame discussions	With human coders	Methodology. Code available at https://github.com/sjdai/LLM-thematic-analysis	GPT-3.5
Qiao *et al*. (2025)	Inductive	Multi-agent LLM system (coders, aggregators and reviewers)	Comparison between single-agent and multi-agent models	Software. Thematic-LM (open-source): https://github.com/sjdai/LLM-thematic-analysis	GPT-4
Gao *et al*. (2025)	Inductive	Descriptive overview of the MindCoder tool	Not specified	Software. MindCoder. Available at https://mindcoder.ai Designed to bridge the gap between professional qualitative tools (e.g. ATLAS.ti, NVivo) and conversational LLMs (e.g. Claude, ChatGPT). Proprietary app (code not available)	GPT-4
Bryda *et al*. (2024)	Inductive	*Zero-shot*	Describes two coding strategies: generative and lexico-semantic	Not specified	GPT-4
Yang *et al*. (2025)	Inductive	Zero-shot (fine-tuning via prompt engineering)	With human coders	Not specified	GPT-4
Mathis *et al*. (2024)	Inductive	Zero-shot (fine-tuning via prompt engineering)	With human coders	Methodology. Code available at https://www.sciencedirect.com/science/article/pii/S0169260724003493?via%3Dihub	Llama 2-70B-Instruct (open-access)

Source: own research.

Given the applied aims of this article, the focus will be placed on specific software tools designed for qualitative coding (as shown in the penultimate column of the previous table), through which LLMs are implemented (as indicated in the final column). Accordingly, tools such as LACA (Chew *et al.*, 2023) and CHALET (Meng *et al.*, 2024) are excluded from this overview (Table 2), as they represent methodological frameworks that integrate LLMs – such as ChatGPT – into the qualitative coding process, whether deductive (LACA) or a combination of deductive and inductive (CHALET).

QualiGPT (Zhang *et al.*, 2024) is a tool based on language models (e.g. ChatGPT), designed to support qualitative data analysis. Although QualiGPT does not operate as a standalone programme – instead using the ChatGPT interface or, alternatively, allowing local installation in a Python environment via GitHub – it offers a customisable approach using language models (such as ChatGPT) tailored specifically to qualitative analysis. It is grounded in methodologies like inductive and deductive coding and draws on core principles of qualitative research, including grounded theory, thematic analysis and reflexive coding. This tool is oriented around principles of *transparency and reflexivity*, offering coding justifications, analytical commentary and decision traceability. Other notable features include its speed and responsiveness – with ChatGPT able to generate codes and themes within seconds or minutes – as well as the availability of automated workflows, which reduce the need for manual configuration. It also includes prompt templates inspired by peer-reviewed research.

MindCoder (Gao *et al.*, 2025) is a web-based application specifically developed for qualitative analysis. Its principal aim is to automate and streamline the qualitative coding process, providing an accessible tool for researchers without programming expertise. Through its intuitive and user-friendly online interface, MindCoder seeks to bridge the gap between professional AI-powered software tools (e.g. ATLAS.ti, NVivo) and conversational language models (like Claude and ChatGPT). It employs automated chains of reasoning based on the chain-of-thought (CoT) prompting technique, which enables structured qualitative analysis across stages such as data reprocessing, automatic open coding, automatic axial coding, conceptual development and report generation.

Table 2

Emerging software tools using LLMs for qualitative coding, according to selected criteria

Software tool	Input formats	Outputs	Interface	Open source	Web source
QualiGPT (Zhang *et al.*, 2024)	.csv, Excel, plain text	Tables, summaries, code lists	ChatGPT interface (OpenAI) or local Python installation	Yes (MIT License)	chatgpt.com/g/g-HtBvl9uXe-qualigpt github.com/KindOPSTAR/QualiGPT
MindCoder (Gao *et al.*, 2025)	.txt, .docx	Cluster diagrams, code labels, conceptual development, visual representations	Proprietary online platform	No (propri-etary)	mindcoder.ai
Thematic-LM (Qiao *et al.*, 2025)	.csv, .json (with agent configura-tion options)	Thematic codes, codebooks, assignment to text segments, the-matic maps, structured reports	Executed in programming environments such as Jupyter Notebooks or directly in Python	Yes	github.com/sjdai/LLM-thematic-anal-ysis
Argument-2Code (Zhao *et al.*, 2024)	Not specified	Not specified	No dedicated interface – integrated into existing analytical workflows	No	Not specified

Source: own research.

Thematic-LM (Qiao *et al.*, 2025) is a computational thematic analysis system designed to perform thematic coding by assigning specialised tasks to individual agents (components of the system), such as coding, code aggregation, and the maintenance and updating of the codebook. This architecture enables a more efficient and scalable approach, capable of handling large volumes of data without compromising performance. It is intended for researchers with programming expertise and access to LLM APIs, as users are required to execute scripts and manually configure system parameters. This involves: a) preprocessing the data; b) defining how agents are invoked (e.g. coder, aggregator); and c) specifying how results are stored and visualised.

Argument2Code (Zhao *et al.*, 2024) is a sophisticated automated system developed to generate inductive codebooks and extract emerging themes without the need for a predefined theoretical framework. It employs a multi-stage process based on chain-of-thought prompting – a technique that guides the model through a series of logical steps to improve coherence and depth in code generation. This approach supports a more open and flexible exploration of the data, enabling the identification of patterns and key concepts directly from the analysed content.

In summary, the use of LLMs in qualitative coding has led to the emergence of a diverse range of approaches and tools that introduce automation – and, consequently, an unprecedented level of speed – into the qualitative analysis process. In particular, software tools such as QualiGPT, MindCoder and Argument2Code illustrate a growing effort to integrate the advanced capabilities of LLMs into environments that are both accessible and methodologically grounded.

Nevertheless, other tools – such as Thematic-LM – present greater technical complexity in their implementation, despite being open-source. In some cases, including Argument2Code, the available information remains limited, as no web or desktop interface is provided and the source code is not publicly accessible. Ultimately, these tools point to a significant transformation in qualitative coding practices, offering new opportunities while also raising challenges related to transparency, human oversight and critical interpretation – issues that will be addressed in the following section.

4. The Challenges of LLMs

The use of LLMs in qualitative research has attracted increasing interest, while simultaneously posing important methodological, ethical and epistemological challenges. As these tools are progressively integrated into data analysis processes, it becomes essential to critically reflect on their limitations – particularly with regard to the transparency of their operations, the reliability of their outputs and the potential influence of biases inherited from their training datasets. Several authors have warned that, although GenAI may offer innovative solutions for task automation and the identification of patterns in large-scale datasets, its use requires caution, both due to the risk of reproducing social and cultural biases, and because of ethical dilemmas related to data privacy and the replicability of research findings. What follows is a summary of some of these limitations, as highlighted in recent studies.

Morgan (2023) identifies several key concerns regarding the application of GenAI to qualitative analysis. The first relates to racist, sexist or other forms of bias that may arise due to limitations in the training datasets, which are typically sourced from internet content (itself replete with various forms of structural bias). For instance, the developers of ChatGPT have reported substantial efforts to train the model to detect and exclude such biases, both in the queries it accepts and in the responses it produces. In theory, these risks could be mitigated through careful prompt formulation. However, Morgan also notes that this bias-filtering capability may itself become problematic when the objective of the research is precisely to examine such biases – a frequent aim in qualitative inquiry.

A second limitation concerns the model's potential to generate inaccurate or nonsensical content, a phenomenon often referred to as "hallucination" (Lakshmanan, 2022). 0As the tool itself acknowledges when questioned about its reliance on probabilistic language prediction: "the model generates responses based on the probability that a sequence of words is coherent and relevant, but not necessarily correct, as this probabilistic approach prioritises fluency and coherence over factual accuracy" (ChatGPT, v4, 2024).

A third concern involves ethical considerations, particularly regarding access to private data, including during model training (Marshall *et al.*, 2024; Head *et al.*, 2023). Unless data are adequately anonymised, the reuse of participant information may compromise individual privacy. Moreover, although there is no documented evidence of ChatGPT or similar systems having been compromised, full guarantees of data protection cannot be assumed under all future scenarios (Morgan, 2023).

Meng *et al.* (2024) identify several challenges that QualiGPT appears to address effectively:

a. Lack of transparency. The "black box" nature of LLMs makes it difficult to understand how data are processed, highlighting the need to improve interpretability and transparency.

b. Issues of consistency and contextual understanding. Responses may vary, and maintaining coherence across multi-turn dialogues is challenging. Structured and precise prompts can help improve consistency.

c. Challenges in prompt design. Crafting effective prompts is time-intensive and lacks a standardised methodology. QualiGPT mitigates this by providing predefined prompts that streamline the process.

d. Difficulties in interpreting LLM outputs (e.g. in the case of ChatGPT). Prompts can be designed to standardise outputs and enhance readability.

e. Data privacy and security concerns. In the digital era, data privacy remains a critical issue, particularly when using language models. There are significant risks of sensitive data exposure, as demonstrated by past incidents.[5]

Meng *et al.* (2024) emphasise the need for robust prompt engineering when deploying LLMs for qualitative coding. OpenAI co-founder Greg Brockman defines prompt engineering as "the art of communicating eloquently to an AI".[6] Rossi explores the implications of prompt variability for the reproducibility of results. The researcher describes prompt engineering as "a process of fine-tuning to obtain optimal outputs from an LLM" (2024, pp. 155–156), but questions the assumption that identical prompts will consistently generate identical – or even similar – outputs. Output instability is a known characteristic of LLMs, as minor variations in prompt wording, or repeated use of the same prompt at different times, can yield divergent results.

One area for future development is the creation of LLMs specifically designed, trained and optimised for research applications (Bail, 2024). In this context, open-source language models are widely regarded as the most viable alternative (Spirling, 2023), as they provide greater transparency, enhanced control and the possibility of training with research-specific datasets.

A significant limitation, however, is the technical barrier posed to researchers – particularly sociologists and other social scientists – with limited computational expertise. While efforts have been made to address this through step-by-step user guides for LLMs (Törberg, 2023), integrating these models into more intuitive interfaces tailored to qualitative research would enhance their accessibility and usability.

Lastly, practical considerations remain regarding the pace at which such technologies are adopted within academia. As Marshall *et al.* observe, the cautious stance taken by some may affect publication outcomes: "[...] there will be some who choose to wait to adopt its use and these individuals will still be reviewers for academic journals or referees for conference submissions – and this will continue to impact qualitative researchers [...] the results of our survey suggest that many reviewers [...] are less likely to accept a paper that reports research using AI" (2024, p. 98). This illustrates one of the unintended consequences of the current transition to AI-assisted methodologies.

In summary, the principal limitations of AI in qualitative research include its inability to interpret the deeper meaning of data – a process that requires thorough contextual and theoretical understanding; its potential to perpetuate rather than correct existing biases; and its lack of transparency, as LLMs cannot provide insight into their decision-making processes, raising important concerns about transparency and accountability.

5. The Role of the Researcher in Relation to LLMs

Although LLMs can process language with impressive fluency by leveraging vast repositories of information, they lack genuine understanding, self-awareness and the capacity to reason about the world in the way that humans do (Mitchell, 2023). Precisely for this reason, human judgement, critical interpretation and subject-matter expertise remain fundamental for guiding and validating research supported by LLMs. In similar terms, Jiang *et al.* (2021) note that while researchers often struggle with the complexity and uncertainty of qualitative analysis, they strongly value full autonomy over the process and insist that this autonomy should not be compromised by AI systems.

While the relationship between researchers and LLMs is generally viewed as one of complementarity, this nonetheless calls for a reconceptualisation of the researcher's role. Researchers are increasingly expected to engage in critical reflection based on computational outputs (Li *et al.*, 2024) and to retain responsibility for interpreting findings, evaluating model-generated suggestions through the lens of disciplinary expertise and the empirical realities under investigation. In this regard, Christou (2023) proposes a relationship grounded in the principles of *rigour, reliability, justification* and *ethics*, ensuring that researchers actively apply their evaluative and cognitive skills to monitor processes, document decision-making and reach substantiated conclusions. Similarly, Schreder *et al.* (2025), drawing on insights from qualitative researchers familiar with LLMs, emphasise the need for tools that support tasks such as coding – while insisting that this support must be accompanied by a *strongly reflective* and interpretative stance.

Following Christou (2023), a series of key recommendations can be made concerning how researchers should engage with data in an LLM-supported context: a) achieving comprehensive familiarity with the dataset to understand it in its entirety and detect any inherent biases or assumptions; b) ensuring that training data are diverse and unbiased, and implementing safeguards to promote transparency and accountability; c) verifying all AI-generated content through cross-referencing to ensure its accuracy and credibility; d) carefully reviewing outputs prior to engaging in theoretical or conceptual analysis; and e) actively contributing to the interpretative process by applying prior and in-depth knowledge of the phenomenon under investigation.

In light of these considerations, a key question emerges regarding the design of tools intended for researchers. In this respect, Schreder *et al.* (2025) argue that tools incorporating LLMs should be developed in accordance with a set of guiding principles, which are presented in Table 3.

Table 3

Design of LLM tools according to purpose, based on Schreder et al. (2025)

Participant privacy	For intentional use	To enhance transparency and validation	For deeper engagement with data
More tools should support local hosting, personalisation and fine-tuning of open-source models.	Control over when and how users are assisted or influenced by AI.	Ability to evaluate performance transparently.	Tools based on LLMs should reinforce researchers' engagement with data (rather than distancing them from it).
Open-source code as a potential solution to privacy issues, and for greater control and transparency.	Flexibility and interactivity. Chat-based LLMs enable, but *do not yet fully support*, intentional use for specific tasks within the research process.	Useful features for positivist approaches include ensuring reproducibility of results, inter-rater reliability (IRR) and annotation error analysis.	Concern regarding the variable performance of LLMs across different contexts, domains, cultures and languages.
Tools should clearly indicate whether they use external APIs, and if so, when and how.	Tools should be designed to guide users in selecting appropriate and intentional uses of LLMs.	Include features to help users interpret and analyse results and suggestions.	LLMs as a means to generate more distinctive approaches to understanding data, helping researchers examine or challenge existing theories using evidence directly drawn from the data.
Obligation to inform users, prior to uploading data, about how privacy is managed: how data can be anonymised, how it can be deleted and whether inputs are used to train models.	Tools should enable researchers to develop their own ideas.		Design should take into account participants' perspectives and interests.
			Provide the option to use models that better reflect a target group.
			Possibility to incorporate the researcher's own context into the analysis, including prior experience, influential texts and theoretical frameworks.

Own research.

Following an in-depth evaluation of various LLMs – including several versions of FLAN-5 and ChatGPT (versions 3 and 4) – Ziems *et al*. (2024) conclude that while these models can enhance traditional processes, they should not be seen as a substitute. They put forward several recommendations aligned with the following principles: a) improving data labelling processes, particularly when managing large datasets; b) ensuring flexibility to adapt and customise models according to specific research needs, while maintaining ethical oversight; c) prioritising fidelity, relevance, coherence and fluency by selecting larger instruction-tuned models aligned with human preferences; and d) favouring the use of open-source LLMs for classification tasks, as opposed to relying on proprietary or closed-source models.

On the importance of open-source tools, Van Dis *et al.* (2023) highlight one of the most pressing challenges facing researchers using LLMs – namely, the quasi-monopolistic environment in which these models operate. LLMs are typically proprietary technologies developed by a small number of large technology companies with the financial and technical capacity to drive AI development. This concentration of control raises considerable ethical concerns (ibid.), particularly regarding the lack of transparency. As the authors warn, "tech companies might conceal the inner workings of their conversational AIs", which "goes against the move towards transparency and open science, and makes it hard to uncover the origin of, or gaps in, chatbots' knowledge [...]" – thereby underscoring the need for "the development and implementation of open-source AI technology" (ibid., p. 225).

Ultimately, the use of LLMs in qualitative research demands a careful balance between automation and human oversight, between efficiency and critical reflection, and between assistance and researcher autonomy. It is crucial, therefore, to acknowledge the active role of the researcher as coder, even within LLM-assisted environments, in order to avoid the illusion of total automation or the fallacy of *algorithmic neutrality*. AI-based tools must not supplant the researcher's judgement, but rather support it within a framework that is methodologically rigorous, ethically sound and transparently documented. As the reviewed literature suggests, the development of such systems should be guided by principles of openness, contextualisation and user-centred design, ensuring that technology enhances – rather than constrains – the interpretive, analytical and creative capacities of researchers.

6. Open Science, LLMs and Qualitative Research in Social Verbatim

It has been suggested that generative AI, in the social sciences as well as in many other disciplines, will open "revolutionary avenues for human reason [...], although the knowledge process starts at the other end to Enlightenment science, which made progress through the logic of induction and patient accumulation of evidence" (Peters *et al.*, 2023, p. 832) – in a manner comparable to the transformative impact of Gutenberg's printing press (Kissinger *et al.*, 2023). This transformation stems from the use of methods that generate results "without explaining why or how their process works based on pregenerated representations of the vast oceans of data on which it was trained" (Peters *et al.*, 2023, p. 832).

Accordingly, limitations in transparency and replicability remain among the most significant challenges in the evolving relationship between science and AI. Analogously, qualitative research has also exhibited certain limitations in these areas (Jiang *et al.*, 2021). For example, in relation to transcription practices, McMullin (2023) found that 41% of the studies analysed made no mention of transcription at all (despite evidence that it had been conducted), 11% referred to the fact that transcriptions were obtained but gave no detail about the process, and 19% included only a minimal statement such as "the interviews were recorded and transcribed". Similarly, Nascimento (2019) noted that qualitative studies – in this case, within the field of management – often describe transcription practices with nothing more than a brief phrase such as "the interviews were transcribed".

Such limitations – not always attributable to a deliberate departure from positivist traditions – have prompted efforts to reinforce the rigour, reliability and validity of qualitative analysis. In this context, AI is increasingly viewed as "an option to support qualitative researchers in their work" (Jiang *et al*., 2021, p. 94), while simultaneously presenting new challenges.

In particular, it is essential to make the workings of LLMs more transparent – to "open the black box" and clarify how their processes function. Wang *et al.* (2019), drawing on interviews with qualitative researchers, reaffirm the importance of AI transparency. Similarly, Yang *et al.* (2019) propose the concept of a "hassle-free AI", based on the idea that interaction with such systems should exhibit an appropriate level of "normality", including a clear understanding of the motivations behind the system's outputs.

A clear example of how the workings of the black box can be made explicit comes from the field of medicine. IBM Watson Health[7] is an artificial intelligence system applied to medical diagnosis. Although the AI used by Watson to analyse medical data is highly complex, physicians are able to access detailed explanations of how the system arrived at a particular conclusion or recommendation. It provides the rationale behind its diagnostic outcomes, including supporting arguments, contextual information and the identification of the most relevant data or symptoms. Analogously, in the social sciences, we should develop tools capable of transforming AI into a reliable and complementary resource – rather than a black box that remains opaque and difficult to interpret.

Advancing in this direction presents a compelling challenge: combining the use of technologies that streamline qualitative research processes with maximum rigour, human oversight and transparency, particularly in workflows involving LLMs. As noted in the introduction, Social Verbatim offers several practical applications of AI designed to address these challenges and contribute to the development of an *open science* model in qualitative research (Breznau, 2021).

The Social Verbatim tool was developed through a systematic study informed by insights gathered from interviews conducted as part of the research project "CS-Transcribe: Research on Needs and Development of a Digital Transcription Tool for the Social Sciences", which involved three profiles of potential users: researchers, transcribers and researcher–transcribers. A total of 15 individual interviews were conducted with 11 women and 4 men, including 7 researchers (4 women and 3 men), 5 researcher–transcribers (4 women and 1 man), and 3 transcribers (all women).

These interviews were based on a first demo version of the tool, which included a core set of functionalities. Informed by feedback collected during the interviews, these features were expanded and refined in the beta version of the application, which included the following components:

- An online platform for automatic transcription of video or audio files, as well as manual transcription.
- An interface for inserting annotations on non-verbal communication using icons.
- A user experience (UX) design prioritising efficiency in the review and correction process.
- A functionality for data anonymisation.
- An interface for coding.
- Automatic detection of silences, missing words and automated time stamps.
- Tools to facilitate collaborative work within research teams.
- Project management features, including detailed information for each transcription (e.g. transcriber identity, progress status).
- A foot pedal interface to support a smoother and more ergonomic experience in both manual transcription and review.
- Reporting interfaces including verbatim lists, statistics, charts and visualisations such as word and code clouds.
- Flexible import and export systems for data input and output, tailored to project needs and compatible with platforms such as ATLAS.ti and NVivo.

In relation to the *open science* paradigm, Social Verbatim offers several key contributions. Firstly, it enables any user to access the content of a transcription cited in a publication, as well as its corresponding audio source (and video, where applicable), once the material has been properly anonymised (by removing identifying references and applying voice and, where necessary, image distortion to ensure participants remain unrecognisable). Such access allows users to examine the "black box" of the transcription process and assess how the transcription was carried out, identifying any inconsistencies that may affect the research findings. Secondly, it permits user access to the broader transcription project in order to obtain general or more detailed information about the interviewees. This is contingent on the researchers' judgement that privacy can be preserved and is facilitated through a pre-anonymised system[8] designed for agile navigation.[9]

In the next phase of development, Social Verbatim aims to integrate LLMs into the coding and analysis process, following the premises outlined below, which align with the *open science* paradigm. Nonetheless, in the context of qualitative research, advancing this paradigm encounters a potential point of conflict – the essential need to protect participant privacy (Gómez *et al.*, 2025), as well as the aforementioned challenges associated with algorithmic opacity. To address these concerns, the following strategies are proposed: a) incorporation of open-source LLMs to ensure transparency in the analytical process; b) development of an interface that enables the traceability of prompts used during the analysis; c) active supervision and review by the researcher throughout the coding and analysis stages, including feedback mechanisms to refine the tool's analytical criteria; d) implementation of high standards of data privacy management, irrespective of user type (premium or general); e) in line with the recommendations of Schreder *et al.* (2025, p. 1), ensuring flexibility and interactivity, thereby overcoming the current limitation whereby *chat-based* LLMs allow, but *do not adequately support*, intentional use for specific research tasks; and f) provision of interface functionalities that allow researchers to develop their own ideas, reinforcing – rather than distancing – their relationship with the data.

7. Conclusions

The theoretical review of the use of large language models (LLMs) in qualitative research has made it possible to identify both significant advances and persistent challenges within the field. The development of LLMs has been driven by the Transformer architecture and advancements in hardware – particularly GPUs and TPUs – which have enabled the efficient processing of large volumes of text. In this context, tools such as QualiGPT, MindCoder and Thematic-LM have been developed to support qualitative analysis by offering intuitive interfaces and advanced functionalities for coding and analysing data. These tools leverage prompt engineering to optimise inputs and generate more accurate and relevant outputs.

Nevertheless, the lack of transparency and the inherent "black box" nature of LLMs remain major concerns. It is therefore essential to improve the interpretability and transparency of these models so that researchers can understand, evaluate and trust the results they produce. Furthermore, the presence of biases in training data – which may reinforce existing social and cultural prejudices – along with concerns regarding data privacy and security, highlight the need to ensure that all data used in LLM training and operation are adequately anonymised and securely managed.

Despite these technological advances, LLMs should be viewed as complementary tools that support – but do not replace – human judgement, interpretation and analytical reasoning. Researchers must remain actively engaged in the coding and analysis processes, applying their disciplinary expertise to guide and critically assess the outputs generated by LLMs.

The integration of LLMs into tools such as Social Verbatim aims to contribute to more transparent and rigorous qualitative research by enabling greater traceability of processes and improved data management practices. Social Verbatim is thus conceived not merely as a means of automating tasks but as a collaborative platform that supports researchers in building more open, reliable and reproducible scientific practices. It facilitates access to and review of each step in the qualitative research workflow, moving towards a research ecosystem in which human–machine collaboration is not only efficient but also transparent and verifiable.

In conclusion, while LLMs offer considerable potential to enhance the depth and efficiency of qualitative analysis, their implementation must be carefully managed to address the associated methodological, ethical and epistemological challenges. Human–machine collaboration, transparency and a commitment to open science must serve as guiding principles in the continued advancement of this field.

8. Funding

"CS-Transcribe: Research on Needs and Development of a Digital Transcription Tool for the Social Sciences" (TED2021-131729B-I00) is a project funded by the Ministry of Science, Innovation and Universities / Spanish State Research Agency (MCIN/AEI) and the European Union's NextGenerationEU package (Recovery, Transformation and Resilience Plan – PRTR). Principal investigators: Juan Miguel Gómez Espino and Carles Xavier Simó Noguera.

9. References

Arlinghaus, C. S., Wulff, C., Maier, G. W., Arlinghaus, C., Wulff, C. & Maier, G. (2024). Inductive coding with chatgpt-an evaluation of different gpt models clustering qualitative data into categories. *OSF Preprints, doi*, 10.

Bail, C. A. (2023). Can generative AI improve social science? https://osf.io/rwtzs/download

Blair, E. (2015). A reflexive exploration of two qualitative data coding techniques. *Journal of Methods and Measurement in the Social Sciences, 6*(1), 14-29.

Carius, A. C. & Teixeira, A. J. (2024). Artificial Intelligence and content analysis: the large language models (LLM) and the automatized categorization. *AI y Soc.*, 40, 2405-2416. https://doi.org/10.1007/s00146-024-01988-y

Chew, R., Bollenbacher, J., Wenger, M., Speer, J. & Kim, A. (2023). *LLM-assisted content analysis: Using large language models to support deductive coding.* arXiv preprint arXiv:2306.14924.

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

Crabtree, B. & Miller, W. (Eds.) (1999). DoingQualitative Research (2nd ed.) [20 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 3*(4), art. 3. http://nbn-resolving.de/urn:nbn:de:0114-fqs020432. © 2002 FQS. http://www.qualitative-research.net/fqs/

Dai, S. C., Xiong, A. & Ku, L. W. (2023). *LLM-in-the-loop: Leveraging large language model for thematic analysis.* arXiv preprint arXiv:2310.15100.

Deterding, N. M. & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods & Research, 50*(2), 708-739.

Dunivin, Z. O. (2024). *Scalable qualitative coding with LLM: Chain-of-thought reasoning matches human performance in some hermeneutic tasks*. arXiv preprint arXiv:2401.15170.

Gao, J., Shu, Z. & Yeo, S. Y. (2025). *Using Large Language Model to Support Flexible and Structural Inductive Qualitative Analysis.* arXiv preprint arXiv:2501.00775.

Glasser, B. G. & Strauss, A. L. (1967). *The development of grounded theory*. Chicago, IL: Alden.

Gómez-Espino, J. M., Simó-Noguera, C. & Carvajal-Soria, P. (2025). *Ciencia abierta y procesos de investigación cualitativa en la app Social Verbatim.* On press.

González-Veja, A. M. D. C., Sánchez, R. M., Salazar, A. L. & Salazar, G. L. L. (2022). La entrevista cualitativa como técnica de investigación en el estudio de las organizaciones. *New trends in qualitative research, 14.*

Hayes, A. S. (2025). «Conversing» with qualitative data: Enhancing qualitative research through large language models (LLM). *International Journal of Qualitative Methods*, 24, 16094069251322346.

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), 1-23.

Kissinger, H., Schmidt, E. & Huttenlocher, D. (2023), February 24. ChatGPT heralds an intellectual revolution. *The Wall Street Journal*. https://www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution

Lakshmanan, L. (2022). *Why large language models like ChatGPT are bullshit artists, and how to use them effectively anyway*. LinkedIn. https://www.linkedin.com/pulse/why-large-language-models-like-chatgpt-bullshit-how-use-lakshmanan/?trk=pulse-article_more-articles_related-content-card

Li, J., Li, J. & Su, Y (2024, Mayo). A map of exploring human interaction patterns with LLM: Insights into collaboration and creativity. In *International conference on human-computer interaction* (pp. 60-85). Springer Nature Switzerland.

Linneberg, S. M. & Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. *Qualitative Research Journal, 19*(3), 259-270.

Marshall, D. T. & Naff, D. B. (2024). The ethics of using artificial intelligence in qualitative research. *Journal of Empirical Research on Human Research Ethics, 19*(3), 92-102.

Mathis, W. S., Zhao, S., Pratt, N., Weleff, J. & De Paoli, S. (2024). Inductive thematic analysis of healthcare qualitative interviews using open-source large language models: How does it compare to traditional methods? *Computer Methods and Programs in Biomedicine*, 255, 108356.

McMullin, C. (2023). Transcription and qualitative methods: Implications for third sector research. *VOLUNTAS: International journal of voluntary and nonprofit organizations, 34*(1), 140-153.

Meng, H., Yang, Y., Li, Y., Lee, J. & Lee, Y. C. (2024). *Exploring the potential of human-LLM synergy in advancing qualitative analysis: A case study on mental-illness stigma*. arXiv preprint arXiv:2405.05758.

Miles, M. B. *et al.* (2015). *Qualitative Data Analysis: A Methods Sourcebook and The Coding Manual for Qualitative Researchers: Matthew B. Miles, A. Michael Huberman, and Johnny Saldaña*. Thousand Oaks, CA: SAGE.

Mitchell, M. (2024). Large Language Models. In M. C. Frank & A. Majid (Eds.), *Open Encyclopedia of Cognitive Science*. MIT Press. https://doi.org/10.21428/e2759450.2bb20e3c

Mitchell, M. & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences, 120*(13), e2215907120.

Molina, M. & Garip, F. (2019). Machine learning for sociology. *Annual Review of Sociology, 45*(1), 27-45.

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

Qiao, T., Walker, C., Cunningham, C. W. & Koh, Y. S. (2025). Thematic-LM: a LLM-based Multi-agent System for Large-scale Thematic Analysis. In *The Web Conference* 2025.

Rossi, L., Harrison, K. & Shklovski, I. (2024). The Problems of LLM-generated Data in Social Science Research. *Sociologica, 18*(2), 145-168.

Schroeder, H., Quéré, M. A. L., Randazzo, C., Mimno, D. & Schoenebeck, S. (2025). Large Language Models in Qualitative Research: Uses, Tensions, and Intentions. *Computer Science*. abril-mayo. https://arxiv.org/abs/2410.07362

Social Verbatim (s. f.). Social Verbatim. https://www.socialverbatim.com. www.app.socialverbatim.com.

Spirling, A. (2023). World view. *Nature*, 616, 413.

Tai, R. H., Bentley, L. R., Xia, X., Sitt, J. M., Fankhauser, S. C., Chicas-Mosier, A. M. & Monteith, B. G. (2024). An examination of the use of large language models to aid analysis of textual data. *International Journal of Qualitative Methods*, 23, 16094069241231168.

Törnberg, P. (2023). *How to use LLMs for text analysis*. arXiv [Preprint] (2023). https://doi.org/10.48550/arXiv.2307.13106

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R. & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, *614*(7947), 224-226.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention Is All You Need. In *Advances in Neural Information Processing Systems* (vol. 30). https://arxiv.org/abs/1706.03762

Wang, Y., Wang, Q., Shi, S., He, X., Tang, Z., Zhao, K. & Chu, X. (2019). Benchmarking the performance and energy efficiency of AI accelerators for AI training. *Proceedings of the 2019 IEEE International Symposium on High Performance Computer Architecture (HPCA)*. https://doi.org/10.1109/HPCA.2019.00015

Wu, Y., Nagler, J., Tucker, J. A. & Messing, S. (2023). Large language models can be used to scale the ideologies of politicians in a zero-shot learning setting. arXiv [Preprint] https://doi.org/10.48550/arXiv.2303.12057.

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R. & Oudeyer, P. Y. (2023, March). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In *Companion proceedings of the 28th international conference on intelligent user interfaces* (pp. 75-78).

Yang, Y. & Ma, L. (2025). Artificial intelligence in qualitative analysis: a practical guide and reflections based on results from using GPT to analyze interview data in a substance use program. *Quality & Quantity*, 1-24.

Zhang, H., Wu, C., Xie, J., Rubino, F., Graver, S., Kim, C., ... & Cai, J. (2024). *When Qualitative Research Meets Large Language Model: Exploring the Potential of QualiGPT as a Tool for Qualitative Coding*. arXiv preprint arXiv:2407.14925.

Zhao, F., Yu, F. & Shang, Y. (2024). A New Method Supporting Qualitative Data Analysis Through Prompt Generation for Inductive Coding. In *2024 IEEE International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 164-169). IEEE.

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z. & Yang, D. (2024). Can large language models transform computational social science? *Computational Linguistics, 50*(1), 237-291. https://doi.org/10.48550/arXiv.2305.03514

Juan Miguel Gómez Espino

Juan Miguel Gómez Espino is a permanent lecturer of sociology at the Universidad Pablo de Olavide (Seville). Since 2001, he has combined teaching and research in the sociology of childhood and education with university management roles (currently as Dean of the Faculty of Social Sciences). He holds a degree in political science and sociology from the University of Granada and earned his PhD in 2009. He has published several papers in peer-reviewed journals such as ***Revista Española de Investigaciones Sociológicas***, ***Empiria***, ***Childhood***, ***International Sociology*** and ***Language and Education***, as well as book chapters with publishers such as Springer. He has also been a visiting lecturer at the University of Sheffield (United Kingdom) and Gramma (Cuba).

Notes

1 This publication forms part of the R&D+i Project "CS-Transcribe: Research on Needs and Development of a Digital Transcription Tool for the Social Sciences", ref. TED2021-130903B-I00, funded by the Ministry of Science, Innovation and Universities, the European Union NextGenerationEU / Recovery, Transformation and Resilience Plan (PRTR), and the Spanish State Research Agency/10.13039/501100011033.

2 The application is accessible at app.socialverbatim.com

3 Social Verbatim (n.d.). Social Verbatim. https://www.socialverbatim.com (accessed 25 April 2025).

4 One of the implications of the emergence of these models is the increasing importance of "prompt engineering", which refers to the practice of designing and optimising model inputs to generate more accurate, useful or contextually relevant outputs.

5 According to OpenAI's policies, user input in ChatGPT may be used to improve its models unless this option is disabled in the settings. In contrast, data submitted via the API is not used to train OpenAI's models. This suggests that using the API offers greater data privacy and security.

6 G. Brockman [@gdb] (11 March 2023). Write your prompt like this: [1] Task: what you want ChatGPT to do [2] Context: extra info that helps set the stage [Post]. X. https://x.com/gdb/status/1634708489078706179

7 https://www.ibm.com/es-es/watson

8 Regarding voice distortion, it is evident that certain nuances of natural speech may be lost, implying a trade-off between linguistic/paralinguistic richness and privacy.

9 When providing access to data, integration with platforms such as Zenodo or similar repositories – with appropriate permissions and restrictions – could be explored. In such cases, *open science* would face a necessary limitation: the safeguarding of participant privacy.