

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

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

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

Juan Miguel Gómez Espino

Universidad Pablo de Olavide (UPO), España
jmgomesp@upo.es

Recibido/Received: 9-6-2025

Recibido/Received: 14-10-2025



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.

CÓMO CITAR: 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>

English version can be read on <https://doi.org/10.54790/rccs.176>

ABSTRACT

This article explores the use of Large Language Models (LLMs) in qualitative coding, highlighting advances and opportunities for the Social Verbatim tool. It reviews the fundamentals of LLMs, their architecture, and the impact of hardware on their development. Additionally, specific applications of LLMs in qualitative research are analyzed, including thematic coding and comparative analysis. Methodological, ethical, and epistemological challenges are addressed, and strategies to mitigate these issues are proposed. Finally, the implications of integrating LLMs into tools like Social Verbatim are discussed, emphasizing the importance of transparency and human-machine collaboration in qualitative research.

KEYWORDS: Large Language Models (LLMs); qualitative coding; Generative Artificial Intelligence (GAI); qualitative research; open science; AI-assisted qualitative analysis.

1. Introducción

En dos años de trabajo de un proyecto denominado CS-Transcribe¹, entre sus principales resultados se encuentra el desarrollo de una herramienta *online* registrada como Social Verbatim. Esta *app* se ha concebido como solución digital para distintas fases del proceso de la investigación social cualitativa y avanza en la dirección de convertirse en una herramienta integral de apoyo a este tipo de investigación, incluyendo, además de utilidades de transcripción, otras relacionadas con la gestión y apoyo al análisis y codificación. En concreto, como se indica en la información publicada en su web (www.socialverbatim.com²), Social Verbatim, además de transcribir automática o manualmente, permite «incorporar comunicación no verbal y contextual mediante iconos, revisar y corregir transcripciones, trabajar en equipo, organizar proyectos, organizar entrevistas, grupos de discusión u otros elementos de análisis, anonimizar fragmentos, insertar comentarios, marcadores y notas de análisis (o) usar *verbatim*s para las publicaciones», entre otras posibilidades³.

En la línea de explorar la ampliación de funcionalidades de esta herramienta, este artículo se propone indagar en los LLM (Modelos de Lenguaje de Gran escala, en inglés, LLM, Large Language Models) para el proceso de codificación cualitativa. A través de esta contribución se pretende conocer los avances en materia de codificación cualitativa por IA, así como los retos y oportunidades que ofrece este tipo de tecnología, aludiendo especialmente al papel de las personas investigadoras como parte activa e inevitablemente reflexiva en el proceso.

Antes de proceder a una revisión de las aportaciones más relevantes en esta materia, a continuación, se introduce el concepto de LLM. Los LLM son modelos de inteligencia artificial diseñados para procesar y generar texto en lenguaje natural a gran escala. Según Mitchell (2024):

[...] un modelo de lenguaje grande (LLM, por sus siglas en inglés) es un sistema computacional, típicamente una red neuronal profunda con un gran número de parámetros ajustables, que implementa una función matemática llamada modelo de lenguaje. [...] Las redes neuronales subyacentes a los LLM se entrenan utilizando amplias colecciones de texto obtenidas de sitios web, libros digitalizados y otros recursos digitales.

Estas *redes neuronales* son un modelo computacional inspirado en el cerebro humano que se compone de «neuronas» (unidades de procesamiento) organizadas en capas, que transforman entradas (como texto o números) en salidas (como predicciones o respuestas).

En los últimos años se ha asistido a un desarrollo extraordinario de esta tecnología, especialmente tras la introducción de la arquitectura de Transformers a partir de un artículo de Vaswani *et al.* (2017). Esta arquitectura es capaz de capturar relaciones a largo plazo en el texto de manera mucho más eficiente que los modelos anteriores como las RNNs (Redes Neuronales Recurrentes) o las CNNs (Redes Neuronales Convolucionales). De hecho, antes de *ese momento*, los modelos de lenguaje se entrenaban desde cero para cada tarea específica, mientras que, con Transformer, se pre-entrenaron modelos con grandes cantidades de datos no supervisados para ajustarlo posteriormente (*fine-tuning*) para tareas específicas⁴.

Esta nueva arquitectura se basa en la *self-attention* (*Attention Is All You Need*), lo que permite procesar grandes volúmenes de texto de manera eficiente e identificar relaciones asignando diferentes pesos a distintas palabras en una oración. Así, pues, los LLM convierten las palabras en representaciones numéricas llamadas *embeddings*. Estas representaciones permiten que el modelo relacione conceptos similares en un espacio matemático, comparándose una palabra con todas las demás para determinar cómo de relevante es en la oración. Para ello, asigna pesos diferentes a cada palabra para entender mejor el significado global del *prompt* que es la instrucción o entrada textual que el usuario proporciona para que el modelo genere una respuesta. Además, con Transformer se consigue procesar todas las palabras al mismo tiempo gracias a la *paralelización*, lo que, a su vez, permitió entrenar modelos con miles de millones de parámetros sin que el tiempo de entrenamiento se disparara, mejorando significativamente su escalabilidad.

Junto al desarrollo de la arquitectura de Transformer, el avance del *hardware* ha sido crucial en el desarrollo de los LLM; especialmente, a) el avance de las unidades de procesamiento gráfico, GPU, que aceleran el cálculo de matrices y tensores esencial en la arquitectura de Transformers; b) de las memorias RAM, ya que los modelos más grandes requieren terabytes de memoria para procesar datos; c) de los chips especializados en IA — como TPUs, de hecho Wang *et al.* (2019) mostraron que las TPUs ofrecen ventajas significativas en términos de rendimiento y eficiencia energética en comparación con las GPUs tradicionales, especialmente en modelos de aprendizaje profundo como los Transformers—. Sin un *hardware* potente, los tiempos de entrenamiento serían prohibitivos y el despliegue de modelos en tiempo real sería inviable.

2. Inteligencia Artificial Generativa (IAG) y los LLM en la codificación cualitativa

Todos estos avances han permitido la introducción de innovaciones metodológicas que pueden transformar (y que, de hecho, están transformando ya) la forma en que las/os científicas/os sociales interactúan con los datos cualitativos (Hayes, 2025; Van Dis *et al.*, 2023). Para Hayes (2025), supone habitar un nuevo «espacio híbrido» interactuando dinámicamente con datos a gran escala, conversando con ellos a través de los LLM en un nuevo modelo de relación a mitad de camino entre las tradiciones cualitativas establecidas y las posibilidades que ofrecen las capacidades computacionales avanzadas.

En cualquier caso, pocas dudas existen sobre los efectos que dicha tecnología tendrá en la manera de concebir nuestro mundo y la ciencia misma. Sin embargo, en sociología, politología, incluso en economía y en otras ciencias sociales, en palabras de Bail, apenas se ha comenzado a explorar cómo la IAG transformará su investigación, a pesar de que «estas herramientas pueden hacer avanzar la escala, el alcance y la velocidad de la investigación en las ciencias sociales —y pueden permitir también nuevas formas de investigación científica [...]» (2024, p. 1).

Las aportaciones de la IAG están siendo ya ampliamente exploradas tanto en el ámbito *experimental o cuasi-experimental* y, por supuesto, en el de la investigación cualitativa. Ziems *et al.* (2024) evaluaron 13 modelos LLM, encontrando niveles aceptables de concordancia con los humanos y concluyeron que, frente a la codificación de texto supervisada y manual, que exige una gran cantidad de datos de entrenamiento anotados por humanos, los LLM ofrecen grandes oportunidades, sin las constricciones de otros métodos incluso no supervisados con resultados ininteligibles. Wu *et al.* (2023) analizaron las declaraciones públicas de los funcionarios electos y demostraron que ChatGPT-3.5 puede producir clasificaciones de su ideología, ya que los resultados se aproximaron mucho al popular método DW-Nominate para medir la ideología.

Hayes (2025) señala algunas posibilidades del uso de los LLM para la investigación cualitativa, entre los que (además de la *orientación básica para la investigación*, dentro de conjuntos de datos extensos y complejos) menciona los siguientes: a) *codificación temática*; b) *análisis comparativo* entre diferentes textos, ya que puede resaltar diferencias en el tono, el énfasis o el marco conceptual; c) *dinámicas internas en los datos*, identificando contradicciones, tensiones o narrativas en evolución dentro del corpus; d) *pruebas de escenarios y ejercicios hipotéticos*; e) *síntesis creativa y estímulo para una mayor indagación*; f) *compromiso reflexivo*; g) otras posibilidades menos convencionales de los LLM como la generación de guiones gráficos, instrucciones o esquemas descriptivos.

A continuación, nos centraremos en los tres primeros, que se insertan en la propia naturaleza del proceso de codificación, para lo que, en primer lugar, nos referiremos a este proceso en un sentido más convencional. En el proceso de investigación cualitativa, la codificación ha ocupado un lugar central, funcionando como puente entre los datos en bruto y la construcción de significado analítico.

Así, pues, un código es un constructo generado por la persona investigadora que simboliza y, por lo tanto, atribuye un significado interpretado a cada dato individual para fines posteriores de detección de patrones, categorización, desarrollo teórico y otros procesos analíticos (Miles *et al.*, 2015, p. 78). La codificación, en este sentido, es el proceso sistemático mediante el cual se organizan, etiquetan y agrupan datos cualitativos (como entrevistas, observaciones o textos) con el fin de identificar patrones, temas o categorías relevantes. Frente a la concepción de la codificación como trabajo técnico y preparatorio, hay cierto consenso en concebirla como una reflexión profunda y, por lo tanto, un análisis e interpretación de los significados de los datos (González-Veja, 2022; Deterding y Waters, 2021). Los códigos se utilizan principalmente, aunque no exclusivamente, para recuperar y categorizar fragmentos de datos similares, para encontrar, extraer y agrupar los segmentos relacionados con una pregunta de investigación, hipótesis, constructo o tema específico.

La codificación, por tanto, es un proceso heurístico que ayuda a la persona investigadora a explorar, descubrir y comprender patrones y temas subyacentes en el conjunto de datos. Es decir, la codificación no solo estructura la información, sino que activa un proceso de reflexión y análisis que lleva a nuevas interpretaciones o hallazgos, funcionando como una guía o estrategia de descubrimiento en el análisis cualitativo.

Cabe distinguir entre dos tipos de codificación: inductiva (Glasser y Strauss, 1967) y deductiva (Crabtree y Miller, 1999). El tipo inductivo consiste en construir patrones y temas desde abajo hacia arriba, organizando los datos en unidades de información cada vez más abstractas. Por su parte, desde la lógica deductiva se comparan los patrones y teorías existentes con los datos (Jiang *et al.*, 2021, p. 94). Si bien desde algunos enfoques se recomienda evitar marcos conceptuales previos respecto de los datos, no parece esta una sugerencia realista y, en la práctica, casi siempre se produce una combinación de ambos métodos (Lindbergh y Korsgaard, 2019), lo que resulta plausible por las limitaciones que tienen cada uno de ellos por separado.

Los/as científicos/as sociales han comenzado a utilizar los LLM para la clasificación de textos y, dentro de estos, las personas investigadoras en sociología. Los LLM pueden ayudar a estas a pasar rápidamente de una visión general de patrones temáticos a aspectos más concretos de la comunicación humana (Hays, 2025). En general, existe un elevado consenso en admitir que estos modelos pueden ser de gran utilidad en la codificación de datos en investigación cualitativa, aunque se advierte de la importancia de utilizarlos con criterio y conciencia sobre sus limitaciones, a las que se aludirá más adelante.

Los LLM realizan tareas de procesamiento de lenguaje natural, es decir, sin datos previos de entrenamiento. A diferencia de otros modelos de IAG como el *machine learning*, aprendizaje automático supervisado, en el que se aportan datos de entrenamiento previamente etiquetados (Molina y Garip, 2019), los LLM funcionan con *zero-shot*, es decir, sin entrenamiento previo (Ziems *et al.*, 2024). En el primer caso, se decide una «etiqueta» o categoría para cada documento (por ejemplo, un tuit etiquetado o anotado, un párrafo de una noticia o un fragmento de discurso), para luego entrenar un modelo que predice automáticamente las etiquetas utilizando caracte-

rísticas de los textos. Una vez entrenado, el modelo puede predecir etiquetas para otros textos similares, codificando automáticamente nuevos documentos.

En cualquier caso, los LLM son más que una promesa. De hecho, se trata de una realidad que está contribuyendo a revitalizar el uso de técnicas computacionales en la investigación cualitativa al aportar: a) *eficiencia*, dado que ayuda a acelerar el proceso de codificación, especialmente cuando se trata de grandes conjuntos de datos; b) *consistencia*, en la medida en que puede garantizar un criterio idéntico en la codificación, frente a los sesgos y errores humanos; c) *análisis de patrones*, al identificar relaciones en los datos que podrían ser difíciles o imposibles de detectar manualmente.

3. Herramientas basadas en LLM usadas para la codificación en análisis cualitativo

A continuación, en el cuadro 1 se muestran algunos ejemplos recientes de estudios (primera columna) en los que se procede a analizar el uso de LLM como medio para la codificación cualitativa; se alude al tipo de codificación empleada, si inductiva deductiva (segunda columna); al tipo de LLM en el que se explicita o no el uso de *fine-tuning* (ajuste fino o posterior del modelo) partiendo del sistema *zero-shot* (como se decía, la *ingeniería de prompts* facilita el afinamiento en los resultados obtenidos en la codificación a través de la práctica de diseñar y optimizar instrucciones o consultas [*prompts*] para obtener respuestas más precisas, relevantes y útiles de modelos de lenguaje, como también se ha indicado) (tercera columna); si la comparación del estudio se da entre diferentes LLM o entre LLM y humanos, habitualmente expertos/a (cuarta columna); se aporta información sobre si se ha generado alguna herramienta propia o alguna metodología específica, y la sexta al LLM concreto utilizado para la codificación (quinta columna); y se indica qué modelo específico se ha utilizado (sexta columna).

Cuadro 1
Uso de LLM en codificación cualitativa según criterios de codificación, comparación, herramientas y modelo LLM

Fuente	Tipo de codificación	Ajuste de LLM	Comparación	Herramienta (app) o metodología	LLM utilizado
Chew <i>et al.</i> (2023)	Deductiva	<i>Zero-shot</i> (<i>fine-tuning</i> mediante <i>prompts</i>)	Con humanos	Metodología. LACA, Análisis de Contenido Asistido por LLM, por sus siglas en inglés. Códigos accesibles en Figshare: https://figshare.com/articles/dataset/LLM-Assisted_Content_Analysis_LACA_Coded_data_and_model_reasons/23291147	GPT 3.5
Ziems <i>et al.</i> (2024)	Inductiva	<i>Zero-shot</i>	Entre modelos	n/c	FLAN -5 (Small, Base, Large, XL, XXL), FLAN UL-2, GPT (3.5, 4), ada-001, babbage-001, curie-001, davinci-001, 002, 003

Fuente	Tipo de codificación	Ajuste de LLM	Comparación	Herramienta (app) o metodología	LLM utilizado
Meng <i>et al.</i> (2024)	Deductiva/ inductiva	Zero-shot	Se compara resultado del modelo sin colaboración humana y con colaboración humana (en codificación deductiva e inductiva)	Metodología CHALET (Collaborative Human-LLM Analysis for Empowering Conceptualization in Qualitative Research). Se trata de una herramienta metodológica sin software específico	GPT-4-1106-preview
Dunivin (2024)	Deductiva/ análisis de contenido	Zero-shot	Entre modelos/ Con humanos	n/c	GPT 3.5 y 4
Xiao <i>et al.</i> (2023)	Deductiva	Zero-shot (fine-tuning mediante prompts)	Con humanos	n/c	GPT 3
Zhang <i>et al.</i> (2024)	Deductiva/ Inductiva	Zero-shot	Con humanos/ Entre modelos	Software. QualiGPT, software open-access: https://github.com/KindOPSTAR/QualiGPT	GPT-4 y Claude 3.5
Zhao <i>et al.</i> (2024)	Inductiva	Zero-shot (fine-tuning (FT) mediante prompts)	Entre modelo sin FT y con FT	Software. A2C (Argument2Code), software diseñado para aprovechar las capacidades de los LLM para mejorar el proceso de análisis cualitativo de datos (no open-source)	Llama-2-13BChat
Tai <i>et al.</i> (2024)	Deductiva	Zero-shot (se analizan efectos de iteraciones después de prompts)	Se comparan resultados después de las nuevas iteraciones (160) a partir de mismos textos de muestra	n/c	GPT 3.5
Arlinghaus <i>et al.</i> (2024)	Inductiva	Zero-shot	Con humanos/ Entre modelos	n/c	GPT-3.5 Turbo and GPT-4o)
Dai <i>et al.</i> (2023)	Deductiva/ Inductiva (análisis temático)	Se proporcionan prompts con marco de discusión (frame discussions)	Con humanos	Metodología. Se ofrece código en: https://github.com/sjdai/LLM-thematic-analysis	GPT 3.5
Qiao <i>et al.</i> (2025)	Inductiva	LLM multiagentes (coders, aggregators y reviewers)	Comparación entre LLM (uniagente y multiagentes)	Software. Thematic-LM. Códigos en open-source: https://github.com/sjdai/LLM-thematic-analysis	GPT-4
Gao <i>et al.</i> (2025)	Inductiva	Se trata de un artículo descriptivo del funcionamiento de Mindcoder	n/c	Software. Mindcoder. En https://mindcoder.ai Herramienta que pretende cerrar la brecha entre las herramientas de IA profesionales (Atlas Ti, NVivo) y los modelos de lenguaje conversacionales (Claude o Chat-GPT). App sin código disponible (derechos reservados)	GPT-4
Bryda <i>et al.</i> (2024)	Inductiva	Zero-shot	Describen dos estrategias para la codificación: generativa y léxico-semántica	n/c	GPT-4
Yang <i>et al.</i> (2025)	Inductiva	Zero-shot (fine-tuning (FT) mediante prompts)	Con humanos	n/c	GPT-4
Mathis <i>et al.</i> (2024)	Inductiva	Zero-shot (fine-tuning (FT) mediante prompts)	Con humanos	Metodología. Pueden obtenerse códigos en: https://www.sciencedirect.com/science/article/pii/S0169260724003493?via%3Dihub	Llama 2- 70B-Instruct (open-access)

Fuente: elaboración propia.

Dadas las pretensiones aplicadas de este artículo, se aludirá a los casos de herramientas específicas de *software* que se proponen para la codificación cualitativa (que aparecen en la penúltima columna) y a través de las cuales se aplica el LLM (que aparece en la última columna). Por tanto, excluimos de este repaso (cuadro 2) herramientas como LACA (Chew *et al.*, 2023) o CHALET (Meng *et al.*, 2024), que son enfoques metodológicos que integran LLM, como ChatGPT, en el proceso de codificación deductiva (LACA) o deductivo/inductivo (CHALET) en investigaciones cualitativas.

QualiGPT (Zhang *et al.*, 2024) es una herramienta basada en modelos de lenguaje (como ChatGPT), diseñada para apoyar el análisis cualitativo de datos. Si bien *QualiGPT* carece de *software* independiente, ya que usa la interfaz de ChatGPT (o, mediante GitHub, permite instalación en local en entorno Python), se trata de un enfoque personalizado que utiliza modelos de lenguaje (como ChatGPT) configurados específicamente para el análisis cualitativo, basado en métodos como codificación inductiva y deductiva y en principios de investigación cualitativa como teoría fundamentada, análisis temático y codificación reflexiva. Se trata de una herramienta orientada por los principios de *transparencia* y *reflexividad*, ya que ofrece justificaciones de codificación, reflexiones analíticas y trazabilidad de decisiones. Otras características destacables son su rapidez y agilidad, con la velocidad de análisis de ChatGPT para generar códigos y temas en minutos o segundos, o la existencia de flujos automatizados, lo que reduce el tiempo de configuración manual o la inclusión de plantillas de *prompts* inspiradas en investigaciones por pares.

Mindcoder (Gao *et al.*, 2025) es una aplicación web diseñada específicamente para el análisis cualitativo cuyo objetivo principal es automatizar y facilitar el proceso de codificación cualitativa, proporcionando una herramienta accesible para investigadores sin necesidad de conocimientos de programación. A través de la interfaz web interactiva intuitiva y accesible, pretende cerrar la brecha entre las herramientas de IA profesionales (Atlas Ti, N Vivo) y los modelos de lenguaje conversacionales (Claude o Chat-GPT). Utiliza cadenas automatizadas de razonamiento basadas en la técnica de Chain-of-Thought Prompting, que permite realizar análisis cualitativos estructurados en pasos como: reprocesamiento de datos, codificación abierta automática, codificación axial automática, desarrollo conceptual y generación de informes.

Cuadro 2
Nuevas herramientas de software que usan LLM para la codificación cualitativa según criterios mencionados

Herramienta software	Entradas	Salidas	Interfaz	Código abierto	Fuente web
QualiGPT (Zhang <i>et al.</i> , 2024)	Formatos csv, formatos de Excel y de texto plano	Puede generar tablas, resúmenes o listas de códigos	Interfaz de ChatGPT (OpenAI) o instalación local en Python	Sí, bajo licencia MIT	chatgpt.com/g/g-HtBvI9uXe-qualigpt github.com/KindOPSTAR/QualiGPT
Mindcoder (Gao <i>et al.</i> , 2025)	Formatos .txt y .docx	Genera cuadros con clústeres, etiquetas de códigos, conceptualización y representaciones visuales	Interfaz propia en plataforma online	No, es una web con derechos reservados	mindcoder.ai
Thematic-LM (Qiao <i>et al.</i> , 2025)	Formatos .csv, .json (con opciones para definir comportamiento de los agentes)	Códigos temáticos, libros de códigos, asignación a fragmentos de entrada, mapa temático e informe estructurado	Se ejecuta en entornos de programación como Jupyter Notebooks o directamente en Python	Sí	github.com/sjdai/LLM-thematic-analysis
Argument2Code (Zhao <i>et al.</i> , 2024)	n/c	n/c	No tiene interfaz específica, sino que es un conjunto de herramientas que se integra en un flujo de trabajo existente	No	n/c

Fuente: elaboración propia.

Thematic-LM (Qiao *et al.*, 2025) es un sistema de análisis temático computacional diseñado para realizar análisis temáticos que asigna tareas especializadas a cada agente (un componente del sistema), como la codificación, la agregación de códigos y el mantenimiento y actualización del libro de códigos, lo que permite un análisis más eficiente y escalable (maneja grandes volúmenes de datos sin perder rendimiento). Está diseñado para investigadores/as con conocimientos en programación y acceso a APIs de LLM, ya que la persona usuaria debe ejecutar *scripts* y configurar los parámetros manualmente. Esto implica: a) preprocesar los datos; b) definir cómo se invocan los agentes (por ejemplo, codificador, agregador); c) especificar cómo se almacenan y visualizan los resultados.

Argument2Code (Zhao *et al.*, 2024) es un sistema automatizado avanzado diseñado para generar libros de códigos inductivos y extraer temas emergentes sin necesidad de un marco teórico predefinido mediante un proceso de múltiples etapas que emplea técnicas de «chain-of-thought prompting» (cadenas de razonamientos) como técnica que guía al modelo a través de una serie de pasos lógicos para mejorar la coherencia y profundidad en la generación de códigos. Este enfoque permite una exploración más abierta y flexible de los datos, facilitando la identificación de patrones y conceptos clave directamente desde el contenido analizado.

En fin, el uso de LLM en la codificación cualitativa ha dado lugar a una variedad de enfoques y herramientas que aportan automatización y, por tanto, una inusitada agilidad al análisis cualitativo. En particular, las herramientas *software* como QualiGPT,

Mindcoder y Argument2Code muestran un esfuerzo creciente por integrar capacidades avanzadas de los LLM en entornos accesibles y metodológicamente informados. Sin embargo, otras herramientas como Thematic-LM presentan mayor complejidad técnica para su implementación a pesar de tratarse de un *software* de código abierto. De algunas herramientas, como Argument2Code, carecemos de suficiente información al no ofrecer interfaz web o de escritorio ni facilitarse el código en abierto. En definitiva, en general, estas herramientas apuntan hacia una transformación significativa de los procesos de codificación, planteando oportunidades, pero también retos, en términos de transparencia, control humano e interpretación crítica, como se mencionará a continuación.

4. Los desafíos de los LLM

El uso de los LLM en la investigación cualitativa ha suscitado un creciente interés, al tiempo que plantea importantes desafíos metodológicos, éticos y epistemológicos. A medida que estas herramientas se incorporan a los procesos de análisis de datos, es fundamental reflexionar críticamente sobre sus limitaciones, especialmente en lo que respecta a la transparencia de sus operaciones, la fiabilidad de sus resultados y el impacto potencial de los sesgos que pueden arrastrar desde sus conjuntos de entrenamiento. Diversos autores han advertido que, si bien la generativa puede ofrecer soluciones innovadoras para la automatización de tareas y la exploración de patrones en grandes volúmenes de datos, su aplicación exige cautela, tanto por la posibilidad de reproducir sesgos sociales y culturales como por los dilemas éticos vinculados a la privacidad de los datos y a la replicabilidad de los hallazgos. A continuación, se presentan algunas de estas limitaciones, señaladas por distintas investigaciones recientes.

Morgan (2023) señala algunos problemas del uso de la IAG para el análisis cualitativo. El primero, los sesgos racistas, sexistas o de otro tipo que se producen debido a límites potenciales en el conjunto de entrenamiento de la IA a través de internet (lleno de sesgos racistas o sexistas y de todo tipo). Por ejemplo, los desarrolladores de ChatGPT han descrito sus considerables esfuerzos para capacitar al programa para detectar este tipo de sesgos y excluirllos tanto de las consultas que acepta como de las respuestas que proporciona, lo que idealmente podría sortearse con una correcta formulación de las preguntas. Sin embargo, también advierte que es necesario contemplar la posibilidad de que esa lucha contra esos sesgos podría ser problemática cuando el objeto sea precisamente identificar esos sesgos, lo que resulta frecuente en la investigación cualitativa.

Una segunda limitación es su capacidad para producir respuestas sin sentido, un proceso que a veces se ha descrito como «alucinación» (Lakshmanan, 2022). Como advierte la propia herramienta cuando se le pregunta en este sentido por su tendencia a la predicción probabilística del lenguaje, «el modelo genera respuestas basadas en la probabilidad de que una secuencia de palabras sea coherente y relevante, pero no necesariamente correcta (ya que) este enfoque probabilístico prioriza la fluidez y la coherencia en lugar de la exactitud factual» (ChatGPT, v.4 2024).

Un tercer problema tiene que ver con cuestiones éticas, especialmente relacionadas con el acceso a datos privados, incluso para el entrenamiento de los modelos (Marshall *et al.*, 2024; Head *et al.*, 2023). Por tanto, a menos que los datos sean lo suficientemente anonimizados, la reutilización de los datos de los participantes podría amenazar su privacidad. Además, si bien no hay evidencia de que el contenido de ChatGPT o de cualquier otra IA similar haya sido amenazado alguna vez, no se asegura la protección de la privacidad en algún posible escenario (Morgan, 2023).

Meng *et al.* (2024) consideran algunos desafíos ante los que QualiGPT estaría respondiendo de forma satisfactoria:

- a. Falta de transparencia. La naturaleza de «caja negra» de los LLM dificulta que se comprenda cómo procesa los datos, para lo que resulta indispensable mejorar la interpretabilidad y la transparencia.
- b. Problemas de consistencia y comprensión del contexto. Las respuestas pueden variar y es difícil mantener la coherencia en diálogos de múltiples turnos, por lo que *prompts* estructurados y precisos pueden mejorar la consistencia.
- c. Dificultad en el diseño de *prompts*. Crear *prompts* efectivos requiere mucho tiempo y carece de un enfoque estandarizado y QualiGPT simplifica este proceso mediante *prompts* predefinidos.
- d. Dificultades en la comprensión de las respuestas del LLM (ChatGPT, en este caso), por lo que se pueden diseñar *prompts* que estandaricen las salidas para mejorar la legibilidad.
- e. Privacidad y seguridad de los datos. En la era digital, la privacidad de los datos es una preocupación fundamental, especialmente al utilizar modelos de lenguaje y existen riesgos significativos de filtración de datos sensibles, como lo han demostrado incidentes históricos⁵.

Meng *et al.* (2024) mencionan la necesidad de introducir una adecuada «ingeniería de prompts» (*prompt engineering*) en los LLM que procedan a la codificación cualitativa. El cofundador de Open-AI, Greg Brockman, definió la «ingeniería de prompt como el arte de comunicarse elocuentemente con una IA»⁶. Rossi profundiza en el problema de los *prompts* en la producción y replicabilidad de los resultados. Comienza describiendo la ingeniería de *prompts* como «un proceso de ajuste necesario para obtener las mejores salidas o respuestas del LLM» (2024, pp. 155–156) y continúa cuestionando la asunción de que, con el mismo *prompt*, se producirían los mismos o datos similares en el futuro. Así pues, la inestabilidad de las salidas es una característica generalizada en el uso de los LLM, dado que los datos generados varían por pequeños cambios en la redacción de los *prompts*, incluso varían con el mismo *prompt* utilizado en diferentes momentos.

Una línea que queda definitivamente abierta es la posibilidad de que en el futuro puedan desarrollarse LLM específicamente diseñados, entrenados y optimizados

para la investigación (Bail, 2024). Y en este sentido señalan que los modelos de lenguaje de código abierto son la mejor alternativa (Spirling, 2023) al ofrecer mayor transparencia, mejor control y la posibilidad de ser entrenados con datos específicos para la investigación.

En este sentido, un inconveniente serio para la utilización de estos sistemas es lo escasamente amigables que resultan para aquellas/os sociólogas/os o investigadores sociales con escasos conocimientos computacionales. Si bien es cierto que se ha tratado de sortear este inconveniente a través de algunas guías prácticas de uso «paso a paso» de LLM (Törberg, 2023), la integración del uso de estos modelos en interfaces de más fácil uso para la investigación cualitativa podría ser de gran utilidad.

Por último, conviene mencionar algunas consideraciones de índole práctica que aportan autores como Marshall *et al.* sobre la prudencia con que se adopta cualquier nueva tecnología y los efectos sobre las publicaciones de artículos: «[...] habrá algunos que optarán por esperar para adoptar su uso y estas personas seguirán siendo revisores de revistas académicas o árbitros de propuestas presentadas en conferencias académicas, y eso seguirá impactando a los investigadores cualitativos [...] los resultados de nuestra encuesta sugieren que muchos revisores [...] serán menos propensos a aceptar un artículo que describa una investigación que utiliza IA» (2024, p. 98), como efecto no deseado en este proceso aquí descrito.

En resumen, las principales limitaciones de la IA vienen en el sentido de su incapacidad de interpretar el significado subyacente de los datos, lo que requiere una comprensión profunda del contexto y la teoría; los sesgos, que la IA puede contribuir a perpetuarlos en lugar de a solventarlos; la falta de transparencia, ya que la IA no puede explicar su proceso de decisión, lo que puede generar preocupaciones sobre la transparencia y la rendición de cuentas.

5. El rol de la persona investigadora respecto a los LLM

Aunque los LLM pueden procesar el lenguaje con una fluidez impresionante a través del acceso a vastos repositorios de información, carecen de comprensión genuina, autoconciencia o la capacidad de razonar sobre el mundo como lo hacen los humanos (Mitchell, 2023). Precisamente por ello, el juicio, la interpretación crítica y la experiencia humana siguen siendo fundamentales para guiar y validar la investigación asistida por LLM. En términos parecidos, Jiang *et al.* (2021) señalan que, si bien los participantes luchan con el desorden y la incertidumbre del análisis cualitativo, también quieren la plena autonomía del proceso e insisten en que la IA no se la debe arrebatar.

Si bien se asume que la relación debe ser de complementariedad, esto no impide que deba reconceptualizarse el rol investigador. Se estima que la persona investigadora debe centrarse en reflexionar críticamente a partir de los análisis compu-

tacionales (Li *et al.*, 2024) y debe seguir siendo quien dé sentido a la información, evaluando las sugerencias del modelo a la luz de su conocimiento disciplinario y de las realidades empíricas en estudio. En este sentido, Christou (2023) propone una relación basada en principios de *precisión, fiabilidad, justificación y ética*, asegurando la participación activa de los investigadores con sus habilidades cognitivas evaluativas para monitorear, documentar procesos y alcanzar conclusiones. Por su parte, Schreder *et al.* (2025), a partir de los discursos recabados por investigadores cualitativos familiarizados con los LLM, apuntan a la necesidad de herramientas que faciliten, entre otras cuestiones, el proceso de codificación, aunque asistiendo de manera *intensamente reflexiva* en dicho proceso.

Siguiendo a Christou (2023), cabe realizar una serie de advertencias sobre cómo la persona investigadora debe intervenir en el proceso de relación con los datos: a) familiarizarse completamente con los datos para comprenderlos en su totalidad e identificar cualquier sesgo o preconcepción inherente; b) garantizar datos de entrenamiento diversos y no sesgados e implementar medidas de transparencia y responsabilidad; c) cualquier contenido producido por la IA debe verificarse mediante referencias cruzadas para garantizar su precisión y credibilidad; d) revisar minuciosamente los resultados antes de llevar a cabo cualquier discusión teórica o conceptual; e) participar activamente utilizando su conocimiento previo y extenso sobre el fenómeno en estudio.

En este contexto, cabe plantearse la pregunta relevante de cómo deben ser las herramientas que se propongan a los y las investigadores/as. En este sentido, Schreder *et al.* (2025) señalan que las herramientas que usen LLM deben diseñarse conforme determinadas claves que pueden visualizarse en el cuadro 3.

Cuadro 3
Diseño de herramientas LLM según finalidad, siguiendo a Schreder et al. (2025)

Privacidad de los participantes	Para un uso intencional	Para mayor transparencia y validación	Para una interacción profunda con los datos
Más herramientas que faciliten el alojamiento local, la personalización y el ajuste fino de modelos de código abierto.	Control sobre cuándo y cómo son asistidos o influenciados por IA.	Posibilidad de evaluar el rendimiento de manera transparente.	Favorecer herramientas basadas en LLM que fortalezcan su relación con los datos (en lugar de que se distancien).
Código abierto como posible solución a problemas de privacidad y para mayor control y transparencia.	Flexibilidad e interactividad. Los LLM basados en chat permiten, pero actualmente <i>no apoyan</i> el uso intencional para tareas específicas en el proceso de investigación.	Para el enfoque positivista, son útiles funciones que garanticen la reproducibilidad de los resultados, la fiabilidad entre evaluadores y el análisis de errores en la anotación.	Preocupación por el rendimiento variable de los LLM en diferentes contextos, dominios de conocimiento, culturas e idiomas.
Las herramientas deben ser explícitas sobre si llaman a APIs externas, cuándo y cómo.	Las herramientas deberían diseñarse para guiar a los usuarios a seleccionar de manera intencional los usos apropiados de los LLM.	Incluir funciones para comprender o analizar los resultados y sugerencias.	LLM como una forma de generar enfoques más únicos para comprender los datos, ayudándolos a examinar o desafiar sus teorías existentes con la evidencia directa de los datos.
Deber de informar a los/as usuarios/as, antes de cargar datos, sobre cómo manejan la privacidad: cómo pueden anonimizar los datos; cómo eliminar los datos; aclarar si los insumos se utilizan para entrenar modelos.	Las herramientas deberían ofrecer funciones que permitan a los investigadores/as desarrollar sus propias ideas.		Diseñar considerando las perspectivas e intereses de los participantes.
			Ofrecer la opción de utilizar modelos que reflejen mejor un grupo objetivo.
			Posibilidad de incorporar su propio contexto al realizar análisis, incluyendo experiencia previa, los textos que influyen en su trabajo y las teorías.

Elaboración propia.

Después de analizar en profundidad diversos modelos LLM, entre ellos FLAN-5 (en diversas versiones) y Chat-GPT (en versiones 3 y 4), Ziems *et al.* (2024) concluyen que estos modelos pueden mejorar, pero no simplemente reemplazar los procesos tradicionales y plantea una serie de recomendaciones que van en el sentido de: a) mejorar el proceso de etiquetado de datos, especialmente cuando se manejan grandes volúmenes de datos; b) flexibilidad para modificar y adaptar el modelo a las necesidades específicas de la investigación, asegurando el control sobre las implicaciones éticas; c) priorizar fidelidad, relevancia, coherencia y fluidez optando por modelos más grandes ajustados a instrucciones que hayan aprendido las preferencias humanas; d) optar por LLM de código abierto para clasificación, en lugar de depender de modelos propietarios o cerrados.

Sobre la importancia de las herramientas de código abierto, Van Dis *et al.* (2023) señalan que uno de los problemas más acuciantes a los que se enfrenta el uso de los

LLM por parte de los investigadores es el cuasi-monopolio tecnológico en el que operan los LLM que tienden a ser productos patentados por un pequeño número de grandes empresas tecnológicas que cuentan con los recursos necesarios para su desarrollo de IA, lo que plantea preocupaciones éticas considerables (*ibid.*), especialmente relacionadas con la falta de transparencia, dado que «las empresas tecnológicas podrían ocultar el funcionamiento interno de sus IAs conversacionales», lo que «va en contra del movimiento hacia la transparencia y la ciencia abierta, dificultando la identificación del origen de la información [...]», haciéndose necesario «el desarrollo e implementación de tecnología de IA de código abierto» (*ibid.*, p. 225).

En definitiva, el uso de LLM en la investigación cualitativa exige un equilibrio cuidadoso entre automatización y control humano, eficiencia y reflexión, asistencia y autonomía. Por tanto, es crucial reconocer el papel activo del investigador/a como codificador/a, incluso en entornos con LLM, para evitar la ilusión de automatización total o de *neutralidad algorítmica*. Las herramientas basadas en IA no deben suplantarse el juicio del o la investigador/a, sino acompañarlo en un proceso riguroso, ético y transparente. Tal como apuntan las investigaciones revisadas, el desarrollo de estos sistemas debe estar guiado por principios de apertura, contextualización y diseño centrado en el usuario, asegurando que la tecnología amplíe —y no limite— las capacidades interpretativas, analíticas y creativas de quienes investigan.

6. Ciencia abierta, LLM e investigación cualitativa en Social Verbatim

Se ha dicho que, tanto en el ámbito de las ciencias sociales como en otros muchos, la IAG abrirá «camino revolucionarios para la razón humana muy distintos a los de la ciencia de la ilustración que se abrió paso a través de la lógica de la inducción y acumulación paciente de evidencias» (Peters *et al.*, 2023, p. 832), de modo análogo a como lo hizo la imprenta de Gutenberg (Kissinger *et al.*, 2023). Y esto ocurre en la medida que emplea métodos que generan resultados «sin explicar por qué o cómo funciona su proceso [...] basándose en representaciones pregeneradas de los vastos océanos de datos en los que fue entrenado» (Peters *et al.*, 2023, p. 832).

Así, pues, los límites en transparencia y replicabilidad constituyen unos de los desafíos más notables de la relación entre ciencia e IA. De manera análoga, la investigación cualitativa ha tendido a adolecer de ciertas limitaciones en este sentido (Jiang *et al.*, 2021). Por ejemplo, respecto a la transcripción, Macmullin (2023) encontró que el 41% de las investigaciones analizadas omitían absolutamente la transcripción (aun habiendo sido realizada), el 11% se referían a que se habían obtenido las transcripciones, pero ignoraban el modo en que se había producido el proceso y el 19% recogían una sencilla frase que advertía que «las entrevistas fueron grabadas y transcritas». Nascimento (2019) afirmó que los estudios cualitativos (en este caso, sobre gestión) no suelen abundar en el modo de la transcripción más allá de una simple frase que dice que «las entrevistas fueron transcritas».

Las limitaciones, no siempre imputables a un pretendido distanciamiento de la tradición positivista, han tratado de ser contrarrestadas mediante propuestas de un mayor rigor, la fiabilidad y la validez de este tipo de análisis, y en este contexto la IA, que se está viendo como «como una opción para apoyar a los académicos cualitativos con su investigación» (Jiang *et al.*, 2021, p. 94), supone un nuevo desafío.

En particular, resulta esencial avanzar en el sentido de explicitar la *caja negra* para dar a conocer *cómo funciona su proceso* de los LLM. Wang *et al.* (2019), que preguntaron a investigadoras/es en cualitativo, reiteraron la importancia de la transparencia de la IA y Yang *et al.* (2019) propusieron la idea de una «IA sin complicaciones», basándose en la idea de que interacción con los humanos debería tener el nivel adecuado de «normalidad» que incluiría precisamente conocer las motivaciones de su acción.

Un ejemplo de cómo es posible explicitar el funcionamiento de la *caja negra* proviene del mundo de la medicina. IBM Watson Health⁷ es un sistema de inteligencia artificial aplicada en el diagnóstico médico. Aunque la IA que emplea Watson para el análisis de datos médicos es compleja, los médicos pueden obtener explicaciones detalladas de cómo la IA llegó a una conclusión o recomendación, proporcionando las razones detrás de sus diagnósticos, argumentos, incluso el contexto o qué datos o síntomas fueron más relevantes para llegar a una conclusión. De forma análoga, deberíamos disponer en ciencias sociales de instrumentos capaces de hacer lograr que la IA se convierta en una herramienta confiable y complementaria en lugar de una *caja negra* difícil de interpretar.

Resulta, por tanto, un reto sugerente avanzar en esta dirección de combinar el uso de tecnologías que faciliten la agilidad en los procesos de la investigación cualitativa con el máximo rigor, control (humano) y transparencia en la investigación cualitativa y en los procesos que utilice (por ejemplo, relacionados con el empleo de LLM). Como se mencionaba en la introducción, en Social Verbatim se presentan algunas posibilidades de utilización de la IA para afrontar de manera práctica estos desafíos que nos permitan avanzar hacia un modelo de *ciencia abierta* en investigación cualitativa (Brezna, 2021).

La herramienta Social Verbatim se desarrolló a partir de un estudio sistemático basado en las aportaciones recabadas mediante entrevistas realizadas en el marco de la investigación «CS-transcribe: Investigación de Necesidades y Desarrollo de una Herramienta Digital de Transcripción para las Ciencias Sociales» a tres perfiles de potenciales usuarios y usuarias: investigadoras/es, transcriptoras e investigadoras-transcriptoras. En total, se realizaron 15 entrevistas individuales en las que participaron 11 mujeres y 4 hombres, incluyendo 7 investigadoras/es (4 mujeres y 3 hombres), 5 investigadoras/es-transcriptoras/es (4 mujeres y un hombre) y 3 transcriptoras (mujeres).

Estas entrevistas se realizaron a partir de una primera versión *demo* de la herramienta que incluía funcionalidades que fueron ampliadas, de acuerdo con las orientaciones facilitadas a través de las entrevistas, en una versión *beta* de la aplicación, en la que se contaba con las siguientes características:

- Plataforma *online* con transcripción automática de vídeos o audios, y con posibilidad de transcripción manual.
- Interfaz para inserción de acotaciones de comunicación no verbal mediante iconos.
- Experiencia de usuario/a (UX) que da prioridad en la agilidad de revisión/corrección.
- Funcionalidad de anonimización de datos.
- Interfaz para la codificación.
- Detección automática de silencios y ausencias de palabras y marcas automatizadas de tiempo.
- Herramientas que ayudan al trabajo colaborativo de los equipos.
- Orientación a la gestión del proyecto, con información específica de cada transcripción (transcriptor, estado de evolución...).
- Interfaz de pedales de transcripción para una experiencia más fluida y ergonómica en la revisión y en la transcripción manual.
- Interfaz de informes que incluyen listado de verbatims, estadísticas, gráficos y nubes de palabras y de códigos.
- Sistemas de importación y exportación flexible: entrada y salida de datos según necesidades de trabajo (como Atlas.ti o NVivo).

En relación con el seguimiento del paradigma de *ciencia abierta*, Social Verbatim aporta aspectos como los siguientes. En primer lugar, permite el acceso de cualquier usuario/a al contenido de una transcripción citada en un artículo, así como a su fuente sonora (y visual, en su caso), una vez que la información ha sido debidamente anonimizada (omitiendo referencias que permitan la identificación y proporcionando voces y, en su caso, imágenes distorsionadas para que resulten irreconocibles). Esto permitirá que la persona usuaria pueda entrar en la *caja negra* del proceso de transcripción y compruebe cómo se ha efectuado dicha transcripción identificando posibles inconsistencias en el proceso que pueda afectar a los resultados. En segundo lugar, se permite el acceso de cualquier usuario/a al proyecto de transcripción para tener información de conjunto o información más precisa sobre aspectos descriptivos de las personas entrevistadas (esto siempre que, a juicio de los/as investigadores/as, haga posible la preservación de la privacidad) a través de un sistema previamente anonimizado⁸ y de navegación ágil⁹.

En la siguiente fase de desarrollo de Social Verbatim, nos disponemos a avanzar en la introducción de LLM en el proceso de codificación y análisis de los resultados, siguiendo las premisas que se indican a continuación y que se sostienen en el paradigma de la *ciencia abierta*. No obstante, en el caso de la investigación cualitativa, avanzar en este paradigma encuentra un posible factor de colisión que se refiere a la indispensable necesidad de proteger la privacidad de los participantes (Gómez *et al.*, 2025), además de las mencionadas dificultades relacionadas con la «oscuridad del algoritmo» que pretendemos minimizar a través de las siguientes estrategias: a) incorporación de LLM de código abierto para ofrecer información del proceso de análisis; b) interfaz que permita la trazabilidad de los *prompts* introducidos en el desarrollo del análisis; c) supervisión y revisión activa de la persona investigadora en

el proceso de codificación, análisis y ofrecimiento de *feed-back* a la herramienta para redefinir los criterios de análisis; d) introducción de elevados estándares de privacidad en la gestión de la información con independencia del tipo de usuario/a (premium o usuarios/as en general); e) en relación con las recomendaciones ofrecidas por Schreder *et al.* (2025): 1) flexibilidad e interactividad, superando el problema de que los LLM *basados en chat* permiten, pero actualmente *no facilitan*, el uso intencional para tareas específicas en el proceso de investigación; 2) interfaz con funciones que permitan a los investigadores/as desarrollar sus propias ideas, ofreciendo una interfaz que fortalezca la relación de la persona investigadora con los datos (en lugar de distanciarlos).

7. Conclusiones

La revisión teórica sobre el uso de Modelos de Lenguaje de Gran Escala (LLM) en la investigación cualitativa ha permitido identificar avances significativos y desafíos persistentes en este campo. La arquitectura de Transformers y el avance del *hardware*, especialmente las GPU y TPUs, han sido fundamentales para el desarrollo de los LLM, permitiendo procesar grandes volúmenes de texto de manera eficiente. En este sentido, herramientas como QualiGPT, Mindcoder y Thematic-LM han sido diseñadas para facilitar el análisis cualitativo, proporcionando interfaces intuitivas y funcionalidades avanzadas para la codificación y el análisis de datos, mediante el mejor aprovechamiento de la llamada ingeniería de *prompts* para optimizar las entradas a los modelos y obtener respuestas más precisas y útiles.

Sin embargo, la falta de transparencia y la naturaleza de «caja negra» de los LLM plantean desafíos significativos, por lo que resulta crucial mejorar la interpretabilidad y la transparencia de estos modelos para que los investigadores puedan comprender y confiar en los resultados. Además, se ha aludido a los sesgos inherentes en los datos de entrenamiento que pueden perpetuar prejuicios sociales y culturales, así como a la privacidad y la seguridad de los datos como preocupaciones fundamentales que garanticen que los datos utilizados para entrenar y operar los LLM estén adecuadamente anonimizados y protegidos.

Por su parte, a pesar de los avances tecnológicos, los LLM deben ser vistos como herramientas complementarias que apoyan, pero no reemplazan, el juicio y la interpretación y análisis humano. Las personas investigadoras deben estar activamente involucradas en el proceso de codificación y análisis, utilizando su conocimiento disciplinario para guiar y validar los resultados generados por los LLM.

La integración de LLM en herramientas como Social Verbatim espera contribuir a una investigación cualitativa abierta y rigurosa en la medida en que estas herramientas faciliten la trazabilidad de los procesos y la transparencia en la gestión de datos. Social Verbatim surge, de hecho, como una propuesta que, más allá de automatizar tareas, busca acompañar a las y los investigadores en la construcción de prácticas científicas más abiertas y rigurosas, facilitando el acceso y revisión de cada paso en

la gestión y análisis de datos cualitativos, avanzando hacia un ecosistema de investigación en el que la colaboración humano-máquina no solo sea eficaz, sino también comprensible y verificable.

En resumen, los LLM representan una oportunidad significativa para mejorar la eficiencia y profundidad del análisis cualitativo, pero su implementación debe ser cuidadosamente gestionada para abordar los desafíos metodológicos, éticos y epistemológicos. La colaboración humano-máquina, la transparencia y la ciencia abierta son principios fundamentales para avanzar en este campo.

8. Financiación

«CS-transcribe: Investigación de Necesidades y Desarrollo de una Herramienta Digital de Transcripción para las Ciencias Sociales» (TED2021-131729B-I00) es un proyecto financiado por MCIN/AEI y por la Unión Europea, Next Generation (PRTR). IPs: Juan Miguel Gómez Espino y Carles Xavier Simó Noguera.

9. Bibliografía

- Arlinghaus, C. S., Wulff, C., Maier, G. W., Arlinghaus, C., Wulff, C. y 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. y 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. y 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. y 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. y Ku, L. W. (2023). *LLM-in-the-loop: Leveraging large language model for thematic analysis*. arXiv preprint arXiv:2310.15100.

- Deterding, N. M. y 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. y Yeo, S. Y. (2025). *Using Large Language Model to Support Flexible and Structural Inductive Qualitative Analysis*. arXiv preprint arXiv:2501.00775.
- Glasser, B. G. y Strauss, A. L. (1967). *The development of grounded theory*. Chicago, IL: Alden.
- Gómez-Espino, J. M., Simó-Noguera, C. y Carvajal-Soria, P. (2025). *Ciencia abierta y procesos de investigación cualitativa en la app Social Verbatim*. Pendiente de publicación.
- González-Veja, A. M. D. C., Sánchez, R. M., Salazar, A. L. y 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. y 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. y Huttenlocher, D. (2023), February 24. ChatGPT heralds an intellectual revolution. *The Wall Street Journal*, 24 de febrero. <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. y Su, Y (2024, Mayo). A map of exploring human interaction patterns with LLM: Insights into collaboration and creativity. En *International conference on human-computer interaction* (pp. 60–85). Springer Nature Switzerland.
- Linneberg, S. M. y Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. *Qualitative Research Journal*, 19(3), 259–270.
- Marshall, D. T. y 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. y 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. y 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. En M. C. Frank y A. Majid (Eds.), *Open Encyclopedia of Cognitive Science*. MIT Press. <https://doi.org/10.21428/e2759450.2bb20e3c>
- Mitchell, M. y 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. y 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. y Koh, Y. S. (2025). Thematic-LM: a LLM-based Multi-agent System for Large-scale Thematic Analysis. En *The Web Conference 2025*.
- Rossi, L., Harrison, K. y 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. y 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. y 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. y 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. y Polosukhin, I. (2017). Attention Is All You Need. En *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. y 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. y 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. y Oudeyer, P. Y. (2023, March). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. En *Companion proceedings of the 28th international conference on intelligent user interfaces* (pp. 75-78).
- Yang, Y. 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., ... y 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. y Shang, Y. (2024). A New Method Supporting Qualitative Data Analysis Through Prompt Generation for Inductive Coding. En *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. y 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

Profesor titular de Sociología de la Universidad Pablo de Olavide (Sevilla). Desde 2001 ha compaginado la docencia y la investigación en sociología de la infancia y la educación con labores de gestión universitaria (actualmente como decano de la Facultad de Ciencias Sociales). Licenciado en Ciencias Políticas y Sociología por la Universidad de Granada, y doctor desde 2009, ha escrito varias publicaciones en revistas de impacto como *Revista Española de Investigaciones Sociológicas*, *Empiria*, *Childhood*, o *International Sociology* o *Language and education* o capítulos libros en editoriales como Springer. Además, ha sido profesor invitado en la Universidad de Sheffield (Reino Unido) y Gramma (Cuba).

Notas

1 Esta publicación forma parte del Proyecto de I+D+i «CS-transcribe: Investigación de Necesidades y Desarrollo de una Herramienta Digital de Transcripción para las Ciencias Sociales», ref. TED2021-130903B-I00, financiado por el Ministerio de Ciencia, Innovación y Universidades, por la Unión Europea NextGenerationEU/PRTR, Agencia Estatal de Investigación/10.13039/501100011033

2 El acceso a la app se hace a través de app.socialverbatim.com

3 Social Verbatim (s.f.). Social Verbatim. <https://www.socialverbatim.com> (consultado el 25 de abril de 2025).

4 Una de las derivadas de la emergencia de este tipo de modelos es la importancia que adquiere lo que se denomina «ingeniería de prompts» que se refiere a la práctica de diseñar y optimizar las entradas que se dan a un modelo para obtener respuestas más precisas, útiles o alineadas con una necesidad específica.

5 Según las políticas de OpenAI, los datos introducidos por los usuarios en ChatGPT pueden utilizarse para mejorar los modelos, salvo que se desactive esta opción en la configuración. En cambio, los datos enviados a través de la API no se utilizan para entrenar los modelos de OpenAI. Esto sugiere que el uso de la API proporciona una mayor garantía de privacidad y seguridad de los datos.

6 G. Brockman [@gdb] (2023, March 11). 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 En cuanto a la distorsión de la voz, es evidente que algunos matices del habla natural se podrían perder, por lo que se trataría de un *trade off* entre riqueza lingüística o paralingüística y privacidad.

9 Al hacer accesibles los datos se podría combinar en el futuro con herramientas como Zenodo o similares con permisos y restricciones, ya que en este caso la *ciencia abierta* tendría un límite prevalente que es la preservación de la privacidad.