Methods in Psychology 12 (2025) 100183

Contents lists available at ScienceDirect

Methods in Psychology

journal homepage: www.sciencedirect.com/journal/methods-in-psychology

Using large language models for narrative analysis: a novel application of generative AI

Sarah Jenner [a,b,c,*], Dimitris Raidos [d], Emma Anderson [e], Stella Fleetwood [d], Ben Ainsworth [a], Kerry Fox [e], Jana Kreppner [a], Mary Barker [b,c,f]

[a] School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK
[b] MRC Lifecourse Epidemiology Centre, Faculty of Medicine, University of Southampton, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK
[c] School of Health Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK
[d] Ipsos, 3 Thomas More Square, London, E1W 1YW, UK
[e] School of Humanities & Social Science, University of Brighton, Mithras House, Lewes Road, Moulsecoomb, BN2 4AT, UK
[f] NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK

ARTICLE INFO

Keywords:
Artificial intelligence
Large language models
Narrative analysis
Story completion
Adolescent health
Health psychology

ABSTRACT

This study, a collaboration between the University of Southampton and Ipsos UK, aimed to develop and test a novel method for analysing qualitative data using generative artificial intelligence (AI). It compared large language model (LLM)-conducted analysis with human analysis of the same qualitative data, explored optimisation of LLMs for narrative analysis and evaluated their benefits and drawbacks. Using existing data, 138 short stories written by young people (aged 13–25 years) about social media, identity formation and food choices were analysed separately three times: by human researchers, and by two different LLMs (Claude and GPT-o1). The method was developed iteratively, combining Ipsos' artificial intelligence (AI) expertise and tools with researchers' qualitative analysis expertise. Claude and GPT-o1 each conducted a narrative analysis of all 138 stories using the same analytic steps as the human researchers. Findings between the humans and both LLMs were then compared. Both LLMs quickly and successfully conducted a narrative analysis of the stories. Their findings were comparable to those of the human researchers and were judged by the researchers to be credible and thorough. Beyond replication, the LLMs provided additional insights into the data that enhanced the human analysis. This study highlights the significant potential benefits of LLMs to the field of qualitative research and proposes that LLMs could one day be seen as valuable tools for strengthening research quality and increasing efficiency. Additionally, this study discusses ethical concerns surrounding responsible AI use in research and proposes a framework for using LLMs in qualitative analysis.

1. Introduction

Recent advances in generative artificial intelligence (AI) have produced excitement and concern in equal measure across the scientific community (The Lancet, 2024). Since the release of ChatGPT (OpenAI, 2025) in November 2022, a rapidly growing body of research has discussed the potential benefits and pitfalls of using AI in research. The focus of many of these early studies has been exploring the use of publicly accessible large language models (LLMs), including ChatGPT, for tasks such as textual data analysis. ChatGPT is a web interface that allows users to communicate with a type of LLM called a 'generative pre-trained transformer', or 'GPT'. *Generative* refers to the model's text generating abilities, *pre-trained* means that the model has been subjected to both supervised and unsupervised training by developers using vast online data resources, and *transformer* refers to the model's ability to examine specific details within the structure and content of text in order to analyse input and produce output (Radford et al., 2018). Other GPT-style LLMs that are free to use (without paying for access to a more advanced version of the software) include Claude (Anthropic, 2025), Gemini (Google, 2025), Llama (Meta, 2025) and DeepSeek (Wenfeng, 2025). The development of these highly competent and intuitive LLMs means that AI is now accessible to anyone with an internet connection, not just those with knowledge and experience of software programming.

LLMs are trained based on principles of natural language processing

* Corresponding author. School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK.
E-mail address: sj@mrc.soton.ac.uk (S. Jenner).

https://doi.org/10.1016/j.metip.2025.100183
Received 21 February 2025; Received in revised form 9 April 2025; Accepted 10 April 2025
Available online 11 April 2025
2590-2601/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(NLP), a field of study which aims to improve the efficiency of human-computer interaction (Jain et al., 2018). LLMs can identify patterns within text and mimic human language, so appear well-placed to assist researchers in conducting qualitative analysis. There are, however, important ethical, methodological and philosophical issues that should be considered before using AI for analysis of research data (Sallam, 2023; Schlagwein and Willcocks, 2023), which will be discussed in this paper.

One well-recognised problem with LLMs is their tendency to 'hallucinate', a phenomenon in which an LLM produces information that is incorrect or does not exist (Verspoor, 2024). This has implications for the reliability and validity of LLM-produced content, and use of the term 'hallucination' itself is also problematic. As Hicks et al. (2024) pointed out, referring to nonsensical responses as 'hallucinations' implies that LLMs are aware and capable of understanding concepts like truth or reality, which is incorrect. Behind the sleek user interface, their 'intelligence' is a clever combination of programming and probability. While using terms like 'intelligence', even when artificial, can lead to associations with capabilities such as sentience or consciousness, there is currently no AI which is considered sentient (Tariq et al., 2022). Consequently, LLMs are thought by some to lack the ability to thoughtfully interpret the latent meaning behind text (Wachinger et al., 2024). However, applying human characteristics to LLMs is unhelpful. It is the interaction between humans and AI that brings value, and analysis generated by LLMs can be thought of as a new form of socially constructed knowledge; their 'intelligence' an extension of human cognition (Zhang et al., 2023). LLMs can provide alternative interpretations of the same data when asked multiple times, and can be instructed to apply a particular lens to their analysis, just as human researchers do (Harper, 2011).

There is growing support from the academic community for the use of LLMs in qualitative analysis, however, before embracing this new approach researchers must understand the importance of thoroughly evaluating and validating all AI generated content (Christou, 2023, 2024; dos Anjos et al., 2024; Jalali and Akhavan, 2024; Mesec, 2023; Turobov et al., 2024). ChatGPT itself, when asked whether AI can conduct reliable qualitative analysis in place of human researchers, stated that "while AI can significantly enhance the efficiency and consistency of qualitative analysis, it cannot fully replace the depth, insight, and contextual understanding that human analysts provide. A hybrid approach that combines the strengths of both AI and human analysis is the most effective way forward." (OpenAI, 2025). LLMs therefore show promise as tools to support rigorous qualitative research, provided that researchers understand how to use them effectively and responsibly and view them as research-assistants rather than entities that will completely replace human researchers.

To date, LLMs have been used in thematic analysis (Christou, 2024; Hitch, 2024), descriptive phenomenological analysis (Hamilton et al., 2023), content analysis (Tabone and De Winter 2023) and grounded theory analysis (Jiang et al., 2021). Hitch (2024) reported that ChatGPT conducted a successful thematic analysis, producing codes that were remarkably similar to those generated independently by human researchers. Subtle differences between the LLM and human coding were observed, which the researchers felt were comparative to the differences occurring between different human coders or in the codes generated by one person over time. Hitch (2024) suggested that LLMs could be used as tools to look for differences between researchers to assess reliability, but also as a supplement to human coding as part of reflexive, collaborative approach to analysis.

LLMs can also save researchers time and resources. One study found that ChatGPT could conduct a thematic analysis in 2 h, whereas human coders spent 23 h on the same task (Morgan, 2023). In another, ChatGPT analysed 470 free text survey responses in 10–15 min, whereas human researchers who each analysed a quarter of the responses each completed their analysis in an average of 27.5 min each (ranging from 20 to 50 min) (Fuller et al., 2024). LLMs could therefore enable researchers to concentrate on more complex tasks such as interpretation and considering the implications and applications of their findings. For those concerned that AI may replace human work, reassurance can be found by exploring the merits of collaborative partnerships between humans and LLMs. Though it is possible that LLMs will eventually be capable of replacing humans completely in some tasks, it is humans who decide what AI will and will not be used for. Many feel that AI works best *with*, rather than *instead of,* humans (De Cremer and Kasparov, 2021), but there is currently limited rigorously tested evidence to support the value of human-AI collaboration in qualitative research; something that this paper aims to provide.

1.1. Aims and objectives

The current study aimed to test the feasibility and efficacy of LLM-assisted narrative analysis of written story data and to demystify the process for researchers who may not have extensive knowledge of LLMs. Narrative analysis is a qualitative method in which researchers develop a set of narratives that summarise and represent the stories told within a textual dataset, which usually consists of interview transcripts or written stories (Clarke and Braun, 2013; Josselson and Hammack, 2021). Because the main goal of narrative analysis is to examine and identify patterns within stories, and LLMs are programmed to identify patterns in text, we hypothesised that LLMs, guided by human researchers, would be well-suited to assist with narrative analysis of story data. This study explored how two different LLMs (Claude 3 Opus and GPT-o1) performed when guided by human researchers to conduct narrative analysis of story data. It aimed to answer the following research questions:

1. How does narrative analysis of short stories conducted by human-guided LLMs compare with a purely human analysis?
2. How can analysis assisted by LLMs be optimised to mimic the subjectivity and transparency of human analysis?
3. What are the benefits and drawbacks of using LLMs as tools to assist with textual analysis?

2. Methods

2.1. Epistemology

This study used a social constructionist epistemology and a relativist ontology. Social constructionism rejects the idea of objective truth and views knowledge as the product of human interactions and sociocultural and historical contexts (Burr, 2015). For narrative analysis, social constructionists view stories and narratives as reflections of shared meaning-making and cultural discourses, rather than of personal beliefs or experiences (Clarke et al., 2017). Viewed through this lens, stories reveal how participants make sense of topics in more creative and indirect ways than traditional methods such as interviews or focus groups (Clarke et al., 2017). Relativism views 'reality' as a construction of human knowledge and experience rather than something which exists independently of human thought. From a relativist perspective, qualitative analysis allows for a range of different interpretations to be drawn from a corpus of text, depending on who has conducted the analysis, and how (Chamberlain, 2015). Taking into account the generative abilities of LLMs, the narrative analyses conducted in the present study aimed to explore alternative and unique interpretations, rather than produce static, reproducible findings.

2.2. Study context

This LLM-assisted analysis was conducted in collaboration with analysts (DR and SF) from Ipsos UK, a multinational market research company operating across the corporate, political, academic and public policy sectors. The collaboration with Ipsos was established because of Ipsos' vast experience in using LLMs for data analysis and began via an

existing relationship with a member of the research team. Researchers from both parties discussed the benefits to each party prior to the study beginning and the responsibilities each party would hold. No financial exchange occurred between the two parties. Ipsos provided access to the LLMs necessary for the analysis as well as expertise in conducting textual analysis using those tools. They provided access to Ipsos Facto, a secure AI chatbot developed by Ipsos and launched in 2023, which gives users access to all commercially available LLMs including ChatGPT, Claude, Gemini and others, within a closed system. All data entered into Ipsos Facto are stored within Ipsos' internal system and no one without the required permissions can access the system or the data. The secure nature of Ipsos Facto is, therefore, comparable to more traditional CAQDAS (computer-assisted qualitative data analysis software) packages – many of which now integrate AI into their software – which store virtual data locally. This security allowed the research team to conduct the analysis whilst adhering to requirements set out by the ethics committee.

The data used for the analysis were provided by the University of Southampton and drawn from an existing story completion study aiming to understand young people's use of social media and dietary choices in developing their identity (unpublished manuscript, Jenner et al. (2025)). Story completion is a creative data collection method where participants write stories about a pre-specified topic, which are then analysed using qualitative methods (Clarke et al., 2017). The story completion task used two pre-written story 'stems' that introduced a character of unspecified gender and a scenario. Participants were given the stems and asked to complete both stories. The story stems (Table 1) were designed by the research team using elements of identity formation theory and insights from experts in story completion, reviewed and edited by panel of young people.

2.3. Participants and sampling

Seventy-seven young people were recruited to the story completion study using a combination of face-to-face and online recruitment methods. Forty-five young people aged 13–21 years were recruited via youth groups in Hampshire, UK and 32 young people aged 18–25 years were recruited from across the UK via online research recruitment platform Prolific (www.Prolific.com). All participants were paid £10 in the form of cash or an Amazon voucher.

2.4. Ethical considerations

Following discussions with the University of Southampton ethics committee, a data sharing agreement was signed by University of Southampton and Ipsos representatives. The agreement included instructions for the secure transfer of data using encrypted file sharing software as well as for storing and deleting the data following completion of the study. It also provided details about the secure nature of Ipsos Facto, the AI system provided by Ipsos for use in this study. Ethical approval was granted by the University of Southampton School of Psychology ethics committee on 03/04/2024 (#78591).

Ethical guidelines associated with use of LLMs in research are limited and evolving, so no standardised guidance currently exists in the UK. Ethical considerations were discussed with the ethics committee from the outset of the project, and though explicit consent was not obtained from participants to use LLMs to analyse their stories, all usual ethical regulations were followed, and special measures were taken to ensure data privacy and security, including the use of Ipsos Facto as a secure, GDPR-compliant LLM interface, which does not use research data as training material for the LLMs it makes use of.

Table 1
Story stems presented to participants.

Stem	Prompts
"Ali has recently started trying to eat healthier. When Ali is scrolling Instagram, an advert for McDonalds comes up …"	Why might eating healthier be important to Ali? What does Ali do next? What does Ali think about the advert? What does Ali look at on social media? How does social media affect Ali's food choices?
"Robin has recently decided to start eating healthier and has been using social media to help them."	Why might eating healthier be important to Robin? How does social media affect Robin's food choices? What does Robin look at on social media? What does Robin post on social media and why?

2.5. Data collection

For the in-person data collection, young people were introduced to the task in small groups and instructed to spend at least 10 min writing their stories. Participants who signed up to take part in the study online via Prolific were provided with task instructions and a link to a Qualtrics page where they completed the story completion task. A total of 154 stories were collected. Fourteen stories were excluded because they contained less than 50 words, a cut-off that was decided in line with previous story completion analyses (Williams et al., 2022). Two further stories were excluded because they contained fantasy narratives that were not relevant to the research question. A final sample of 138 stories, ranging in length from 57 to 772 words, was therefore analysed.

2.6. Data analysis

The stories were analysed using narrative analysis, a method where narratives are identified from data through a multi-stage process of reading and grouping stories based on their common themes and types (Josselson and Hammack, 2021; Williams et al., 2022). The purpose of narrative analysis is to explore texts that tell stories and how people construct meaning through the stories they tell (Josselson and Hammack, 2021). It is a reflexive, inductive method that focuses on producing rich interpretation of a given topic by examining both explicit and implicit meaning, often aiming to read between the lines to discover the unconscious or the unspoken. Reflexivity is a key element of narrative analysis, where analysts reflect on their own internal dialogue and self-critique the processes they use to examine the data and draw out their own interpretations.

The human narrative analysis was carried out first to ensure that findings were independent of those produced by the LLM-assisted analysis which followed.

2.6.1. Human analysis

The four-step analysis process (see Table 2) was developed based on

Table 2
Human narrative analysis steps.

Narrative analysis step	Tasks
1	Read and annotate all stories. Note down elements of story content and structure that are relevant to the research question/topic.
2	Read each story as one unit of meaning (rather than 'coding' separate words or phrases, the overarching message or narrative 'type' of each story should be considered). Generate narrative 'groups' and categorise each story into one (or more) group(s).
3	Read through all stories within each narrative group one at a time and summarise the key themes, messages and ideas of each group in a short paragraph.
4	Use the summary paragraphs to generate short (one- or two-word) descriptive labels for each group. This may involve discussions amongst the research team, going back to the read the stories themselves, and using tools such as mind maps to finalise the narrative labels.

published guidance (Josselson and Hammack, 2021; Williams et al., 2022) and discussions with researchers experienced in narrative analysis.

2.6.2. LLM-assisted analysis

2.6.2.1. Data preparation. To answer the first and third research questions, analyses generated by two different LLMs were compared to each other and to the human analysis. The LLM-assisted analysis (see Table 3) was conducted across several sessions (June–October 2024) by SJ, DR and SF at the Ipsos offices in London, UK and remotely using Microsoft Teams video call. The first LLM used was Claude 3 Opus (hereafter referred to as 'Claude'), Anthropic's most advanced LLM to date. The researchers then repeated the analysis using OpenAI's GPT-o1 model, the most advanced iteration of 'ChatGPT' at the time of analysis.

The two models chosen for this analysis were identified due to their advanced analytic capabilities, large context windows and high token limits (a token limit is a number indicating the amount of information an LLM can process at any one time, including both input and output (Briganti, 2024)). The extensive token limits - 200k tokens for Claude Opus and 128k tokens for GPT-o1 - meant that large portions of data could be analysed by each model. Specifically, tokens represent words or word-parts, so larger token limits make analysis of textual data using LLMs quicker and easier.

2.6.3. Reflexivity

The first author (SJ) is a young, white British woman who, at the time of writing, is a PhD student and lecturer at a UK university. SJ is an experienced qualitative researcher but was not familiar with narrative analysis prior to carrying out this study. SJ conducted the analysis whilst reflecting on the biases that might have influenced her interpretation of the data. For example, SJ does not use social media but appreciates the importance of social media to young people and believes it can be a useful tool to promote health and facilitate social and identity development. Using LLMs to assist with the analysis allowed SJ to explore alternative views which may have challenged her own assumptions and biases.

In order for this study to remain aligned with the principles of narrative analysis, the researchers spent time thinking about how reflexivity could be maintained in LLM-assisted analysis. Therefore, as part of the LLM-assisted analysis and to answer the second research question, Claude and GPT-o1 were each asked to write a short reflexive paragraph to consider and explain how their own biases may have influenced their interpretation of the data. It is important to note that though these paragraphs may seem convincing, the models are not able to 'understand' these biases. Therefore, these paragraphs do not represent true reflexivity. They can, however, produce a coherent appraisal of how their training data may have contributed to their interpretations of the story data.

Table 3
LLM-assisted analysis.

LLM-assisted analysis step	Tasks
1	Prepare the story data for LLM-assisted analysis. Collate all 138 stories into a PDF document, formatted using markdown notation (a commonly used data format that can be easily read by LLMs (Auer et al., 2024)).
2	Use prompt engineering to develop clear, detailed instructions for the LLM. Create an initial prompt describing the method used by the human researchers to conduct the narrative analysis. Instruct the LLM, using simple and concise language, to *"Act as a prompt engineer. Review the following prompt. Optimise it to make it better. Ask me any questions before proceeding …"*. (Details of the prompt engineering process are available in supplementary materials.)
3	Use the optimised prompt (instructions for how to 'do' the analysis) to instruct the LLM to conduct a narrative analysis of all 138 stories.
4	Review the output from the LLM (the interpretation/ 'findings') and ask further questions of the model if needed. Use simple but specific language, for example: *'Explain to me your reasoning for how you created the story groupings. Describe the process you used in numbered steps. Give me 3 examples from the 'Ali' stories and 3 examples from 'Robin' stories you used to develop each group. For each example provide the participant number and their age and a text excerpt from the story. The examples should follow the following format: Participant number: Participant age: Text excerpt from the story:'*
5	Repeat steps 2–4 for each model used.

2.6.3.1. Claude's paragraph

> "As an AI language model, I must acknowledge the potential biases in my training data that may have influenced my analysis of the stories. Societal biases, stereotypes, and dominant cultural narratives could have skewed my understanding of healthy eating habits, social media use, and identity formation. My pre-existing knowledge and biases may have subconsciously shaped my interpretations, prioritizing certain narratives over others. Throughout the process, my understanding likely evolved as I encountered diverse experiences. It is crucial to consider the limitations of AI in qualitative research, as we lack the lived experiences and contextual understanding of human researchers. To mitigate biases, AI models should be trained on diverse datasets, incorporate expert feedback, and collaborate with human researchers. By engaging in reflexivity, I aim to transparently acknowledge my biases and limitations as an AI researcher, striving for accurate, meaningful, and ethically sound insights through continuous self-awareness and improvement."

2.6.3.2. GPT-o1's paragraph

> "During the analysis of stories, I recognized that inherent biases in my training data may have influenced my understanding and interpretation. My training data predominantly reflects Western cultural perspectives on social media, identity formation, and dietary choices, leading me to emphasize themes like body image issues, mental health struggles, and negative influences of social media. This affected my grouping by prioritizing familiar patterns, potentially overlooking positive influences, unique cultural experiences, or individual resilience. As the research progressed, I became more aware of the need to consider each story's context and actively seek diverse experiences. I acknowledge that as an AI language model, my reliance on patterns can perpetuate biases and limit my ability to fully capture complex human experiences, leading to oversimplifications. In the future, I will strive to incorporate broader perspectives, seek more diverse data, and collaborate with human researchers to enhance research validity and inclusivity."

3. Findings

The findings of the human analysis are reported in detail elsewhere (unpublished manuscript; Jenner et al. (2025)). The steps described in Table 2 initially produced six narratives, and after merging several narratives due to their similarity – following discussions amongst the researchers – resulted in four final narratives. Following completion of the human analysis the LLM-assisted analysis was conducted iteratively, as described below.

Both Claude and GPT-o1 analysed the same data following the same steps as the human researchers. In step 1, the LLMs produced notes summarising the key content and structure points from the stories (Table 4). Claude took the same approach as the human researchers, analysing all the data at once, whereas GPT-o1, which has a smaller

Table 4
Comparison between human, Claude and GPT-o1 interpretations in Step 1 of the narrative analysis process.

Human analysis	Claude analysis	GPT-o1 analysis
For Ali stories: - Healthy eating linked to feeling better, sports performance, weight loss, and appearance. - Disordered eating mentioned across all gender versions of Ali. - Health knowledge varied; some valued moderation and a balanced diet. - Some stories noted 'big food' marketing tactics like McDonald's. - McDonald's ads triggered cravings, physical reactions, and nostalgia. - 'Tempt' appeared 49 times; 'crave' 21 times. - Fast food brought short-term happiness but often regret. - Seen as acceptable if part of a 'cheat day' or occasional treat. For Robin stories: - Healthy eating linked to weight loss for body image or health. - Robin used social media to help others and be a positive influence. - Stories showed awareness of social media's good and bad sides. - Influencers were inspiring when promoting health but harmful when pushing unrealistic beauty standards. Structure notes: - Stories followed different styles and trajectories. - Many fit within one of the seven basic narrative structures that are common in literature. - Ali stories often followed an 'overcoming the monster' theme (resisting fast food ads). - Robin stories often followed a 'rebirth' theme (escaping influencer pressure with support).	**Content Notes:** - Seeing fast food ads while trying to eat healthy is a common temptation - Reasons for eating healthy: improve health, lose weight, feel better, set good example, live longer - Strategies to resist temptation: scroll past quickly, block ads, seek healthy content, remind self of goals - Giving in leads to regret, shame, feeling of failure - Social media has positive and negative influences (provides ideas and support but also unrealistic standards and ads) - Social media can fuel unhealthy comparisons and restrictions if unchecked - Curating feed and connecting with real friends is helpful - Some share their journey to healthy eating to inspire others **Structural Notes:** - Many follow a temptation > decision > outcome arc, either resisting and feeling accomplished or giving in and regretting - Some depict a full transformation from unhealthy habits to positive changes inspired by social media - Others are more cautionary, showing a spiral into unhealthy restrictions fuelled by social media - A few focus on sharing one's own journey to now help others	**Example: Ali's Story ([P001 - Ali])** **Content Notes:** **Positive Changes and Feelings:** "Ali was enjoying her life now that she was eating healthier." "She felt way more respected and felt more agile than ever." "She started wearing fewer layers of clothes at school making her feel way more relaxed." **Initial Disinterest in Fast Food:** "Fast food was non-existent to her eyes anymore." **Encounter with Social Media Advertisement:** "While scrolling through her phone she came across an advert on her Instagram … A McDonald's advert." "Her mouth started to water but she was trying her hardest to resist the temptation of looking at it any longer." **Internal Conflict and Temptation:** "She had a brief flashback of her previous body and health and used that as fuel to resist the advert." **Succumbing to Temptation:** "Reluctantly, Alice ordered it and once it came she took a slow bite out of it." "She was in love. She ordered another. Then another." **Negative Outcome:** "All that hard work working out was wasted." "She became what she most disliked in the past." **Structure Notes:** - Positive Beginning: Ali is enjoying the benefits of healthy eating. - Conflict/Dilemma: Temptation triggered by a McDonald's advert on Instagram. - Climax: Internal struggle leading to giving in to temptation. - Negative Ending: Reversion to old habits and loss of progress."

Table 5
Comparison between human, Claude and GPT-o1 interpretations in Step 2 of the narrative analysis process.

Human analysis	Claude analysis	GPT-o1 analysis
Narrative A) Social media is good and bad. Characters make a conscious effort to seek out the good and ignore the bad.	**Narrative A)** - Finding healthy meal ideas and recipes - Feeling inspired by fitness influencers and transformations - Connecting with others on a health journey for support - Documenting own progress and meals to stay accountable	**Narrative A)** Positive transformation and empowerment facilitated by social media **Narrative B)** Balanced approach to health and wellness **Narrative C)** Navigating misinformation and conflicting information on social media **Narrative D)** Self-Acceptance, Resilience, and Critical Awareness Against Negative Influences
Narrative B) Characters use social media to help themselves and others by building a supportive online community. Narrative D) Social media is seen by the characters as positive, helpful and inspirational.	**Narrative B)** - Realizing social media "health" content was unhealthy, disordered - Deciding to be a voice of moderation and listening to body's needs - Sharing balanced approach to inspire healthy changes in others"	**Narrative E)** Empowerment through social support and community **Narrative F)** From personal struggle to advocacy and influencing others
Narrative C) Social media makes the characters feel bad. There's too much pressure and misinformation.	**Narrative C)** - Creating unrealistic standards of "healthy" that are unattainable - Fueling comparisons and feelings of inadequacy - Promoting unsustainable restrictions and "clean eating" - Focusing more on appearance than true health and enjoyment	**Narrative G)** Negative impact of social media on body image and mental health
Narrative E) Characters can overcome the temptation of unhealthy food on social media and can stick to their health goals.	**Narrative E)** - Scrolling past fast food ads, blocking/ unfollowing - Seeking out healthy recipes and fitness inspiration instead - Reminding self of goals and why health is important	**Narrative H)** Overcoming Temptation and Internal Conflict to Maintain Healthy Choices
Narrative F) It is almost impossible for characters to stick to their health goals whilst using social media.	**Narrative F:** - Seeing ad and feeling unable to resist - Ordering fast food, often in secret and larger amounts than intended - Immediate regret, shame, feeling of ruining progress	**Narrative I)** Negative influence of social media leading to temptation and relapse
No corresponding narratives	No corresponding narratives	**Narrative J)** Unique Narratives

token limit than Claude took a different approach. The token limit meant that the story dataset needed to be split up into smaller documents before being fed into the model. Seven different documents were created and GPT-o1 analysed each 'batch' of stories one at a time. For each batch, GPT-o1 analysed each story individually (Table 4 contains an example from participant one's 'Ali' story).

In Step 2, the LLMs grouped the stories into narratives based on the key messages, ideas and themes from each story. Claude used the same approach as the human researchers and its narratives were an almost exact replication of the initial narratives developed by the human researchers. GPT-o1 created narratives for each story individually, resulting in 32 different narratives across all 138 stories, many of which were similar but labelled differently. The researchers then asked GPT-o1 to merge all similar narratives, leading to an initial batch of ten narratives which mapped onto the human- and Claude-generated narratives (Table 5). It is of note that the analysis using GPT-o1 was significantly quicker, however this was because the researchers had already fine-tuned the method using Claude.

Because the human researchers had combined narratives E and F in their analysis, they asked Claude to: "Explain your reasoning for keeping groups E and F separate rather than combining them. Explain how you would combine groups E and F." Claude grouped the two narratives together in the same way that the human researchers had. Table 6 shows how each of the combined LLM-generated narratives mirrored the combined human-generated narratives.

GPT-o1 generated one narrative (J), which did not reflect any of the human- or Claude-generated narratives. This narrative remained part of the analysis to allow the researchers to explore differences in how GPT-o1 approached the narrative analysis compared with the human researchers and with Claude.

In Step 3, both LLMs summarised each of their narratives in a short descriptive paragraph (Table 7). GPT-o1 also merged similar narratives

Table 6
Human, Claude and GPT-o1 groups, merged and compared.

Human analysis	Claude analysis	GPT-o1 analysis
New narrative A) (A from Table 5) Characters show awareness of both the good and bad sides of social media. The good things can be inspirational e.g. seeking out content about self-improvement or healthy recipe ideas. A fair amount of the self-improvement content relates to moderation/balance and dismissing 'diet culture' and restriction.	**New narrative A)** (A from Table 5) finding healthy meal ideas and recipes, feeling inspired by fitness influencers and transformations, connecting with others on a health journey for support, documenting own progress and meals to stay accountable.	**Narrative A)** Positive transformation and empowerment facilitated by social media **Narrative B)** Balanced approach to health and wellness **Narrative C)** Navigating misinformation and conflicting information on social media **Narrative D)** Self-Acceptance, Resilience, and Critical Awareness Against Negative Influences
New narrative B) (B + D from Table 5 combined) Characters use social media to successfully find healthy recipes and inspiration … Social media is seen as a form of positive social support.	**New narrative B)** (B from Table 5) realizing social media "health" content was unhealthy, disordered, deciding to be a voice of moderation and listening to body's needs, sharing balanced approach to inspire healthy changes in others.	**Narrative E)** Empowerment through social support and community **Narrative F)** From personal struggle to advocacy and influencing others
New narrative C) (C from Table 5) Social media is a place where bullying and trolling happens, there are temptations such as unhealthy food adverts and excessive pressure to look 'perfect'.	**New narrative C)** (C from Table 5) creating unrealistic standards of "healthy" that are unattainable, fueling comparisons and feelings of inadequacy, promoting unsustainable restrictions and "clean eating", focusing more on appearance than true health and enjoyment.	**Narrative G)** Negative impact of social media on body image and mental health
New narrative D) (E + F from Table 5 combined) Fast food adverts on social media are extremely tempting and hard to resist but it is possible to overcome cravings and stick to health goals if you are able to overcome the power of the marketing techniques.	**New narrative D)** (E + F from Table 5 combined) This combined group would first describe the common experience of encountering fast food ads while trying to eat healthier, then delve into the internal struggle that ensues. It could then branch out into the two different paths characters take - resisting or succumbing - and explore the emotions, consequences, and lessons learned from each path. Combining the groups would shift the focus to the overarching experience of facing temptation and the various ways individuals respond to it, rather than emphasizing the specific outcomes. It could provide a more comprehensive look at the role of temptation in shaping the characters' journeys towards healthier lifestyles.	**Narrative H)** Overcoming Temptation and Internal Conflict to Maintain Healthy Choices **Narrative I)** Negative influence of social media leading to temptation and relapse
No corresponding narratives	No corresponding narratives	**Narrative J)** Unique Narratives

from the previous step to create a set of four new narratives, three of which (B, C, D) mapped on to the human- and Claude-generated narratives.

The final stage of the analysis involved asking the LLMs to generate descriptive labels for each of the narratives, which mirrored the narratives produced in the human analysis (Table 8).

Having observed Claude produce the initial analysis presented in Tables 4–8 in approximately 60 s, the researchers began to consider more advanced analytic tasks that they could ask of the model. They asked Claude to provide a list of stories corresponding to each narrative, including participant ID number and illustrative excerpts from each story. Claude quickly returned a response that initially seemed accurate, but which on further inspection contained hallucinations. Claude did not consistently assign the correct participant ID number to each story excerpt. When asked why it was not able to correctly assign text excerpts to participant IDs, Claude explained the mistake it had made and gave the researchers some suggestions to improve communication with the model, which included refinements to how the stories were formatted before being entered into the model.

Following these hallucinations, the researchers experimented with re-formatting the data document using JavaScript Object Notation (JSON) (see supplementary materials). JSON is a data interchange format; a simple way of formatting information so that it can easily be read by both computers and humans (Bassett, 2015). It is an independent universal language, meaning that it can be used to communicate information to any software, regardless of its native programming language. JSON proved to be a better format to use to communicate with Claude, but the model still struggled to assign all 138 stories to one of the four narratives. A new approach, which involved feeding each story, in JSON format, into the model one by one, was the most effective way of ensuring Claude assigned every single story to a narrative. The same process was used with GPT-o1, following which the human researchers reviewed all the analytic 'decisions' made by the two LLMs. The researchers used Excel spreadsheets to compare the analyses conducted by Claude and GPT-o1 to the human analysis. This exercise revealed that both LLMs were able to group all 138 stories into four narratives in a fraction of the time taken for the human analysis. The total time taken to conduct the analysis using Claude was estimated at approximately 35 h. This included preparing and cleaning the raw story data before starting the analysis, conducting the analysis with Claude, and then reviewing Claude's findings and interpretations. For GPT-o1 the process was quicker because the researchers had already prepared the data and refined the method with Claude. It is estimated that the GPT-o1 analysis took approximately 12 h. It took the human researchers approximately 64 h (across 16 weeks) to conduct the full narrative analysis of all 138 stories whilst documenting findings throughout the process.

Table 7
Comparison between human, Claude and GPT-o1 interpretations in Step 3 of the narrative analysis process.

Human analysis	Claude analysis	GPT-o1 analysis
Narrative A: characters understand that they have a personal responsibility to be discerning about the content they consume through social media, to avoid harmful or negative content and seek out positive content. They feel confident and able to do this by the end of their story.	**Narrative A:** stories where characters leverage social media platforms to support their journeys towards healthier lifestyles. In these narratives, individuals find inspiration, motivation, and valuable resources through engaging with health-focused content, connecting with like-minded communities, and sharing their own experiences. Social media serves as a positive tool, providing access to nutritious recipes, fitness tips, and encouragement from others on similar paths. These stories highlight the potential of social media to facilitate positive change and empower individuals in their pursuit of wellness.	No corresponding narrative
Narrative B: characters have used social media to successfully achieve their own health goals and are proud of what they have achieved. They want to use social media to help others achieve their goals, and to portray identities as good, moral, successful people.	**Narrative B:** stories where characters undergo transformative journeys, often learning from their own experiences or the experiences of others. These narratives showcase individuals who have overcome challenges related to unhealthy eating habits or negative body image and have emerged as role models or sources of inspiration for others. Through their stories, these characters share valuable lessons, provide guidance, and encourage others to prioritize their well-being. This group highlights the power of personal growth, self-discovery, and the potential to positively influence others through shared experiences and wisdom.	**Narrative B** (A, E, F from Table 6 combined): This group focuses on the positive role of social media and community support in facilitating personal growth and healthy lifestyle changes. Individuals leverage these platforms to find inspiration, following health influencers, accessing resources, and discovering new healthy recipes. Characters engage with like-minded individuals who provide encouragement, accountability, and shared experiences. Some individuals use their experiences to educate and inspire others, becoming advocates for healthy living. The stories emphasize the empowerment that comes from external support systems and the potential of social media to positively influence identity formation and dietary choices.
Narrative C: characters feel pressure to look like the influencers they see on social media. They are insecure and often get bullied online or in real life. They feel inadequate and idealise the unachievable bodies and lifestyles they see online.	**Narrative C:** stories that illustrate the negative impact of social media on characters' self-perception, body image, and eating habits. In these narratives, individuals are exposed to unrealistic standards, harmful comparisons, and the pressure to conform to certain body types or diets. Social media becomes a source of anxiety, self-doubt, and unhealthy obsessions, leading characters down paths of restrictive eating, overexercising, or disordered thinking. These stories underscore the potential pitfalls of social media and its capacity to exacerbate existing insecurities or create new ones.	**Narrative C** (G, D, C from Table 6 combined): this group delves into the negative effects of social media on mental health, body image, and self-perception. Individuals face challenges such as exposure to unrealistic standards including idealized images and content leading to distorted self-perception and unhealthy behaviors. Characters must navigate conflicting dietary information, leading to confusion and potential harm. Characters experience low self-esteem, anxiety, and mental health struggles. Importantly, the narratives also highlight the journey toward resilience and self-acceptance, including recognizing negative influences and developing strategies to mitigate their impact, prioritizing personal values and well-being over external judgments and educating and supporting others.
Narrative D: characters are tempted by fast food adverts on social media and feel that social media has the power to divert them from their health goals. They are faced with a choice: to maintain control over their own choices (positive endings), or to concede control and 'give in' to the pressures of social media (negative endings).	**Narrative D:** stories where characters encounter triggers or temptations, particularly in the form of fast food advertisements on social media, while trying to maintain a healthy eating lifestyle. This group highlights the internal struggle between desire and willpower that individuals face when attempting to make positive changes in their lives. Whether the characters resist or succumb to the temptation, these stories demonstrate the challenges and setbacks people experience when adopting healthier habits in a world filled with enticing triggers.	**Narrative D** (I, H, B from Table 6 combined): This group encapsulates the journey of individuals as they navigate the temptations posed by negative social media influences, particularly advertisements for unhealthy food options. It highlights the internal struggles faced when trying to maintain or adopt healthier eating habits. Some individuals give in to cravings induced by social media, leading to relapse into unhealthy eating habits, accompanied by feelings of guilt and disappointment. Others manage to resist temptations, demonstrating resilience and self-discipline. They often adopt strategies such as mindful eating, seeking support, and redefining their relationship with food. A portion of individuals find a middle ground by integrating healthy habits with occasional indulgences, emphasizing moderation and a sustainable approach to wellness.
No corresponding narrative	No corresponding narrative	**Narrative E** (J from Table 6): This group comprises stories that offer unique perspectives or encompass elements not prevalent in other groups. These narratives may include supernatural or metaphysical elements, stories that use unconventional themes to convey messages about intuition, societal critique, or personal transformation. Stories make general observations on the influence of social media without focusing on specific personal experiences. Innovative storytelling methods or unique angles on the interplay between social media, identity formation, and dietary choices.

Table 8
Comparison between human, Claude and GPT-o1 interpretations in Step 4 of the narrative analysis process.

Human analysis	Claude analysis	GPT-o1 analysis
Narrative A: Personal responsibility	**Narrative A:** Social Media for Good	No corresponding narrative
Narrative B: Role models	**Narrative B:** Learning and Inspiring	**Narrative B:** Empowerment and Positive Transformation through Social Media and Community
Narrative C: Social comparison	**Narrative C:** Social Media for Bad	**Narrative C:** Overcoming Negative Impacts of Social Media on Mental Health and Body Image
Narrative D: Control	**Narrative D:** Facing temptation	**Narrative D:** Navigating Temptation and Adopting Healthy Choices
No corresponding narrative	No corresponding narrative	**Narrative E:** Unique Narratives and Reflective Commentaries

Table 9
Four steps for successful LLM-assisted qualitative analysis.

Steps	Tasks	Details/examples
Step 1: Develop a structured analysis plan	Ensure there is clear understanding amongst the human researchers of the method(s) to be used and of the role that LLMs will play in the analysis.	Think about: • Whether the human or LLM analysis will be conducted first, or whether they will happen concurrently • Who will conduct each stage of the analysis • The extent to which LLMs will assist with the analysis
Step 2: Select the best model for the task	Identify the LLM(s) that are the best fit for the purpose of each analysis that is conducted, taking into account the type of data, chosen analysis method, and desired outputs.	Take into account: • The model's analytic and reasoning capabilities • Differences between free models and paid subscriptions • Context windows/token limits • Speed of model • Ability to read different document types (word documents, PDFs, web links etc.)
Step 3: Format data appropriately	Structure data in the most clear and simple format that LLMs are able to understand and work with.	Markdown and JSON (JavaScript Object Notation) are effective ways of communicating with both Claude and GPT-o1. Experiment and choose one that works best for the data.
Step 4: Use prompt engineering and optimisation	Provide LLMs with a clear and simple outline of the aims, methods and expectations of the analysis. Create clear, simple instructions that are then reviewed and edited by the model to maximise comprehension and ensure efficient communication between humans and LLMs.	*"I am a researcher in the psychology department of a large UK university working on my PhD. Below you will find my research question and the analysis method I followed."* *"Act as a prompt engineer. Review the following prompt. Optimise it to make it better. Ask me any questions you need to before proceeding."*

4. Discussion

4.1. Summary of findings

A narrative analysis of 138 short stories conducted by human qualitative researchers was compared to narrative analyses conducted by two commercially available chatbot-style LLMs: Anthropic's *Claude 3 Opus* and OpenAI's *GPT-o1*. This novel LLM-assisted analysis method was developed iteratively as a collaboration between qualitative researchers and data analysts with expertise in AI. The comparison between the human analysis and the LLM analyses aimed to assess whether each model could conduct a thorough, sound analysis using the same steps as the human researchers. The researchers acknowledged that different humans were likely to have grouped the stories slightly differently, or that the same researcher could feasibly make different decisions upon revisiting the data at a later date. Therefore, it was not realistic or appropriate to define the LLMs 'success' based on them producing an exact replica of the human analysis. Accordingly, the researchers evaluated each of Claude's and GPT-o1's analytic decisions manually to decide whether the models' justifications were coherent and rational. As a result of this validation process, it was determined that both models had been able to conduct a systematic, credible narrative analysis of the story data.

Research question 1. How does narrative analysis of textual data in the form of short stories conducted by LLMs compare with a human analysis?

Claude generated four narratives which mirrored the four narratives developed in the human analysis. GPT-o1 also generated four narratives, three of which reflected narratives from the human analysis, the fourth of which was unique. The researchers felt that this fourth narrative did not adequately describe any of the stories included in the analysis. In comparison to Claude and the human researchers, GPT-o1 conducted a more simplified analysis which the researchers felt lacked nuance. It is therefore possible that Claude is more suited to interpretive methods such as narrative analysis and that GPT-o1 may be more suited to other types of textual analysis, something that could be explored by further research to compare the differences between models and identify the analytic strengths and weaknesses of different models.

Research question 2. How can analysis using LLMs be optimised to mimic the subjectivity and transparency of human analysis?

This work provided helpful methodological insights into how narrative analysis might be enhanced by LLMs. The LLM analysis built on the findings of the human narrative analysis and helped the researchers develop more well-rounded conclusions. Having to explain the analysis steps to the LLMs and review every stage of their analytic process meant that the researchers became closer to the data and to the analysis than they had been previously. Additionally, thinking about the LLMs' capabilities for reflection and reflexivity also prompted the researchers to consider their own biases in greater detail and think about how these may have contributed to their interpretations and shaped their approach to the analysis.

Research question 3. What are the benefits and drawbacks of using LLMs as tools to assist with textual analysis?

This study presents a rigorously tested method for using LLMs as tools to accelerate and enhance narrative analysis. The method is accessible to researchers with qualitative expertise who may have limited knowledge or experience in using AI. In particular, the study provides evidence to support the use of Claude to analyse textual data in the form of written stories. GPT-o1 conducted the analysis with similar rigour as Claude, but did not generate findings that were as closely aligned with those of the human researchers. It is important that

researchers who wish to utilise LLMs for analysis of textual data understand the importance of selecting the most appropriate LLM for their analysis, and of validating all AI-generated content.

4.2. Proposed guidelines for LLM-assisted analysis

During the iterative development of the method used in the current study, the researchers developed some general guidelines for successful and efficient LLM-assisted analysis of qualitative data (Table 9).

As part of step 1 in the above guidelines, researchers should also consider how their own personal and philosophical views will impact the ways in which they choose to incorporate LLMs into analyses. For some, the growing presence of LLMs in research has sparked anxieties about a loss of meaningful work for researchers (Bankins and Formosa, 2023) and the demise of individuality in academic writing (Nakadai et al., 2023). For example, anecdotal evidence suggests that occurrences of certain words and phrases in academic publications have increased exponentially since the release of LLMs such as ChatGPT. A PubMed search revealed that the occurrence of the word 'delve' in citations and abstracts increased by 654 % from 2020 to 2024 and similar trends have emerged in other databases (Stokel-Walker, 2024). Because of the way LLMs are trained, any text that is published online will inevitably make its way back into the model, which may lead to a gradual but noticeable reduction in diversity within scientific writing and LLM-generated ideas. This has potential implications for how LLMs analyse data and express their interpretations of qualitative data. Individual researchers will, therefore, hold their own beliefs about the extent to which LLMs should be used in qualitative analysis, and these roles and expectations should be discussed prior to beginning any analysis.

4.3. Strengths and limitations

4.3.1. Strengths

The standout strength of the current study is the rigorous comparison between human and LLM-assisted narrative analysis. The human analysis was conducted prior to the LLM analysis to ensure that findings from the LLM analysis did not influence the human researcher's interpretations. Much of the published research on the use of LLMs for qualitative analysis has, so far, taken relatively 'quantitative' approaches to evaluating the capabilities of AI (Bano et al., 2023; Tai et al., 2024; Törnberg, 2024; Zhao et al., 2023). Studies often discuss researcher subjectivity, bias, positionality and backgrounds as limitations that prevent rigorous research from being conducted (Tai et al., 2024; Törnberg, 2024); something that many qualitative researchers will disagree with. The current study approached the task of validating LLM-assisted analysis by working with the subjectivity and individuality of human researchers to acknowledge and celebrate the nuances of human-AI collaboration and consider the extent to which LLMs embody both objectivity and subjectivity. The subjective nature of human qualitative analysis is considered to be one of its strengths, providing that researchers engage in reflexivity to examine the impact of their own characteristics on the research process (Braun et al., 2023; Mauthner and Doucet, 2003). Just as human researchers can never truly operate outside of the influence of their own experiences and biases (Hamilton et al., 2023), neither can AI.

Some see LLMs in their current form as objective but ponder whether more advanced models could be programmed with capabilities like imagination and reflexivity, allowing them to diverge from their programming and become responsive and adaptive to different contexts (D'Amato, 2024). Other researchers, however, already see LLMs as subjective tools. The data LLMs are trained on impact their responses, just as the experiences and ideas that humans are exposed to influence their interpretations of data. So, the question should not be 'How can we ensure that LLMs are not biased?' but rather 'How can we ensure that LLMs are reflexive and transparent analysts?'. The findings of the current study confirm what others have written about LLMs' current capabilities for reflexivity. Hamilton et al. (2023) explored whether ChatGPT could engage in 'bracketing' - a type of reflexivity where qualitative researchers identify and then set aside their own biases – and found that ChatGPT could describe the biases that may have been present in the information it was trained on but was not able to evaluate how these biases may have affected its analysis of the data. Similarly, Balmer (2023) found that ChatGPT could suggest how it might itself engage in reflexivity and why examining its own biases and assumptions might be important, but was not able to produce substantive, insightful or unique examinations of its own impact on the research process. Though LLMs seem not to have the ability to be sufficiently reflexive yet, they can support human creativity and meaning-making, blurring the lines between objectivity and subjectivity (Yan, 2024). The fact that Claude and GPT-o1 were both able to produce coherent 'reflexive' paragraphs as part of the current analysis does not yet demonstrate that LLMs can act as reflexive researchers, but it does provide additional evidence to support the models' capabilities for providing accurate research-relevant output and mimicking human cognition.

The collaboration between the University of Southampton and Ipsos represented a unique opportunity for researchers to explore the capabilities of two different LLMs for narrative analysis whilst adhering to important ethics regulations around participant confidentiality and data protection. This was made possible by Ipsos' state-of-the-art Ipsos Facto system, which is powered by industry-leading generative AI technology from Ipsos partners OpenAI, Google and Anthropic. Ipsos Facto contains built-in enterprise-level security features, including Two-Factor Authentication (2FA), to ensure secure access for authorized users (Ipsos staff), which also makes it compliant with General Data Protection Regulation (GDPR) (European Parliament, 2016) and university ethics guidelines. Ipsos and their technology partners do not use data entered into the models to train or improve the LLMs utilised by Ipsos Facto.

4.3.2. Limitations

Though LLMs can conduct high-quality narrative analysis, researchers should still be cautious when publishing findings of LLM-assisted analyses. AI technology is far from perfect, and advancements are being made with tremendous speed; models used for analysis may already be out of date once papers are written and published. The current study presents a novel exploration of the capabilities of LLMs specifically for narrative analysis of story data, within the context of the models available in 2024 and acknowledges that models such as Claude and GPT-o1 will be superseded by those with far greater capabilities, which may come with additional risks and challenges. The authors also acknowledge that choosing which LLMs are most appropriate for use in textual analysis depends on the type of analysis being conducted and the qualitative principles associated with that analysis. These principles include, among others, reflexivity, transparency and epistemology.

Access to secure, GDPR-compliant LLMs is an issue that will impact the ability of academics to carry out analysis using AI. Without access to secure systems such as Ipsos Facto, researchers are left with the alternative of accessing commercially available online LLMs that are developed by companies that may not be transparent in disclosing where data entered into their models are stored or how they are used, making responsible AI use more difficult. There are also limitations for users who only have access to the free versions of models such as ChatGPT or Claude, where there are significant differences in model capabilities and security features between free and paid versions. In terms of security, data entered into the free version of ChatGPT leaves its country of origin to be stored on servers in the US, which has implications for researchers outside of the US. Many institutional ethics boards, particularly those covered by GDPR, require data to be stored internally and then deleted after a specific amount of time, and further permissions are often required to store data abroad. The current research benefitted from the collaboration with Ipsos, and use of the Ipsos Facto system at no cost, but this is usually something that is only accessible to researchers who are

able to purchase licenses from Ipsos.

There are also wider issues associated with use of LLMs that must not be ignored, such as the social and environmental cost of developing and maintaining the models. In 2023, an article in The Guardian reported that content moderators – people who help to train models like ChatGPT by reviewing harmful, explicit and graphic material - employed by OpenAI in Kenya had called for their government to investigate the company's exploitative and unethical working conditions (The Guardian, 2023). The article reported that many AI companies outsource their moderation and data labelling work to countries in East Africa, where workers can be paid as little as $1.46 per hour to review disturbing, often traumatising, content. In 2024, the European Trade Union Confederation (ETUC) called for widespread adoption of two directives which aim to maintain and enforce corporate accountability to prevent exploitation and poor treatment of workers conducting AI-related labour (European Trade Union Confederation, 2024). It is, however, unclear how many companies have enforced such directives.

Furthermore, the environmental impact of LLMs is known to be extensive: the computational infrastructure required to develop, host and maintain LLMs relies on extensive cooling systems requiring massive water and energy consumption, which in turn increases carbon emissions (Singh et al., 2025). Whether these negative impacts are equal to or outweigh the potential positive impacts of LLMs for society and the planet is not yet known, so more research is needed to evaluate the true costs and benefits of these novel technologies (Ren et al., 2024).

This study focussed specifically on the application of LLMs to narrative analysis of short stories. It is therefore not possible to make general conclusions about the efficacy of LLMs for other methods of qualitative analysis such as thematic analysis. There is, however, emerging evidence to suggest that LLMs and other forms of AI can conduct thorough, high-quality thematic analysis (Christou, 2024; Hitch, 2024; Towler et al., 2023). Researchers should appraise a range of evidence when deciding how to incorporate AI into their qualitative research, and should seek out evidence to support the use of LLMs for the particular type of analysis they plan to conduct, including the different sub-types of thematic analysis (Braun and Clarke, 2021).

4.4. Implications and future research

This study highlights the extensive potential benefits of LLMs to the field of qualitative research. LLMs can save researchers time, provide alternative interpretations, and when viewed through a social constructionist lens, can be thought of as co-constructors of socially and culturally meaningful knowledge (Yan, 2024). There is, therefore, a growing need for higher education institutions to embrace the potential for AI to enhance qualitative analysis and to support their staff and students to embrace new technologies (Neumann et al., 2023). Given that LLMs are trained to identify patterns and meaning within text, they are ideal tools to assist researchers in conducting analyses of textual data using existing frameworks such as thematic analysis or narrative analysis (Christou, 2024). LLMs have already been widely adopted by qualitative analysts in the commercial sector (e.g. Ipsos) and AI is now embedded within many CAQDAS packages designed to support researchers to conduct qualitative analyses (e.g. MAXQDA; VERBI software (2024), Atlas.TI; Scientific Software Development GmbH (2024)) and systematic reviews (Khalil et al., 2022). Consequently, institutions that do not equip their researchers with the skills to utilise these tools risk being left behind. Investing in secure and trustworthy AI systems like Ipsos Facto will allow qualitative researchers from a range of backgrounds and disciplines to conduct rigorous, responsible and ethical LLM-assisted research.

For qualitative researchers to feel confident incorporating LLMs into their analyses, higher education institutions and funding bodies will need to produce clear, comprehensive guidance and training to support researchers to learn to use AI in ethical and responsible ways. Preliminary guidelines have emerged (Russell Group, 2023), but whether these will be implemented at institutional level remains to be seen. Implementation of such guidance will reassure ethics committees, researchers and most importantly, participants, that responsibly and ethically conducted LLM-assisted research poses no threat to people.

4.5. Conclusion

There are a range of issues that are important for qualitative researchers to understand and consider before using LLMs for qualitative analysis. A basic understanding of the fundamental principles of AI is essential for any qualitative researcher wishing to explore LLM-assisted analysis. This includes being critical of the roles LLMs play in the research process and of any content that is produced by LLMs, which is encompassed by reference to 'responsible AI', part of which is maintaining transparency in how AI is used in research. LLMs are highly capable of assisting researchers in conducting narrative analysis of textual data, from data summarisation to interpretation of narratives and 'thoughtful' analytic commentary and discussion. Applications of LLMs across a range of research fields will vary, but one thing remains certain: AI is here to stay. Many researchers will benefit from exploring how LLMs can improve the quality and efficiency of their research, therefore maximising the positive impact that research can have on people and communities.

CRediT authorship contribution statement

Sarah Jenner: Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Dimitris Raidos:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Emma Anderson:** Writing – review & editing, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Stella Fleetwood:** Writing – review & editing, Supervision, Software, Resources, Methodology, Conceptualization. **Ben Ainsworth:** Writing – review & editing, Methodology, Conceptualization. **Kerry Fox:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Jana Kreppner:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Mary Barker:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Funding

This work was conducted as part of a PhD studentship funded by the Economic and Social Research Council (ESRC) South Coast Doctoral Training Partnership (SCDTP). The SCDTP had no direct involvement in any part of this study other than providing funding for the studentship.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sarah Jenner reports financial support was provided by UK Research and Innovation Economic and Social Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thank you to Toni Williams for her advice and guidance on developing the story stems and the initial framework for the human narrative analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.metip.2025.100183.

Data availability

Data will be made available on request.

References

Anthropic, 2025. Claude 3. In: Anthropic. https://claude.ai/.

Auer, C., Lysak, M., Nassar, A., Dolfi, M., Livathinos, N., Vagenas, P., Ramis, C.B., Omenetti, M., Lindlbauer, F., Dinkla, K., 2024. Docling technical report. arXiv preprint arXiv:2408.09869.

Balmer, A., 2023. A sociological conversation with ChatGPT about AI ethics, affect and reflexivity. Sociology 57 (5), 1249–1258.

Bankins, S., Formosa, P., 2023. The ethical implications of artificial intelligence (AI) for meaningful work. J. Bus. Ethics 185 (4), 725–740.

Bano, M., Zowghi, D., Whittle, J., 2023. Exploring Qualitative Research Using LLMs. *arXiv preprint arXiv:2306.13298*.

Bassett, L., 2015. Introduction to JavaScript Object Notation: a To-the-point Guide to JSON. O'Reilly Media, Inc.

Braun, V., Clarke, V., 2021. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. Counsell. Psychother. Res. J. 21 (1), 37–47.

Braun, V., Clarke, V., Hayfield, N., Davey, L., Jenkinson, E., 2023. Doing Reflexive Thematic Analysis. Springer.

Briganti, G., 2024. How ChatGPT works: a mini review. Eur. Arch. Otorhinolaryngol. 281 (3), 1565–1569.

Burr, V., 2015. Social Constructionism. Routledge.

Chamberlain, K., 2015. Epistemology and qualitative research. Qualitative Res. Clinical and Health Psy. 9–28.

Christou, P.A., 2023. How to use artificial intelligence (AI) as a resource, methodological and analysis tool in qualitative research? Qual. Rep. 28 (7).

Christou, P.A., 2024. Thematic analysis through artificial intelligence (AI). Qual. Rep. 29 (2), 560–576.

Clarke, V., Braun, V., 2013. Successful Qualitative Research: A Practical Guide for Beginners.

Clarke, V., Hayfield, N., Moller, N., Tischner, I., 2017. Once upon a Time…: Story Completion Methods.

D'Amato, K., 2024. ChatGPT: towards AI Subjectivity. AI & SOCIETY, pp. 1–15.

De Cremer, D., Kasparov, G., 2021. AI should augment human intelligence, not replace it. https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it.

dos Anjos, J.R., de Souza, M.G., de Andrade Neto, A.S., de Souza, B.C., 2024. An analysis of the generative AI use as analyst in qualitative research in science education. Revista Pesquisa Qualitativa 12 (30), 1–29.

European Parliament, 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation).

European Trade Union Confederation, 2024. Artificial Intelligence for Workers, Not Just for Profit: Ensuring Quality Jobs in the Digital Age. ETUC Strategic Discussion on AI at Work.

Fuller, K., Morbitzer, K.A., Zeeman, J.M.M., Persky, A.C., Savage, A., McLaughlin, J.E., 2024. Exploring the use of ChatGPT to analyze student course evaluation comments. BMC Med. Educ. 24 (1), 423.

Google, 2025. Gemini. In: Google. https://gemini.google.com/.

Hamilton, L., Elliott, D., Quick, A., Smith, S., Choplin, V., 2023. Exploring the use of AI in qualitative analysis: a comparative study of guaranteed income data. Int. J. Qual. Methods 22, 16094069231201504.

Harper, D., 2011. Choosing a qualitative research method. Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners, pp. 83–97.

Hicks, M.T., Humphries, J., Slater, J., 2024. ChatGPT is bullshit. Ethics Inf. Technol. 26 (2), 38.

Hitch, D., 2024. Artificial intelligence augmented qualitative analysis: the way of the future? Qual. Health Res. 34 (7), 595–606.

Jain, A., Kulkarni, G., Shah, V., 2018. Natural Language processing. Int. J. Comput. Sci. Eng. 6, 161–167. https://doi.org/10.26438/ijcse/v6i1.161167.

Jalali, M.S., Akhavan, A., 2024. Integrating AI language models in qualitative research: replicating interview data analysis with ChatGPT. Syst. Dyn. Rev.

Jenner, S., Anderson, E., Fox, K., Ainsworth, B., Barker, M., Kreppner, J., 2025. Narratives of Identity, Social Media and Food Choice: a Story Completion Study with Young People.

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 Inter. 5 (CSCW1), 1–23.

Josselson, R., Hammack, P.L., 2021. Essentials of Narrative Analysis. American Psychological Association.

Khalil, H., Ameen, D., Zarnegar, A., 2022. Tools to support the automation of systematic reviews: a scoping review. J. Clin. Epidemiol. 144, 22–42.

Mauthner, N.S., Doucet, A., 2003. Reflexive accounts and accounts of reflexivity in qualitative data analysis. Sociology 37 (3), 413–431.

Mesec, B., 2023. The language model of artificial intelligence chatGPT-a tool of qualitative analysis of texts. Authorea Preprints.

Meta, 2025. Llama3. In: Meta. https://llama.meta.com/llama3/.

Morgan, D.L., 2023. Exploring the use of artificial intelligence for qualitative data analysis: the case of ChatGPT. Int. J. Qual. Methods 22, 16094069231211248.

Nakadai, R., Nakawake, Y., Shibasaki, S., 2023. AI language tools risk scientific diversity and innovation. Nat. Hum. Behav. 7 (11), 1804–1805.

Neumann, M., Rauschenberger, M., Schön, E.-M., 2023. "We need to talk about ChatGPT": the future of AI and higher education. In: 2023 IEEE/ACM 5th International Workshop on Software Engineering Education for the Next Generation (SEENG), OpenAI. (2025). ChatGPT. In OpenAI. https://chat.openai.com/.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., 2018. Improving Language Understanding by Generative Pre-training.

Ren, S., Tomlinson, B., Black, R.W., Torrance, A.W., 2024. Reconciling the contrasting narratives on the environmental impact of large language models. Sci. Rep. 14 (1), 26310.

Russell Group, 2023. Russell Group principles on the use of generative AI tools in education. https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf.

Sallam, M., 2023. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare.

Schlagwein, D., Willcocks, L., 2023. ': The ethics of using (generative) artificial intelligence in research and science 38 (3), 232–238. ChatGPT et al.

Scientific Software Development GmbH, 2024. ATLAS.ti. In: Scientific Software Development GmbH.

Singh, A., Patel, N.P., Ehtesham, A., Kumar, S., Khoei, T.T., 2025. A survey of sustainability in Large Language models: applications, economics, and challenges. 2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC).

Stokel-Walker, C., 2024. AI Chatbots Have Thoroughly Infiltrated Scientific Publishing. https://www.scientificamerican.com/article/chatbots-have-thoroughly-infiltrated-scientific-publishing/.

Tabone, W., De Winter, J., 2023. Using ChatGPT for human–computer interaction research: a primer. R. Soc. Open Sci. 10 (9), 231053.

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. Int. J. Qual. Methods 23, 16094069241231168.

Tariq, S., Iftikhar, A., Chaudhary, P., Khurshid, K., 2022. Examining some serious challenges and possibility of AI emulating human emotions, consciousness, understanding and 'self'. Journal of NeuroPhilosophy 1 (1).

The Guardian, 2023. 'It's destroyed me completely': Kenyan moderators decry toll of training of AI models. Guardian.

The Lancet, 2024. Rethinking research and generative artificial intelligence, 404 (10447), 1.

Törnberg, P., 2024. How to use large-language models for text analysis. https://doi.org/10.4135/9781529683707.

Towler, L., Bondaronek, P., Papakonstantinou, T., Amlôt, R., Chadborn, T., Ainsworth, B., Yardley, L., 2023. Applying machine-learning to rapidly analyze large qualitative text datasets to inform the COVID-19 pandemic response: comparing human and machine-assisted topic analysis techniques. Front. Public Health 11, 1268223.

Turobov, A., Coyle, D., Harding, V., 2024. Using ChatGPT for thematic analysis. arXiv preprint arXiv:2405.08828.

VERBI software, 2024. MAXQDA. VERBI software.

Verspoor, K., 2024. 'Fighting fire with fire'—using LLMs to combat LLM hallucinations. Nature.

Wachinger, J., Bärnighausen, K., Schäfer, L.N., Scott, K., McMahon, S.A., 2024. Prompts, pearls, imperfections: comparing ChatGPT and a human researcher in qualitative data analysis. Qual. Health Res., 10497323241244669

Wenfeng, L., 2025. DeepSeek. In: Williams, T.L., Lozano-Sufrategui, L., Tomasone, J.R. (Eds.), Stories of Physical Activity and Disability: Exploring Sport and Exercise Students' Narrative Imagination through Story Completion. Qualitative Research in Sport, Exercise and Health, vol. 14, pp. 687–705, 2022.

Yan, D., 2024. Posthuman creativity: unveiling cyborg subjectivity through ChatGPT. Qual. Inq., 10778004241231923

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., Carroll, J.M., 2023. Redefining qualitative analysis in the AI era: utilizing ChatGPT for efficient thematic analysis. arXiv preprint arXiv:2309.10771.

Zhao, F., Yu, F., Trull, T., Shang, Y., 2023. *A new method using LLMs for keypoints generation in qualitative data analysis* (9798350339840). In: (2023 IEEE Conference on Artificial Intelligence (CAI), Issue.